Analysing Customer Engagement of Turkish Airlines Using Big Social Data

Fie Sternberg[1], Kasper Hedegaard Pedersen[1], Niklas Klve Ryelund[1], Raghava Rao Mukkamala[1,2], Ravi Vatrapu[1,2]

[1]*Centre for Business Data Analytics, Copenhagen Business School, Denmark*
[2]*Westerdals Oslo School of Arts, Comm & Tech, Norway*
{*rrm.digi, rv.digi*}*@cbs.dk*

Abstract

Companies started taking advantage of the unlocked potential of Big Social Data, however, research on airlines' use of social media is limited. This research aims to investigate to what extent Turkish Airlines can utilize their Facebook page to improve performance metrics. This study will exploit the concepts of Big Social Data, customer satisfaction, sentiment analysis to answer the research questions by employing data- and text mining, machine learning. The results showed a weak relationship between the business data and Facebook data, however, the findings provided explanations to customer behavior and showed that most of the company's Facebook users were likely to purchase a Turkish Airline ticket. Therefore, Turkish Airlines could utilize their Facebook page in the short-term to improve revenue-generating indicators such as customer satisfaction and likelihood of purchase.

Keywords-**Big Social Data, Facebook Data, text analytics**

I. INTRODUCTION

"Although we often hear that data speak for themselves, their voices can be soft and sly" [1]. Big data has become an important driver for businesses across industries. Therefore, companies need a strategic plan for collecting, organizing and using data to create value for their customers and improve their business. Data production is increasing at an exponential rate, especially posts and comments on social media sites. Big data pledges to help companies understand the purchasing behavior of the customers and use this understanding to create more efficient marketing [2]. Although the implementation of big data cannot guarantee sustainable competitive advantage in terms of understanding consumer behavior, it might improve short-term revenue. To build a lasting advantage, companies need to use big data to understand what will foster customer loyalty [2].

One industry that could benefit from the advantages of Big Data Analytics is the airline industry. The airline industry is interesting because of its importance to the global economy, international presence and fierce competitive environment. Due to globalization and the easy access to air transport most people in the developed world have a relation to the airline industry, and the industry experiences growth in demand. However, the intense competition has led to small profit margins, which results in constant overbookings as an attempt to improve revenues [3]. Airlines are highly affected by perishability; their service capacity, cannot be stored for sale in the future, and therefore management are forced to yield the capacity as a lost passenger seat is a loss of money forever. Thus, airlines seek to fill up each aircraft to reach the highest possible revenue [3]. Airlines are therefore looking for sources of competitive advantage that will make them stand out from competition; enter big data analytics.

An interesting airline to examine is Turkish Airlines, which is the national flag carrier airline of Turkey with a total fleet of more than 300 aircrafts[1], and flies to more countries than any other airline in the world with 120 countries. Up until now, Turkish Airlines seems to have been performing well, and they have been in constant growth from 2013-2016 with the number of passengers carried increasing from 48.3 million to 62.8 million[2]. The airline aims to achieve its mission statement "to become the preferred leading European air carrier with a global network of coverage"[1], an achievement that could be reached by optimizing their utilization of big data and exploit the predictive advantages that social media has proven to serve for other companies.

Information on social media are starting to reveal a significant portion of human social life, and emotional content of social media streams is being used to estimate economic quantities in country-wide economies, and predict stock market movements [4]. Social media actions such as liking and commenting are proxies for consumer's attention to a particular product and the shared digital artefact can create social influence [5]. This paper will utilize the content of Turkish Airlines Facebook page, and investigate whether business data such as passenger numbers, revenue and stock prices can be estimated by looking at this content. This will be done by examining the size, activity and sentiment of Turkish Airlines' Facebook page; more specifically, the likelihood of ticket purchase. For this purpose, the subsequent research question has been devised:

To what extent can Facebook data be utilized to improve the performance metrics of Turkish Airlines?

In order to answer the research question, the following sub questions will be investigated:

---

[1]https://www.turkishairlines.com/
[2]http://investor.turkishairlines.com/en

IEEE
computer
society

1) How are Turkish Airlines currently using their Facebook page?
2) To what degree are the active users on Turkish Airlines' Facebook page 'likely' to purchase a ticket?
3) Is it possible to identify a correlation between Facebook data and revenue, stock prices, and passenger numbers?

In section II, we present conceptual framework and related work. Methodology and data analysis results will be presented in sections III and IV respectively. Finally we will conclude in section V.

II. Conceptual Framework and Related Work

A. Big Social Data

Big data can reveal fascinating insights into various fields, and allow us to observe social phenomena at another level [6]. According to conceptual model of Social data [5], [7], the Social data consists of two types: Social Graph and Social Text. Social graph consists of the structure of the relationships emerging from the appropriation of social media affordances such as posting, linking, tagging, sharing, liking etc. It focuses on identifying the actors involved, the actions they take, the activities they undertake, and the artifacts they create and interact with. Social text consists of the communicative and linguistic aspects of the social media interaction such as the topics discussed, keywords mentioned, pronouns used and sentiments expressed. Big social data is generated by social media engagement, and it is key to crucial insights into human behavior [6], [8] Social networks consisting of actors and social relations appear everywhere across social science disciplines. Social Network analysis is neither a theory nor a methodology, but rather a perspective or paradigm [9]. It is rooted in the belief that social life is created primarily by relations and the patterns they form [9]. Social media data possess unique characteristics; besides rich textual content, explicit information about the originating social context is generally available. Information often includes timestamps, emotions and interpersonal ties [4]. Information on social media are starting to reveal a significant portion of human social life, and the emotional content of social media streams is being used to estimate economic quantities in country-wide economies, and predict stock market movements [4]. This paper will utilize the content of Turkish Airlines Facebook page in the same way, and investigate whether business data such as passenger numbers, revenue and stock prices can be estimated by looking at Facebook content.

B. Customer Satisfaction

According to [10], the most significant factor for success measures is the customer satisfaction of the service quality of a specific airline. Customer satisfaction is a vital goal for businesses as this measure will greatly improve the likelihood of future purchases [11]. Reichheld and Sasser [12] state that customer satisfaction forms loyalty

to a company. If loyalty is formed, customers are inclined to re-buy products despite attractive offers on competing products, endorse the product line to others, as well as offer constructive and truthful feedback back to the firm [12]. It is important to note that satisfaction may not result in repurchases and long-term relationships; there is a distinction between transaction-specific and cumulative customer satisfaction [13], where cumulative satisfaction refers to the overall evaluation based on the total experience with a service over time. High customer satisfaction is achieved when expectations are exceeded, and highly positive emotions are identified [13]. The ultimate goal is customer retention, which can result from overall satisfaction or highly positive emotions. Overall satisfaction occurs because of repurchases and indicates a long-term relationship between company and customer, but not necessarily highly positive emotions for each other [13]. Thus, retention may be the consequence of lasting relationships without emotional bonds or emotional bonds without a long-term connection [13]. All in all, social media presents opportunities to help form close relationships with customers; social media has an interactive nature which entails the ability to establish communication between sellers and customers and involves customers in content generation and value creation hereby resulting in firms being better able to satisfy customer needs [13].

C. Sentiment Analysis

Sentiment analysis or opinion mining is the computational study of people's opinions, appraisals, attitudes, and emotions toward entities, individuals, issues, events, topics and their attributes [14]. Sentiment analysis is based on natural language processing (NLP), which tries to produce meaningful insights by mining text data [15]. Based on NLP different sentiment analysis techniques can be applied, an example includes the dictionary based feelings meter: negative, neutral, positive [15]. Topics of sentiment analysis can take different variations, this project focuses on likelihood of purchase in order to track public viewpoints of Turkish Airlines and examine whether such viewpoints can perform trend prediction in revenue streams, passenger numbers and stock prices. Facebook pages often contain huge amounts of opinionated text that is not always easy to comprehend. The average human reader will have difficulty accurately summarizing information and opinions contained on the Facebook page, and furthermore, it is known that human analysis of text information is subject to considerable biases, e.g., people often pay more attention to opinions that are consistent with their own preferences [14], [16]. A fundamental technology in sentiment-analysis and opinion-mining applications is text classification. Opinion mining from text data can reveal and summarize opinions about a topic from consumers for optimizing decisions and business intelligence and reveal the consumers subjective feelings and thoughts [14], [16].

D. Research related to Airline Industry

Research on the use of data in aviation is limited, but it is being used as proved by Larsen [17]. The industry makes use of primary data sets that come from many different parameters such as flight tracking data, airport operations data, weather conditions, airline information, market information, passenger information, aircraft data and air safety reports. Challenges to the data sources in the aviation industry are size, scale and formats. Data is used to optimize operations and scheduling [17]. McAfee and Brynjolfsson [18] describe how profitable it can be for an airline to switch to the use of big data in predicting estimated arrival times and Belcastro et. al. [19] describe how big data such as weather conditions and flight information can more accurately predict arrival times and how flight information data can improve flight scheduling. According to Kasturi et. al. [20] airlines should take more advantage of the substantial amounts of data collected by the operational parts of the airplane such as the engine to improve efficiencies and reduce operational costs. Data on social media is a valuable tool that should be used as parameter for optimization and increase customer satisfaction as proved by [21]. Social media is according to Hvass and Munar [21] a valuable tool that can be used to interact directly to consumers and monitor customer opinions and evaluations. They describe how social media is underutilized by the airline industry in both interacting and monitoring activity. Some airlines were depicted to mainly post negative posts on weather delays, while other airlines mostly spread positive information like there were no daily delay system wide [21]. Furthermore, Hvass and Munar [21] show how airlines have different strategies on the social media platforms, where some airlines appeared to respond to most posts made by consumers. This would have a positive effect on engagement of the site.

With the advancement of the Internet, purchasing airline tickets online is now the norm instead of going to the local travel agency. This technology has led to the consumer benefits of convenience, price reductions and product diversity [22]. As a contribution to explaining consumer behavior towards purchasing flight tickets the study by [23] discusses different drivers of online airline ticket purchasing. The study uses data from questionnaires, collected in Spain, to conclude that online purchase intention depends on individual habits of using the website, price savings, performance expectancy by the consumer in completing the online transaction, and facilitating conditions [23]. Finally, it was found that "the greater the perceived online purchase intention, the greater the chance of online purchase use" [23]. Research in [24] employed Twitter data in a sentiment analysis to investigate the possibilities to use textual analysis of social media to identify, monitor, and analyze customer experiences for the purpose of optimizing service performance. Airline customers share both their positive and negative experiences

with the airline, and often directly addressing the customer service with their problems [24]. The positive experiences were mostly found to be related to online check-in services, cheap prices, and flight experiences (Misopoulos et al., 2014). On the other hand, the negative experiences were mostly related to usability of the website, delays, and lost luggage [24]. Another study [25] used Twitter data from several commercial airlines to evaluate customers' opinion of the service provided and it suggested a generally negative consumer sentiment towards the airlines [25]. However social media data can be used effectively to identify consumers' preferences, to identify dissatisfaction with the services and to correct unintended errors [25].

III. RESEARCH METHODOLOGY

A. Dataset Description and Collection

The data consists of stock prices, revenue, passenger numbers, and likelihood of purchase model/based on social media activity and two type of datasets: business and Facebook were used. The business data consists of Turkish Airlines' stock price, quarterly revenue, and monthly passengers carried. The daily stock price between June 2016 and November 2017 was collected from Yahoo Finance[3] as Turkish Airlines is a publicly traded company. The revenue is reported on a quarterly basis and can be found in the quarterly reports. Fortunately for the paper, the number of passengers carried is reported monthly, resulting in more reliable data[4]. The raw Facebook data consisted of 5,488,066 data points, where each row is equivalent to an action on the Facebook page. The raw data on the Facebook page is unstructured data; the term unstructured does not mean that the data are without pattern, but merely that they are expressed in a way that only humans can interpret meaningfully [26]. The Facebook data is fetched by SODATO (social data analytics tool) [27] and the data extracted from the Facebook page is both numeric and textual. The textual data includes posts, comments and comment replies that will be analyzed in this project. The processing methodologies used in this project include textual, predictive and visual analytics. The extraction process pertains monitoring, modeling and comparing patterns in the data [9]. The data mined in this project includes social media data, particularly text data from posts, comments and comment replies. The methodological process of text mining entails classification, which aims to produce an output set of categorized information using some degree of human intervention in the analysis and is a form of supervised machine learning [9]. As part of text mining, this research paper has applied two different techniques: text classification by customized model and keyword analysis.

[3]https://finance.yahoo.com/
[4]https://investor.turkishairlines.com/

B. Textual Data Processing using Mutato

This project applied supervised machine learning to classify the posts and comments on the Facebook page. Naïve Bayes Classifier using machine learning approach were utilized with the help of our custom in-house tool Multi-dimensional Text Analytics Tool (Mutato) [16], on Facebook text; e.g. posts, comments and comment replies. The comments were mostly written in English; however, some Turkish and French posts were also present, which were excluded in data pre-processing phase. A domain specific model for text classification with five labels indicating the likelihood of purchase by the customer was developed. The Naïve Bayes classifier is a generative classifier that works on the probabilistic measure of likelihood of word features that are occurring in the given text documents pertaining to the class labels. We prepared a training set containing 4,000 text pieces that were manually trained and labeled by the authors, based on predefined categories constructed on a Likert scale on *likelihood of purchase*: as shown in Tab. I and used the training set to train the Naïve Bayes classifier. The *likelihood of purchase* classifier was built based on the premise of the concept of customer satisfaction. Theory states that high customer satisfaction is achieved when expectations are exceeded, and positive emotions are identified [13], furthermore, it has been found that customer satisfaction could potentially improve likelihood of future purchases [11]. Thus, the likelihood of purchase model was created on the assumption that a positive attitude in the text piece was equivalent to a satisfied customer and therefore likelihood of purchase.

Text Classification Model: Likelihood of purchase	
Label	**Definition**
Very likely	If the text piece expressed an extreme liking towards Turkish Airlines. E.g. "I will always fly with you guys... One of my best airlines"
Likely	If the text piece expressed positive sentiment towards Turkish Airlines it was categorized as likely of purchasing. E.g. "TURKISH AIRLINES IS VERY GOOD"
Unlikely	If the text piece expressed a negative sentiment towards Turkish Airlines, it was categorized as unlikely of purchasing.. E.g. "How can we cancel and get a refund?"
Very Un-likely	If the text piece expressed extreme negativity towards Turkish Airlines. E.g. "This has been a nightmare travel experience"
Neutral	Irrelevant for this topic.

Table I
TEXT CLASSIFICATION MODEL: LIKELIHOOD OF PURCHASE

Mutato was also used to conduct a keyword analysis of the most occurring words on Turkish Airlines' Facebook page. Keyword analysis is a method of automatically identifying the frequent occurring words from a text corpus [28]. The keywords were compared to the likelihood of purchase model, to see if there was an alignment. Mutato was used for the data transformation and a Wordle word cloud was generated for visual representation of the text data by using weighted frequency of words. The analysis enabled us to get an overview of the general feeling towards Turkish Airlines.

C. Predictive Analytics

Simple predictive analytics were performed with the use of p-value and R-squared, and scatter plots were chosen to visualize it to show how much one variable was affected by the other. It was tested whether the null hypothesis could be rejected for different variables. Ideally in regression analysis one would like to have their regression model to have significant variables and produce a high R-squared value, as this combination indicates that changes in the predictors (independent variables) are related to changes in the dependent variables [29]. Yet the interpretations of the significant variables are the same for both high and low R-squared models and a significant p-value indicates that the null hypothesis can be rejected; however, low R-squared values are problematic when precise predictions are needed [29]. Often with data analysis it is assumed observations are registered independent from each other, but in time series this assumption is often false, and time series analysis would have helped to account for the correlation in the analysis [30]. However, this paper only dealt with simple predictive analysis without accounting for time series data.

D. Visual Analytics

After the datasets had been transformed in Mutato and Excel, Tableau was utilized for visualization. Visual analytics was applied in this project as an attempt to envision to what extent the Facebook page of Turkish Airlines can be used for evaluation purposes and decision-making processes. A major challenge of visual analytics is to find the most useful way to match data transformations with interactive visual representations and analytical reasoning techniques [26]. The structure of the Facebook data contained 5,488,066 data points and is incomprehensible without proper visualization. The data was transformed into lower-dimensional representations for better visualization and analysis. By using Tableau, key features in the data were illuminated to obtain meaningful facts from the data and produce relevant visual representations. The data representation method must facilitate the analytical reasoning methods and capture the intermediate and final results of the reasoning [26]. The data visualizations used scatter plots to represent the relationship among multiple variables. These matrixes provided valuable insights into the data biases and the quality challenges.

IV. RESULTS

A. Text Analytics Results

First of all, the majority of text on Turkish Airlines' Facebook page is categorized as "likely" to purchase. Posts and comments can reveal perceptions of a brand, in this case perception shows whether an actor would be likely or unlikely to purchase a ticket. An interesting finding was that most of the actors who commented or posted on the Facebook page were considered 'likely' purchasers as shown in Fig. 1. The model can be said to be more of

a customer satisfaction model measuring the likelihood of future purchases and retainment of customers and customer loyalty.
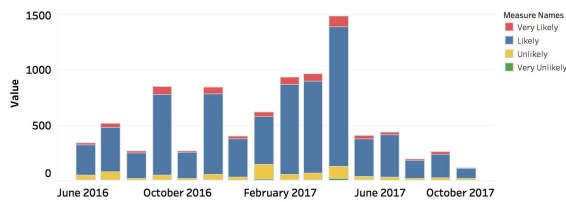

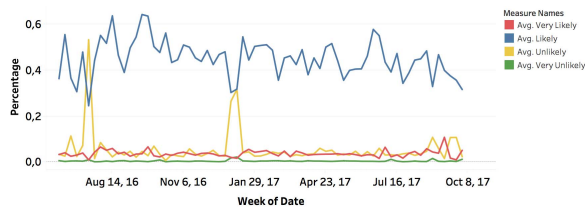Figure 1.   Distribution of Likelihood of purchase model

Figure 2.   Average percentage distribution of likelihood of purchase

The keyword analysis of the entire text corpus from the Turkish Airlines Facebook page, showed an overall positive sentiment. In case of *likely* and *most likely* categories, words like 'best', 'love' and 'great' were the most frequent occurring words, where as wordcloud for *unlikely* and *very unlikely* posts contained words like 'cancelled', 'time', 'refund' and 'information'. It clearly indicates that reasons for their unlikelihood of purchase could be due to cancellations, lack of information, poor customer service and unsatisfactory experiences leading to refunds.
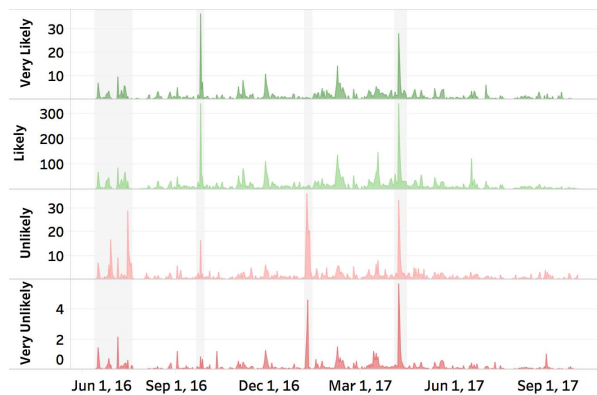

Figure 3.   Likelihood of Purchase model over temporal dimension

1) The four peaks in the likelihood data can be matched with special events: Figure 3 highlights four periods in the Facebook data showing interesting peaks at certain times. As shown in figure 3, the four likelihood labels contain different scale of counts (noted on the Y-axis). However, by visualizing the highlights of Facebook likelihood in the time span of 1.5 years, we can provide an understanding of the customers' perceptions and feelings towards various

aspects of the business and/or unique events. The four highlights shown in figure 3 shows how the Turkish Airlines's Facebook page is used by their customers. The first highlight starts in mid-June and ends in mid-July 2016, showing an increase in the *unlikely* to purchase category. The highest peak in the *unlikely* category happened at the same time as 300 people were killed, and more than 2,000 people injured in Turkey as a result of an attempted coup to overthrow the government on July 15, 2016[5]. More than 30 Turkish Airlines flights were cancelled because of the closure of Atatrk Airport in Istanbul the following two days after the coup attempt[6]. This makes sense and it can be stated that cancellations might possibly have caused some negative responses on Turkish Airlines' Facebook page, this relates well to the wordcloud created on the basis of unlikely and very unlikely labels. The second highlight is caused by activities unrelated to Turkish Airlines' core business. The peak in *likely* and *very likely* is related to a status posted on September 27, 2016 regarding the 2016 Turkish Airlines Euro League Basketball Championship Game Press Conference. One of the teams playing in the tournament was the Turkish team Fenerbahce Istanbul. Turkish Facebook users may have caused an increase in the greater likelihood of purchase around this date. The post is unrelated to the core business and therefore may not affect financial results. On January 16, 2017, a Turkish Airlines operated cargo flight crashed into a residential area in Kyrgyzstan upon attempted landing. The tragic incident killed the four crew members and around 34 people on the ground, which caused the third highlight. On Facebook, Turkish Airlines referred to the plane crash as 'an ACT Airlines cargo aircraft', hence taking distance to their responsibility. Nevertheless, the aircraft was operated under a Turkish Airlines flight number and the Facebook comments that day are almost only in the *unlikely* and *very unlikely* categories. The fourth highlight happened on April 7, 2017, when a baby was unexpectedly born onboard a Turkish Airlines flight. This event lead to thousands of positive responses on Facebook, but also a fair share of complaints to the airline about letting the pregnant woman enter the plane in the first place. These highlights show that most comments on Turkish Airlines' Facebook page are not even related to the core business of transporting passengers.

B. Predictive Analytics Results

We have tried to predict various parameters from business data such as revenues, stock prices and passengers carried over the period of the time using the independent variables from social data such as interactions, textual data from posts, comments etc. We found very weak correlations between various independent and dependent variables and therefore

[5]http://www.bbc.com/news/world-europe-36816045

[6]As flights resume at Istanbul airport, U.S. warns travelers away

we only report the most significant results in view of the space limitations.

1) Love reaction is explanatory to stock prices: Facebook likes were analyzed as well, with focus on the specific type of reaction of the like button. There were eight different reactions Facebook users had made use of in the data, made for each specific post. The eight different reactions registered for posts were 'like', 'love', 'sad', 'angry', 'wow', 'haha', 'thankful' and 'pride'. Looking at the different emotions one can react to a post, the standard 'like' reaction was by far the most used on Turkish Airlines' Facebook posts, with 94% of reactions being a standard 'like'. According to Molloy[7] a person reacting with other than like is an indicator that the person is feeling more strongly about that post and finds the subject of the post relevant to a person's interests. Considering this, this paper looked at the different reactions made for Turkish Airlines' posts to see if a focus on users with more feelings towards posts would have an impact on business data.
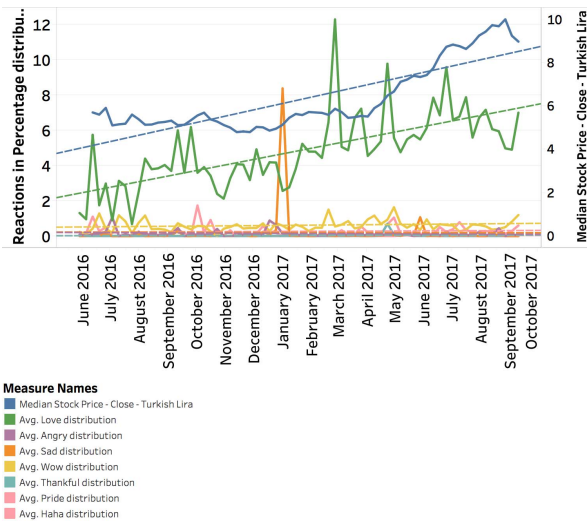

Figure 4.   Weekly distribution of like reactions with stock price

By taking the standard 'like' out of the graph and breaking the remaining seven reactions up we saw in figure 4 that the 'love' reaction was by far the most used of the reactions, with an increasing frequency over time observed. Since Facebook's introduction to the different emotions in 2016, the most used reaction is the 'love' reaction accounting for more than half of reactions other than like, matching the behavior of Turkish Airlines' Facebook users. The increasing 'love' reaction may be a result of an increase of similar posts featuring on a person's news feed. A person hitting 'love' once to a Turkish Airlines post may be inclined to hit 'love' again for future posts. Facebook has revealed that hitting other than 'like' to a post will be analyzed by Facebook as how relevant the post is to a person, and hence be used as a mean to decide what should be shown

in the news feed[7]. Thus, a person hitting 'love' to a Turkish Airlines post may be exposed to an increasing number of posts from Turkish Airlines. By comparing 'love' reaction
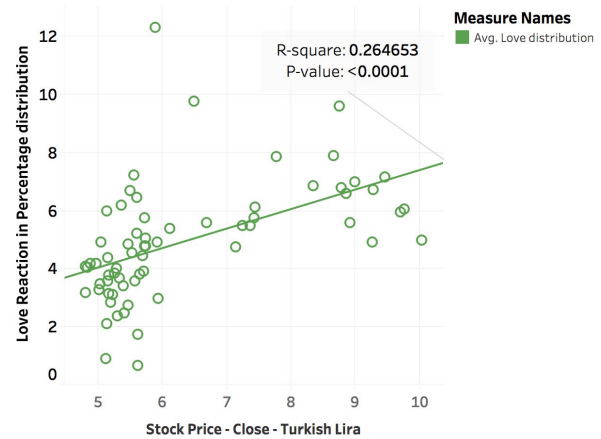

Figure 5.   Scatter plot between love reactions and stock prices

to stock prices in figure 5, we saw a fit over time with similar trend lines, both increasing approximately at the same rate. Diving deeper into the fit in figure 5, we saw that the coefficient was statistically significant, with a p-value well below 0.05. Thus, we rejected the null hypothesis and accepted that Facebook data may influence stock prices. This may stem from persons who have stronger feelings towards posts from Turkish Airlines may be inclined to purchase Turkish Airlines stocks, resulting in pushing prices upwards, or people who already own stocks may be reacting stronger to positive news from Turkish Airlines. However, this assertion is vague which can be seen in the goodness of fit, reflected in a R-square of 0.26, which again suggests that the data may suffer from omitted variable bias. All other 'like' reactions had very low R-squared and an insignificant p-value suggesting 'love' being the only somewhat reliable value explaining stock prices.

V. DISCUSSION AND CONCLUSION

The discussion will answer our research questions by using the findings described earlier.

1) How is Turkish Airlines currently using their Facebook page?: Turkish Airlines does not seem to have an established social media strategy, and their postings seem to be rather arbitrary. However, when they do post most of the reactions are positive. Furthermore, they use photos and videos and visual representations on their Facebook page, which has proven to generate the most activity, also mainly the positive kind. Another study has shown that social media is underutilized by the airline industry in interacting and monitoring activity of their customers [21], this also applies to Turkish Airlines as they do not seem to have a clear social media strategy for customer interaction, as their

[7]Why you should 'love' instead of 'like' the Facebook posts that really matter to you

engagement seems arbitrary. On the other hand, airlines are perceived to utilize their platforms for informative purposes on relevant circumstances such as delays, etc. [21]. This is also the case for Turkish Airlines and they become very active when customers start complaining about flight delays. Furthermore, airlines have different strategies on social media when it comes to responding to customers. Interaction with customers on a Facebook page increases the activity on the site and has a positive effect on comments and future customer posts [21]. Turkish Airlines engage with customers and reply to comments; however, they also encourage customers to send them private messages when the questions become of a negative nature; resulting in the Facebook page having a more positive sentiment than negative.

2) To what degree are the active users on Turkish Airlines' Facebook page likely to purchase a ticket?: Most of posts and comments on Turkish Airlines' Facebook page indicate a likelihood of purchase. This correlates well with the overall positive attitude on the page found in the keyword analysis, and the 'love' reaction on the 'like' parameter being the most utilized. There were some instances of users showing an unlikelihood of purchase, most likely due to cancellations, lack of information and customer service. However, this relates more to a lack of customer retention than unlikelihood of purchase, as these users were already customers. Likelihood of purchase classifications were evaluated, in the training set, on the scale of customer satisfaction based on the positivity of the post or comment. This can be related to the ultimate goal of customer retention, which scholars proclaim results from overall satisfaction and positive emotions [13]. The likelihood of purchase model can be concluded to be a likelihood of customer retention model. Thus, the result from the data transformation shows an overall likelihood of customer retention.

3) Is it possible to identify a correlation between Facebook data and revenue, stock prices, and passenger numbers?: It is not currently possible to utilize Turkish Airlines' Facebook page to predict business data. The findings showed significance between some of the variables; likely to purchase and stock price, unlikelihood of purchase and stock price, and love reactions and stock price. The significance of likely to purchase and stock price did not make economic sense, and even though the latter two significant findings did make economic sense, their low R-squared values leads us to advise Turkish Airlines not to act on the predictions from this analysis. More data points and perhaps different variables are needed to forecast business data using Turkish Airlines Facebook data. In some cases, additional predictors can increase the explanatory power of the model, and hence the model may suffer from omitted variable bias. In other cases, the data contains a higher amount of unexplainable variability. According to veracity-based value, gathering a lot of data does not mean the data becomes clean and accurate;

data on customers' needs to remain consolidated, cleansed, consistent, and current to make the right decisions. Big data veracity refers to the noise, abnormalities and biases in data. It asks questions regarding whether the data mined is meaningful to the problem being analyzed . In the case of this analysis, it can be said that the data mined has made sense in answering the overall research question. However, this particular sub-question could not be answered in a meaningful way, and therefore it is advised to make the same analysis later on, after recommending Turkish Airlines to make better use of their Facebook page to see if the predictive capabilities of their Facebook data become better. One important variable that may explain more of the stock price movements is the cost of jet fuel as it has a major impact on the profitability of any airline by representing around 20% of average operating costs of airlines[8]. In the period of focus of this paper, from June 2016 to November 2017, fuel prices have been increasing again after a few years of constant fuel cost reduction[8]. However, the fuel prices are to some degree offset by the continuous improvement in fuel efficiency, which also leads to a decoupling of CO_2 emissions from expanding air transport services[8].

VI. CONCLUSION

This research paper attempted to visualize and explain how Turkish Airlines can utilize their Facebook data to improve their performance metrics. The findings have shown that there is no reliable relationship between Turkish Airlines Facebook data and their business data. However, Turkish Airlines' Facebook data can be used to assess customer satisfaction of the service quality, a valuable aspect, as this measure could potentially improve likelihood of future purchases. This will be possible in the short-term, and by utilizing big social data to understand customer satisfaction; Turkish Airlines' might be able to improve their sales figures. Understanding how people feel about the brand is one of the key elements for social media monitoring, and Turkish Airlines' Facebook page offers the possibility to gather such information and undertake the appropriate actions needed. Noteworthy is that the data highlights that social media can be used effectively to identify consumers' preferences, to identify dissatisfaction with the services and to correct unintended errors. It must be noted, however, that the possibility of the Facebook data as a forecaster cannot be rejected, as some of the relations were found to be significant. Nevertheless, the data is not precise enough to be utilized for this purpose in the short-term. Turkish Airlines should optimize the use of their Facebook page for building relationships with their customers, and potentially utilize it for predictive purposes in the long-term if the precisions of the predictors improve.

[8]Economic Performance of the Airline Industry

First, future studies could make a similar analysis by using a longer time period and using time lag. By completing a similar study with a longer time horizon a few years afterwards, future work might benefit from an improved social media strategy implemented by Turkish Airlines, which potentially could lead to better predictive capabilities of financial performances using Facebook data. Moreover, future studies can benefit from using a different model with new labels to train the Facebook data. A suggestion for such a new model is focusing on the different topics that the Facebook users are talking about, hence, making four to six labels of general topics that will identify what is commonly talked about in a given period of time. This suggestion is due to the finding that the likelihood of purchase model entailed a lot of neutral classification; hence exposing that another model may have been of more relevance. Finally, Twitter could be used to collect the social media data instead of Facebook, as Twitter has proven to be a better platform than Facebook for modeling stock prices due to the higher volume of data and immediacy.

REFERENCES

[1] F. Mosteller, S. E. Fienberg, and R. E. Rourke, *Beginning statistics with data analysis*. Courier Corporation, 2013.

[2] N. Dawar, "Use big data to create value for customers, not just target them," Harvard Business Review, Aug. 2016, https://hbr.org/2016/08/use-big-data-to-create-value-for-customers-not-just-target-them.

[3] P. Belobaba, A. Odoni, and C. Barnhart, *The global airline industry*. John Wiley & Sons, 2015.

[4] C. A. Davis, G. L. Ciampaglia, L. M. Aiello, K. Chung, M. D. Conover, E. Ferrara, A. Flammini, G. C. Fox, X. Gao, B. Gonçalves *et al.*, "Osome: the iuni observatory on social media," *PeerJ Computer Science*, vol. 2, p. e87, 2016.

[5] R. Vatrapu, R. R. Mukkamala, A. Hussain, and B. Flesch, "Social set analysis: A set theoretical approach to big data analytics," *IEEE Access*, vol. 4, pp. 2542–2571, 2016.

[6] Z. Tufekci, "Big questions for social media big data: Representativeness, validity and other methodological pitfalls." *ICWSM*, vol. 14, pp. 505–514, 2014.

[7] R. R. Mukkamala, A. Hussain, and R. Vatrapu, "Towards a set theoretical approach to big data analytics," in *proceedings of 3rd International Congress on Big Data (IEEE Big-Data 2014)*, June 2014, http://www.itu.dk/people/rao/pubs_accepted/2014_IEEE-BigData-socialdata-set-theory.pdf.

[8] D. Boyd and K. Crawford, "Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon," *Information, communication & society*, vol. 15, no. 5, pp. 662–679, 2012.

[9] C. Cioffi-Revilla, *Introduction to Computational Social Science: Principles and Applications*. Springer Science & Business Media, 2013.

[10] R. A. Ganiyu *et al.*, "Customer satisfaction and loyalty: A study of interrelationships and effects in nigerian domestic airline industry," *Oradea Journal of Business and Economics*, vol. 2, no. 1, pp. 7–20, 2017.

[11] V. Mittal and W. A. Kamakura, "Satisfaction, repurchase intent, and repurchase behavior: Investigating the moderating effect of customer characteristics," *Journal of marketing research*, vol. 38, no. 1, pp. 131–142, 2001.

[12] F. F. Reichheld and J. W. Sasser, "Zero defections: Quality comes to services." *Harvard business review*, vol. 68, no. 5, pp. 105–111, 1990.

[13] C. Sashi, "Customer engagement, buyer-seller relationships, and social media," *Management decision*, vol. 50, no. 2, pp. 253–272, 2012.

[14] C. C. Aggarwal and C. Zhai, *Mining text data*. Springer Science & Business Media, 2012.

[15] B. Pang and L. Lee, "Opinion mining and sentiment analysis," *Foundations and trends in information retrieval*, vol. 2, no. 1-2, pp. 1–135, 2008.

[16] J.-R. Reichert, K. L. Kristensen, R. R. Mukkamala, and R. Vatrapu, "A supervised machine learning study of online discussion forums about type-2 diabetes," in *IEEE Healthcom 17International Conference on E-health Networking, Application & Services*. IEEE, 2017, pp. 1–7.

[17] T. Larsen, "Cross-platform aviation analytics using big-data methods," in *Integrated Communications, Navigation and Surveillance Conference (ICNS), 2013*. IEEE, 2013, pp. 1–9.

[18] A. McAfee, E. Brynjolfsson, T. H. Davenport, D. Patil, and D. Barton, "Big data: the management revolution," *Harvard business review*, vol. 90, no. 10, pp. 60–68, 2012.

[19] L. Belcastro, F. Marozzo, D. Talia, and P. Trunfio, "Using scalable data mining for predicting flight delays," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 8, no. 1, p. 5, 2016.

[20] E. Kasturi, S. P. Devi, S. V. Kiran, and S. Manivannan, "Airline route profitability analysis and optimization using big data analyticson aviation data sets under heuristic techniques," *Procedia Computer Science*, vol. 87, pp. 86–92, 2016.

[21] K. A. Hvass and A. M. Munar, "The takeoff of social media in tourism," *Journal of Vacation Marketing*, vol. 18, no. 2, pp. 93–103, 2012.

[22] S. Forgas, R. Palau, J. Sánchez, and R. Huertas-García, "Online drivers and offline influences related to loyalty to airline websites," *Journal of Air Transport Management*, vol. 18, no. 1, pp. 43–46, 2012.

[23] T. Escobar-Rodríguez and E. Carvajal-Trujillo, "Online drivers of consumer purchase of website airline tickets," *Journal of Air Transport Management*, vol. 32, pp. 58–64, 2013.

[24] F. Misopoulos, M. Mitic, A. Kapoulas, and C. Karapiperis, "Uncovering customer service experiences with twitter: the case of airline industry," *Management Decision*, vol. 52, no. 4, pp. 705–723, 2014.

[25] M. M. Mostafa, "An emotional polarity analysis of consumers? airline service tweets," *Social Network Analysis and Mining*, vol. 3, no. 3, pp. 635–649, 2013.

[26] J. J. Thomas and K. A. Cook, "A visual analytics agenda," *IEEE computer graphics and applications*, vol. 26, no. 1, pp. 10–13, 2006.

[27] A. Hussain and R. Vatrapu, "Social data analytics tool (sodato)," in *DESRIST-2014 Conference (in press)*, ser. Lecture Notes in Computer Science (LNCS). Springer, 2014.

[28] A. Kinra, R. R. Mukkamala, and R. Vatrapu, *Methodological Demonstration of a Text Analytics Approach to Country Logistics System Assessments*. Cham: Springer International Publishing, 2017, ch. 11, pp. 119–129.

[29] B. B. McShane and D. Gal, "Blinding us to the obvious? the effect of statistical training on the evaluation of evidence," *Management Science*, vol. 62, no. 6, pp. 1707–1718, 2015.

[30] W. A. Woodward, H. L. Gray, and A. C. Elliott, *Applied Time Series Analysis with R*. CRC press, 2017.