Spring 2018

Effectiveness of Social Media Analytics on Detecting Service Quality Metrics in the U.S. Airline Industry

Xin Tian
Old Dominion University, xtian@odu.edu

EFFECTIVENESS OF SOCIAL MEDIA ANALYTICS ON DETECTING SERVICE QUALITY METRICS IN THE U.S. AIRLINE INDUSTRY

by

Xin Tian

B.S. July 2008, Beijing Forestry University
M.S. August 2011, Old Dominion University

A Dissertation Submitted to the Faculty of
Old Dominion University in Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY

BUSINESS ADMINISTRATION – INFORMATION TECHNOLOGY

OLD DOMINION UNIVERSITY

May 2018

Approved by:

Wu He (Chair)

Chuanyi Tang (Member)

Ling Li (Member)

David D. Selover (Member)

ABSTRACT

EFFECTIVENESS OF SOCIAL MEDIA ANALYTICS ON DETECTING SERVICE QUALITY METRICS IN THE U.S. AIRLINE INDUSTRY

Xin Tian
Old Dominion University, 2018
Director: Dr. Wu He

During the past few decades, social media has provided a number of online tools that allow people to discuss anything freely, with an increase in mobile connectivity. More and more consumers are sharing their opinions online with others. Electronic Word of Mouth (eWOM) is the virtual communication in use; it plays an important role in customers' buying decisions. Customers can choose to complain or to compliment services or products on their social media platforms, rather than to complete the survey offered by the providers of those services. Compared with the traditional survey, or with the air travel customer report published by U.S. Department of Transportation (DOT) each month, social media offers features that can spread information quickly and broadly. This dissertation offers a novel methodology that, by utilizing emotional sentiment analysis, can help the airline industry to improve its service quality. Longitudinal data, retrieved from Twitter, are collected from twelve U.S.-based airline companies, in order to represent airline companies in different levels and categories. The data covers three consecutive months in Quarter 2 of 2017. Applied alongside the service quality metrics of the airline industry, the benchmark datasets for each metric are created. The purpose of this dissertation is to bridge the gap in traditional methodology for a service quality measurement in the airline industry and to demonstrate the way in which socialized

textual data can measure the quality of the service offered by airline service providers. In addition, sentiment analysis is applied, in order to get the sentiment score of each tweet. Emotional lexicons are used to detect the emotion expressed by the tweet in two emotional dimensions: each tweet's Valence and Arousal are calculated. Once the SERVQUAL model is applied and the keywords to find the corresponding social media data are created for each dimension, the results show that responsiveness, assurance, and reliability are positively correlated to the AQR score that measures the service quality of airline industry. This study also finds that a large amount of negative social media data will negatively affect the AQR score. Finally, this study finds that the interaction of the sentiment score and the arousal score of textual social media data play the important role in predicting the service quality of the airline industry. Finally, an opinion-oriented information system is proposed. In the last, this study provides theory verification of SERVQUAL.

Keywords: social media, service quality, SERVQUAL, emotional sentiment, text mining, natural languages processing (NPL), airline, sentiment analysis, eWOM

DEDICATION

This dissertation is dedicated to both of my parents for their endless love, support, and encouragement. My father, Fusheng Tian, not only raised and nurtured me but also taxed himself dearly over the years to provide for my education and intellectual development. My mother, Jinghua Gu, has been a source of encouragement and support during moments of despair and discouragement. Her motherly care has recently been shown in incredible ways. This dissertation is also dedicated to my husband, Xiaotao Jing, for his patience, love, friendship, and humor, for cooking delicious meals for me and the kids, and for supporting me to complete the incredible work. It is also dedicated to my son Braydon and to my daughter, Ellie, for letting me experience the love and joy of motherhood. Thanks for coming into my life and along with me in the journey to complete this work.

ACKNOWLEDGEMENTS

There are many people to thank for their support during my doctoral training. First and foremost, I am extremely grateful to my dissertation chair, Dr. Wu He, and to my dissertation committee members, Drs. Chuangyi Tang, Ling Li, and David Selover. Not only were they incredibly supportive and helpful in all of the aspects of my dissertation work, but they have also served as mentors and sounding boards during my time as doctoral student. I would have been lost without their advice on navigating doctoral studies, publishing, and finding a job. They have influenced me greatly as exemplary scholars who pursue bold ideas, work hard, and act with integrity.

I am also very thankful for the general support of the faculty and staff at Old Dominion University. The faculty of the Strome College of Business has been very helpful, even those who were not directly involved in the doctoral program. In particular, I want to thank Dr. Li Da Xu, who introduced me to the Information Technology Ph.D. program and who helped me a great deal on my journey. I also want to thank Dr. Harris Wu and Dr. Lan Cao for their support, and I salute Katrina Davenport, who is always willing and able to assist doctoral students with anything.

Finally, I am incredibly grateful to my husband, Xiaotao, for his patience and kindness. It would not have been possible to complete my doctoral studies without his support for the family.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Tables **Page**

CHAPTER

1 INTRODUCTION

With the emergence of the social media online platform, more and more consumers and companies are communicating and sharing experiences and product/service reviews online. Social media provides online tools and allows people to discuss anything freely, with the increase in mobile connectivity. According to research conducted by IBM Big Data, more than 2.5 quintillion bytes of data were generated daily in 2012 (Zikopoulos et al., 2012). Business analysis and big data mining were developed in order to extract insightful information from the collected massive data (He et al., 2015). In order to survive in the intense competition of today's business world, business can be based not only on the competition of lower prices, but also on service quality delivery (Zeithaml, Parasuraman, & Malhotra, 2000). Hearing about superior service offered by businesses becomes more and more important for customers, since customer loyalty, satisfaction levels, word-of-mouth behavior, deal-seeking behavior, and customers' behavioral intentions are highly affected by the perceived service quality of firms with which they do business (Park, Gretzel, & Sirakaya-Turk, 2007).

In the traditional way, businesses, including airline companies, use conventional survey-based techniques to conduct the measurements of service quality and customer satisfaction. Those survey-based techniques are AHP, SERVQUAL, and SERVPREF. And the most important technique to study service quality is SERVQUAL. SERVQUAL has been verified by past literatures in Table 1. Tsaur, Chang, and Yen (2002) studied the service quality in the airline service industry, using fuzzy MCDM. Tsaur et al. (2002)

found that many intangible attributes are difficult to measure. A survey was conducted and, by applying AHP, Tsaur et al. found that the most critical aspects of service quality are tangible, and the least critical aspect is empathy, in airline service. In addition, Tasur et al. (2002) found, from survey results, that courtesy, safety, and comfort are the most important attributes. Another study performed about service quality in the airline industry identified several SERVQUAL and industry-based items that significantly influence consumers' perceptions of overall service quality and their intention to re-patronize (Young, Cunningham, & Lee, 1994). In addition, the results suggest that the Air Travel Consumer Report has not been properly disseminated, nor has it been used by most consumers (Young, Cunningham, & Lee, 1994). However, survey-based methodology has its limitations, such as sample size, group of participants, and its need for respondents' recall of past events. These limitations can constrain the scalability of the measurements of service quality.

As increasing social media tools earn more usage and prevalence, an exponentially increasing number of customers are posting their life experiences (shopping, service, deal-seeking, and problem-solving experiences) on social media platforms like Facebook, Twitter, Blogs, YouTube, and review websites. Electronic Word of Mouth (eWOM) is the virtual communication that plays an important role in customers' buying decisions; negative WOM has more impact on a variety of aspects of business than positive WOM (Park & Nicolau, 2015). Pan and Crotts (2012) define social media as "the digital version of word-of-mouth", since social media "represents the materialization, storage, and retrieval of word-of-mouth content online." Customers share their opinions, ideas, suggestions, and complaints freely on their social media platforms

via their online community. Microblogs, like Twitter, are especially popular tools. Because mobile applications allow people to post information everywhere at any time, Twitter has become a popular and convenient way for consumers to complain about their perceived service or about products.

Rather than using a survey-based approach, consumer-generated social media content contains a variety of valuable pieces of information, like opinions, experiences, and viewpoints. The valuable information makes social media an important source to use in analyzing consumers' decision-making about purchases. Compared to the traditional method of conducting a market survey and social media data analytics, Leung, Lee, and Law (2012) suggest that social media content analysis may be more trustworthy and more reliable than information provided by the marketing departments within a company. Social media has characteristics that include its ability to spread messages more quickly and broadly than the use of any other methods. Thus, examining social media content is becoming important to businesses who want to pursue superior service quality and gain competitive intelligence with improved market performance. Social media content analysis is required to understand consumers' perceived service and to evaluate service quality. The trend is to analyze the massive amount of structured and unstructured data available from the social media platform. Social media data posted by consumers about appointed businesses can be retrieved and analyzed by those businesses as well as by their rivals, in order to better understand service quality (Zikopoulos et al., 2012). Businesses can learn how to improve their service and their products in order to achieve a sustainable competitive advantage. Previous studies (Parasuraman, Zeithaml, & Berry, 1988; Ramanathan & Karpuzcu, 2011) have already established multiple dimensions of

service quality. Specifically, for airline service quality, Young et al. (1994) measured the service quality of passenger airlines based on SERVQUAL and discussed it in a U.S. Air Travel Report released and published by Department of Transportation (DOT).

This dissertation seeks to develop a framework that uses social media analytics (with an emphasis on emotional sentiment analysis) to help to study the service quality perceived by consumers in the airline industry. According to SERVQUAL, the service quality will be investigated based on a benchmark dataset for each dimension of service quality. Furthermore, emotional sentiment (Lexicon-based) analysis will be applied, in order to examine the tweets for each service quality dimension. Compared with the Air Travel Report published by the DOT, the effectiveness of a social media analysis of service quality will be inspected. Twelve U.S.-based commercial airline carriers (United Airlines, Southwest Airlines, Frontier Airlines, Alaska Airlines, Express Jet, SkyWest Airlines, Delta Airlines, American Airlines, Hawaiian Airlines, Spirit, Virgin America and JetBlue) were researched on Twitter for the second quarter of 2017. My research questions are as follows:

What dimensions of airline service quality from textual social media data are associated with Airline Consumer Report by DOT?

What is the effect of massive negative textual social media data on an airline carrier and its rivals?

What is the pattern of emotional sentiment for each dimension of airline service quality?

What is the relationship between textual social media data and the Airline Consumer Report by DOT?

CHAPTER

2 MOTIVATIONS OF RESEARCH

Since an airline's service is highly related to its customer service, more and more customers are sharing their experiences on social media -- regardless of whether they had good or bad experiences with the airline. The most influential incident in the U.S. airline industry in the past five years happened in 2017. The United Express Flight 3411 oversale incident occurred on April 9, 2017 at O'Hare International Airport. A passenger, Dr. David Dao, was forcibly removed from his seat by airport police officers after having boarded the fight (Wahba, 2017). This was among the most serious customer service incidents in the history of the airline industry, and the incident was widely spread and shared by tweets through Twitter from the consumers on the airplane. The tweets, along with the videos, were soon picked up by mainstream media agencies like CNN and BBC and received a huge amount of attention on that day. The consequence of this incident was predictable: the public relationship and stock price of United were affected, and consumers revealed widespread dissatisfaction on their social media platforms.

On that day, over 60,000 tweets (retrieved from Twitter with "@united") mentioned United Airlines and either discussed or commented on the incident. Many people expressed that they would boycott United Airlines or tried to find an alternative airline, if that was possible. Even more, customers who held the Chase United Credit Card destroyed the card and posted pictures on Twitter to boycott United Airlines. Tweets about this incident were retweeted and were commented upon much more frequently than regular tweets; consumers were influenced, and the plight of the man in

the video evoked sympathy. Obviously, the power of social media is strong, and those negative tweets set off a chain reaction about this incident. Competitor airlines (among them Southwest Airlines) also took advantage of the opportunity to gain more attention by persuading consumers about their better service and by offering promises about never overselling in the future (Wahba, 2017). Even beyond United Airlines, other airline carriers were affected after the incident. The number of tweets that mentioned airline carriers increased in the following days. Wong et al. (2017) studied the way in which the event gained traction on social media and found that mimesis behavior encourages homogeneous behavior and reactions during times of crisis.

Luo (2007) reinforced that negative reviews or comments can be more powerful in decreasing sales than positive reviews can be in increasing sales (Chevalier & Mayzlin, 2006). From this, we can note that social media can play an important role in contemporary business, particularly in bridging the gap of traditional research methodology for airline industry. Even though the DOT has the official data for all of the operational airline companies in the U.S., only reports with statistics for the two months prior to the current date can only be accessed on the DOT website. Due to this time lag, customers cannot get the real-time data for decision-making and airline companies cannot immediately respond to customers regarding the report provided by DOT in order to use it to improve their service quality or to maintain their customer relationship. This dissertation will introduce a novel methodology and will utilize emotional sentiment analysis to help the airline industry to improve its service quality and respond to a public crisis in a short time. Bowen and Headley (2017) published the Air Quality Rating (AQR) to measure the service quality of U.S. airline carriers. Alaska Airlines and Delta

Airlines were announced as No.1 and 2 in the AQR report in 2017. However, the data used for generating the AQR report was mostly based on that of the 2016 U.S. Airline Consumer Travel Report published by DOT. Bowen and Headley (2017) reported that nine airline companies showed improvement in the AQR report during 2016. In this study, all of the U.S.-based airline companies in the DOT airline consumer report are included: Southwest Airlines, Alaska Airlines, American Airlines, United Airlines, Delta, Express Jet, Hawaiian Airlines, Jetblue, Skywest, Spirit, Virgin America, and Frontier Airlines. The reason for choosing these companies is that each company can represent one type of airline carrier. For example, Southwest Airlines is known to be flexible by allowing passengers to choose their seats and by charging no baggage fee. Frontier Airlines, JetBlue, and Spirit are known by their cheap airfares; United Airlines, Delta, and American Airlines have the largest air route networks, and United Airlines suffered from the incident in April 2017; and Alaska Airlines is listed as No.1 in the AQR report for its excellent service. AQR has been cited to make an airline industry standard and to allow for a comparison of airline companies' performance (Bowen & Headley, 2017). In order to calculate the AQR score, DOT data is used and is applied to the formula, based on the on-time rate (OT), denied boarding (DB), mishandled baggage (MB), and customer complaints (CC). Only the on-time rate has a positive impact; the other three have a negative impact on AQR. The formula to calculate the AQR is listed below:

$$AQR = \frac{(+8.63*OT)+(-8.03*DB)+(-7.92*MB)+(-7.17*CC)}{(8.63+8.03+7.92+7.17)} \tag{1}$$

With its ability to measure service quality in the airline industry, AQR has been designed for the industry, and uses the data from DOT report. The results can be used by airline carriers, can help newly nominated airline services, and can assist DOT in modifying the rules. However, since the AQR is based on the Air Travel Report from DOT, the data in the report has at least a two-month time lag. Consumers cannot get real-time information when they need it to make purchase decisions, and airline companies are unable to improve their service right away.

Sometimes, consumers have reported incidents on their social media platforms and this information has spread rapidly. Then, airline companies can respond to it officially and can regain their reputations using social media tools. To monitor and analyze social media effectively, airline companies need to adapt multiple technologies and to hire data analysts and data scientists to mine tweets from Twitter, reviews on Tripadvisor.com, and comments from their websites and from other social media platforms. The most effective and feasible method is mining Twitter data, since Twitter has a lot of users and it provides an API to connect with the server and to retrieve the specific data. To better use these data, businesses can develop an information system to monitor their mentions on social media. For service companies, emotional sentiment and data mining can be applied to investigate service quality. Since service quality is the backbone of business, the earlier that any negative issues are found, the easier it is to improve service quality. If the sentiment has changed considerably, an alert can be sent to the appropriate person and then the proper operations and responses can be performed. Businesses can then use this information system to improve their service quality, to monitor their market performance, and to adjust their marketing strategies.

CHAPTER

3 LITERATURE REVIEW

Service quality assessment is an important area for multidisciplinary research. Operational management, marketing, and management information systems have had many research articles written about measuring service quality. Prior work about measuring quality has mainly focused on physical products and on tangible goods (Palese & Piccoli, 2016). In the late 20th century, the famous model -- SERVQUAL-- for measuring quality of service was proposed by Parasuranman et al. (1985, 1988). According to Parasuranman et al. (1985), service quality is hard to measure because of intangibility. Most research papers cover the application of SERVQUAL and its ability to conduct a survey to evaluate consumers' perceived service. This method is feasible, but it is limited by the sample size, response rate, and reliability of the responses. With the emergence of social media and the prevalence of mobile platforms (mobile applications), socialized textual data has been found to be a boon to business. Customer experience is the basis for the effective measurement of service quality (Petter et al., 2012). Social media analysis provides the ability to retrieve socialized textual data and to analyze them through text mining, clustering, and sentiment analysis. Competition among airline carriers is becoming intense, and a competitive advantage can be discovered by airline carriers who take service quality into account. Most especially, with the emergence of technology and social media, airline carriers can now communicate with consumers in multiple ways: via online chat, Twitter, Facebook, official websites, phone calls, online surveys, and so on. Airline carriers may change their marketing strategies based on the

results of their social media analytics, and can gain a reputation from the opinions of their consumers.

3.1 Service Quality in the Airline Industry

Studies about service quality have been done since the 1980s (Grönroos, 1984; Rust & Oliver, 1994; Cronin, Brady, Thomas & Hult, 2000). As defined by Grönroos (1984), service quality is "the outcome of an evaluation process, where the consumer compares his expectations with the service he perceives he has received." Grönroos suggested that, when considering service quality, technical quality, functional quality, and corporate image also must be considered. Further, other researchers have studied service quality and have suggested that customer satisfaction has a positive relationship with service quality (Mukherjee, Nath, & Pal, 2003; Ramanathan & Karpuzcu, 2011). Since service quality is highly related to customer satisfaction and firm performance, it is important for companies of the service industry to measure and to evaluate customer satisfaction. Sasser, Olsen, and Wyckoff (1978) mentioned, back in the '70s, that service quality could be measured by materials, facilities, and personnel. From the 1980s until the 2010s, service quality measurement has been developing, and the research method has improved over these thirty years. In order to measure service quality accurately, the SERVQUAL model is the fundamental model for assessing service quality, since it compares expected service and perceived service from consumers (Parasuraman et al., 1985). Ten dimensions are covered in the SERVQUAL model: access, communication, competence, courtesy, credibility, responsiveness, security, tangibility, and understanding/knowing the customer.

Several years later, Parasuraman et al. (1988) refined the ten dimensions into five major dimensions: reliability, responsiveness, assurance, empathy, and tangibility. Martin (1986) argued that only two main dimensions should be assessed for service quality: service procedure and consumers' conviviality. Martin (1986) considered the managerial portion and the communication with consumers. During the following twenty years, researchers expanded and modified the SERVQUAL five dimensions across different service industries. Ramanathan and Karpuzcu (2011), in their study, suggested using seven metrics to measure service quality: responsiveness, flexibility, availability, assurance, personnel contact quality, reliability, and tangibles (as shown in Figure 1). And Novack, Rinehart, and Langley (1994) argued that personnel of the company and the traits of executives can be used to measure service quality. Another aspect was proposed by Parasuraman et al. (1988), who said that customers' opinions can strengthen the understanding and the measurement of service quality. According to previous research (Parasuraman, Zeithaml and Berry, 1988; Ramanathan and Karpuzcu, 2011), service quality metrics can be defined as follows:

- Reliability: The ability to perform the promised service, both dependably and accurately
- Responsiveness: Willingness to help customers and to provide prompt service
- Flexibility: Flexibility to allow for different transaction options and methods
- Availability: The availability of products in stock
- Personnel Contact Quality: The knowledge and courtesy of employees, as well as their ability to ease communication with customers

- Tangibles: The appearance of the physical facilities, the equipment, the appearance of personnel, and the communication materials.
- Assurance: The ability to convey trust and confidence to customers and to make them feel that they are receiving good service

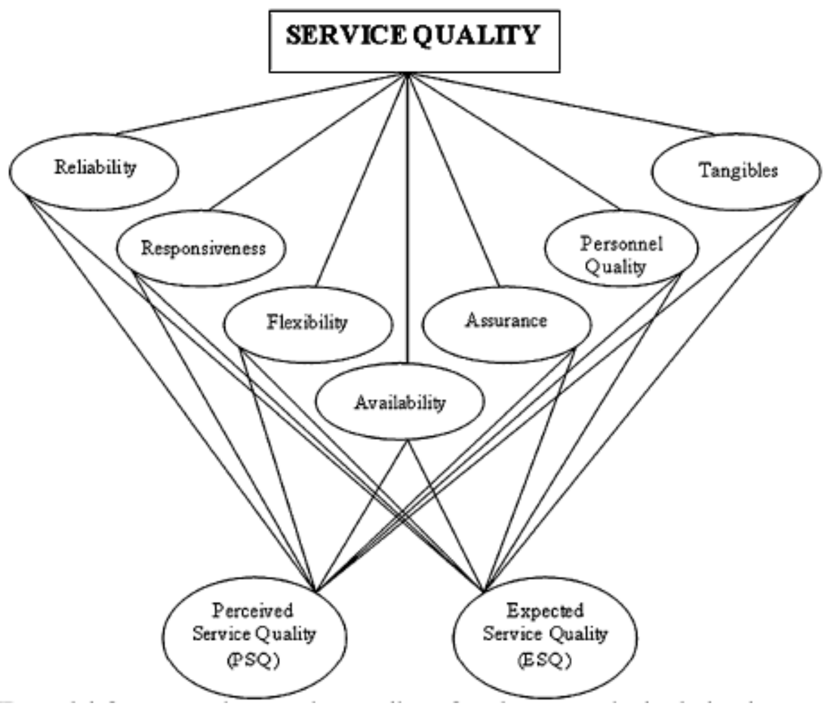


Figure 1. Seven Metrics for Measuring Service Quality

(Adapted by Ramanathan & Karpuzcu, 2011)

Ladhari (2009) reviewed most of the measurements of service quality and found that SERVQUAL is the most popular model, one that used by many researchers. SERVQUAL has better reliability and validity and can evaluate the service expectations

and perceptions of consumers (Parasuraman et al., 1988). In the past thirty years, prior research has applied SERVQUAL as the guideline and theoretical background, and then has developed survey questions to measure each dimension of service quality. To compare the expectations of customers and their perceptions after they received the service, researchers have researched the differences between expected service and perceived service across a number of service industries. Customer expectations and customer perceptions are the two essential parts of the SERVQUAL model.

The SERVQUAL model has been applied to evaluate service quality in a variety of industries, as among them education, banking, insurance, airline services, and health care. Table 1 shows SERVQUAL in different research contexts and in modified dimensions for evaluating service quality in different industries. Yang and Fang (2004) identified eight dimensions and sub-dimensions of online service quality. They are responsiveness, service reliability, ease of use, access, system reliability, timeliness, security, and competence. These eight dimensions derive from SERVQUAL; modifications were based on context. In another paper, El Saghier and Nathan (2013) found that only four dimensions (reliability, responsiveness, empathy, and assurance) can influence service quality in banking services. Bansal and Taylor (2015) examined switching intentions, service quality, and customer satisfaction, and argued that service quality is one antecedent to customer satisfaction, while service quality is the key for switching intentions.

TABLE 1 A BRIEF REVIEW OF THE DIMENSIONS OF SERVICE QUALITY

Author(s)	**Dimensions**	**Context**
Yang & Fang (2004)	responsiveness, reliability, credibility, competence, access, courtesy, communication, information, responsiveness, and website design	E-service
Akbaba (2006)	tangibles, adequacy in service supply, understanding and caring, assurance, and convenience	Hotel industry
Polyorat & Sophonsiri (2010)	tangibles and empathy, reliabilities, responsiveness, and assurance	Chain restaurant
El Saghier and Nathan (2013)	reliability, responsiveness, empathy, and assurance	banking services
Kitapci, Taylan Dortyol, Yaman & Gulmez (2013)	empathy, tangibility, responsiveness, and assurance	Supermarket
Thaichon, Lobo, Prentice & Quach (2014)	network quality, customer service and technical support, information quality, and security and privacy	Internet service providers
Saeedpoor et al. (2015)	tangibility, reliability, knowledge and skill of staff in maintaining mutual trust	Life insurance firms

	between customers and service providers, willingness to deliver services in a timely manner and to help customers; empathy	
Ali, M., & Raza, S. A. (2017).	compliance, assurance, reliability, tangibles, empathy, and responsiveness	Service industry: Islamic bank

To evaluate perceptions of airline service, prior studies were based on survey questions and on the SERVQUAL model. Ostrowski et al. (1993) examined the relationship between service quality and retained preference, which measured customer loyalty in the commercial airline industry. The data was collected from two air carriers; the researchers found a positive relationship between service quality and customer loyalty in the commercial airline industry. And AHP methodology was applied in measuring service quality in the airline industry by Tsaur et al. (2012). Applying the fuzzy set theory to evaluate the service quality of the airline, Tsaur et al. (2012) found that many intangible attributes are difficult to measure. Applying the AHP-based survey showed that the most concerning aspects of service quality are tangible, and that the least concerning aspect is empathy. Courtesy, safety, and comfort were the most concerning attributes for Tsaur et al. (2012). Mazzeo (2003) also found that being on-time plays an important role in service quality in the airline industry; flight delays are significantly related to weather conditions, air congestion, and scheduling decisions (U.S. Bureau of

Transportation Statistics, 2000). The time lag is an obvious issue when analyzing service quality in the airline industry. Consumers need a service quality model to evaluate the service quality, and two dependent variables – perceived service and expected service – are essential for measuring service quality (Grönroos, 1984). To maintain a business' market share, Mazzeo (2003) also argued that when customers have more choices, companies have more incentive to improve service quality by offering lower prices and better service. However, gathering the opinions of consumers is not easy. Surveys can be conducted with limited sample sizes, but the respondents may not represent all of the consumers. This dissertation will bridge the gap in the traditional survey method for service quality measurement in airline industry.

Hussain et al. (2015) investigated the relationships among service quality, service provider image, customer expectations, perceived value, customer satisfaction, and brand loyalty in a Dubai-based airline. Questionnaires were conducted based on the SERVQUAL model and identified six dimensions: reliability, responsiveness, assurance, tangibility, security and safety, and communication. Another research paper regarding service quality in the airline industry was written by Tsaur et al. (2002), who established five aspects and fifteen service quality criteria. I will use both of these research papers to design the benchmark dataset for assessing service quality in the airline industry using social media data. Table 2 shows the six evaluation dimensions of airline service quality, as studied in this dissertation.

TABLE 2 SIX DIMENSIONS OF AIRLINE SERVICE QUALITY

Dimensions of Service Quality	**Attributes and Key Words**
Responsiveness	Willingness to help passengers; providing prompt service; keeping passengers informed about delivery of service; keeping passengers updated if any modified schedule; quickly response customer's requirements.
Assurance	Providing service actively; language skill or translation help of crew members; pilots' informative announcement in different contexts of culture; employee's skillfulness; courtesy towards customers
Tangibility	Comfortable seats and the cleanliness of the cabin; cleanliness of the aircraft interior and exterior; variety of food, food service and food quality; on-board entertainment: movie and music; the appearance of the crew; complimentary pillow or blankets
Reliability	Efficiency of the check-in process, flight punctuality, timeliness (arrival in promised time), handling of missing luggage complaints.
Security and Safety	Personal safety; luggage safety; animal safety
Communications	communication between cabin crew and passengers; the ability to communicate with passengers in different languages; the communication between pilot and passengers; informative announcement during the flight.

In this model, security and safety are the greatest and most important assets of the airline industry; all airline carriers are ensuring their passengers the required security and safety. Especially after the events of 9/11, passengers consider the most important factors --

security and safety -- before they make purchase decision or any travel decision. Since Tsaur et al. (2002) argued safety is one of the most concerned attributes, and I have included the security and safety as one metric of service quality. Sum Chau and Kao (2009) point that communication plays an important role in service quality. They found communication style is critical for measuring service quality. This was confirmed by Hussain et al. (2015) and pilot's informative announcement during the flight are important factors leading service quality.

3.2 Social Media Analysis

Since social media has massive consumer-generated content, mining social media data, in recent years, has become a highly economical and efficient way for businesses to understand consumer needs (Crotts, Mason, & Davis, 2009; Duan et al., 2013). Big data technology enables organizations to conduct deep analyses of their business data to both measure and understand service quality. Bates et al. (2014) discussed six use cases that leverage big data analytics to identify and manage high-risk and high-cost patients. A recent study by Xiang, Schwartz, Gerdes, & Uysal (2015) used a text analytical approach to analyze a large quantity of consumer reviews extracted from Expedia.com and demonstrated the utility of big data analytics to better understand the relationship between hotel guests' experience and satisfaction. Based on the SERVQUAL model and its variations, we can reasonably use big data technology to mine a large volume of social media data in order to potentially identify customer expectations for a service and the perceptions of those customers after they receive that service.

In this dissertation, I propose a framework for big social data comparative analytics (Figure 2) to help interested businesses to leverage big data solutions mine social media data in order to contextually compare the service quality among peers. The proposed framework adopts the seven service quality measurement metrics proposed by the five dimensions of service quality: reliability, responsiveness, empathy, assurance and tangibility (See Table 2). Prior studies found that customer experience or customer satisfaction can used to measure the perceived service quality and the expected service quality from customers. In SERVQUAL model is not crucial for measuring service quality of the company (Taylor & Baker, 1994; Olorunniwo, Hsu, & Udo, 2006).

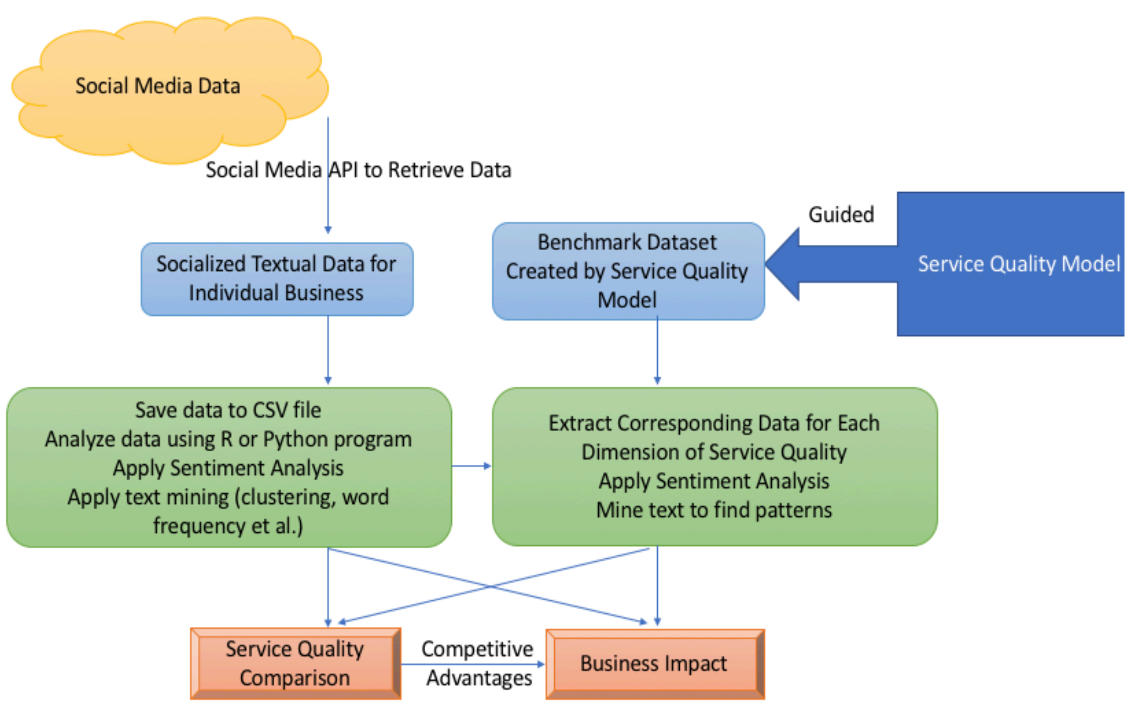


Figure 2. A Proposed Framework for Using Social Media Analytics to Study Service Quality

In the proposed framework, big data technology is used as a solution to analyze social media data from targeted business and their peers, in order to visualize and benchmark comparisons among peers across different service quality measurement metrics that may impact customer satisfaction. Text classification algorithms can be used to mine consumer-generated social media content based on the specified service quality measurement metrics. Then, sentiment analysis, which is the computational detection and study of opinions, sentiments, emotions, and subjectivities in text, can be conducted on the texts associated with each metric in order to identify consumer perceptions (including expected and perceived), generating a score from -1 (the most negative opinion) to 1 (the most positive opinion). The overall sentiment score of Dimension i can be calculated using the following formula provided by Duan et al. (2013):

$$S_i = \frac{N_{pi} - N_{ni}}{N_{pi} + N_{ni}} \tag{2}$$

where N_{pi} denotes the number of positive sentences in Dimension i and N_{ni} denotes the number of negative sentences in Dimension i. To make the process of conducting sentiment analysis easier, researchers can use existing popular sentiment analysis tools or services such as Lexalytics, SentiWordNet, SentiStrength, Social Mention, Trackur, Sysomos, and Viralheatto extract positive or negative sentiment scores from text (Pang & Lee, 2004). These sentiment analysis tools mainly rely on machine learning techniques such as Support Vector Machine (SVM), Naive Bayes, Maximum Entropy, and Matrix Factorization to classify texts into positive or negative categories, and they have been used in many studies in the sentiment analysis literature (Pang, Lee,

& Vaithyanathan, 2002; Li & Wu, 2010). In this dissertation, I will use Lexicon-based sentiment analysis rather than machine learning. The applications of sentiment analysis are powerful, and the results could be broad and valuable. Past research has found that a fluctuation in sentiment on social media has been identified to correlate with fluctuations in the stock market (Ranco et al., 2015). For example, the result of the presidential election can be predicted by gauging public opinion on social media when policy announcements are made. Wong et al. (2017) studied event-gained traction on social media and found that mimesis behavior encourages homogeneous behavior and reactions during the times of crisis. However, they do not consider the sentiment that is contained in the tweets and they do not examine the contribution of sentiment to information spread. Anitsal et al. (2017) investigated the top ten airline carriers in the U.S. using sentiment analysis and found that the customer relationship can be investigated in detail, and that Delta, Southwest, Alaska, and SkyWest Airlines have the most positive sentiments expressed about their cabin crews and their attitude to their passengers. Customer-employee collaboration in the co-creation of the airline service industry is vital to the success of airline companies. Waguespack and Rhoades (2014) argued that airline carriers have established a social media center in effort to avoid service quality failures, and that viral incidents can significantly affect publicity. In psychology, emotional sentiment has two major dimensions: Valence and Arousal (Mäntylä et al., 2016). Valence is the direction of the emotion, like positive and negative. The higher the Valence score, the more positive the emotion. Another dimension, Arousal, describes the intensity of emotion, or the psychological state of being reactive to stimuli. The Arousal

score will indicate the action of the increased alert and the readiness of responses (Preoţiuc-Pietro, 2016). Based on the prior literature, I offer the following hypothesis:

Hypothesis:

H1: Three dimensions (Responsiveness, Assurance, and Reliability) of service quality will be positively associated with score of AQR.

H2: The sentiment analysis results of tweets will align with the DOT Air Travel Consumer Report.

H3: The larger volume of negative social media data for the service quality dimension, the more complaints received and shown on the DOT Air Travel Consumer Report (lower AQR).

H4: Emotional direction (valence) of textual social media data is correlated with AQR.

H5: Emotional intensity (arousal) of textual social media data is correlated with AQR.

H6: The higher score of interaction of valence and Arousal, the higher AQR.

H7: The higher sentiment score and higher arousal, the higher AQR

IBM, SAP, Oracle, and Microsoft are some of existing data analytic platforms. They can be integrated to store, manage, analyze, and compare data from different companies across numerous social media sources in order to generate detailed service

quality reports, for better understanding and suggestions. As they compare consumers' expectations and perceptions of service quality for each business, the businesses generate a data analysis report containing concept maps, word clouds, and sentiment scores (positive and negative) for each service quality measurement metric, as well as a weighted score for the overall service quality of each business. Using such a report, I can then conduct a service quality comparison for different businesses against individual service quality metrics. Such comparisons are supported by advanced analytics, text mining, sentiment analysis, business intelligence, and contextual data generated from the big data. This proposed framework is especially applicable to assessing service quality for consumer service-oriented sectors such as retail, insurance services, financial services, healthcare, and e-business. Customer service is important to the companies in these sectors because it can help to differentiate a company from its competitors. The proposed analysis is particularly relevant for very large businesses or for businesses with a large population of active social media users. Even when the user-generated content is small, companies can still do a manual analysis. However, if managers feel that they have too much social media data to analyze efficiently, they should consider a more automated analysis. My proposed framework, the social media analysis method and the proposed information system, will be helpful in this regard. An information system will be proposed, as well.

CHAPTER

4 DATA COLLECTION

4.1 Reported Data from the Department of Transportation (DOT)

This paper will use the 2017 Air Consumer Report from U.S. Department of Transportation, from the 2nd quarter (April, May, June) to the 3rd quarter (July, August, September). Since the report was published two months later, some data should be collected in the later report. For example, the July report has the published data regarding May.

Data included in the report are flight delays, mishandled baggage, oversales, consumer complaints, airline animal incident reports, and the customer service report to the Department of Homeland Security. All of the data are officially published in the airline consumer report. The report includes information from 12 U.S. carriers: Spirit, ExpressJet, JetBlue, SkyWest, Frontier, American, United, Virgin America, Alaska, Southwest, Delta, and Hawaiian. In this study, I will investigate all of the U.S.-based airline carriers in this dissertation and will focus on twelve U.S. based airline carriers: Alaska Airlines, the No.1 U.S. carrier (Bowen and Headley, 2017), United Airlines, Frontier, and Southwest Airlines. The report is divided into six sections:

- Flight Delays
- Mishandled Baggage
- Oversales

- Consumer Complaints (flight problems, baggage, reservation/ticketing/boarding, customer service, fares, refunds, oversales, disability, discrimination, advertising, animals, other)
- Customer Service Reports to the Transportation Security Administration
- Airline Reports of the Loss, Injury, or Death of Animals During Air Transportation

The sections that deal with flight delays, mishandled baggage, and oversales are based on data collected by the Department's Bureau of Transportation Statistics. The section that deals with consumer complaints is based on data compiled by the OAEP's Aviation Consumer Protection Division (ACPD). The section that deals with customer service reports to the Department of Homeland Security's Transportation Security Administration (TSA) is based on data provided by TSA. The section that deals with animal incidents during air transport is based on reports required to be submitted by the airlines to the ACPD. Each section of the report is preceded by a brief explanation of how to read and understand the information provided.

The report is usually issued during the second week of each month. Oversales are reported at a quarterly (rather than a monthly) rate, and the oversales figures may be slightly older than the other data, in certain months.

TABLE 3 DOT REPORT COMPLAINT CATEGORIES

Flight Problems	Cancellations, delays, or any other deviations from schedule, whether planned or unplanned.
Oversales	All bumping problems, whether or not the airline complied with DOT oversales regulations.
Reservations, Ticketing, Boarding	Airline or travel agent mistakes made in reservations and ticketing; problems in making reservations and in obtaining tickets, due to busy telephone lines or waiting in line, or delays in mailing tickets; problems boarding the aircraft (except oversales).
Fares	Incorrect or incomplete information about fares, discount fare conditions and availability, overcharges, fare increases, and level of fares in general.
Refunds	Problems in obtaining refunds for unused or lost tickets, fare adjustments, or bankruptcies.
Baggage	Claims for lost, damaged or delayed baggage, charges for excess baggage, carry-on problems, and difficulties with airline claims procedures.
Customer Service	Rude or unhelpful employees, inadequate meals or cabin service, treatment of delayed passengers.
Disability	Civil rights complaints by air travelers with disabilities.
Advertising	Advertising that is unfair, misleading, or offensive to consumers.

Discrimination	Civil rights complaints by air travelers (other than disability); for example, complaints based on race, national origin, religion, etc.
Animals	Loss, injury, or death of an animal during air transport provided by an air carrier.
Other	Frequent flyer, smoking, tour credit, cargo problems, security, airport facilities, claims for bodily injury, and other issues not classified above.

According to Airline Consumer Report published by the DOT for June, July, and August of 2017, I can derive the reported data of April, May, and June (Second Quarter) of 2017 for these criteria: on-time rate, involuntary denied boarding (per 10,000 passengers), mishandled baggage (per 1,000 passengers), and customer complaints (per 100,000 passengers). Since denied boarding incidents are reported quarterly, the data that shows in the Table 4 is from 2017 August report. Involuntary denied boarding uses the same numbers for all three-consecutive months in Quarter 2 of 2017. Table 4 shows the reported data by the twelve airline carriers in this study, from April to June of 2017.

TABLE 4 DOT AIRLINE CONSUMER REPORT (APR - JUN 2017)

Airline Carriers	**Month**	**On-time**	**Involuntary Denied Boarding**	**Mishandled Baggage**	**Customer Complaints**
Alaska	Apr	81.6%	0.42	1.41	1.00
	May	82.6%	0.42	1.60	0.44
	Jun	82.9%	0.42	1.85	0.47
Frontier	Apr	79.5%	0.49	2.31	2.42
	May	76.6%	0.49	2.57	3.01
	Jun	73.1%	0.49	2.39	1.86
Southwest	Apr	79.5%	0.64	2.43	0.50
	May	77.3%	0.64	2.90	0.59
	Jun	73.3%	0.64	3.35	0.50
United	Apr	81.9%	0.44	2.12	3.04
	May	82.3%	0.44	2.12	2.02
	Jun	79.4%	0.44	2.47	2.09
Hawaiian	Apr	**88.8%**	0.08	2.52	1.58
	May	**89.7%**	0.08	3.00	1.05
	Jun	**90.4%**	0.08	2.47	0.60
SkyWest	Apr	80.0%	0.26	3.03	0.81
	May	82.4%	0.26	2.61	0.70
	Jun	81.0%	0.26	3.08	0.47
American	Apr	78.7%	0.56	2.81	2.68
	May	80.1%	0.56	2.56	2.15
	Jun	73.2%	0.56	3.20	2.09
Spirit	Apr	77.0%	**1.25**	1.46	**7.20**
	May	69.0%	**1.25**	1.65	**11.39**
	Jun	68.3%	**1.25**	1.82	**7.38**
ExpressJet	Apr	75.7%	0.63	**4.67**	1.79
	May	76.8%	0.63	**3.45**	1.18
	Jun	75.1%	0.63	**4.00**	0.83
VirginAmerica	Apr	64.6%	0.53	1.42	2.94
	May	58.7%	0.53	1.57	3.12
	Jun	67.2%	0.53	1.73	2.70
Delta	Apr	76.9%	0.09	3.04	2.52
	May	82.8%	0.09	1.67	1.21
	Jun	82.8%	0.09	2.08	0.80
JetBlue	Apr	72.4%	0.04	1.50	1.19
	May	67.2%	0.04	1.66	1.44
	Jun	60.6%	0.04	1.65	1.27

In all three reports, the top four complaint categories are flight problems, baggage, reservation/ticketing/boarding problems, and customer service. Figure 3 illustrates that Hawaiian has the highest on-time rate in the three consecutive months, while Virgin America has the lowest on-time rate in the first two months and the second lowest in the following month. The good weather may be the reason for the on-time rate. Since the flights do not be affected much by weather problems in Hawaii, that leads to the higher on-time rate than flights operated much in the East Coast. Overall on-time rates for all airline carriers show that the April has the better on-time rate than the other two months (averaging approximately 76%). Regarding mishandled baggage, ExpressJet performed the worst in all three months; most airline carriers reported about 2.6 incidents of mishandled baggage per 1,000 passengers, with American Airlines and Southwest Airlines having the poorest performance in baggage handling (Figure 4). Figure 5 shows that Spirit Airlines had the most complaints across the three months, followed by Virgin America, United, and Frontier. The rest of the airline carriers received few complaints.

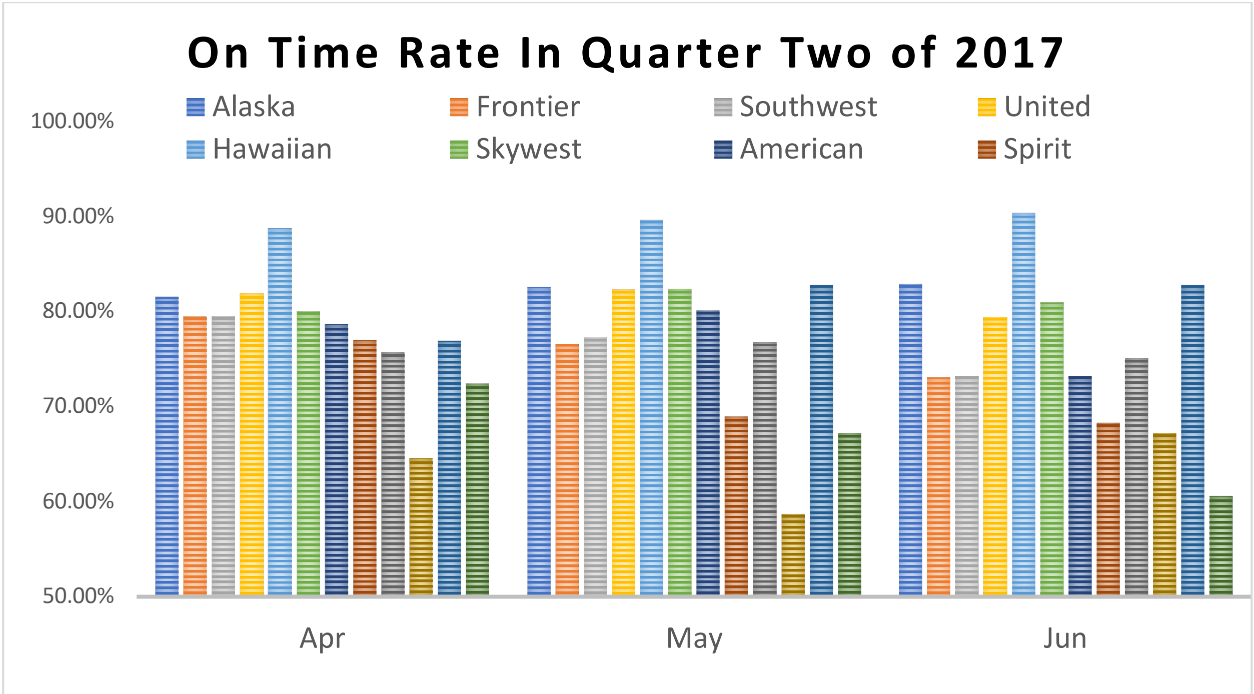


Figure 3. On Time Rate in Quarter Two of 2017

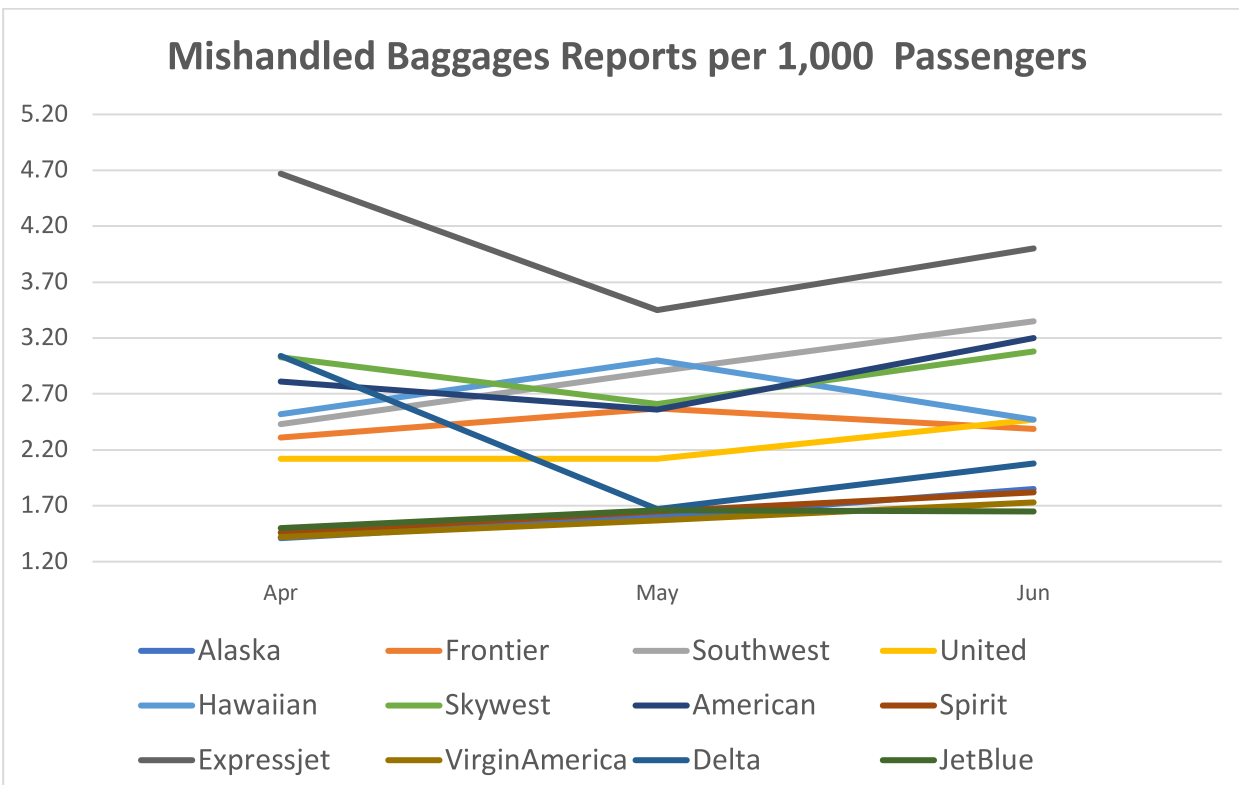


Figure 4. Mishandled Baggage Reports per 1,000 Passengers

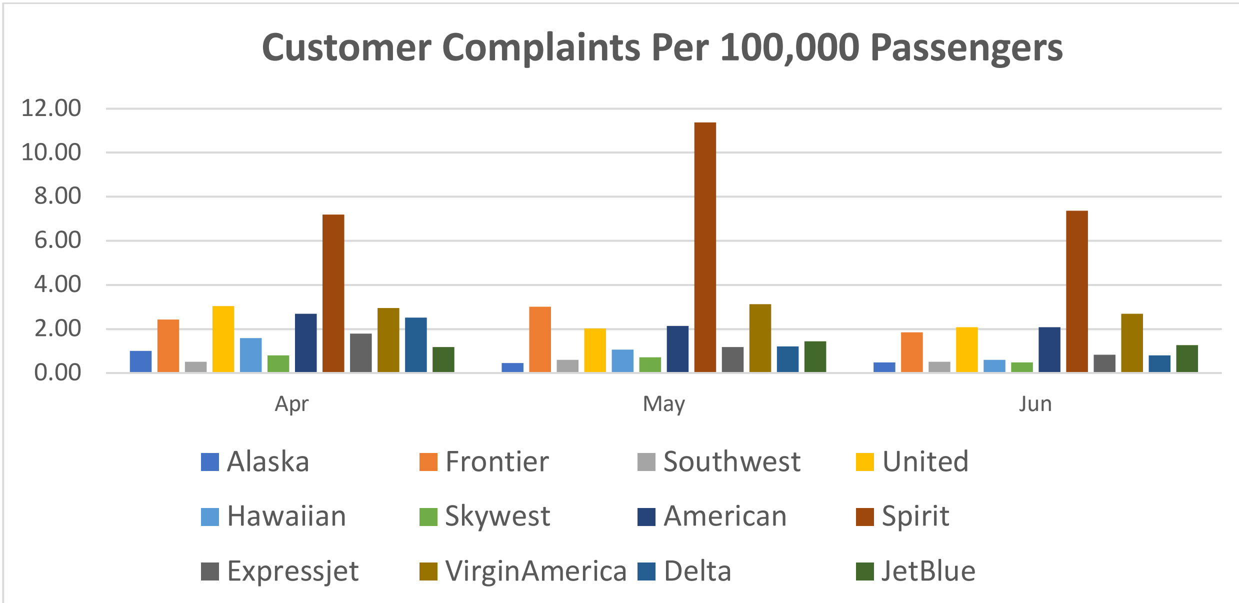


Figure 5. Customer Complains Per 100,000 Passengers

After applying the formula to calculate Airline Quality Ratings (AQR), the twelve airline carriers have the AQR showed in Table 5. Table 6 shows the number of enplanements (passengers who boarded on the plane) of each of the airline carriers over the three months, separately. Southwest Airlines (No.1), American Airlines (No.2), and Delta (No.3) carried the most passengers in this period. Virgin America and Hawaiian carried the least passengers compared to other airline carriers. Since Virgin America and Hawaiian have limited routes for customer to choose, the number of their scheduled flights is much lower than that of larger airlines.

TABLE 5 AQR SCORE FOR ALL AIRLINE CARRIERS FROM APR-JUN 2017

	AQR Score		
Airline Carriers	**April**	**May**	**June**
Alaska	**-0.46**	**-0.38**	**-0.45**
Frontier	-1.01	-1.24	-0.94
Southwest	-0.61	-0.81	-0.91
United	-1.10	-0.87	-0.98
Hawaiian	-0.85	-0.76	-0.53
SkyWest	-0.83	-0.65	-0.72
American	-1.20	-1.05	-1.21
Spirit	**-1.89**	**-3.11**	**-2.25**
ExpressJet	-1.47	-1.08	-1.14
Virgin America	-0.95	-1.07	-0.99
Delta	-1.22	-0.49	-0.50
JetBlue	-0.55	-0.57	-0.54

TABLE 6 NUMBER OF ENPLANEMENTS OF EACH AIRLINE CARRIERS

Airline Carriers	**April**	**May**	**June**
Alaska	2,008,750	2,193,846	2,237,360
Frontier	1,248,519	1,361,806	1,449,410
Southwest	**13,331,080**	**13,567,828**	**14,090,883**
United	6,734,304	7,214,180	7,715,405
Hawaiian	819,181	883,544	931,026
SkyWest	2,783,144	2,952,995	3,161,131
American	10,070,814	10,637,844	10,762,186
Spirit	1,813,630	1,888,028	1,916,923
ExpressJet	1,324,149	1,361,065	1,369,080
Virgin America	**653,156**	**683,482**	**715,225**
Delta	9,953,754	11,122,894	11,330,703
JetBlue	2,913,464	2,905,641	2,899,515

4.2 Data from Social Media: Twitter

Social media data (Twitter data) were collected regarding twelve U.S. carriers from Quarter 2, 2017 (April 1 to June 30). Twitter is one of most popular social media platforms; on it, consumers post their opinions, complaints, concerns, and compliments, rather than reporting them to the DOT website. Twitter is the online social networking service that enables users to send and read short 140-character messages called “tweets.”

Only registered users can read and post tweets; unregistered visitors are not allowed to post tweets but can read others' tweets. Hence, Twitter has become a popular public platform to mine the opinions of people all over the world and across all age categories.

From this social media platform, one can not only find the number of complaints, concerns, opinions, and compliments; one can also reflect the content of those posts. It is not possible for consumers, even the airline companies, to know the details of the complaints in the DOT report. They cannot know the issues within the golden time frame in which they could solve the problem; the DOT report lags by several months after the dates when the issues occurred. The social media platform could be an alternative way to reflect any issues in more real time, and the airline companies can be aware of the problem within a reasonable time frame. One of the specialties of Twitter is a feature called "retweet", which is not the same as "reply." Users on Twitter who follow the person who posted the messages on Twitter can forward those same messages and/or their own opinions and post to Twitter again. Sometimes, this activity requires less than a minute, considering Twitter's large number of users, and the fact that each user has lots of followers. The speed of spreading messages, then, is tremendous and quick. The majority of the U.S. airline carriers have official accounts on Twitter and Facebook. Their goal is to use social media as a marketing tool to improve their reputations or their impressions from consumers. Table 7 shows when each airline joined Twitter and its official Twitter account name. Twitter also offers a feature that enables a business to monitor the posts from other users (customers). If the business has used this feature, it will show "Responsive 24/7" on the left hand of the Business' Twitter home page.

Among the twelve airline carriers, United Airlines and Alaska Airlines are the only two to have utilized this feature to improve their communication with their customers.

TABLE 7 DECRIPTION OF TWITTER DATA

Airline Carrier	Twitter Account	Time to Join Twitter	Responsive 24/7 on Twitter
Hawaiian Airlines	@HawaiianAir	April 2009	No
United Airlines	@United	March 2011	Yes
Alaska Airlines	@AlaskaAir	February 2008	Yes
SkyWest Airlines	@SkyWestAirlines	January 2011	No
Frontier Airlines	@FlyFrontier	March 2008	No
Southwest Airlines	@SouthwestAir	July 2007	No
American Airlines	@AmericanAir	March 2009	No
Spirit Airlines	@SpiritAirlines	February 2009	No
Delta Airlines	@Delta	May 2007	No
ExpressJet Airlines	@expressjet	January 2012	No
Virgin America	@VirginAmerica	January 2008	No
JetBlue Airways	@JetBlue	May 2007	No

A Python program was written with the Twitter API to retrieve the tweets for the twelve airline companies from April 1st to June 30th, 2017. This program used the tweet.py to retrieve the old tweeter data. The data was saved in a JSON file with all useful information: username, posting date, geo-location, the number of retweets, and the number of favorites, hashtags, and permanent links.

```
To use this script, you can pass the following attributes:
```

```
     username: Username of a specific twitter account (without @)
        since: The lower bound date (yyyy-mm-aa)
        until: The upper bound date (yyyy-mm-aa)
  querysearch: A query text to be matched
         near: A reference location area from where tweets were generated
       within: A distance radius from "near" location (e.g. 10mi)
    maxtweets: The maximum number of tweets to retrieve
    toptweets: Only the tweets provided as top tweets by Twitter (no parameters
required)
       output: A filename to export the results (default is "output.csv")
```

To retrieve the old tweets about one particular screen name (such as United), the Python code was written like this:

```
# Get tweets by username and bound dates [united, '2015-09-10', '2015-09-12']
python Exporter.py --username "@united" --since 2015-09-10 --until 2015-09-12
```

Each airline company in this study has an official Twitter account (see Table 7). The following screenshots (Figure 6) were taken on Jan. 18, 2018. On the main page, each airline carrier has some Twitter information: description of the airline company, the number of followers, the number of followings, the number of tweets, the total number of likes, the geo-location of the business, the data to join Twitter.

#iflyalask
Tweets
199K
Following
5,583
Followers
282K
Likes
13.6K
Lists
2
Alaska Airlines
@AlaskaAir
Official account of Alaska Airlines. Our Social Care team is here for you 24/7. For a formal response: bit.ly/ContactAK. Tag photos with #iFlyAlaska
Seattle, WA
alaskaair.com
Joined February 2008
Responsive 24/7
2,764 Photos and videos
Tweets
Tweets & replies
Media
Alaska Airlines @AlaskaAir · 6h
Before we unveil our new uniforms tomorrow, we're flashing back to some old classics. #iFlyAlaska #iAmAlaska #FlyFashion bit.ly/2mHmN4Z

Frontier Airlines
Tweets
9,430
Following
1,866
Followers
95K
Likes
4,589
Frontier Airlines
@FlyFrontier
We fly to 80+ destinations in the US, Mexico, Puerto Rico and more! If you need further assistance please contact us here: bit.ly/contactFA
Denver, CO
flyfrontier.com
Joined March 2008
2,094 Photos and videos
Tweets
Tweets & replies
Media
Pinned Tweet
Frontier Airlines @FlyFrontier · Jan 16
We're looking forward to providing millions of people access to affordable fares between the U.S. and Mexico! Read more about our codeshare agreement with @flyvolaris: bit.ly/2ENl3if

Figure 6. Screenshots of Airlines' Twitter Account

4.3 Searching Tweets

Python code has been developed to retrieve old tweets from Twitter; the username and keywords can be specified, and it can set a maximum number of tweets, limit the

location, and set the time period. The program will save requested tweets in the CSV format after retrieval. For each airline company, I use "since" and "until" to limit the lower bound date and the upper bound date. The data is for three months and includes all of the tweets that mentioned each airline company, from Twitter users all over the world. There are many users who mentioned one of the airline companies. The tweets were collected from any users who mentioned one of the specific airline companies on Twitter during the specified periods. The original code examples for retrieving tweets are listed below (Figure 7):

```
python Exporter.py --querysearch "@united" --since 2017-04-01 --until 2017-06-30

python Exporter.py --querysearch "@AlaskaAir " --since 2017-04-01 --until 2017-06-30

python Exporter.py --querysearch "@ SouthwestAir" --since 2017-04-01 --until 2017-06-30

python Exporter.py --querysearch "@ FlyFrontier" --since 2017-04-01 --until 2017-06-30
```

Figure 7. Example Code to Retrieve Twitter Data

Table 8 shows the Twitter data for twelve airline carriers from April 1 until June 30, 2017. Delta has the most mentioned tweets (363,775) from the largest number of unique users (127,902) on Twitter, while ExpressJet has the lowest number of mentions on Twitter, from 50 unique users. ExpressJet carried twice the passengers that Virgin America carried in the three-month period. However, ExpressJet only received 77 mentions on Twitter in the three months. The reason might be that the customers who purchased the cheap airline tickets may not have a high expectation of service quality

from ExpressJet. American Airlines received the most mentioned tweets from individual users (3.01 times per user).

TABLE 8 THE DESCRIPTIVE DATA FOR THE FOUR AIRLINE CARRIERS (TWITTER DATA FROM APR.1 TO JUN. 30, 2017)

Airline Carriers	**Number of Tweets from Apr.1 to Jun. 30, 2017**	**Number of Unique Users**	**Average Tweets Per User**
United Airlines	283,045	**151,135**	1.87
Alaska Airlines	42,126	15,486	2.72
Southwest Airlines	111,584	46,341	2.41
Frontier Airlines	8,341	4,941	1.69
American Airlines	235,958	78,328	**3.01**
Delta Airlines	**363,755**	127,902	2.84
ExpressJet	77	50	1.54
Hawaiian Airlines	4,745	2,219	2.14
SkyWest Airlines	275	160	1.72
Spirit	28,664	14,285	2.01
JetBlue Airlines	74,490	26,833	2.78
Virgin America	19,669	7,961	2.47

CHAPTER

5 RESEARCH METHODOLOGY

5.1 Preprocessing the Retrieved Twitter Data

After three months of Twitter data for all twelve airline companies was retrieved, the data was saved in the CSV file. In order to process it easily, I converted it from a CSV format into an Excel format. Now, I had 12 Excel files for each airline company in this study. In each Excel file, there are ten fields: username, date, retweets, favorites, text, geo, mentions, hashtags, id, and permalink. The username, or “screen name” is unique on Twitter. By username, all of the tweets (Twitter posts) posted by this user can be selected in the dataset. The date contains the date and time when the tweets were posted. Retweets shows the number of retweets of each tweet. Retweets are similar to the forward function in the email; if retweeted, other users will get the same tweet that is shown on the user’s Twitter account. Favorites shows the number of favorites, just as “like” does on Facebook. Text shows the textual message only. Geo is an optional field. It shows the geographical location of the users when they posted the tweets. Most of tweets don’t include the geo information, and, because the geo information is shared by the users when they posted the tweets, it is not mandatorily reported. Mentions may contain zero or multiple objects. If the users use “@” symbol followed by other usernames, the field of mentions will show the mentioned users, for example “@VirginAmerica.”Hashtag is the topic used by users who post the tweets. A hashtag typically begins with “#” and is followed by a topic. Sometimes, users post more than one hashtag, like “#VirginiaAmerica #PPC #marketing #paidads #paidadvertising.” The field of id

corresponds to the username, but in numerical format. Each Twitter user has unique username and id. Permalink is the permanent link that directs users to each specific tweet. For instance,

https://twitter.com/helloreneerod/status/874494528396316672

This link will direct you to the see the original post showed in Figure 8.

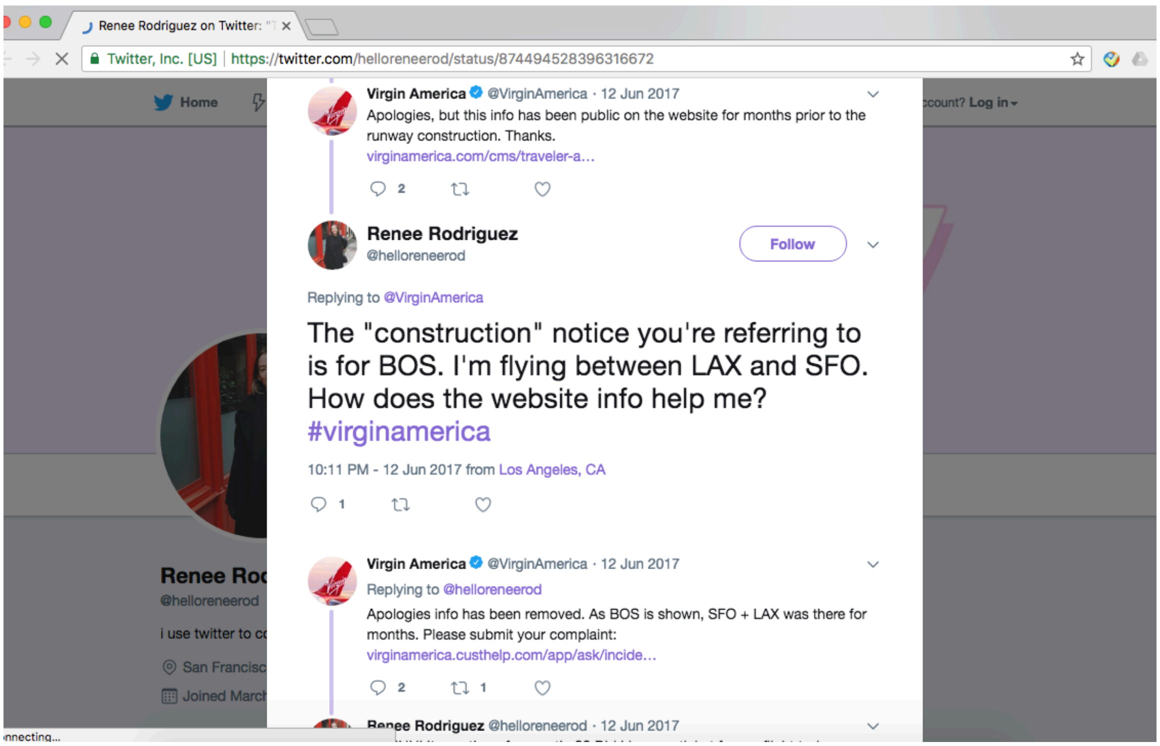


Figure 8. Example of Original Twitter Post

Secondly, in order to be able to use the Twitter data to conduct the statistical analysis and sentiment analysis at a later time, I preprocessed the Twitter data. Since this study focuses on the service quality of airline industry, I would like to focus on the tweets posted from real customers, not from the airline staffs. The tweets posted by airline's username (See Table 8) were removed. In addition, because some tweets do not reveal any useful information that can be used to measure the service quality of airlines, those

tweets were removed from the dataset. If a tweet didn't contain any words, or the tweet only contained one word, such as "Ok", "thanks", "hi", "awesome", or "yes", it was removed.

Thirdly, I removed the keywords that were used to retrieve the Twitter data. I used "@VirginAmerica" to search, then show, tweets containing this keyword through the Excel file. For example, the original tweet *"@VirginAmerica some of the best customer service I have every had. companies could learn a lot from their operations*", were processed as "*some of the best customer service I have every had. companies could learn a lot from their operations.*" There are some typos in the tweets, but they are normal and frequently occur in textual data on social media platforms.

Fourthly, all the tweets (in text field of the Excel file) were converted to lower case. The reason is simple. It was used to match the words listed in the Lexicons database (all lower case). This is an essential step before I do the sentiment analysis. In computer languages, lower case and upper case are different values. In order to compare words, they must in the same case, lower case or upper case. For example, if a tweet contains word "Great", the sentiment score was calculated by the matching word in the Lexicons database. However, the Lexicons database only contains the word "great." If this is the case, the sentiment score of this tweet will not be calculated correctly, and neither will the value of its valence or its arousal.

Finally, I needed to remove all of the meaningless information in the tweets. For example, "http" or "https" is always shown in tweets as the beginning of the URL. I would remove "https://" first, and then remove "http://". The order is important here; if I removed http first, https will remain "s" in the text field. "www" is another word that

needed to be removed. After that, I trimmed the white spaces in the tweets. The preprocessing workflow is illustrated in Figure 9.

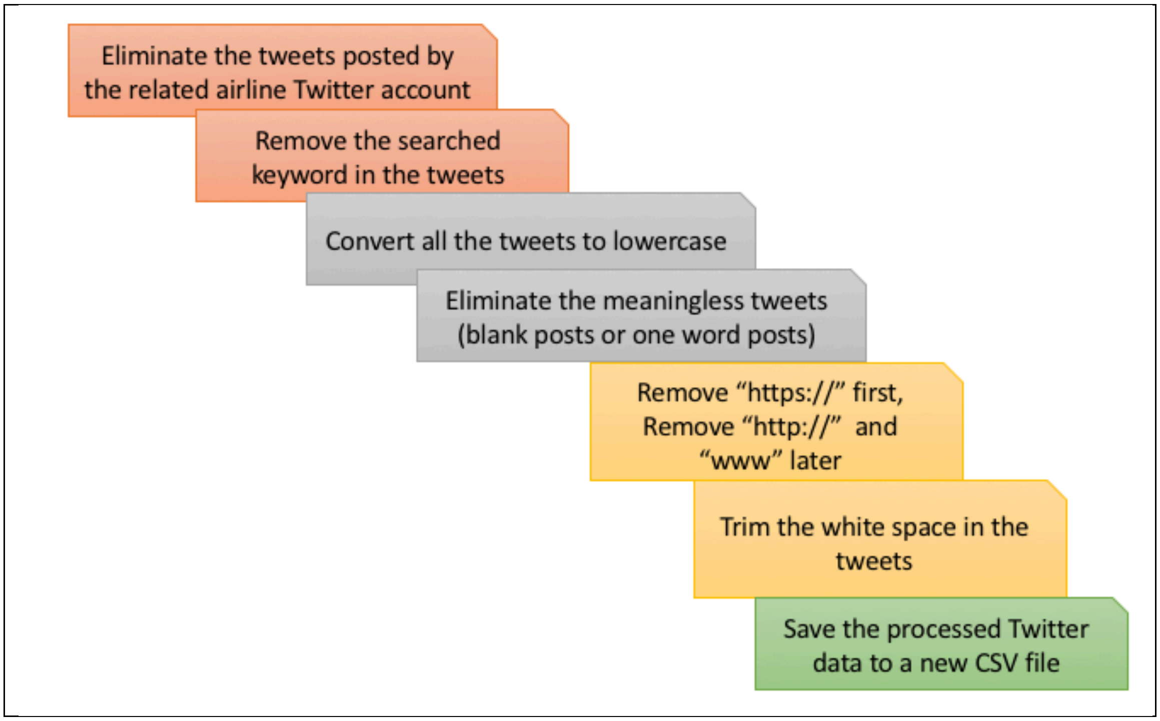


Figure 9. The Workflow of Processing the Twitter Data

After processing the Twitter data as shown in the steps in Figure 9, the final Twitter datasets could be described in Table 8. United Airlines was mentioned by the largest number of unique users (135,569 unique users). And American Airlines received the highest number of average tweets per user (2.09). This can be interpreted to show that the customers of American Airlines would like to communicate with the airline company on Twitter. The highest number of tweets was received by Delta Airlines; it was mentioned by Twitter users (or its customers) over 227,000 times during this period (April - June 2017).

TABLE 9 THE DESCRIPTIVE OF PROCESSED TWITTER DATA

Airline Carriers	**Number of Tweets from Apr.1 to Jun. 30, 2017**	**Number of Unique Users**	**Average Tweets Per User**
United Airlines	224,789	**135,569**	1.66
Alaska Airlines	26,616	13,969	1.91
Southwest Airlines	68,522	42,561	1.61
Frontier Airlines	7,301	4,574	1.60
American Airlines	150,454	72,071	**2.09**
Delta Airlines	**227,121**	116,165	1.96
ExpressJet	58	45	1.54
Hawaiian Airlines	3,401	2,025	1.68
SkyWest Airlines	231	141	1.64
Spirit	23,045	13,095	1.76
JetBlue Airlines	49,503	24,693	2.00
Virgin America	11,817	7,270	1.63

5.2 Text Mining and Supervised Classification by SERVQUAL Dimensions

Text mining has been used a lot in recent years by researchers and businesses. Text as input information can be retrieved from the Internet. For example, comments about products, reviews of a movie or a restaurant, complaints about customer services, discussions about a hot topic, can all be mined. Text mining was defined by Hearst (1999) as, "the most use of large online text collections to discover new facts and trends about the world itself." Text mining has been investigated by multiple disciplines: linguistics, computer science, computational statistics, and information technology. Standard techniques in text mining are text classification, text clustering, word frequency and co-occurrence words, document summarization, and latent corpus analysis (Meyer, Hornik & Feinerer, 2008). R is a powerful text mining; it is an open source tool with lots of text mining packages. In particular, the tm package developed by Meyer et al. (2008) is the core for text mining. In order to do text mining, the dataset should be processed again to meet the requirements of creating the word corpus. All of the punctuation and stop words in English, and then the white spaces in the tweets must be removed. The following word frequency cloud (Figure 10) shows the frequency of the words used in the tweets for each airline company. The larger the word, the higher the frequency of this word's mention in the Twitter data.

Alaska Airline

American Airline

Delta Airlines

ExpressJet

Frontier Airlines

Hawaiian Airlines

JetBlue

SkyWest Airlines

SouthWest Airlines

Spirit

United Airlines

Virgin America

Figure 10. Word Clouds of Tweets for Each Airline Carrier

As shown in Figure 10, the word clouds indicate the emphasis on specific words in the Twitter data for each airline companies. For example, Virgin America and Alaska Airlines received more positive words than others, like “thanks”, “like”, “love”, “great”, “best”, and so on. However, United Airlines received lots of negative words, like “never”, “dont”, and “cant.” If one looks at the tweets about Spirit Airlines, one can see that most of its tweeting customers complained about bad service, delayed flights, and cancelled flights. Text mining can initially check the most discussed topics on Twitter and can show the most frequent discussions through word clouds.

In Table 2, I have listed the six dimensions of Service Quality. In text mining, the technology named “supervised classification” can be used to classify the text into different categories. Supervised classification uses the Naïve Bayes theorem to classify the categories. I used this method to test the six dimensions of service quality. In this study, three graduate students reviewed most of the Twitter data and picked up some key words that can represent the dimensions of service quality. The three graduate students were trained on a knowledge about airline service quality before they searched for the keywords from Twitter dataset. When they had completed the keywords, they shared the keywords with each other. To get the common sense of the keywords for each dimension, they exchanged their understanding of the attributes of dimensions and then finalized the keywords list, as shown in Table 10.

TABLE 10 KEY WORDS OF EACH DIMENSION OF SERVICE QUALITY

Dimensions of Service Quality	**Key Words**
Responsiveness	Nice, kind, call, phone, cell, reschedule, schedule, information, info, notify, handle, rude, counter agent, minor, unaccompanied, help, refund, track, prefer, hotel, helpful, wonderful, agent, gate, callback, support, website, site, love, great, app, club, status, mile, 1k, frequent, lounge, mvp, best, attentive
Assurance	translate, English, service, rude, customer service, manager, favorite, win, treat, landing, awesome, perfect, issue, solve, elevate points, care, remove, poor, behavior, thrown, dumped, drag, cost, change, culture, skill, courtesy, upgraded, upgrade, professionalism, complaint, refund, hotel, reimbursed, reimbursement
Tangibility	Food, beverage, drink, wine, clean, dirty, meal, choice, music, movie, comfortable, uncomfortable, pillow, blanket, repair, game, sit, water, amenity, cabin, crew, bathroom, seat, air conditioner, wifi, breakfast, dinner, window, aisle, staff, neck, body, wi-fi
Reliability	check in, delay, on time, luggage, miss, baggage, handle, depart, arrival, lost, mechanical, failure, bags, bag, checked, late, wait, hour, board, cancel, inconvenience, cancelled, delayed, departed, departure, arrivals, arrive, carry on
Security and Safety	safe, safety, safely, animal, cat, dog, cage, security, secure

Communications	open, foreigners, communicate, communication, talk, announcement, announce, attendants, attendant, pilot, contact, answer, response

Table 11 shows some tweet examples that are related to each dimension of service quality. These examples may have some overlaps across the different dimensions. Since customers may complain multiple issues in one tweet, it can be difficult to split the corresponding service quality in one tweet. This is the one of the limitations of this research. However, this limitation does not change the consequence of the service quality or the sentiment. In the Twitter dataset, many tweets discuss one major dimension of service quality, and the mood of this customer can be tested by sentiment analysis. Due to the large amount of data, scientists want to find a way to figure out the sentiment of the textual social media data. In this case, I used the open source packages in the R program to conduct the sentiment analysis. Three major packages are used: "sentimentr", "ANEW" dictionary, and "NRC."

TABLE 11 SERVICE QUALITY DIMENSION RELATED TO TWEETS

Dimensions of Service Quality	**Tweets Examples**
Responsiveness	• I love alaskaair left my wallet on the plane at O'Hare last night and their staff went above and beyond to find it. you guys rock. • many thanks to your pdx crew coming from lax. they had to deal with some awful folks & they handled it like champs. #iflyalaska • had a great flight as always and just wanted to say thanks you providing a way to see family at an affordable price • gonna make me miss my flight home? you pay for hotel • you can move your united status to alaska - do it! • on flight 380 from nashville to las vegas. dave was the best flight attendant i've ever had! so attentive
Assurance	• trying to convert elevate points. get world's most generic error message with no actual contact info!! • looks like this flight to sd is paaacked! hopefully i don't get thrown off. • we just got married and taking to our honeymoon! we got the special treatment!! thanks you so much dawn and linda from alaska 31

	• I just experienced the perfect landing in by the captain of flight 2221. awesome job!!! • @delta $150 bucks each way for unaccompanied minor fee? whew! I guess we'll be flying alaska @$25 each way. thx #smarttravel
Tangibility	• cabin crew on flight 934 today was awesome. • in flight entertainment is awesome.
Reliability	• baggage handle destroyed • we're all ready to go now. it was more funny than an inconvenience and was handled well by the staff! • almost an hour and still no luggage. have a little pity on the weary travelers :-(• app says flight 975 arrived at 10:08. funny we're still taxiing for a gate. guess i'll get my baggage guarantee at 10:28
Security and Safety	• nice to visit family but even better to get home. safe travels • thank god great & safe flight to #seattle . special thanks to for making it a smooth one! • was there any explanation for why that happened? safe travels to you and the dog. • alaska airlines you just lost a customer. if you won't accommodate a service dog when the seat is paid for is ridiculous. cassius you rock.

Communications	• thanks for the update - patiently waiting • gate agent at koa boarded me and walked up to the plane & pilot said no boarding. communicate people! • please figure out how to communicate better with your customers. need better planes and some kind of hotel options if no flights. • flight attendant made an announcement that we were clear to fly and continued boarding at 930ish. but pilots couldn't be found.

5.3 Sentiment Analysis and Data Visualization

Opinion mining and sentiment analysis contribute to the development of an opinion-oriented information system for service or products providing companies (Pang & Lee, 2008). If service providing companies want to get feedback or opinions from customers, the traditional way is for them to conduct surveys and to distribute the surveys to the customers who have used the services in that company. However, a survey is not enough to get the customers' feedback in this digital world, because a survey needs time to be collected and analyzed. The issues of service cannot be addressed by the company quickly. Fortunately, customers like to post their service experiences on social media platforms rather than completing a survey or complaining to the company directly. Twitter and Facebook are the two of the most popular social media platforms.

Twitter Sentiment Analysis is the process of determining the emotional tone behind a series of words. It is used to gain an understanding of the attitudes, opinions, and emotions expressed within an online mention. Having a solid understanding of current public sentiment can be a great tool for any business. When deciding if a new marketing campaign is being received warmly, or if a news release about the CEO is causing customers to become angry, the people in charge of handling a company's public image need these answers quickly. And social media can deliver those answers quickly. One simple, yet effective, tool for testing the public waters is to run a sentiment analysis.

There are many ways to do sentiment analysis. Many approaches use the same general idea. Here are the three steps to do the sentiment analysis:

1. Create or find a list of words associated with strongly positive or negative sentiment.

2. Count the number of positive and negative words in the text.

3. Analyze the mix of positive to negative words. The use of many positive words and few negative words indicates a positive sentiment, while the use of many negative words and few positive words indicates a negative sentiment.

To perform the sentiment analysis, download the positive and negative words and evaluate the tweets with those positive and negative lexicons. The list of positive words contains 2,003 words and the list of negative words contains 4,782 words (Hu & Liu, 2004; Liu, Hu & Cheng, 2005). These word lists include some misspelled words that are possible appear frequently in social media content.

After having been run through the sentiment words lists, the tweets should be preprocessed before the sentiment scores can be calculated. First, the data must be

cleaned. This step involves the removal of stop words, all of the numbers, and all of the white spaces, and the conversion of all of the words or letters to the lower case.

The first step, creating or finding a word list (also called a Lexicon), is generally the most time-consuming. In this study, I use the existing Lexicons and made some modifications to those Lexicons. For example, social media popular acronyms or slang are not included in the existing Lexicons. “omg”, “lol”, “thx”, and “wtf” are very common acronyms on social media platforms. I included these types of words in the Lexicon for in this study. Researchers can edit the Lexicons when they study specific topics. However, some words have double emotional meaning. For instance, “sick" is an example of a word that can have positive or negative sentiment depending on what it's used to refer to. If you're discussing a pet store that sells a lot of sick animals, the sentiment is probably negative. On the other hand, if you're talking about a skateboarding instructor who taught you how to do a lot of sick flips, the sentiment is probably very positive. Sentiment analysis uses machine learning algorithms. In this research, R Studio was used to load the sentiment analysis packages and to analyze the processed tweets for all twelve of the airline companies.

A sentiment analysis works like this: I first take a bunch of tweets about whatever I am looking for. I then parse those tweets out into their individual words, and then count the number of positive words and compare it to the number of negative words. I use the open source R program to calculate the sentiment score of the sentence from https://github.com/exploratory-io/exploratory_func. This function first maps the predefined sentiment type (positive or negative) or the value (how positive or how negative). And then it considers the intensity of the sentiment. If I am using the positive

or negative to determine the sentiment, it is not enough to learn the mood or emotion of the customers. For example, here are two tweets in the dataset: “I’m feeling so good!” and “I’m feeling much better!” They are both positive. However, the two sentences express the intensity that influences the different results in emotion. “so” and “much” are the intensifiers in these sentences and score differently. In order to make sure of the accuracy of sentiment analysis, another popular package is used. Affective Norms of English Words (ANEW) dictionary has classified the words and has assigned them two elements to measure the emotions: **Valence and Arousal (**Russell, 1980**)**. Emotions are a subjective thing, and while we can measure their magnitude to a certain degree by monitoring people’s physical response, it can be difficult to tell the difference between a “good” emotion and a “bad” one. For example, when people face the same problem, they can have very different emotions. Even if the directions of emotion are same, their level of physical response will be different. Valence is the direction of the emotion, and arousal is the level/amount of the physical response. Figure 12 illustrates the relationship between valence and arousal. For example, the positive words “happy” and “relax” have different arousal levels. “happy” has a higher arousal level than “relax.” We can tell the intensity of the sentiment using those words.

Sentiment scores are calculated by the sum of positive minus negative, then they are divided by the number of words of the tweet. So, the sentiment score ranges from -1 to +1. Minus 1 means that the tweet is completely negative in sentiment, while positive 1 means that this tweet is completely positive in sentiment. In this study, I used dictionary-based methods, also called Lexicon, which all center around the determination of text *T*’s

average sentiment (referred as valence) with sentiment dictionary *D* through the equation (3):

$$h_D^T = \frac{\sum_{w \epsilon D} h_D(w) \cdot f^T(w)}{\sum_{w \epsilon D} f^T(w)} = \sum_{w \in D} h_D(w) \cdot p^T(w) \quad (3)$$

Where I denote each of the words in a given sentiment dictionary *D* as *w*, word sentiment scores as $h_D(w)$, word frequency as $f^T(w)$, and normalized frequency of *w* in *T* as $P^T(w) = f^T(w) / \sum_{w \in D} f^T(w)$. In this way, I can measure the sentiment of a text in a manner analogous to taking the temperature of a room. Analyzing individual word contribution is important to the tweets and this equation allows for a meaningful interpretation. For example,

"funny that you tweeted this after a cancellation of a flight from dca > sfo because of lack of crew"

Sentiment score: 0.01147079, Valence: 6.25, Arousal:7.92

In this case, the tweet expressed a strong emotion. However, the valence score is a little bit higher (6.25 out of 9); it is not that accurate in this tweet. The sentiment score shows the neutral to positive sign. If we read this tweet, it should be a complaint about the flight cancellation. So, typically, we can think about it as a negative. Arousal can reveal strong emotional intensity. In addition to the sentiment score, marketers or business practitioners can also use arousal as a sign to determine whether the textual data should be flagged.

To calculate the valence and arousal score for the sentences or paragraphs, I use the ANEW dictionary and apply that into Python code. First, the ANEW terms should be defined in a Python file with words, stem-words, average arousal score, and average valence score (Figure 11).

```
"abortion": {
  "dict": "anew", "word": "abortion", "stem": "abort",
  "avg": [ 3.50, 5.39 ], "std": [ 2.30, 2.80 ], "fq": 6
},
"absurd": {
  "dict": "anew", "word": "absurd", "stem": "absurd",
  "avg": [ 4.26, 4.36 ], "std": [ 1.82, 2.20 ], "fq": 17
},
"abundance": {
  "dict": "anew", "word": "abundance", "stem": "abund",
  "avg": [ 6.59, 5.51 ], "std": [ 2.01, 2.63 ], "fq": 13
},
"abuse": {
  "dict": "anew", "word": "abuse", "stem": "abus",
  "avg": [ 1.80, 6.83 ], "std": [ 1.23, 2.70 ], "fq": 18
},
"acceptance": {
  "dict": "anew", "word": "acceptance", "stem": "accept",
  "avg": [ 7.98, 5.40 ], "std": [ 1.42, 2.70 ], "fq": 49
},
"accident": {
  "dict": "anew", "word": "accident", "stem": "accid",
  "avg": [ 2.05, 6.26 ], "std": [ 1.19, 2.87 ], "fq": 33
},
"ace": {
  "dict": "anew", "word": "ace", "stem": "ace",
  "avg": [ 6.88, 5.50 ], "std": [ 1.93, 2.66 ], "fq": 15
},
"ache": {
  "dict": "anew", "word": "ache", "stem": "ach",
  "avg": [ 2.46, 5.00 ], "std": [ 1.52, 2.45 ], "fq": 4
},
"achievement": {
  "dict": "anew", "word": "achievement", "stem": "achiev",
  "avg": [ 7.89, 5.53 ], "std": [ 1.38, 2.81 ], "fq": 65
},
```

Figure 11. ANEW Terms Definition in Python Code

Second, I divide the tweet into individual terms and then I try to match the ANEW dictionary. If find that any words have matched, then I use the following procedure to calculate the arousal and valence score. Partial code written in Python is shown in Table 12.

TABLE 12 PARTIAL PYTHON CODE TO CALCULATE VALENCE AND AROUSAL

```python
def valence( term ):

#  Return the average valence for a term
#
#  term:  Term to check (can be string or list of strings)

    if isinstance( term, str ):
        return valence_raw( term )[ 0 ]

    elif not isinstance( term, list ):
        return 0.0

    #  At this point we know we're working with a list of terms

    c = 2.0 * math.pi
    prob = [ ]
    prob_sum = 0.0
    v_mu = [ ]

    for t in term:
        if exist( t ):
            v = valence_raw( t )

            p = 1.0 / math.sqrt( c * math.pow( v[ 1 ], 2.0 ) )
            prob.append( p )
            prob_sum = prob_sum + p

            v_mu.append( v[ 0 ] )

    val = 0.0
    for i in range( 0, len( v_mu ) ):
        val = val + ( prob[ i ] / prob_sum * v_mu[ i ] )

    return val

# End function valence
```

```python
def arousal( term ):

#  Return the average arousal for a term
#
#  term:  Term to check (can be string or list of strings)

    if isinstance( term, str ):
        return arousal_raw( term )[ 0 ]

    elif not isinstance( term, list ):
        return 0.0

    #  At this point we know we're working with a list of terms

    c = 2.0 * math.pi
    prob = [ ]
    prob_sum = 0.0
    a_mu = [ ]

    for t in term:
        if exist( t ):
            a = arousal_raw( t )

            p = 1.0 / math.sqrt( c * math.pow( a[ 1 ], 2.0 ) )
            prob.append( p )
            prob_sum = prob_sum + p

            a_mu.append( a[ 0 ] )

    arousal = 0.0
    for i in range( 0, len( a_mu ) ):
        arousal = arousal + ( prob[ i ] / prob_sum * a_mu[ i ] )

    return arousal
```

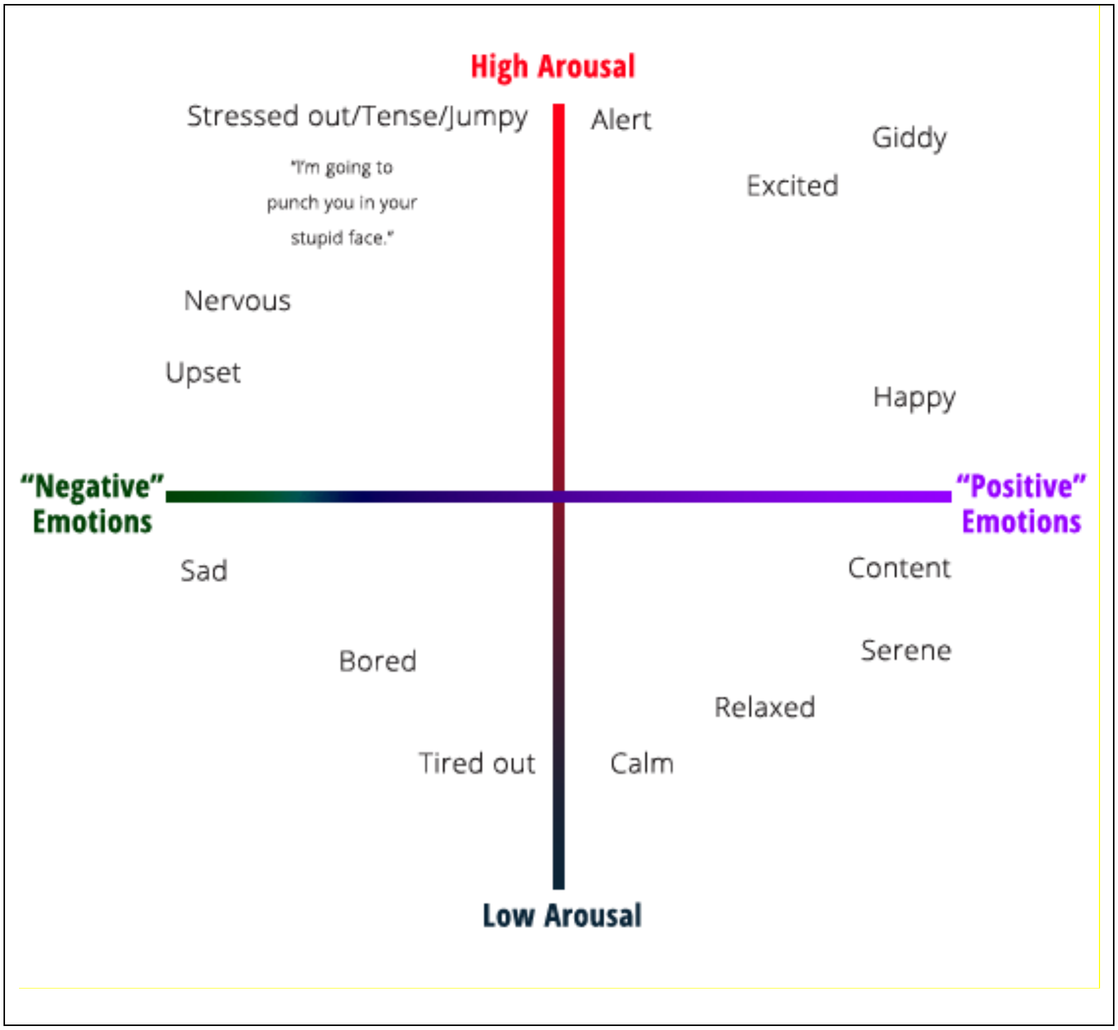


Figure 12. Valence and Arousal Illustration

In this study, I used three major Lexicons to calculate the sentiment for each tweet. The ANEW dictionary measured the valence and arousal for the tweets. To apply the ANEW dictionary, each tweet was given a score for valence and for arousal. Both valence and arousal are positive scores. The higher the valence score, the more the positive direction of this tweet. The higher the arousal score, the more intensive the emotion expressed by the tweet.

The second Lexicon is Sent140Lex that was developed by National Research Council from Canada (NRC) and was created from the “sentiment140” corpus of tweets. It uses Pairwise Mutual Information with emoticons as positive and negative labels. The third Lexicon, EmoLex, was used to find emotion words like happy, sad, surprised, angry, and so on. EmoLex was developed by NRC and can calculate the sentiment of common words and phrases using Mechanical Turk.

The third Lexicon, EmoLex, was created by experts from NRC of Canada, It uses the eight basic emotions (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust) in psychology (Plutchik, 1962). Figure 13 illustrates the wheel of emotion in four opposing pairs: joy-sadness, anger-fear, trust-disgust, and anticipation-surprise. EmoLex is a list of English words and also provides translations to other languages words including Spanish, Traditional Chinese, Simplified Chinese, Japanese, French, German, and Arabic (Mohammad & Turney, 2010; Mohammad & Turney, 2013). EmoLex can be used to identify emotions and sentiment, as well as to analyze hashtags, emoticons, and word-color associations. Matching the words in the lexicons can be used to analyze English texts, and then scientists can gather the emotional sentiment about textual data.

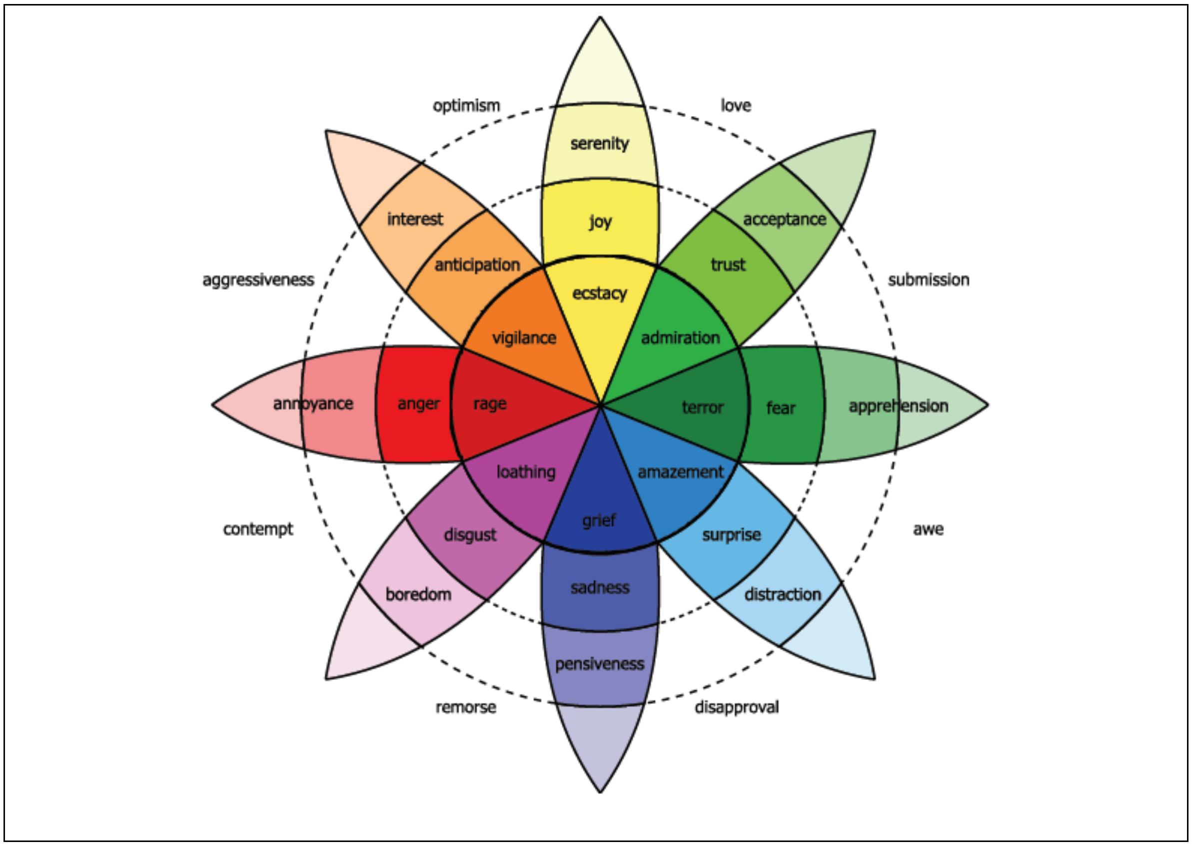


Figure 13. Plutchik Wheel of Emotion (Plutchik, 1980)

The way people use words conveys information about themselves and the situation or status they are in (Pennebaker et al, 2003). The words that people use are diagnostic of their mental, social, and physical appearance. Individuals' choice of words can reveal their social status, age, sex, and motives (Pennebaker et al, 2003). In the digital era, people share their life experiences on social media platforms so frequently and this information is valuable to analyze the textual social media data. For example,

"*I'm so excited. I just booked my tickets! #flyfrontier #frontierairlines #cheapairfare*"

anger	anticipation	disgust	fear	joy	sadness	surprise	trust
0	1	0	0	1	0	1	1

The above example shows the customer was excited to book an inexpensive fare from Frontier. Through emotional sentiment analysis, having matched the word list in the lexicons, the emotions “anticipation”, “joy”, “surprise” and “trust” were identified from this tweet. So, in most cases, emotional sentiment can be mixed in with a variety of basic emotions. Mixed emotions have been discussed intensively by psychologists (Aman & Szpakowicz, 2007). Sometimes, one positive word in joy and one negative word in fear can be uttered at the same time. For instance, if this is the the customer’s first time to take the plane, he or she will have mixed emotions. The customer may anticipate something in the brand-new experience, and may also fear the safety of air transportation. Such situations happens frequently in our lives, like the first time we go to school, the first time we shop online, the first time we get a job offer, and so on. People always express mixed emotions in those moments, and a sentiment score cannot detect the mixed emotion from textual data. So, I used the EmoLex to identify the emotional sentiment.

Sentiment analysis is in demand because of its efficiency. Thousands of text documents or tweets can be processed for sentiment in seconds, compared to the hours it would take a team of people to manually complete this task. Because it is so efficient (with an 80% accuracy for English content) many businesses are adopting text and sentiment analysis and incorporating it into their processes.

5.4 Pearson Correlation and Multivariable Regression Model

In order to test the relationship between social media Twitter data and DOT-reported data, I used Pearson correlation to find the relationship between each service quality metric and AQR score. I tested the correlation between the AQR score and each dimension of service quality discussed in Chapter 3. The three dimensions (responsiveness, assurance, and reliability) of service quality were expected positively associated with AQR score.

The multivariable regression model was used to create a prediction model and to find which variables would play important role to the AQR score. To answer my research questions in Chapter 1, I tested the variables as discussed in Chapter 5.3 (sentiment analysis). Since AQR scores are calculated by the DOT monthly report, and it has a two month lag for customers and airline companies for decision making, I tested to see if social media data could be used to monitor airline service quality instead of having the airlines wait for the DOT report released two months later.

The AQR score is the dependent variable that can be calculated by the data from DOT monthly report. The independent variables come from social media data: the volume of tweets, the sentiment score, the 8-emotion sentiment score (joy, sadness, anger, fear, trust, disgust, anticipation, and surprise), the valence score, and the arousal score (using the ANEW dictionary, the scores range from 1 to 10). A higher valence score means a more positive response, and a higher arousal score means a more excited and more intense emotion.

In the regression mode, the control variables were introduced to control the model fit. Since different airline companies use their social media marketing differently and since customer support varies, this may affect the number of followers on their social media platforms. The fewer followers means the fewer mentions on Twitter. The control variables are: the time that each airline company began to use Twitter, whether they have someone to respond 24/7 on Twitter (dummy variable), the number of followers of each airline for each month. Equation (3) shows the initial regression model:

$$\text{AQR} = \alpha + X_1\beta_1 + X_2\beta_2 + X_3\beta_3 + \cdots + X_k\beta_k + \varepsilon \quad (3)$$

In this study, I tested the effectiveness of multiple variables on the AQR score. I wanted to find out about the relationship between the sentiment indicators and the AQR score. The control variables were investigated as well, in order to determine the regression model fit. The dataset contained 12 U.S.-based airlines companies' data for three consecutive months in 2017.

To examine the effectiveness of social media data analytics on service quality of airline industry, I specified the following empirical models:

The first model (equation 4) was used to test the sentiment score's effect on the AQR score. The sentiment score, the valence, and the arousal were used as independent variables to find the relationship between AQR score. I expected that, the lower the sentiment scores of tweets, the more complaints would be listed in the DOT report; this would result in a lower AQR score.

In the first model, AQR was monthly data and then the sentiment score could use the value of mean to represent the entire month. In terms of valence and arousal, these scores were highly dispersed in the dataset. Hence, I used the sum of the valence and the sum of

the arousal for each month to represent the valence and arousal for this month. Equation (4) presents the first predicted model.

AQR = f (constant, the volume of tweets, the average sentiment score, sum of valence, sum of arousal) (4)

The explicit form of equation (4) is represented as follows:

$$AQR = \alpha_0 + \alpha_1(volume\ of\ tweets) + \alpha_2(average\ sentiment\ score) + \alpha_3(\ valence) + \alpha_4(arousal) + \varepsilon \quad (5)$$

In the next model, I would test the interaction of valence and arousal in the regression model. Since valence and arousal are the most critical dimensions of emotion-related behavior (Lang, 1995), Robinson, Storbeck, and Meier (2004) found that valence-arousal interactions had a significant effect on evaluation processing. The new model is shown in equation (6).

AQR = f (constant, the volume of tweets, the average sentiment score, sum of valence, sum of arousal, valence * arousal) (6)

The explicit form of equation (6) shows:

$$AQR = \alpha_0 + \alpha_1(volume\ of\ tweets) + \alpha_2(average\ sentiment\ score) + \alpha_3(\ valence) + \alpha_4(\ arousal) + \alpha_5(valence * arousal) + \varepsilon \quad (7)$$

If I considered all of datasets, I put all of the data into one dataset and annotated the airline carrier's name on a new column. I was able to test the effectiveness of sentiment analysis on measuring service quality in the airline industry. I also needed take control variables into account. I expected that I would find that the larger the volume of negative tweets, the lower the AQR score would be. In terms of valence and arousal, they were expected positively related to the AQR score. Equation (8) represents the model.

AQR = f (constant, the volume of tweets, sentiment score) (8)

The explicit form of equation (8) shows:

$$AQR = \alpha_0 + \alpha_1(volume\ of\ tweets) + \alpha_2(average\ sentiment\ score) + \varepsilon \qquad (9)$$

Next, having included all of the control variables in the model, the model is shown in equation (10). Control variables are the number of followers, the number of passengers, the number of schedule flights,

AQR = f (constant, the volume of tweets, sentiment score, control variables) (10)

The explicit form of equation (10) shows:

$$AQR = \alpha_0 + \alpha_1(volume\ of\ tweets) + \alpha_2(average\ sentiment\ score) +$$

$$\alpha_3(time\ to\ join\ Twitter) + \alpha_4(Twitter\ response\ feature) +$$

$$\alpha_5(the\ number\ of\ followers) + \alpha_6(the\ number\ of\ passengers) +$$
$$\alpha_7(the\ number\ of\ scheduled\ flights)\ +\ \varepsilon \tag{11}$$

To make the model more fit and more efficient, I introduced the control variables into the regression mode (shown in equation (12)).

AQR = joy, sadness, anger, fear, trust, disgust, anticipation, and surprise

f (constant, the volume of tweets, emotional sentiment score, control variables) (12)

$$AQR = \alpha_0 + \alpha_1(volume\ of\ tweets) + \alpha_2(joy) + \alpha_3(sadness) + \alpha_4(anger) +$$
$$\alpha_5(fear) + \alpha_6(trust) + \alpha_7(disgust) + \alpha_8(anticipation) + \alpha_9(surprise) +$$
$$\alpha_{10}(time\ to\ join\ Twitter) + \alpha_{11}(Twitter\ response\ feature) +$$
$$\alpha_{12}(the\ number\ of\ followers) + \varepsilon \tag{13}$$

Due to the big incident that occurred with United Airlines on April 2017, I was interested in looking at time as the variable in the regression model. I expected to find that big incidents or accidents cause a lot of negative discussions on social media platforms and may affect other competitors as well, for a short time after the incidents. The model is shown in equation (14).

AQR = f (constant, the volume of tweets, emotional sentiment score, month, control variables) (14)

The explicit form of equation (14) above is represented as follows:

$$AQR = \alpha_0 + \alpha_1(volume\ of\ tweets) + \alpha_2(joy) + \alpha_3(sadness) + \alpha_4(anger) + \alpha_5(fear) + \alpha_6(trust) + \alpha_7(disgust) + \alpha_8(anticipation) + \alpha_9(surprise) + \alpha_{10}(time\ to\ join\ Twitter) + \alpha_{11}(Twitter\ response\ feature) + \alpha_{12}(the\ number\ of\ followers) + \alpha_{13}(month) + \varepsilon \quad (15)$$

I expected that the month would have influence on the AQR score, as well. Since the Twitter data was collected for a three month period, and the United Airlines incident happened in early April of 2017, the volume and sentiment score should be expected to vary due to the different months. The big incident not only affected the perception of service quality for United Airlines, but it also affected the entire airline industry. The Twitter data was expected to have a similar pattern for the majority of the airline carriers over the three months.

CHAPTER

6 RESULTS AND DISCUSSION

In this chapter, I discuss the results for sentiment analysis, Pearson correlation analysis, and regression analysis. Data visualization is provided in Section 6.1. Pearson correlation is done in Section 6.2, followed by descriptive data from regression analysis in Section 6.3 and testing results of different regression models in Section 6.4.

6.1 Data Visualization

In order to find the answers to my research questions, and to test the hypothesis, I used social media analysis to find the patterns or the trends of social media data during the research period. And then I had several numeric variables or columns in the final dataset for all U.S.-based airline companies: 8 columns of emotional signs (anger, anticipation, disgust, fear, joy, sadness, surprise, and trust), the volume of Tweets, the average sentiment score, the number of passengers, the number of scheduled flights, the number of months on Twitter, the number of followers on Twitter, whether they airline had an online response on Twitter, the sum of the valence score, and the sum of the arousal score. All the variables are numeric, and I used R to perform a quick analysis about the relationships and distances among them. This method helped me to discover an interesting relationship among some of the variables and may help me to exploit and build a better statistical model, later.

Other than looking for the correlations among the variables, I wanted to understand the distances among these numeric variables. I calculated the distances using

"Euclidean" distance that is a straight-line distance between two points presented in (q_1, q_2), and then the distance is calculated by the formula (16):

$$d(p, q) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2} \quad (16)$$

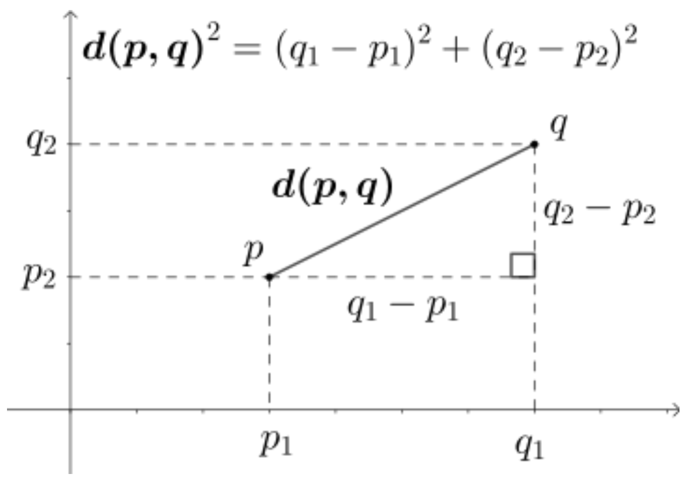


Figure 14. Euclidean distance in R2

Figure 15 illustrates that there are three clusters for all of the numeric variables. The number of passengers has the largest distances among other variables. Other two major clusters are in green dots and orange dots. On one hand, the volume of tweets, arousal, and valence are in one cluster. They are close to each other. On the other hand, all of the variables related to emotional sentiment are close, with less distance.

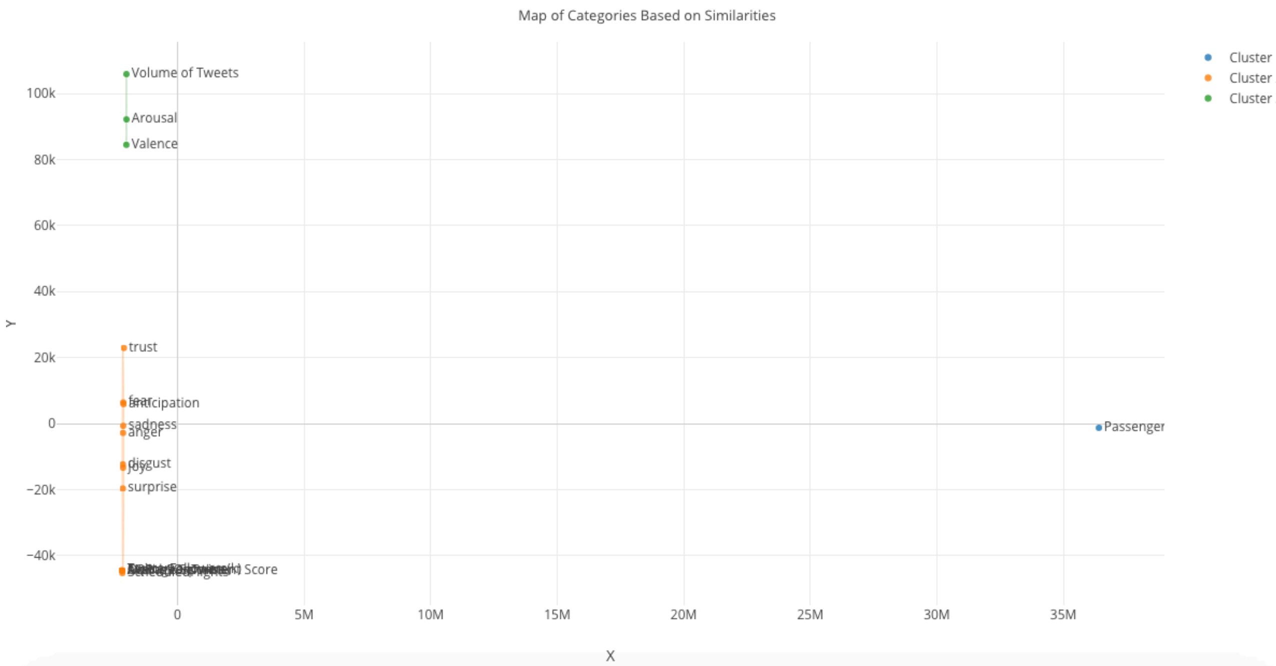


Figure 15. Map of Categories Based on Similarities

Table 13 shows the top three shortest distance pairs of similarity of variables, with the value of distances between each pair. A smaller value of distance means that the pair is close to each other, and a larger value means that the pair is far away from each other. As Table 13 shows, the average sentiment score is very close to the online response (Distance = 2.365).

Then, we can assume that the sentiment score would be correlated to the online response feature on Twitter. But we are not sure of the positive or negative relationship between the two variables. Another two pairs of variables are AQR with average sentiment score, and AQR with online response. So, if the airline companies used an online response, then AQR should be infected by this. The sentiment score was related to AQR, as I had predicted. And it looks as if the other emotional sentiment indicators are far away from the AQR score in the distance analysis.

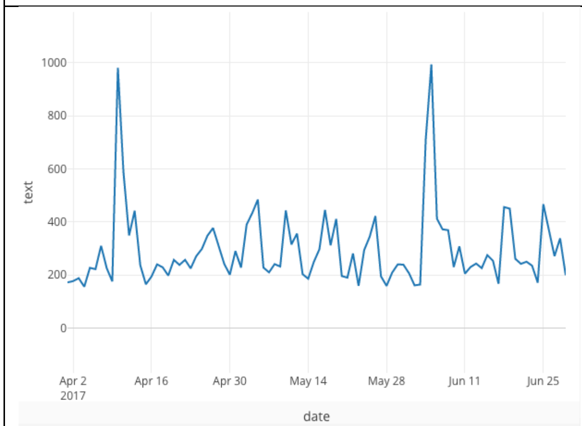
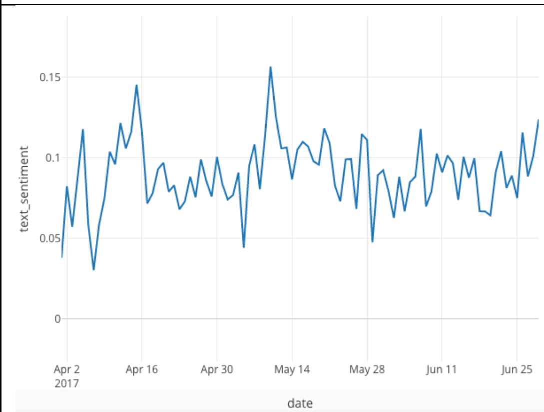
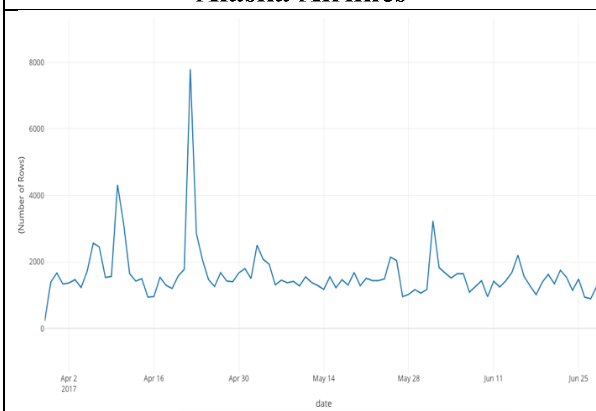
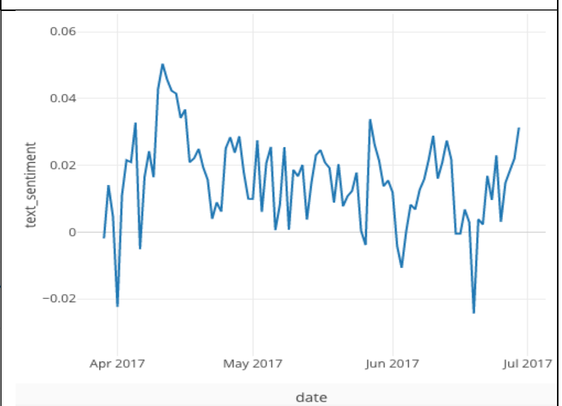
TABLE 13 SIMILAR PAIRS WITH DISTANCES

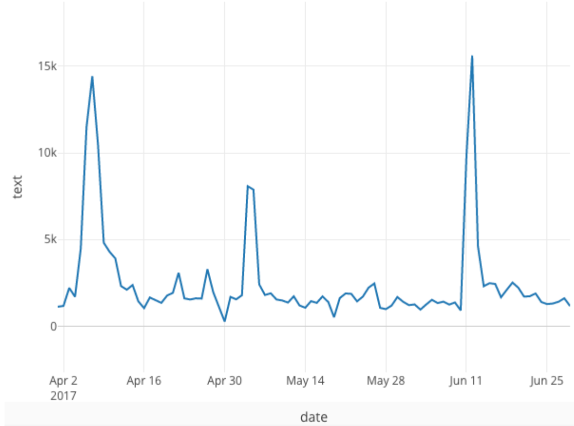
Column 1	Column 2	Distance ▲
Average Sentiment Score	OnlineResponse	2.3651321904
AQR	Average Sentiment Score	6.9409906109
AQR	OnlineResponse	7.7506451344

After I analyzed the distance among all the numeric variables, I ran the overall sentiment analysis for all airline companies during the three month period. In the following figures, all of the airline companies' Twitter data are shown in volume and in sentiment scores. I then put them into three categories: large airline carriers, cheap airline carriers, and other airline carriers. "Large airline carriers" comprises those that carry the largest number of passengers, including four airline companies: United Airlines, Southwest Airlines, American Airlines, and Delta Airlines. "Cheap airline carriers" means that the air fares on those are cheaper than on other major airlines. This category includes JetBlue, Spirit and Frontier. Those airline carriers always have their air fare deals online. For example, Frontier Airlines offers one-way air ticket for only $20 from ATL (Atlanta airport) to Orlando, FL. "Other airline carriers" include Hawaiian Airlines, Alaska Airlines, SkyWest Airlines and Virgin America. These airline carriers mostly have limited routes compared with the large airline carriers and the cheap airline carriers. Hawaiian Airlines mainly offers roundtrip fares from the U.S. mainland to Hawaii, and Alaska Airlines offers roundtrip fares from the U.S. Mainland to Alaska, as well as roundtrip fares among major cities in west coast in the U.S. Table 14 shows the trend of

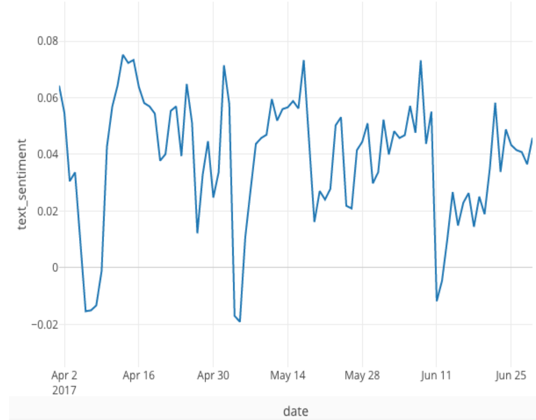
sentiment score and the trend of volume of tweets for each airline carrier in the three consecutive months.

TABLE 14 OVERALL VOLUMES AND MEAN OF SENTIMENT SCORE FOR EACH AIRLINE CARRIERS IN THREE MONTHS

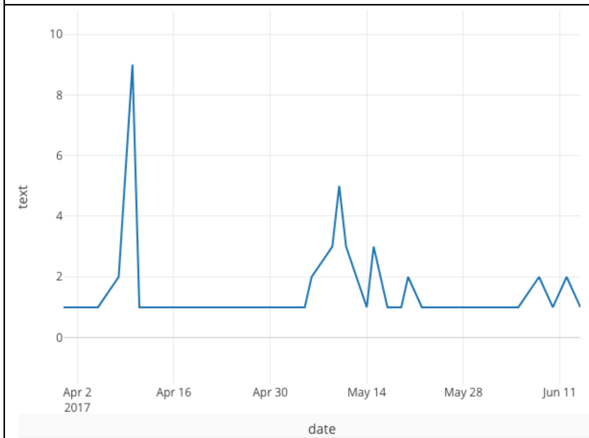
Overall Volume of Tweets	**Overall Mean of Sentiment Score**
 **Alaska Airlines**	 **Alaska Airlines**
 **American Airlines**	 **American Airlines**



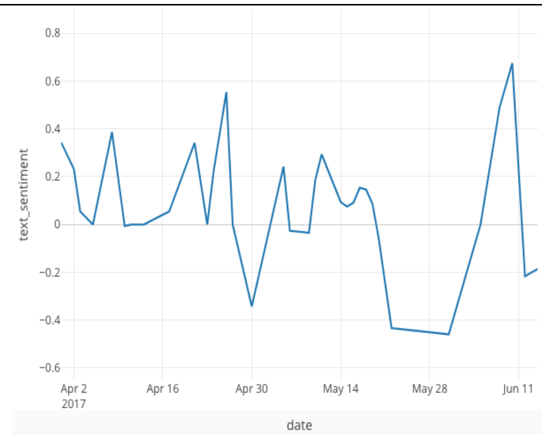
Delta Airlines



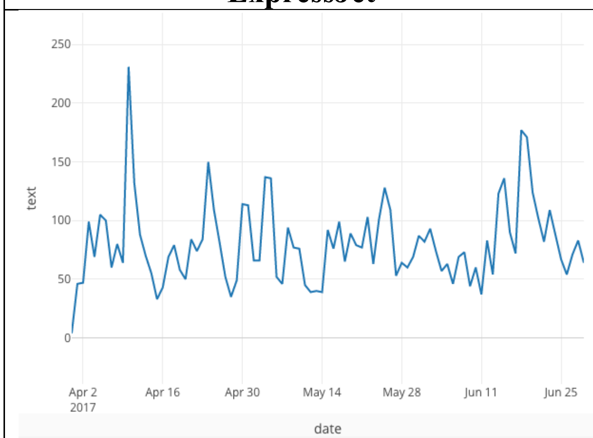
Delta Airlines



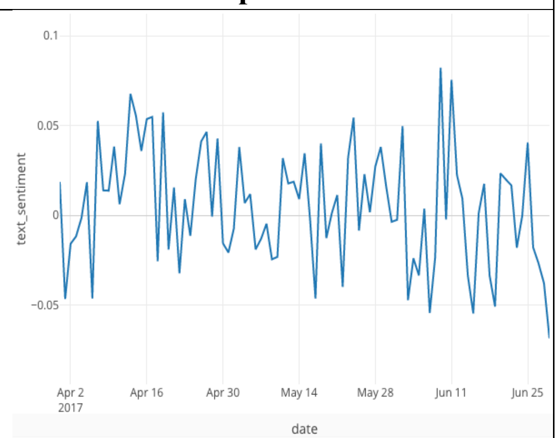
ExpressJet



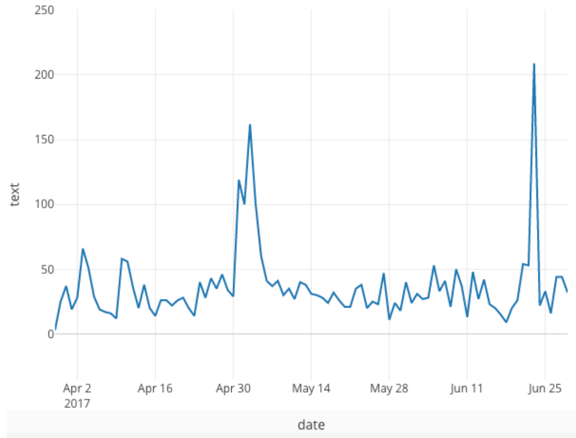
ExpressJet



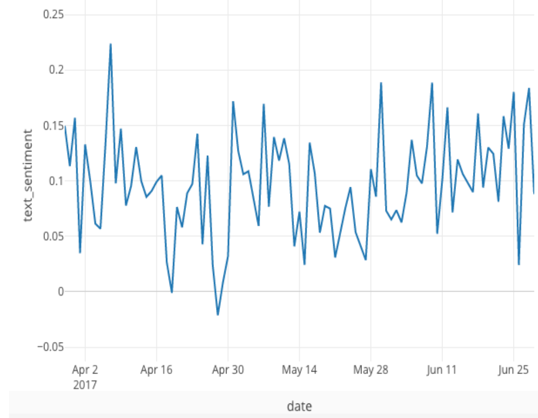
Frontier Airlines



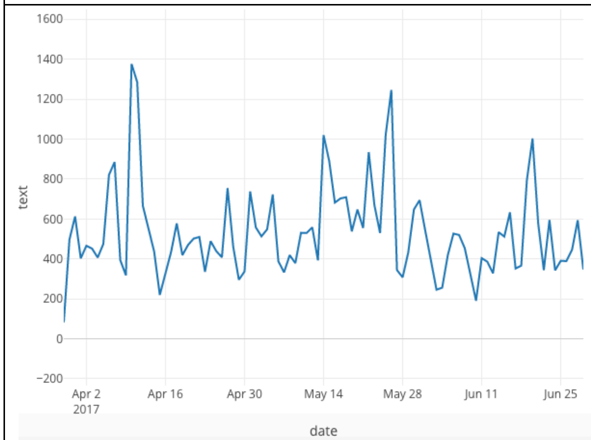
Frontier Airlines



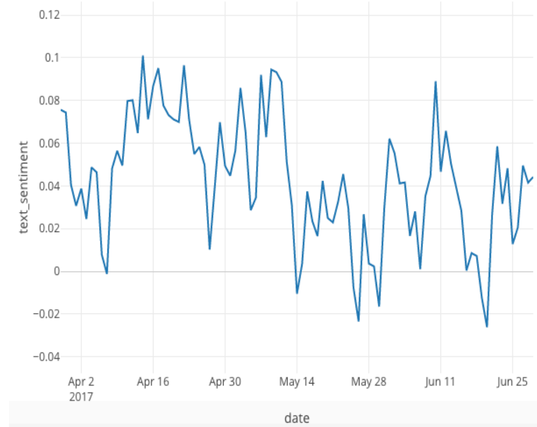
Hawaiian Airlines



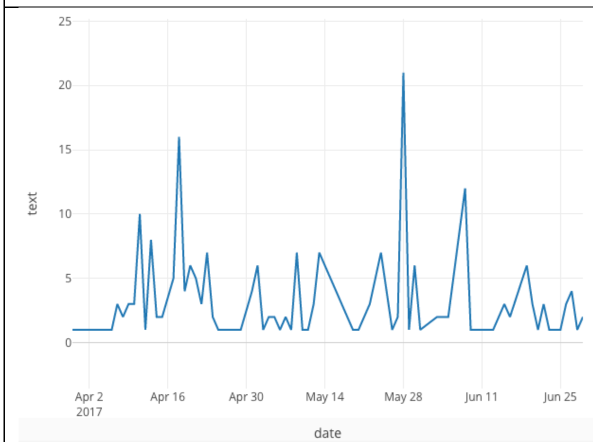
Hawaiian Airlines



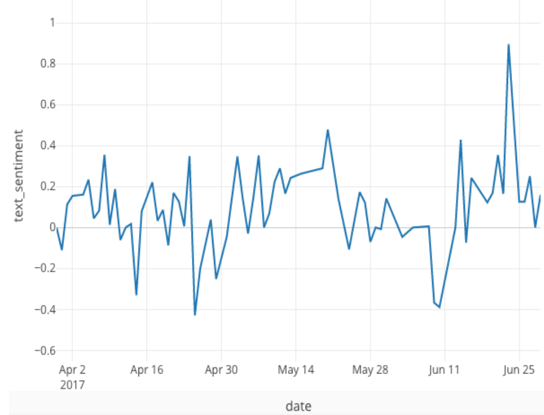
JetBlue



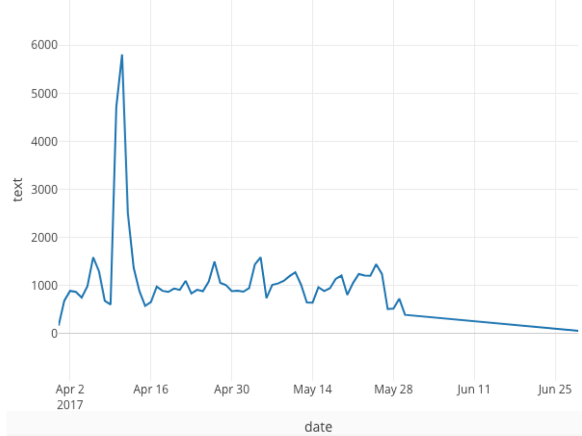
JetBlue



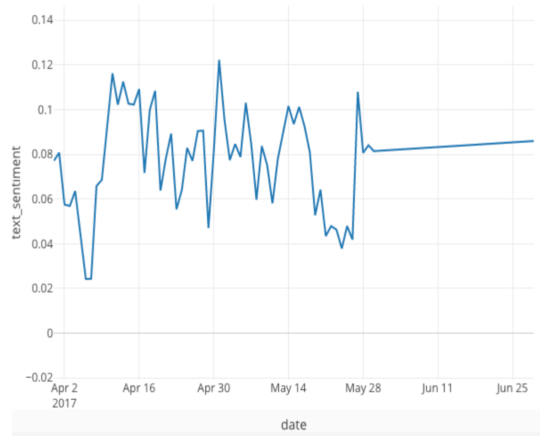
Skywest Airlines



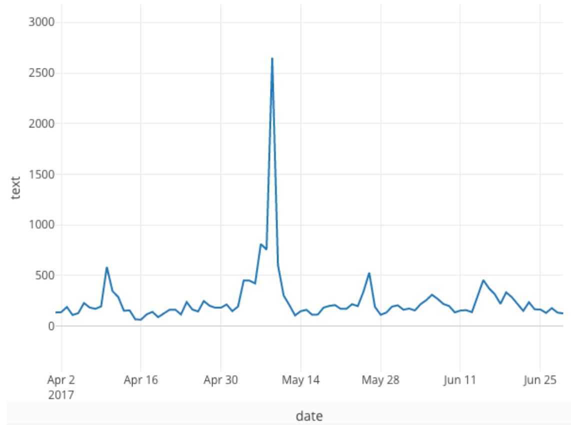
Skywest Airlines



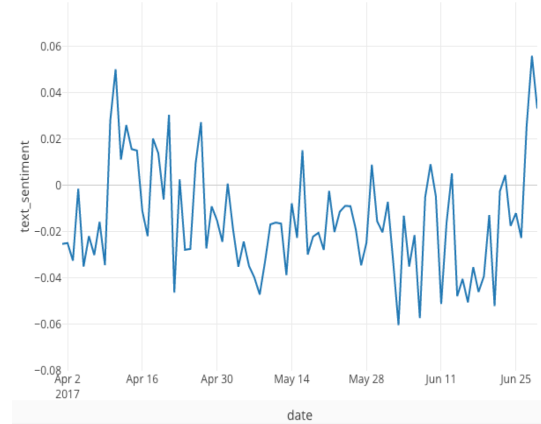
Southwest Airlines



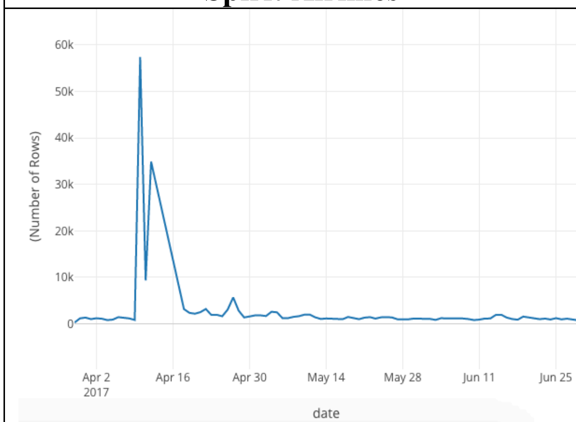
Southwest Airlines



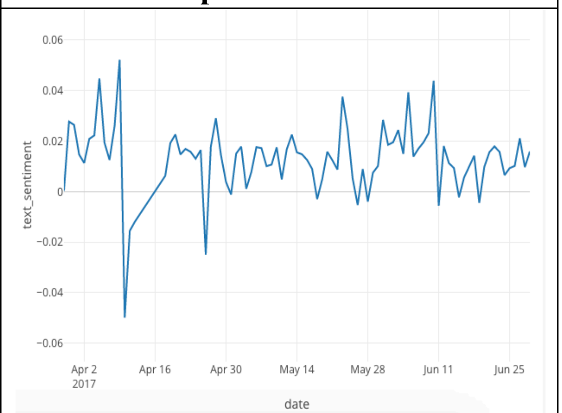
Spirit Airlines



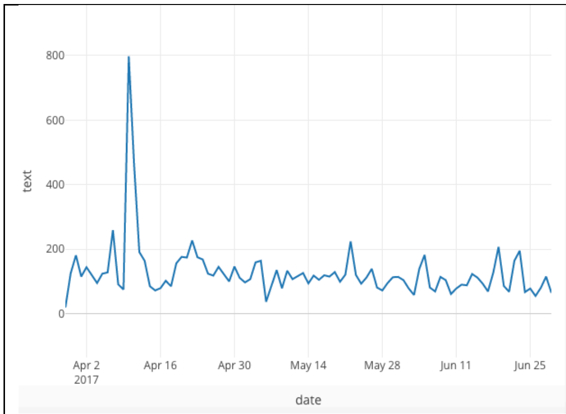
Spirit Airlines



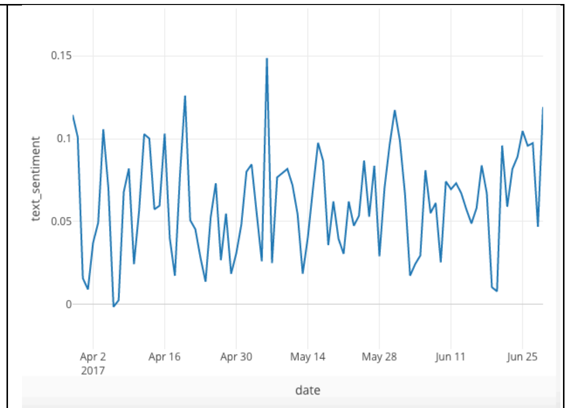
United Airlines



United Airlines



VirginAmerica



VirginAmerica

From Table 14, we can see that United Airlines and Delta Airlines received the most tweets each day. If we look at the volume of the tweets, most of the airline carriers received a high volume of tweets during the period from Apr. 9 to Apr. 12. Since United Airlines had the critical incident on Apr. 9, and news of it was spread widely on social media, many people discussed and mentioned the incident on social media and expressed their indignant emotion and anger towards United Airlines. Southwest Airlines has a much lower volume of tweets compared to the large airline carriers and the other airline carriers. Due to the size of airline companies and limited airline routes, Spirit and SkyWest Airlines received only fewer than 100 mentioned tweets per day.

When we look at the mean of the sentiment score for each airline, surprisingly, Alaska Airlines, Southwest Airlines, and Virgin America scored all positive sentiments during the period. Customers were satisfied as they perceived the service of these airline carriers. Compared to the DOT report, this is consistent with the results of sentiment analysis. The impactful incident resulted in a tremendous drop in sentiment score (the lowest sentiment score during the study period), as shown in the mean of the sentiment score of United Airlines. Compared to other airline carriers in the same time frame (Apr.

9 to Apr. 11), several airlines showed a lower sentiment score than at a normal time. Customers expressed their worries about the service quality and lost confidence in the perceived services by airline carriers.

There was a chain reaction after the United Airlines incident. The smaller issues met by other airlines could be amplified by customers during this especially hard time, even though these issues might have been tolerated by customers during a normal time. However, the competitors took advantage of this incident to draw more customers' attention and to offer good service in order to improve service quality. Spirit Airlines had a hard time during this study period and scored lots of negative sentiment. And the result is not surprising: Spirit received the lowest score on AQR. That means that Spirit received the most complaints, baggage mishandling, and involuntary denied boarding. Spirit is one of the cheap fare airline carriers, and it may sacrifice its maintenance time and its on-time rate to lower the cost of its operations. The problem will be there, and customers who choose the lowest fare should have an expectation that they will receive lower service quality, but they still complain about that, especially when the cancellation issue ruins the customers' vacations and any safety and security issues that occur during the flights are complained about on Twitter.

In the next section, the descriptive statistics of all of the variables are presented. Before I analyzed the data and performed the regression analysis, all the variables should be checked to avoid endogeneity problems. I checked for auto-correlated errors, simultaneous causality, and omitted variables. Hence, I needed include all of the possible control variables in the regression model to predict the service quality of airline carriers.

6.2 Descriptive Statistics

Table 15 presents descriptive statistics of all of the variables in this study, including the dependent variable: the AQR score, independent variables, and a set of control variables in proposed models in Chapter 5. There are 36 observations for three months of data for 12 airline carriers. The sentiment score was calculated by average of the sentiment score for one month. There are eight emotional sentiments, and each one was the sum of all of the numbers for all the tweets in one month. The same method was used to calculate the monthly arousal and valence value. Since not all of the tweets matched words in the ANEW dictionary, this shows that the larger the level of arousal, the more excited or angry customers. Months on Twitter is the indicator that shows the number of months since the airline carriers opened their official accounts on Twitter. If the airline carrier has 24/7 online response on Twitter, then I used a dummy variable to represent it on the dataset. Twitter followers could be another control variable in the model. The number of followers is shown by 1000; 36 means 36K (36,000 followers). Other control variables, like the number of passengers and the scheduled flights may affect the customers' use of Twitter to complain the perceived service quality of the airline carriers. Different months are coded as 1, 2, and 3 to represent April, May, and June 2017.

TABLE 15 DESCRIPTIVE STATISTICS OF ALL VARIABLES

Variables	Min	Max	Mean	Std. Deviation
AQR	-3.11	-0.380	-0.9853	0.55148
Volume of Tweets	7	144939	21785.33	31912.341
Average Sentiment Score	-0.026	0.148	0.046	0.043
Anger	1	43493	4907.61	8425.617
Anticipation	4	47999	7483.42	10870.964
Disgust	1	33243	3794.56	6487.302
Fear	4	53495	5964.06	10261.335
Joy	3	29815	5112.58	7022.114
Sadness	4	43351	5686.86	9020.981
Surprise	1	24555	3551.28	5271.454
Trust	6	67179	9193.25	14078.762
Sum of Valence	15	118636.98	21074.838	29113.552
Sum of Arousal	17.5	127637.64	21858.771	30472.918
Control Variables				
Passengers	653156	14090883	4694609.58	4460462.60
Scheduled Flights	198	3965	1374.75	1173.339
Months on Twitter	63	121	99.62	18.652
Twitter Followers (x1000)	2.5	2019.00	764.181	754.284
Online Response on Twitter (dummy variable)	0	1	0.17	0.378

Table 15 shows the statistical data of all the variables. The mean of AQR is -0.9853, with the standard deviation of 0.55148, the minimum value is -3.11, and the maximum value is -0.380. The average sentiment score has a mean of 0.046 with the standard deviation of 0.043; the minimum value is -0.026, and the maximum value is 0.148. This result indicates that the average sentiment score for a month on each of the airline carriers on Twitter will be around neutral. Looking at the sentiment score of entire month would dilute the impact of the sentiment score on the service quality of the airline. So, in the prior Chapter 6.1, I present the daily changes of average sentiment scores for each airline carrier on its Twitter data. The mean of the volume of tweets is 21785.33, with the standard deviation of 31912.341. The mean of the sum of valence is 21074.838, with the standard deviation of 29113.552. The mean of the sum of arousal is 127637.64, with the standard deviation of 21858.771. All of the eight emotional sentiment variables vary in both mean and standard deviation. The large standard deviation in those eight emotional sentiment indicators indicates that some airline carriers were mentioned much less than others.

In the next section, I will perform both the Pearson correlation analysis between the AQR and the six dimensions of service quality and an ANOVA test to check the difference among the three months.

6.3 Pearson Correlation Analysis

Pearson correlation is the method used to measure two or more variables that are related to each other. To test whether the sentiment of Twitter aligns with the AQR score that used the DOT's reported data, I used Pearson correlation Analysis to test the

relationship between the six dimensions of service quality and the AQR scores. Next, I present the results of the Pearson correlation Analysis.

I considered the key words of each of the six dimensions of service quality in airline industry, and I then saved all of the data to the SQL server database and applied the query to select all of the tweets in each dimensional category. Each airline carrier had six sub-datasets for each dimension of service quality. To do the Pearson correlation analysis, I consolidated the same dimension for all the twelve airline carriers into one data file. Now,I had six data sets for all the dimensions of service quality. In the data set, there were five columns: volume of tweets, average of sentiment score, sum of valence, sum of arousal, and AQR score.

Table 16 provides the results of the Pearson correlation of the variables of dimension - assurance. The AQR score was positively correlated to the average of sentiment score at a significance level of 0.05. This could be interpreted to show that the sentiment score of dimension-assurance is positively correlated with the AQR score ($r = 0.241$ weak - moderate relationship). Looking at other variables, I found that valence and arousal were highly correlated at a significance level of 0.01. This result did not surprise me, since valence and arousal can be highly correlated in the sentiment analysis and with the defined scores in the ANEW dictionary. Another finding is that the volume of tweets about assurance was negatively correlated with valence and arousal, at a significance level of 0.01. That means that the more the tweets discussed assurance of service quality, the more negative the sentiment (lower valence) and the higher the arousal score, with a strong intensity of emotion.

TABLE 16 PEARSON CORRELATION - DIMENSION TYPE: ASSURANCE

	1	2	3	4	5
1 Volume of Tweets	1.000				
2 Average of Sentiment Score	-0.242	1.000			
3 Valence	-0.993**	-0.226	1.000		
4 Arousal	-0.995**	-0.228	1.000**	1.000	
5 AQR	-0.036	0.241*	0.064	0.062	1.000

*. Correlation is significant at the 0.05 level (2-tailed)

**. Correlation is significant at the 0.01 level (2-tailed)

In Table 17, the second dimension, communication, is examined. Looking at the AQR first, I found that the average of sentiment score is positively correlated with AQR score ($r = 0.412$) at a significance level of 0.05, when people discussed communication issues of service quality in the airline industry. And valence and arousal were highly correlated with each other in this dimension, as well. The volume of tweets that mentioned communication problems was negatively correlated with arousal and valence, at a significance level of 0.05. If more tweets about communication were received on Twitter, there would be a higher chance to get the negative sentiment with complains of service quality.

TABLE 17 PEARSON CORRELATION - DIMENSION TYPE: COMMUNICATION

	1	2	3	4	5
1 Volume of Tweets	1.000				
2 Average of Sentiment Score	-0.302	1.000			
3 Valence	-0.994**	-0.276	1.000		
4 Arousal	-0.995**	-0.278	1.000**	1.000	
5 AQR	-0.009	0.412*	0.019	0.019	1.000

*. Correlation is significant at the 0.05 level (2-tailed)

**. Correlation is significant at the 0.01 level (2-tailed)

Regarding reliability of service quality, Table 18 shows that the AQR score is positively correlated to the average of sentiment score ($r = 0.293$ weak relationship) at a significance level of 0.10 (p-value is 0.083). Interestingly, the more the tweets discussed reliability, the more positive the sentiment, with a strong intensity of emotion. Customers consider reliability as an important indicator of service quality and feel happy and confident to fly with the airline carriers.

TABLE 18 PEARSON CORRELATION - DIMENSION TYPE: RELIABILITY

	1	2	3	4	5
1 Volume of Tweets	1.000				
2 Average of Sentiment Score	0.099	1.000			
3 Valence	0.994**	0.106	1.000		
4 Arousal	0.995**	0.105	1.000**	1.000	
5 AQR	0.005	0.293†	0.019	0.018	1.000

†. Correlation is significant at the 0.10 level (2-tailed)

*. Correlation is significant at the 0.05 level (2-tailed)

**. Correlation is significant at the 0.01 level (2-tailed)

Responsiveness is another dimension of service quality. Table 19 shows the Pearson correlation results for this dimension. The AQR score was strongly and positively correlated with the sentiment score ($r = 0.518$) at significance level of 0.01. The volume of tweets was negatively correlated with the average of sentiment score at a significance level of 0.05. When customers posted tweets about responsiveness, the more tweets, the lower the sentiment score that the airline received. Valence and Arousal were still positively correlated to the volume of tweets, at a significance level of 0.05. That means that the more tweets about responsiveness, the more intensive emotions were expressed by the customers. In addition, the direction of valence was different from the average sentiment score, in this case. I cannot say which one is more accurate, but the level of arousal might explain the intensity of emotions and the trends of the emotions of customers.

TABLE 19 PEARSON CORRELATION - DIMENSION TYPE: RESPONSIVENESS

	1	2	3	4	5
1 Volume of Tweets	1.000				
2 Average of Sentiment Score	-0.399*	1.000			
3 Valence	0.993**	-0.373*	1.000		
4 Arousal	0.996**	-0.385*	0.998**	1.000	
5 AQR	0.053	0.518**	0.077	0.083	1.000

*. Correlation is significant at the 0.05 level (2-tailed)

**. Correlation is significant at the 0.01 level (2-tailed)

Security and safety are important to airline carriers. Customers should have confidence and should feel peaceful about taking flights with the airline carriers. Other dimensions of service quality are considered less important than security and safety. Table 20 shows that the AQR score is positively correlated to the sentiment score of the tweets ($r = 0.375$ moderate relationship) in this dimension. If the sentiment score is lower than normal, that means that customers are worried about the safety of their flights. So, monitoring the sentiment score of Twitter data can be used to find the potential risk of safety issues. The earlier the airline carriers find out this information, the higher their chance to avoid any fatal incidents. For example, the airline carriers might schedule the maintenance of their aircrafts more frequently after getting an alert from social media data.

TABLE 20 PEARSON CORRELATION - DIMENSION TYPE: SECURITY AND SAFETY

	1	2	3	4	5
1 Volume of Tweets	1.000				
2 Average of Sentiment Score	-0.236	1.000			
3 Valence	0.985**	-0.220	1.000		
4 Arousal	0.988**	-0.222	0.999**	1.000	
5 AQR	0.055	0.375*	0.080	0.077	1.000

*. Correlation is significant at the 0.05 level (2-tailed)

**. Correlation is significant at the 0.01 level (2-tailed)

In terms of tangibility, this dimension may not be as important as other five dimensions from customers' perspective, as shown in Table 21. Since the AQR score doesn't have a significant relationship with any variables (sentiment score, volume of tweets, valence and arousal), the tangibility may be not suitable to monitor the service quality of the airline industry on social media.

TABLE 21 PEARSON CORRELATION - DIMENSION TYPE: TANGIBILITY

	1	2	3	4	5
1 Volume of Tweets	1.000				
2 Average of Sentiment Score	-0.185	1.000			
3 Valence	0.991**	-0.173	1.000		
4 Arousal	0.992**	-0.175	1.000**	1.000	
5 AQR	0.077	0.203	0.080	0.113	1.000

*. Correlation is significant at the 0.05 level (2-tailed)

**. Correlation is significant at the 0.01 level (2-tailed)

The relationships among all of the six dimensions of service quality and AQR were tested by Pearson correlation analysis. Tables 16 to 21 show the results of the correlation. Except for tangibility, the average sentiment score of other dimensions were correlated with the AQR score. Three dimensions especially - responsiveness, assurance and reliability - were positively associated with the AQR score at a significance level of 0.05, 0.05, 0.10 respectively. Hence, *H1* was supported. Sentiment scores were highly correlated to the AQR, then *H2* was supported too. Using a sentiment score of tweets, one can estimate the number of complaints on the DOT Air Travel Consumer Report.

6.4 OLS Regression Models

Figure 16 shows the correlation matrix and the different colors show the correlations among all of the variables for all of the airline carriers in the model. In detail,

the Table shows the positive relationship related to the dependent variable, AQR score. Obviously, as seen in Table 14, the highest positive correlationship was the average sentiment score in a month. The variables that most positively correlated with the AQR score are the average sentiment score (r = 0.494), the number of months on Twitter (r = 0.243), and whether the airline carriers have 24/7 responses on Twitter (r = 0.235). In addition, the number of followers on Twitter was relatively highly correlated with the AQR, with r = 0.318. The more Twitter followers of airline companies, the more accurate the prediction of the service quality that was reflected on the AQR score. Of the other eight emotional variables, only six of them (fear, sadness, trust, disgust, surprise, and joy) were positively related to the AQR score. So, the most tweets expressing those six emotions resulted in the AQR score. When analyzing the emotional sentiment, business practitioner should focus on the six emotions. If the emotional score changed a lot, the airline carrier should be aware of the service quality and should read the tweets in detail to improve their service quality. In this way, airline companies can avoid big incidents in the market and can also mitigate their influence on the public.

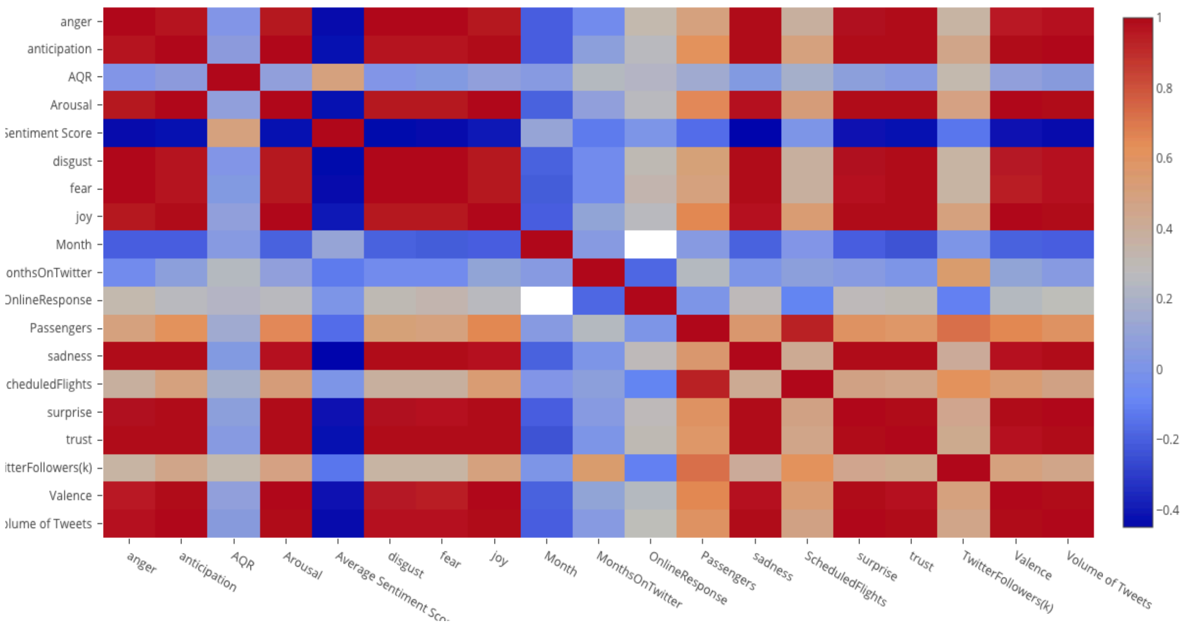


Figure 16. Correlation Matrix

TABLE 22 POSITIVE CORRELATION RELATED AQR SCORE

Column 1 ▲	Column 2	Correlation
AQR	Volume of Tweets	0.057300219025
AQR	ScheduledFlights	0.17889385577
AQR	fear	0.026416915287
AQR	sadness	0.034041916601
AQR	Average Sentiment Score	0.49457856057
AQR	Month	0.047100821877
AQR	TwitterFollowers(k)	0.31774598232
AQR	trust	0.054062992976
AQR	disgust	0.019765004598
AQR	surprise	0.066803713698
AQR	Arousal	0.080459836912
AQR	Valence	0.083692214087
AQR	joy	0.093012621665
AQR	Passengers	0.1513253372
AQR	MonthsOnTwitter	0.24316579389
AQR	OnlineResponse	0.23505010831

6.4.1 Sentiment Score and AQR

In this section, I report the results of having used OLS regression to examine the relationship between AQR and sentiment score. The regression results are reported in Table 23. To measure the service quality of airline industry, the dependent variable is the AQR score that is calculated by the formula using the data from DOT monthly report. In column 1, I wanted to test whether the volume of tweets would affect the AQR. The volume of tweets was converted to natural logarithm as log (volume). Other variables were included as control variables, such as the characteristics of the airline carrier and the characteristics of their Twitter account. The number of passengers and the number of

scheduled flights are the characteristics of the airline carrier. I used a natural logarithm to convert these two variables, as well. They were converted to Log (passengers) and Log(flights). The number of Twitter followers, months on the Twitter (from the account creation), and whether using online response features on Twitter were the characteristics of Twitter account of each airline carrier. In this model, the R squared value was 0.389, and the *p*-value was .019, which is less than .05. That means that 38.9% can be explained by this model; this model is statistically significantly in predicting airline service quality - AQR score. The coefficient of log (volume of tweets) was -1.20 and the *p*-value is .402 which is not statistically significant. Only two variables - months on Twitter and online response feature on Twitter - had a *p*-value less than 0.1 and 0.05, respectively.

However, in this model, I found two high variance inflation factors (VIF). Since the suggested threshold of multicollinearity problem is 10 (by Gefen, Straub, and Boudreau, 2000), the number of passengers (VIF: 25.234) and the number of scheduled flights (VIF: 25.214) were not tolerant in this model. They had the multicollinearity problems. If I included both of them in the regression model, it became problematic. Hence, I decided to eliminate the number of scheduled flights and keep the number of passengers, in this model.

Column 2 in Table 23 shows the model without variable - the number of scheduled flights. The VIFs in the second model were not over 3.5, which is acceptable; this indicates that the model doesn't have multicollinearity problem. The largest VIF was 3.214 of the number of followers on Twitter. In the second model in Column 2, the R squared value was 0.335 and the *p*-value was .025, which is less than 5%. That means that this model can explain 33.5% and is statistically significant at a level of 5%. After

removing the scheduled flights in the model, I found that the coefficient of log (volume of tweets) was - 0.272 with a *p*-value of .018 (significant at the level of 5%). So, I could interpret that when the log (volume of tweets) increased by 1, then AQR would be negatively affected by that and decrease 0.272. In addition, the coefficient of online response feature on Twitter was 0.775 with a *p*-value of 0.006 (significant at the 1% level). Using the online response feature on Twitter will significantly improve the service quality of the airline carrier. This feature enables communication with customers on social media and leaves the responses on the airline's official Twitter account. This interaction can be read by other customers and can significantly benefit the airline's service quality. I would recommend all the airline carriers who would like to improve their service quality take advantage of this impressive feature on Twitter. Attracting more followers on Twitter may help to improve theAQR score in this case (*p*-value is .049). Because the coefficient of the number of followers is fairly small, it may not have a strong association with AQR, but it is still helpful for an airline to attract more followers on Twitter.

Next, I would like to find the association between the sentiment score and AQR. To check whether social media data can be used to measure service quality as DOT consumer reports do, the monthly data of Twitter was used to calculate the mean of the sentiment score that represents the monthly sentiment score for each airline carrier. Column 3 of Table 23 represents that the R squared value was 0.500 of this model with *p*-value 0.001 (significant at level of 5%). That means that this model can explain 50% and is a statistically significantly predictor of the AQR score. The coefficient of the average sentiment score was 7.349, and the *p*-value is .000, which is significant at a 1% level. The

sentiment score was positively and highly associated to AQR. This can be interpreted in this way: when the sentiment score increases by 1, the AQR will increase 7.349. The sentiment score can be used to measure the service quality and to predict the AQR score. The online response feature also significantly impacted the AQR, with a coefficient 0.443 at a significant level of 5%. This indicates that the sentiment score of tweets aligned with the DOT Air Travel Consumer Report. Hence, *H2* was supported.

In addition, I wanted to test the hypothesis about whether the larger amount of negative social media data about the airline companies would result in more complaints on the DOT report with the lower AQR score. Column 4 in Table 23 presents the relationship between the AQR and the interaction of average sentiment score and the volume of tweets. The R square was .511 with a *p*-value of 0.000, which is less than 1%. That means that this model had a better explanation when considering the interaction of the sentiment score and the volume of tweets. If the product of the sentiment score and the volume of tweets increases by 1, the AQR will increase 2.538 (*p*-value is .000, which is a 1% level of significance). Hence, the larger number of tweets with negative sentiment score can be used to predict the AQR score. It can be interpreted that the service quality can be measured by the number of negative tweets. It is easy to monitor the sentiment score by using R or Python; airline companies can react quickly and can mitigate the influences of any negative news. Thus, *H3* was supported.

As shown in Table 23, I found that the number of passengers or the number of scheduled flights did not impact on AQR score. The online response feature on Twitter can be used by airline companies to improve their service quality. Analyzing the

sentiment score and the volume of tweets was enough to measure the service quality of the airline industry and to predict the AQR score and complaints on DOT report.

TABLE 23 REGRESSION MODELS RESULTS (VOLUME, SENTIMENT SCORE AND AQR)

Dependent Variable	**AQR (1)**	**AQR (2)**	**AQR (3)**	**AQR (4)**
(Constant)	3.091	-.841	-2.247	-1.409
Log (Volume of Tweets)	-1.20	-.272 *		
	(-.253)	(-.572)		
Average Sentiment Score			7.349**	
			(.579)	
Log (Volume of Tweets) *				2.538**
Average Sentiment Score				(.622)
Log (The number of	-1.507	-.092	.020	.017
passengers)	(-1.183)	(-.072)	(.016)	(.014)
Log (The number of	1.396			
scheduled flights)	(1.121)			
Months on Twitter	.016 *	-.010	.005	-.002
	(.552)	(.335)	(.186)	(-.078)
Twitter Followers	.0002	.0002 *	.0002	.0002
	(.392)	(.545)	(.300)	(.268)
Twitter Online Response	.822 **	.775 **	.433*	.229
	(.563)	(.531)	(.297)	(.157)
R-squared	.389	.335	.500	.511
Adjust R-squared	.262	.224	.416	.429

Standard deviation (in parentheses)

*. Significance at the 5% level

**. Significance at the 1% level

***. Significance at the 0.1% level

6.4.2 Emotional Sentiment (eight basic emotions) and AQR

I tried to determine whether the eight basic emotions would affect the AQR score independently. However, the model did not significantly predict the AQR score, with a p-value of 0.520. All of the variables of the eight emotional sentiments were not statistically significantly, associated with AQR. Unfortunately, the eight emotions of the sentiment score did not have an impact on AQR, in this case. Hence, I did not test the measurement of service quality using eight emotional sentiment analyses.

6.4.3 Valence/Arousal and AQR

Next, I wanted to use valence and arousal to check whether social media data associated with AQR. AQR was still a dependent variable. The valence and the level of arousal were calculated based on monthly tweets. The sum of valence of each month of each airline carrier was calculated and was then converted to natural logarithm as log (sum of valence). The same method was applied to calculate arousal, as log (sum of arousal). In Table 24, Column 1 presents the regression model, with valence and arousal as independent variables. The R squared value was 0.409 at a significant level of 5% (*p*-value is .012). That means that this model was statistically significant to explain 40.9% of the relationship among those variables. Except for the number of passengers, other variables were significant at a level of 5% (valence, arousal, months on Twitter, the number of followers on Twitter) and 1% (Twitter online response). Even though this model could explain the prediction of service quality in airline industry, I found that it also had the multicollinearity problem. Since valence and arousal words came from the ANEW dictionary, when I calculated those two values, it showed the extent to which they

correlated with each other. So, this model needs improvement; I decided to keep arousal as the independent variable. Because valence is somewhat similar as sentiment score to show positive and negative sentiment, I introduced the sentiment score into the model to replace the valence in the model shown in Column 4.

Column 2 is the model without valence. Having an R squared value 0.310 and being statistically significant at a level of 5% (p-value is .040), this model can explain 31%. If the log (sum of arousal) increases by 1, the AQR score will decrease 0.261. As we know, the arousal represents the intensity of emotion. If tweets were detected that had a high arousal score, airline companies could know the service failure in a short time and could fix this in a timely manner. It also indicates that the negative tweets with a higher arousal score will lead to a lower AQR score. Hence, business practitioners can use the two variables to predict the service quality and, perhaps, service failure. *H5* was supported.

In order to test *H4*, I needed to introduce valence, but not include arousal in this model. Column 3 shows the model with the R squared 0.318 and significance at a level of 5% (*p*-value is .035). The Log (sum of valence) was statistically significant at a 5% level (*p*-value is .019) in this model with a coefficient of -0.274. This can be interpreted in this way: when the valence increases by 1, the AQR score will decrease -0.274. This result is not what was expected, but it shows the association between valence and AQR. I expected that they would be positively related to each other. Another explanation could be that the more customers that use higher valence words, the lower the service quality they received. Sometimes, customers may express sarcasm on Twitter and result in a special relationship. Hence, *H4* is partially supported. The interaction of valence and

arousal could not be used as one variable in the model due to multicollinearity issue, so *H6* is not supported.

After introducing sentiment score in the model, as shown in Column 4 in Table 24, I found that arousal was not statistically significant (*p*-value is .948) - only the sentiment score with a coefficient of 7.409 at a level of 1% (*p*-value .002). Other variables were not significantly affected by the AQR. So, this model did not well predict service quality in the airline industry. The next step is to investigate the interaction of the sentiment score and the arousal score. The results of the model are shown in Column 5 in Table 24.

In the Column 5 model, we can see that the interaction of the sentiment score and the level of arousal seemed significantly important to predict the AQR score (coefficient was 2.467 and *p*-value was .000). That means that, when the interaction of the sentiment score and the arousal increased by 1, the AQR will increase 2.467. This positive relationship was expected and *H7* is supported. Other variables were not important factors in predicting the AQR score. This finding can benefit the airline industry as they monitor their service quality. Since they can only focus on the sentiment score and the arousal score, it is not necessary to monitor the other variables listed in this study. The model had R squared at 0.514 and was significant at a level of 1% (*p*-value is .000). Compared to other models in this study, the model in Table 24, Column 5 had the best fit to predict the service quality. 51.4% is explained by this model. That means the interaction of the sentiment score and the arousal had a moderate effect on AQR score. That is acceptable, because there are so many other outside factors for the service quality of airline industry. For example, the following factors could affect the service quality in

the airline industry: the flight time, the length of a flight, a weather problem, the specific route, the food on board, and so on. Using social media data analysis can benefit the service industry to detect any service failure immediately and to help them to fix the problem in a short time.

TABLE 24 REGRESSION MODELS RESULTS (VALENCE, AROUSAL, AND AQR)

Dependent Variable	**AQR (1)**	**AQR (2)**	**AQR (3)**	**AQR (4)**	**AQR (5)**
(Constant)	-1.923	-1.634	-1.643	-2.488	-1.871
Average Sentiment Score				7.409**	
				(.584)	
Log (Sum of Valence)	-6.626*		-.274*		
	(-13.346)		(-.553)		
Log (Sum of Arousal)	6.237*	-.261*		.008	
	(12.602)	(-.528)		(.016)	
Average Sentiment Score*					2.467***
Log (Sum of Arousal)					(4.560)
Log (The number of	.037	.064	.065	.093	.122
passengers)	(.029)	(.052)	(.052)	(.074)	(.098)
Months on Twitter	.016*	.010	.011	.007	.000
	(.552)	(.353)	(.367)	(.220)	(.008)
Twitter Followers	.0002*	.0002	.0002	.0002	.000
	(.604)	(.425)	(.437)	(.242)	(.199)
Twitter Online Response	.835**	.732**	.746**	.425	.245
	(.572)	(.502)	(.511)	(.291)	(.168)
R-squared	.409	.310	.318	.502	.514
Adjust R-squared	.287	.195	.204	.339	.433

Standard deviation (in parentheses)

*. Significance at the 5% level

**. Significance at the 1% level

***. Significance at the 0.1% level

The three months of Twitter data could not be tested the differences in SPSS. Since the model was not valid if I chose to see the time effect during this period. Thus, the equation (15), proposed in Chapter 5, is not valid.

Finally, I used the variable importance method (random forest) to double-check the regression models above. Figure 17 shows all of the variables with relative importance to the AQR score. Obviously, the average sentiment score had the highest importance to AQR, followed by emotion: disgust, the number of scheduled flights, months on Twitter (airline carrier official account), the number of passengers, and emotion: fear. This important variable test indicates that simply getting the sentiment score of tweets is good enough for predicting service quality in the airline industry. Since the sentiment score is aligned with AQR score that is generated from the DOT consumer report, researchers and marketers in the airline industry can get an alert about service quality in a short time. They don't have to wait to see the DOT consumer report, two months later. Another interesting finding is that the number of scheduled flights and the number of passengers also affect the AQR score. That means that more consumers use social media tools to complain about the experiences they have met aboard airline carriers. Most complaints have the three expressed emotions that are important to AQR: disgust, fear, and anger. In terms of valence and arousal, valence seems more important than arousal to the AQR score. Looking only at arousal may not contribute to a strong prediction about service quality in the airline industry.

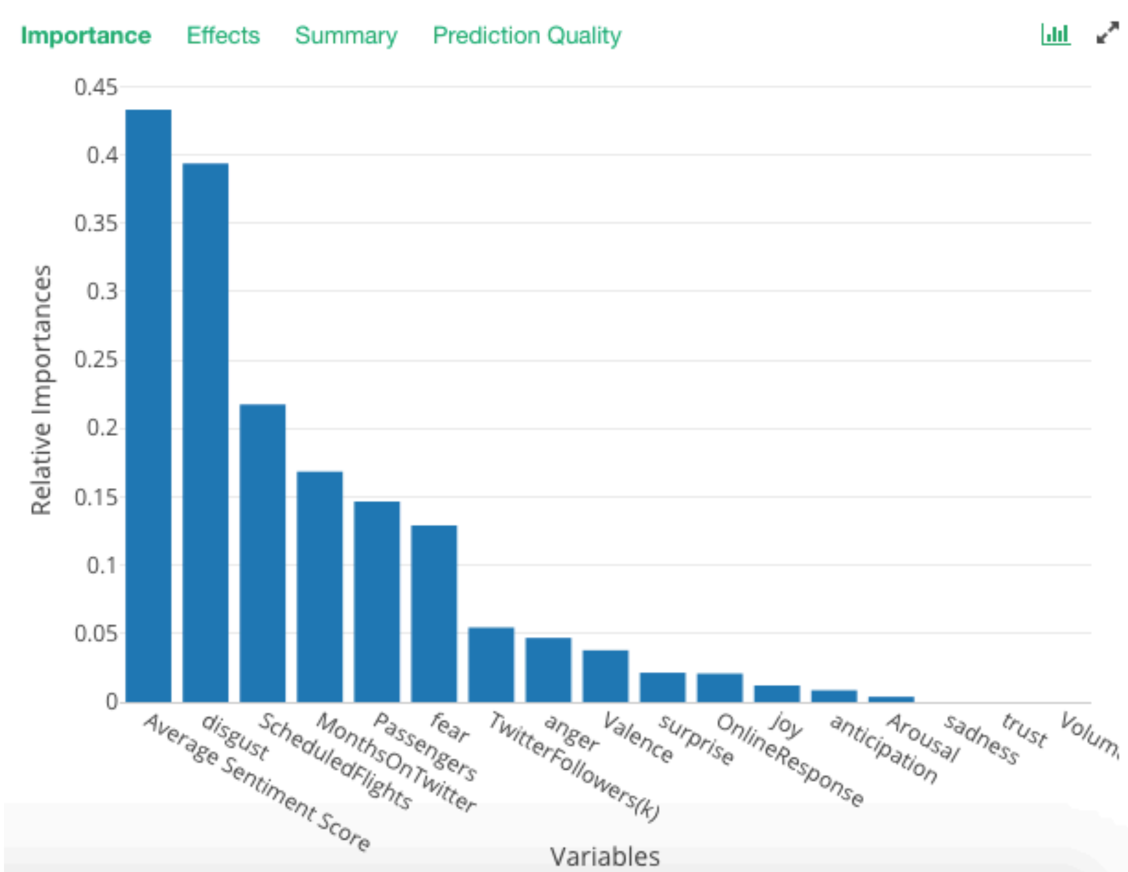


Figure 17. Variable Importance Test Results

CHAPTER

7 PROPOSED BUSINESS ANALYTICS FRAME WORK FOR SERVICE INDUSTRY

Pang and Lee (2008) pointed out the challenges and opportunities engendered by the growing popularity of opinion-rich resources, such as online comments, blogs, social media platforms, and online forums. By investigating the value of those unstructured data, businesses are able to gather competitive intelligence and improve their products and services in a short time. In order to actively use information technology to seek and retrieve the data and to understand consumer's opinions, computational treatment of subjective texts like reviews, comments, and opinions have suddenly erupted in the past ten years (Pang and Lee, 2008). With sentiment analysis, the information system could be developed for information retrieval, opinion mining, and sentiment analysis. This new system can provide data visualization and immediate responses to negative comments, and can reduce the risk of public crisis. Business practices use the opinion-oriented information system to make decisions and to improve their products and services.

In this case, the airline industry indeed needs such an opinion-oriented information system to monitor the quality of their service to the public. The following steps and technologies can be used for creating the system. The entire system could be integrated, with all of its features, into one piece. Then, each step can be connected to the others, automatically. All of the data retrieved from social media should be saved in a CSV file and then saved to the database for later processing. A script language can be used to create the opinion-oriented system, to mitigate the work load for the employees and to provide the ultimate support from computing 24/7. Keywords can be created for each dimension,

in order to track service quality; they can be saved into different tables in a database. And then all of the sentiment analysis and statistical analysis can be performed, in order to support business decision making. The concrete steps are listed below:

1. Use Python code using Twitter API to get the tweets for one's own company or for competitors
2. Gather all of the tweets or parts of tweets from past time or real time. The company can limit the tweets to specified numbers
3. Use the sentiment analysis packages of R to get the sentiment score and the emotion for each tweet
4. Apply the proposed service quality index
5. Save all of the tweets and the sentiment scores to Microsoft SQL Server Databases
6. Manipulate the data and find the pattern of the current data
7. Visualize the date, create report (weekly, monthly, quarterly, annual), respond to the public
8. Make decisions, adjust business strategy, respond to public if necessary.

Table 18 illustrates the conceptual layer of the opinion-oriented information system proposed by this study. With this information system, businesses in service industries can have analyze results every day to help them to better understand and connect to their customers. When something goes wrong, the system will send an alert to the proper departments and will direct them to perform some actions, depending on the

type of alert. This information system has good scalability due to its expandable ability. This system should be able to monitor multiple popular social media platforms, such as Twitter, Facebook, and Instagram. When retrieving social media data from those social media platforms, Python can be used to connect the API provided by the social media or to web-scrap the social media data. In addition, the keywords in each dimension of service quality can be modified at any time. Since more customers will use social media platform to communicate and stay tuned with businesses, the keywords may be changed by time. More Internet slang is likely to emerge in the future. For example, “LOL” means “laugh out loud”; it is a popular element of Internet slang. And “bump” is always used in the forums that move the posts or comments to the top on the first page. “Troll” is another word in Internet slang that used a lot today. “Trolls” are the people who want to take pleasure from starting disagreements and angering other people online. “I can’t even” is another way to say, “I’m speechless.” This phrase is used when you have no words to express to respond to incredible or unbelievable things. There is much more slang that is used on the Internet today, and there is likely to be more and more in the future. Thus, updating the keywords for each dimension of service quality will be necessary.

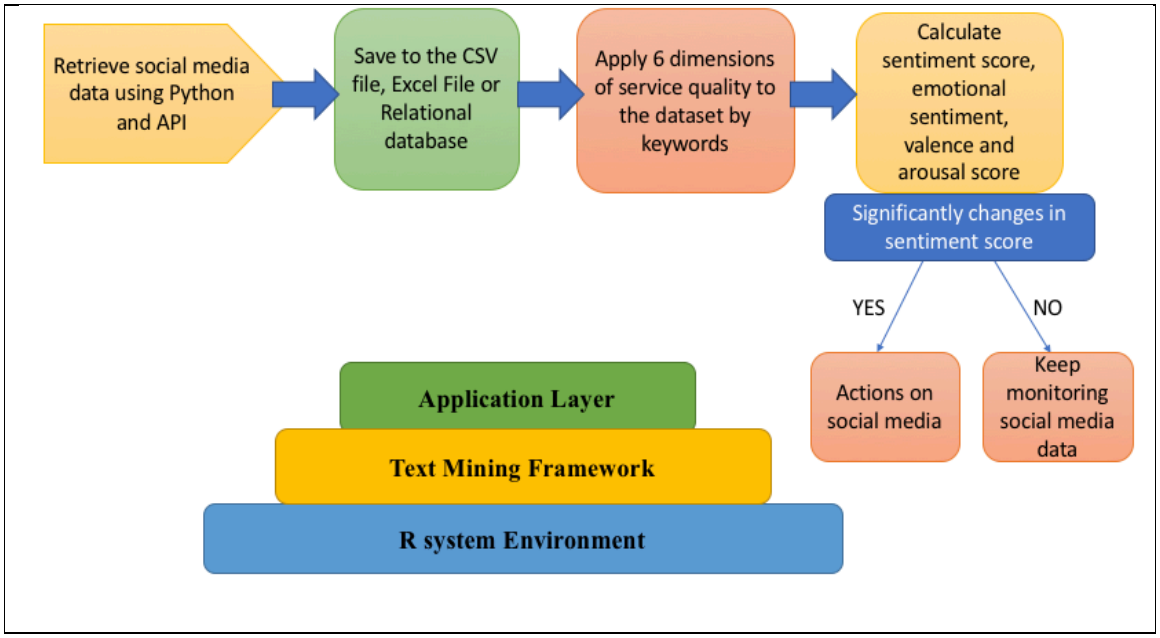


Figure 18. Conceptual Layer of Opinion-Oriented Information System

CHAPTER

8 RESEARCH LIMITATIONS

This study is limited in several ways. First, this study uses only one social media platform to test the effectiveness of social media data analysis. Some other popular social media websites such as Google+, Facebook, and Instagram are not included in the data set. And the social media data was retrieved from Twitter for consecutive three months in 2017. The results from this study may not fully reveal and explain the effectiveness of social media analysis to test the service quality. As the service industry has the characteristic of peak season, the results of this study may not explain the situation from off season.

Second, the six dimensions of service quality use defined keywords in this study. The keywords may not 100% accurate for each dimension. Due to it was generated by three graduate students, the keywords may have differences by other researchers. The ANEW dictionary should be updated accordingly because the more slangs are emerging out in recent years.

Third, Sentiment analysis was based on Lexicons. The dictionary-based “bag-of-words” approaches suffer from various drawbacks. Sentiment analysis are applied the Lexicons and words may be evaluated out of context or with the wrong sense. Another limitation is the language. I analyze the English tweets only from Twitter. But there are many other languages are used in the Twitter and other social media platform. It is better to include other languages in the dataset and do the same sentiment analysis on them. To do this, businesses in service industry can use social media to predict their perceived

service quality by customers. And the tweets are relative short and contain a few words in one post. This may also limit the results. Sriram, Fuhry, Demir, Ferhatosmanoglu & Demirbas (2010) state that using authors' information and other features like the number of followings, the number of followers will achieve higher quality of sentiment analysis.

CHAPTER

9 CONCLUSION AND FUTURE RESEARCH

9.1 Conclusion

The purpose of this study is to determine the effectiveness of social media analysis in detecting service quality in the airline industry. Having applied the SERVQUAL model and having adapted the model to propose the six dimensions of service quality in the airline industry, I used the social media data from Twitter to gather the tweets from three months in 2017, and then pre-processed and analyzed the sentiment after the data collection.

I cleaned the data, using several methods in this study, and removed all of the data that were not related to customers' comments. A word cloud from text mining was used to find the frequency of words used for each airline company. A word cloud is more intuitive and is considered to be more attractive by non-researchers and business practices.

I found that the overall volumes and mean of sentiment score showed different trends, from Table 14. In terms of valence and arousal, the six dimensions were tested by Pearson correlations. Responsiveness, assurance, and reliability were the top three, and were highly correlated to the AQR score. In these regression models, the sentiment score aligned with the DOT data on report. And the sentiment score and arousal interactions offered the best independent variable and the best model fit, as shown in Table 14. Sentiment score and emotional sentiment change were found to be consistent with the conclusion made by Wahba (2017). Negative sentiment could have a short-term effect on company's market value, but in the long run, may not have influence. Analyzing the

social media data also can avoid the huge incident affects in public relationship. Business practitioners should put more investments in the digital marketing, data analysis and hire more professionals to assist top managers in the companies to make the right decision in a timely manner. This study also provides verification of SERVQUAL model in airline industry.

9.2 Future Research

There are many possibilities to extend the research in this dissertation. First, future research can consider multiple social media platforms to check the service quality for any businesses in the service industry. For example, the Internet provider that offers Internet and cable services to the customers could use social media analysis to predict and measure their service quality. Government can also use social media analysis to determine its citizens' opinions and to improve its service by simplifying processes.

Comparing different categories, i.e., which category has the average higher speed of retweet, as well as the number of retweets, can be investigated. Retweet is the feature on Twitter that allows users to share the original post on their own post. It is possible to get the first retweet time on the Twitter dataset and to find how long it was retweeted on Twitter, and to find the top five or top ten users who mentioned the airline carriers' name frequently, and then get the related information on Twitter, such as the number of their followers or the number of their followings.

In terms of sentiment analysis, I am also interested in analyzing emotions from images posted by customers. As we know, pictures can be included to users' social media

profiles or posts. A study of these will be helpful to improving service quality in business. Then, even more resources from other media can be applied. For example, it might be helpful to extract data from newspapers as another source, and, if possible, to include more languages: Korean, Russian, Japanese, and Mandarin are the best choices.

Fake comments are another issue; it is necessary to detect sarcasm in tweets. Sarcasm is not easy to detect. It may need more time to configure out after the algorithm is released, to make it capable of detecting sarcasm in tweets.

And remove these fake comments before we process the data. Researchers could be able to use the same methodology to analyze data from social media to measure service quality for all the service providers in multiple industries.

AQR is calculated by using the report data from DOT. It only can be calculated by month. In the future, if there are official data about the daily performance of airline carriers, the dependent variable should be more suitable than AQR.

And more, social media data is powerful, and it can be related to many elements in the business world. For example, ranks, reputations, ratings, revenue, stock price, sales and so on. I will include Bloomberg data as variables in the future research and social media data is expected to impact the business in different aspects.

APPENDICES

APPENDIX A: RETRIEVE OLD TWEETS FROM TWITTER API CODE

```
def main(argv):
    if len(argv) == 0:
        print('You must pass some parameters. Use \"-h\" to help.')
        return
    if len(argv) == 1 and argv[0] == '-h':
        f = open('exporter_help_text.txt', 'r')
        print (f.read())
        f.close()
        return
    try:
        opts, args = getopt.getopt(argv, "", ("username=", "near=", "within=",
"since=", "until=", "querysearch=", "toptweets", "maxtweets=", "output="))
        tweetCriteria = got.manager.TweetCriteria()
        outputFileName = "output.csv"
        for opt,arg in opts:
            if opt == '--username':
                tweetCriteria.username = arg
            elif opt == '--since':
                tweetCriteria.since = arg
            elif opt == '--until':
                tweetCriteria.until = arg
```

```
        elif opt == '--querysearch':
            tweetCriteria.querySearch = arg
        elif opt == '--toptweets':
            tweetCriteria.topTweets = True
        elif opt == '--maxtweets':
            tweetCriteria.maxTweets = int(arg)
        elif opt == '--near':
            tweetCriteria.near = '"' + arg + '"'
        elif opt == '--within':
            tweetCriteria.within = '"' + arg + '"'
        elif opt == '--within':
            tweetCriteria.within = '"' + arg + '"'
        elif opt == '--output':
            outputFileName = arg
    outputFile = codecs.open(outputFileName, "w+", "utf-8")

outputFile.write('username;date;retweets;favorites;text;geo;mentions;hashtags;id;
permalink')
    print('Searching tweets...\n')
    def receiveBuffer(tweets):
        for t in tweets:
```

```
        outputFile.write(('\n%s;%s;%d;%d;"%s";%s;%s;%s;"%s";%s' % (t.username, t.date.strftime("%Y-%m-%d %H:%M"), t.retweets, t.favorites, t.text, t.geo, t.mentions, t.hashtags, t.id, t.permalink)))
                outputFile.flush();
                print ('More %d saved on file...\n' % len(tweets))


            got.manager.TweetManager.getTweets(tweetCriteria, receiveBuffer)
        except arg:
            print('Arguments parser error, try -h' + arg)
        finally:
            outputFile.close()
            print('Done. Output file generated "%s".' % outputFileName)
```

APPENDIX B: SENTIMENT ANALYSIS CODE IN R

```
> pos = scan('positive-words.txt', what='character', comment.char=';')
Read 2006 items
> neg= scan('negative-words.txt', what='character', comment.char=';')
Read 4783 items
#Adding words to positive and negative databases
pos = c(pos, 'Congrats', 'prizes', 'prize', 'thanks', 'thnx', 'Grt', 'gr8', 'plz', 'trending', 'recovering', 'brainstorm', 'leader')
neg = c(neg, 'Fight', 'fighting', 'wtf', 'arrest', 'no', 'not', 'fight')
```

#Score Sentiment

```
score.sentiment = function(tweets, pos.words, neg.words)
{
 require(plyr)
 require(stringr)
 scores = laply(tweets, function(tweet, pos.words, neg.words) {
  tweet = gsub('https://','',tweet) # removes https://
  tweet = gsub('http://','',tweet) # removes http://
  tweet=gsub('[^[:graph:]]', ' ',tweet) ## removes graphic characters
  #like emoticons
  tweet = gsub('[[:punct:]]', '', tweet) # removes punctuation
  tweet = gsub('[[:cntrl:]]', '', tweet) # removes control characters
  tweet = gsub('\\d+', '', tweet) # removes numbers
  tweet=str_replace_all(tweet,"[^[:graph:]]", " ")
```

```
    tweet = tolower(tweet) # makes all letters lowercase
    word.list = str_split(tweet, '\\s+') # splits the tweets by word in a list
    words = unlist(word.list) # turns the list into vector
    words.matches = match(words,words)
    pos.matches = match(words, pos.words) ## returns matching
    #values for words from list
    neg.matches = match(words, neg.words)
    pos.matches = !is.na(pos.matches) ## converts matching values to true of false
    neg.matches = !is.na(neg.matches)
    words.matches =!is.na(words.matches)

    score = (sum(pos.matches) - sum(neg.matches))/(sum(words.matches)) # true and false
are
    #treated as 1 and 0 so they can be added    #Sentiment scores are calculated by the sum
of positive minus negative, then divided by the number of words of tweet.
    return(score)
  }, pos.words, neg.words )
  scores.df = data.frame(score=scores, text=tweets)
  return(scores.df)
}
```

APPENDIX C: EMOTICONS WITH SENTIMENT

:-) :) :o) :] :3 :c) :> =] 8) =) :} :^) Positive

:D C: Extremely-Positive

:-D :D 8D xD XD =D =3 <=3 <=8 Extremely-Positive

<=3 <=8 8===D 8===B Negative

--!-- Negative

:-(:(:c :< :[:{ Negative

D: D8 D; D= DX v.v Extremely-Negative

:-9 Negative

;-) ;) *) ;] ;D Positive

:-P :P XP :-p :p =p :-Þ :Þ :-b :b Positive

:-O :O O_O o_o 8O OwO O-O 0_o O_o O3O o0o ;o_o; o...o 0w0 Positive

c.c C.CNegative

:-/ :/ :\ =/ =\ :S Negative

:| Neutral

d:-) qB-) Positive

:)~ :-)>.... Neutral

:-X :X :-# :# Positive

O:-) 0:3 O:) Negative

:'(;*(T_T TT_TT T.T Q.Q Q_Q ;_; Negative

:-* :* Positive

^o) Negative

>:) >;) >:-) Neutral

B) B-) 8) 8-) Neutral

^>.>^ ^<.<^ ^>_>^ ^<_<^ Negative

D:< >:(D-:< >:-(:-@[1] ;(`_´ D< Negative

<3 <333 Positive

</3 Negative

=^_^= =>.>= =<_<= =>.<= Positive

\,,/ \m/ Extremely-Positive

\m/\>.</\m/ Extremely-Positive

\o/ Extremely-Positive

\o o/ Positive

d'-' d'_' d'-'b d'_'b Positive

o/\o Positive

:& Extremely-Negative

:u Neutral

@}-;-'--- Positive

3:00 Positive

=]:-)= Neutral

d^_^b d-_-b Positive

(^_^) (^-^) (^ ^) (^.^) Positive

(~_^) (^_~) ~.^ ^.~ Positive

(>_<) (>.<) Negative

(>_>) (¬_¬) Negative

(-_-) Negative

(^o^) Positive

(^3^) Positive

(^_^') ^_^_^') ^^" ^^^_.^') ^^_^^; ^&^^.^;& ^^^; ^^^7 Negative

d(>w<)b Extremely-Positive

q(;^;)p Extremely-Negative

9(x.x)9 (;.;)9 Negative

(._.) (,_,) Negative

［(－－)］ZZzzz... Neutral

(X_X) x_x Negative

^///^ >///< >///> o///o -///- =///= Positive

_|_ (-.-) _|_ t(>.<t) Negative

(V)!_!(V) Neutral

ʘ‿ʘ Extremely-Positive

\m/*.*\m/ Extremely-Positive

O?O Negative

&_& Positive

0-0 Positive

(^^^) Negative

(_/) Positive

B) Positive

B(Negative

X3 Extremely-Positive

:3 Positive

x_O O_x	Negative
O|¯|_ orz	Positive
m(_ _)m	Positive
^__^	Neutral
^^	Positive

APPENDIX D: TEXT MINING IN R

Packages Used in This Study

- twitteR: Provides an interface to the Twitter web API
- stringr: String operations in R
- ROAuth: Provides an interface to the OAuth 1.0 specification allowing users to authenticate via OAuth to the server of their choice.
- RCurl: Provides functions to allow one to compose general HTTP requests

and provides convenient functions to fetch URIs, get & post forms, etc. and process the results returned by the Web server.

- ggplot2: An implementation of the grammar of graphics in R. It combines the advantages of both base and lattice graphics: conditioning and shared axes are handled automatically, and you can still build up a plot step by step from multiple data sources.
- reshape: Flexibly restructure and aggregate data using just two functions: melt and cast
- tm: A framework for text mining applications within R.
- RJSONIO: This is a package that allows conversion to and from data in Javascript object notation (JSON) format. This allows R objects to be inserted into Javascript/ECMAScript/ActionScript code and allows R programmers to read and convert JSON content to R objects
- wordcloud: visual representation in the form of word cloud where size of the word is proportional to the frequency of words used in the tweets
- gridExtra: Provides a number of user-level functions to work with "grid"

graphics, notably to arrange multiple grid-based plots on a page, and draw tables.

- plyr: Tools for Splitting, Applying and Combining Data
- syuzhet: this package helps to extract sentiment and sentiment-derived plot arcs from text using three popular sentiment lexicons: AFINN, BING, NRC
- sentiment: classifies the emotions of text
- sentimentr: calculate text polarity sentiment at the sentence level and optionally aggregate by rows or grouping variable(s)
- RColorBrewer: color schemes for the plots and word cloud
- e1071: Functions for latent class analysis, short time Fourier transform, fuzzy clustering, support vector machines, shortest path computation, bagged clustering, naive Bayes classifier
- SparseM: Some basic linear algebra functionality for sparse matrices is provided: including Cholesky decomposition and backsolving as well as standard R subsetting and Kronecker products.

REFERENCES

Akbaba, A. (2006). Measuring service quality in the hotel industry: A study in a business hotel in Turkey. *International Journal of Hospitality Management*, 25(2), 170-192.

Ali, M., & Raza, S. A. (2017). Service quality perception and customer satisfaction in Islamic banks of Pakistan: the modified SERVQUAL model. *Total Quality Management & Business Excellence*, *28*(5-6), 559-577.

Aman, S., & Szpakowicz, S. (2007, September). Identifying expressions of emotion in text. In International Conference on Text, Speech and Dialogue (pp. 196-205). Springer, Berlin, Heidelberg.

Anitsal, M., Anitsal, I., & Anitsal, S. (2017). A Sentiment Analysis of Air Passengers of Top Ten US Based Airlines. (pp. 37-50).

Bansal, H. S., & Taylor, S. (2015). Investigating the relationship between service quality, satisfaction and switching intentions. In *Proceedings of the 1997 Academy of Marketing Science (AMS) Annual Conference* (pp. 304-313). Springer International Publishing.

Bowen, B.D. & Headley, D. E. (2017). *The Airline Quality Rating (AQR) Report*. Retrieved from https://airlinequalityrating.com/

Chevalier, J. A. & Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. *Journal of marketing research*, *43*(3), 345-354.

Cronin, J, Brady, M., Thomas, G., & Hult, M. (2000), Assessing the Effects of Quality, Value and Customer Satisfaction on Consumer Behavioral Intentions in Service Environments, *Journal of Retailing*, 76(2), 193-217.

Duan, W., Cao, Q., Yu, Y., & Levy, S. (2013, January). Mining online user-generated content: using sentiment analysis technique to study hotel service quality. In *System Sciences (HICSS), 2013 46th Hawaii International Conference on* (pp. 3119-3128). IEEE.

El Saghier, N., & Nathan, D. (2013). Service quality dimensions and customers' satisfactions of banks in Egypt. In *Proceedings of 20th International Business Research Conference* (pp. 4-5).

Estelami, H. 2000. Competitive and procedural determinants of delight and disappointment in consumer complaint outcomes. *Journal of Service Research*, 2: 285–300.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2013). G*Power Version 3.1.7 [computer software]. Uiversität Kiel, Germany. Retrieved from http://www.psycho.uni-duesseldorf.de/abteilungen/aap/gpower3/download-and-register

Fernandez, N., & Marin, P. 1998. Market power and multimarket contact: Some evidence from the Spanish hotel industry. *Journal of Industrial Economics*, 46:301–315.

Gefen, D., Straub, D., & Boudreau, M.-C. (2000). Structural equation modeling and regression: Guidelines for research practice. *Communications of the Association for Information Systems*, 4(1), 7.

Grönroos, C. (1984). A service quality model and its marketing implications. *European Journal of marketing*, *18*(4), 36-44.

He, W., Shen, J., Tian, X., Li, Y., Akula, V., Yan, G., & Tao, R. (2015). Gaining competitive intelligence from social media data: Evidence from two largest retail chains in the world. *Industrial Management & Data Systems*, *115*(9), 1622-1636.

He, W., Tian, X., Chen, Y., & Chong, D. (2016). Actionable Social Media Competitive Analytics for Understanding Customer Experiences. *Journal of Computer* Information Systems, 56(2), 145-155.

Hearst, M (1999). Untangling Text Data Mining. In *Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics*, pp. 3-10. Association for Computational Linguistics, Morristown, NJ, USA. ISBN 1-55860-609-3.

Hennig-Thurau, T., Gwinner, K. P., Walsh, G., & Gremler, D. D. (2004). Electronic word-of-mouth via consumer-opinion platforms: what motivates consumers to articulate themselves on the internet?. *Journal of interactive marketing*, *18*(1), 38-52.

Hu, M. & Liu, B. (2004), Mining and Summarizing Customer Reviews. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2004)*, Aug 22-25, 2004, Seattle, Washington, USA

Hussain, R., Al Nasser, A., & Hussain, Y. K. (2015). Service quality and customer satisfaction of a UAE-based airline: An empirical investigation. *Journal of Air Transport Management*, *42*, 167-175.

Jurafsky, D., & Martin, J. H. (2014). *Speech and language processing* (Vol. 3). London: Pearson.

Kitapci, O., Taylan Dortyol, I., Yaman, Z., & Gulmez, M. (2013). The paths from service quality dimensions to customer loyalty: An application on supermarket customers. *Management Research Review*, 36(3), 239-255.

Ladhari, R. (2009). A review of twenty years of SERVQUAL research. *International Journal of Quality and Service Sciences*, 1(2), 172-198.

Lang, P. J. (1995). The emotion probe: Studies of motivation and attention. *American psychologist*, *50*(5), 372.

Leung, D., Lee, A. & Law, R. (2012). Examining hotel managers' acceptance of web 2.0 in website development: a case study of hotels in Hong Kong. In: Sigala, M., Christou, E. & Gretzel, U., eds., 2012. *Social media in travel, tourism and hospitality – Theory, practice and cases.* Surrey, England: Ashgate Publishing Limited. pp. 53-65

Liu, B., Hu, M. & Cheng, J.S. (2005), Opinion Observer: Analyzing and Comparing Opinions on the Web. *Proceedings of the 14th International World Wide Web conference (WWW-2005)*, May 10-14, 2005, Chiba, Japan.

Luo, X. (2007). Consumer negative voice and firm-idiosyncratic stock returns. *Journal of Marketing*, 71(3), 75-88.

Martin, W. B. (1986), Defining What Quality Service is for You, *Cornell Hotel and Restaurant Administration Quarterly*, 26(4), 32-38.

Mäntylä, M., Adams, B., Destefanis, G., Graziotin, D., & Ortu, M. (2016, May). Mining valence, arousal, and dominance: possibilities for detecting burnout and productivity?. In *Proceedings of the 13th International Conference on Mining Software Repositories* (pp. 247-258). ACM.

Meyer, D., Hornik, K., & Feinerer, I. (2008). Text mining infrastructure in R. Journal of statistical software, 25(5), 1-54.

Mazzeo, M. (2003). Competition and service quality in the US airline industry. *Review of Industrial Organization*, 22: 275–296.

Mohammad, S. M. & Turney, P. D. (2010, June). Emotions evoked by common words and phrases: Using Mechanical Turk to create an emotion lexicon. In Proceedings of the

NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text (pp. 26-34). Association for Computational Linguistics.

Mohammad, S. M. & Turney, P. D. (2013). Crowdsourcing a word–emotion association lexicon. *Computational Intelligence*, 29(3), 436-465.

Mukherjee, A., Nath, P., & Pal, M. (2003), Resource, Service Quality and Performance Triad: a Framework for Measuring Efficiency of Banking Services, *Journal of the Operational Research Society*, 54(7), 723–735.

Novack, R. A., Rinehart, L. M., & Langley, C. J. (1994), An Internal Assessment of Logistics Value, *Journal of Business Logistics*, 15(1), 113-152.

Ostrowski, P. L., O'Brien, T. V., & Gordon, G. L. (1993). Service quality and customer loyalty in the commercial airline industry. *Journal of travel research*, *32*(2), 16-24.

Olorunniwo, F., Hsu, M. K., & Udo, G. J. (2006). Service quality, customer satisfaction, and behavioral intentions in the service factory. *Journal of services marketing*, *20*(1), 59-72.

Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1985). A conceptual model of service quality and its implications for future research. *the Journal of Marketing*, 41-50.

Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1988). Servqual: A Multiple-Item Scale for Measuring Consumer Perceptions of Service Quality, *Journal of Retailing*, 64(1) 12-40.

Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual review of psychology*, 54(1), 547-577.

Palese, B., & Piccoli, G. (2016). Online Reviews as a Measure of Service Quality. In *2016 Pre-ICIS SIGDSA/IFIP WG8. 3 Symposium, Dublin 2016*.

Pan, B. & Crotts, J.C. (2012). Theoretical models of social media, marketing implications, and future research directions. In: Sigala, M., Christou, E. & Gretzel, U., eds., 2012. *Social media in travel, tourism and hospitality – Theory, practice and cases*. Surrey, England: Ashgate Publishing Limited. pp. 73-85

Pang, B. & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends® in Information Retrieval*, *2*(1–2), 1-135.

Pang, B., Lee, L., & Vaithyanathan, S. (2002, July). Thumbs up?: sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10* (pp. 79-86). Association for Computational Linguistics.

Park S. & Nicolau J.L. (2015). Asymmetric effects of online consumer reviews, *Annals of Tourism Research*, 50: 67-83.

Park, J. W., Robertson, R., & Wu, C. L. (2004). The effect of airline service quality on passengers' behavioural intentions: a Korean case study. *Journal of Air Transport Management*, *10*(6), 435-439.

Park, Y.A., Gretzel, U. & Sirakaya-Turk, E. (2007). Measuring web site quality for online travel agencies. *Journal of travel & tourism marketing*, 23(1), pp. 15-30.

Petter, S., DeLone, W., & McLean, E. R. (2012). The past, present, and future of" IS Success". *Journal of the Association for Information Systems*, *13*(5), 341.

Pournarakis, D. E., Sotiropoulos, D. N., & Giaglis, G. M. (2017). A computational model for mining consumer perceptions in social media. *Decision Support Systems*, 93(1), pp. 98-110.

Plutchik, R. (1962). The emotions: Facts, theories, and a new model. Random House.

Plutchik, R. (1980). A general psychoevolutionary theory of emotion. In Plutchik, R. and Kellerman, H. (Eds.), Emotion: Theory, Research, and Experience, Volume 1, pp. 3–33. Academic Press.

Preoţiuc-Pietro, D., Schwartz, H. A., Park, G., Eichstaedt, J., Kern, M., Ungar, L., & Shulman, E. (2016). Modelling valence and arousal in facebook posts. In *Proceedings of the 7th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis* (pp. 9-15).

Prince, J. T., & Simon, D. H. (2009). Multimarket contact and service quality: Evidence from on-time performance in the US airline industry. *Academy of Management Journal*, *52*(2), 336-354.

Ramanathan, R. & Karpuzcu, H. (2011), Comparing Perceived and Expected Service Using an AHP Model: an Application to Measure Service Quality of a Company Engaged in Pharmaceutical Distribution, Opsearch: Indian Journal of Operational Research, 48(2), 136-152.

Ranco, G., Aleksovski, D., Caldarelli, G., Grčar, M., & Mozetič, I. (2015). The effects of Twitter sentiment on stock price returns. *PloS one*, *10*(9), e0138441.

Robinson, M. D., Storbeck, J., Meier, B. P., & Kirkeby, B. S. (2004). Watch out! That could be dangerous: Valence-arousal interactions in evaluative processing. *Personality and Social Psychology Bulletin*, *30*(11), 1472-1484.

Russell, J. A. (1980). A circumplex model of affect. *Journal of personality and social psychology*, 39(6), 1161–1178.

Rust, R. T. & Oliver, R. L. (1994), Service Quality: New Directions in Theory and Practice, Thousand Oaks, Sage Publications, CA, 1-19.

Saeedpoor, M., Vafadarnikjoo, A., Mobin, M., & Rastegari, A. (2015, January). A servqual model approach integrated with fuzzy AHP and fuzzy topsis methodologies to rank life insurance firms. In *Proceedings of the International Annual Conference of the American Society for Engineering Management.* (p. 1). American Society for Engineering Management (ASEM).

Sasser, W. E., Olsen, R. P., & Wyckoff, D. D. (1978). *Management of service operations: Text, cases, and readings*. Allyn & Bacon.

Sriram, B., Fuhry, D., Demir, E., Ferhatosmanoglu, H., & Demirbas, M. (2010). *Short text classification in twitter to improve information filtering.* Paper presented at the Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval.

Sum Chau, V., & Kao, Y. Y. (2009). Bridge over troubled water or long and winding road? Gap-5 in airline service quality performance measures. *Managing Service Quality: An International Journal*, *19*(1), 106-134.

Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-based methods for sentiment analysis. *Computational linguistics*, *37*(2), 267-307.

Taylor, S. A., & Baker, T. L. (1994). An assessment of the relationship between service quality and customer satisfaction in the formation of consumers' purchase intentions. *Journal of retailing*, *70*(2), 163-178.

Thaichon, P., Lobo, A., Prentice, C., & Quach, T. N. (2014). The development of service quality dimensions for internet service providers: Retaining customers of different usage patterns. *Journal of Retailing and Consumer Services*, 21(6), 1047-1058.

Tsaur, S. H., Chang, T. Y., & Yen, C. H. (2002). The evaluation of airline service quality by fuzzy MCDM. *Tourism management*, *23*(2), 107-115.

Waguespack, B. P., & Rhoades, D. L. (2014). Twenty five years of measuring airline service quality or why is airline service quality only good when times are bad?. *Research in Transportation Business & Management*, *10*, 33-39.

Wahba, G. (2017). Measuring the effect of Viral Negative Sentiment on Market Value: Case Study on United Airlines Crisis 2017.

Wong, L., Palaniappan, S., & Hong, J. L. (2017). Social Media Analytics Towards Understanding Mimetic Behavior in Crisis Situations. *Advanced Science and Technology Letters*, 146, 102-113. http://dx.doi.org/10.14257/astl.2017.146.19

Xiang, Z., Schwartz, Z., Gerdes Jr, J. H., & Uysal, M. (2015). What can big data and text analytics tell us about hotel guest experience and satisfaction?. *International Journal of Hospitality Management*, *44*, 120-130.

Yang,Z. & Fang, X. (2004). Online service quality dimensions and their relationships with satisfaction: A content analysis of customer reviews of securities brokerage services. *International Journal of Service Industry Management*, Vol. 15 No. 3, pp. 302-326.

Young, C., Cunningham, L., & Lee, M. (1994). Assessing service quality as an effective management tool: the case of the airline industry. *Journal of Marketing Theory and Practice*, *2*(2), 76-97.

Zeithaml, V.A., Parasuraman A. & Malhotra, A. (2000). *A conceptual framework for understanding e-service quality: implications for future research and managerial practice.* Working paper, Report Summary No. 00-115, Marketing Science Institute, Cambridge, MA.

Zikopoulos, P., & Eaton, C. (2011). *Understanding big data: Analytics for enterprise class hadoop and streaming data*. McGraw-Hill Osborne Media.

Zikopoulos, P., Parasuraman, K., Deutsch, T., Giles, J., & Corrigan, D. (2012). *Harness the Power of Big Data the IBM Big Data Platform*. McGraw Hill Professional.

VITA

Xin (Shirley) Tian

Dept. of Information Technology and Decision Science, Old Dominion University, Norfolk, VA, 23529

EDUCATION

Old Dominion University Norfolk, VA
Ph.D., Information Technology May 2018

- Research Interest: Social Media Analytics, Cyber Security, Information System, Database Management, Business Intelligence, Business Analytics

Old Dominion University Norfolk, VA
M.S., Computer Science August 2011

Beijing Forestry University Beijing, China
Bachelor of Management, Information Management and Information System July 2008

SCHOLARSHIP & HONORS

Outstanding Doctoral Student in Information Technology, Strome College of Business, Old Dominion University, 2017

Productive Member of International Review Board for service to the Informing Science Institute and the I^{n}SITE 2016 conference.

The 4th National Women in Cybersecurity conference student travel award (NSF Award# 1303441) held on March 31- April 1, 2017, Tucson, AZ

IEEE Security Privacy 2017 student travel award, San Jose, CA

GREPSEC III 2017 workshop student travel award, San Jose, CA

Conference on Cybersecurity Education, Research and Practice's NSF Sponsored Graduate Student Workshop. NSF Funded $1500 Scholarship. Oct.29-Oct.30 2016 Kennesaw, GA

CIS Student Service Award, Graduate Information Technology Student Society, Old Dominion University (2011)

JOURNAL PUBLICATIONS

He, W., **Tian, X.,** Tao, R., Zhang, W., Yan, G., Akula, V. (2017). Application of Social Media Analytics: A Case of Analyzing Online Hotel Reviews. *Online Information Review.*

Dai, M., He, W., **Tian, X.**, Giraldi, A., Gu, F. (2017). Working with Communities on Social Media: Varieties in the Use of Facebook and Twitter by Local Police. *Online Information Review.*

He, W., & **Tian, X.** (2017). A Longitudinal Study of User Queries and Browsing Requests in a Case-Based Reasoning Retrieval System. *Journal of the Association for Information Science and Technology (JAIST)*. [A* Journal according to Australian Business Deans Council Journal Quality List]

He, W., **Tian, X**., Hung, A., Akula, V., & Zhang, W. (2017) Measuring and comparing service quality metrics through social media analytics: a case study. *Information Systems and e-Business Management*, 1-22.

He, W., **Tian, X**., Chen, Y., & Chong, D. (2016). Actionable Social Media Competitive Analytics for Understanding Customer Experiences. *Journal of Computer Information Systems*, *56*(2), 145-155.

He, W., Shen, J., **Tian, X**., Li, Y., Akula, V., Yan, G., & Tao, R. (2015). Gaining competitive intelligence from social media data: Evidence from two largest retail chains in the world. *Industrial Management & Data Systems*, *115*(9), 1622-1636.

CONFERENCE PROCEEDINGS

Jiang, P., **Tian, X**., Xin, C. & He, W. (2017). Teaching Hands-On Cyber Defense Labs to Middle School and High School Students: Our Experience from GenCyber Camps. In J. Johnston (Ed.), *Proceedings of EdMedia 2017* (pp. 640-644). Washington, DC: Association for the Advancement of Computing in Education (AACE). Retrieved March 3, 2018 from https://www.learntechlib.org/p/178491/.

Tao, R., Luo, Y., Liu, G., Shi, Y., & **Tian, X**. (2017, August). A GRADE Requirements Engineering Management Model Based on the Co-creation of Value. In *International Workshop on Learning Technology for Education in Cloud* (pp. 59-71). Beijing, China. Springer, Cham.

He, W., Yan, G., Shen, J. & **Tian, X** (2017). Developing a Workflow Approach for Mining Online Social Media Data. IEEE International Conference on Cloud and Big Data Computing, Silicon Valley, CA, August 4 to 8, 2017.

Tian, X. (2016) Investigating Cyberbullying in Social Media: The case of Twitter. Conference on Cybersecurity Education, Research and Practice (CCERP).

Tian, X., He, W., Tao, R., & Akula, V. (2016). Mining Online Hotel Reviews: A Case Study from Hotels in China. In *AMCIS (2016).*

He, W., **Tian, X**., & Anwar, M. (2016) Developing and Using Evidence-based E-learning Videos for Cybersecurity Education. Conference on Cybersecurity Education, Research and Practice (CCERP).

He, W., Yang, J., **Tian, X**. (2016) Comparing Social Media Use by Small Businesses in China and the US: A Case Study. In *GITMA (2016).*

He, W., **Tian, X**., & Shen, J. (2015). Examining Security Risks of Mobile Banking Applications through Blog Mining. In *MAICS* (pp. 103-108).

Tian, X., & Schwartz, S. (2015). The Use of Analytical Platform to Identify Valuable Interventions in Retail Pharmacies. In *SAIS (2015).*

He, W., **Tian, X**., Shen, J., & Li, Y. (2015). Understanding Mobile Banking Applications' Security risks through Blog Mining and the Workflow Technology. In *ICIS (2015).*

Tian, X., & Bai C. (2014). The Impact of Information Overload on Decisional Makings of MNEs: A Culture Perspective. *International Conference on Technology and Industrial Management (2014).*

He, W., Yuan, X., & **Tian, X**. (2014, August). The self-efficacy variable in behavioral information security research. In *Enterprise Systems Conference (ES), 2014* (pp. 28-32). IEEE.

Funded Grant PROJECTS

- Improving Security Behavior of Employees in Cyberspace through Evidence based Malware Reports and E-Learning Materials. (NSF funded) Research Assistant 2013-2017
- Enhancing Cyber Security Education Using POGIL. (NSF funded) Research Assistant 2017- present
- He, W., Yan, G., Shen, J. & Tian, Xin (2017). Developing a Workflow Approach for Mining Online Social Media Data. IEEE International Conference on Cloud and Big Data Computing, Silicon Valley, CA, August 4 to 8, 2017.
- He, Wu (PI) & Xin, Chunsheng (Co-PI), "Cybersecurity + Computer Programming Summer Camp", funded by NASA Virginia Space Grant Consortium, $11,000. (summer 2017).
- He, Wu (PI); Watson, Silvana; Major, Debra; Pribesh, Shana, & Xu, Li, "Investigating the Effectiveness of Pair Programming for Students with Learning Disabilities", funded by the U.S. National Science Foundation (NSF). September 2017 to September 2020. (amount: $299,999)

TEACHING EXPERIENCE Teaching Evaluation 4.22/5.00 (5 is the highest score)

- IT 150G Basic Information Literacy and Research (primary instructor)

Spring 2017, Fall 2017
- IT 325 Website and Webpage Design (teaching assistant)
- IT 360T Principles of Information Technology (teaching assistant & lab instructor)
- IT 430 Objected-Oriented Programming with Java (teaching assistant)
- IT 473 Systems Design/Implementation (teaching assistant)
- IT 360T Principles of Information Technology (primary instructor) Spring 2018
- IT 325 Website and Webpage Design (primary instructor) Spring 2018

SERVICE & OUTREACH ACTIVITY

- Coordinator in 2016 GenCyber summer camp (NSA & NSF sponsored) http://securitybehavior.com/gencyber-camp/
- Coordinator in 2017 Cybersecurity and programming summer camp (NASA sponsored) http://securitybehavior.com/summercamp2017/
- Instructor and co-advisor for cybersecurity after school club at Tallwood High School, Virginia Beach, VA (as part of GenCyber program)
- Website design and maintenance for The Hampton Roads Cybersecurity Education, Workforce and Economic Development Alliance (HRCyber) http://securitybehavior.com/hrcyber/
- Journal Reviewer, Information Discovery and Delivery
- Journal Reviewer, InformingSciJ
- Journal Reviewer, Journal of the Brazilian Computer Society
- Journal Reviewer, Interdisciplinary Journal of e-Skills and Lifelong Learning (IJELL)
- Conference Reviewer, Americas Conference on Information Systems (AMCIS) 2015, 2016, 2017
- Conference Reviewer, Southern Association for Information Systems (SAIS) 2015, 2016, 2017
- Conference Reviewer, WMSCI 2017
- Conference Reviewer, EdMedia 2017
- Conference Reviewer, SIGCSE 2017
- Conference Reviewer, InSITE 2015, 2016, 2017
- Conference Reviewer, Conference on Cybersecurity Education, Research and Practice (CCERP) 2016, 2017
- Conference Reviewer, 2016 International Joint Conference on Neural Networks (IJCNN)

PROFESSIONAL AFFILIATIONS

- Association of Information Systems
- Decision Sciences Institute

PROFESSIONAL SKILLS

- Professional Certification: Oracle Database 11*g* Administrator Certified Associate (OCA)
- Operating System： windows 7, windows 10, Linux, MAC
- Software: Microsoft Office (with advanced Word, Excel, PowerPoint, ACCESS, VISIO, Project, etc.) Adobe Dreamweaver, Adobe Fireworks, Adobe Flash , Microsoft Visual Studio, Adobe Photoshop CS4, Adobe ImageReady CS4, Adobe Premiere Pro, Adobe Illustrator, Adobe Captivate, CoolEdit Pro, Microsoft SQL Server, Camstasia, Blackboard, Joomla, Wamp, MYSQL, Postrge SQL, Oracle 11g, NVIVO, Audacity, Respondus 4.0/StudyMate Author
- Languages: PHP, ASP, Visual Basic. NET, ASP.NET, Javascript, SQL, JQuery，JAVA, Ruby, R, Python