Final Project Report: Analysis Based on Tweet Reviews from Airline Customers

Executive Summary

This project report tends to provide airline company leadership management team insights of what customers care about most to render improved service in the future. This project will also incorporate important methodology of text mining which mainly includes sentiment analysis and topic modeling.

While Sentiment analysis can recognize positive, neutral, and negative reviews, assisting in distinguishing consumer groups for corresponding airlines, topic modeling will extract detailed information for each category and further explain it, based on the categorical result after sentiment analysis. For the results of the analysis, customers are satisfied with the airline service of VirginAmerica, JetBlue and Southwest Airlines. They complain most about Airways, American Airlines and United Airline, which provides valuable business insights for airline companies to enhance aspects that they excel at and to improve aspects that customers are most dissatisfied with.

Project Purpose

According to the Bureau of Transportation Statistics, U.S. airlines and foreign airlines serving the U.S. carried an all-time high of 1.1 billion systemwide (domestic and international) scheduled service passengers in 2019 and this number keeps building up every year. However, the number of air travelers' complaints soared to a new high each year as well. According to Forbes, airline customers filed 102,550 grievances last year, but only 1% to 2% of overall airline complaints are sent to the government, which means that there were roughly 10 million consumer complaints about airline service in 2020. This number is astonishing and the continuing high number of complaints will have a significant negative impact on brand image even followed by dropped stock price which could further damage shareholders' benefits, if no effective actions were placed towards this situation.

On the one hand, our analysis based on tweet reviews from airline customers tends to incorporate valid text mining methods including sentiment analysis and topic modeling to help airline companies to use the index generated as a benchmark to compare with other companies, especially with competitors targeting the same group of customers. For example, for low-cost carriers who target price-sensitive air travelers such as Southwest and Spirit Airlines, they ought to pay attention to other positive comments besides "affordable" that the competitor gets more as well as negative tweets that itself gets more in order to stay competitive in the market.

On the other hand, Airlines companies may use this report to troubleshoot aspects that did not bring enough attention. With sentiment analysis and topic modeling, the leadership management team is able to Identify the aspects that customers care most during their flights and then tell whether they perform well or not in those aspects, among other companies and between different aspects within the company. This analysis will use the help of text mining to unrelease these problems hidden in the pile of data and bring them back to the spotlight in order to make further adjustments accordingly to improve the flight experience for customers.Overall, This project is intended to assist benchmarking analysis and service improvements for internal use of different airlines.

Data Description

This Kaggle dataset consists of tweet contents of 6 major U.S airlines from Crowdflower's Data for Everyone library, scraped the twitter data from February of 2015. One of the major potential problems of this dataset is sample bias because most of the words in this dataset are negative, as the data collected is mainly specific to each airline's problems. Therefore, it is likely that it cannot be beneficial when it comes to the analysis of customers' attitude because customers who have positive feedback on the airline's service may not send prasing tweets. However, since one of the most significant purposes of this project is to chase down the aspects that need to be improved, this dataset is still able to provide valuable business insights.There is another minor mistake in the dataset where the column named Delta should be JetBlue since all text is about JetBlue Airlines services.

Dataset Link: https://www.kaggle.com/code/nursah/airline-tweets-sentiment-analysis/data

Methodology and Results

Sentiment Analysis

Our first methodology is sentiment analysis, and we try two models, which are VADER Lexicon and SVM model. We have 14639 distinct customers' reviews, among which 10000 reviews are classified into a training data frame, and the rest into a testing data frame. We then split the training and testing data frames into reviews and polarity data frames respectively, and normalize reviews data frames.

First, we use VADER Lexicon to perform sentiment analysis. Since our dataset has human-created labels, we can compute the accuracy rate. Through plotting the relationship between accuracy rate of sentiment polarity prediction and threshold for VADER Scores, we find the optimal threshold as 0.724, which achieves an accuracy rate of 75.9%. (See Appendix 1.)

Second, we try to train our own sentimental classifier using SVM model, since we have human-created sentiment labels. We first vectorize the normalized training data frame, using TF-IDF approach and a mix of uni-grams and bi-grams as features. We then vectorize test data frame using features created based on training data. We then use SVM model to train our custom sentiment classifier, and label each review's sentiment. The accuracy of SVM model is 81.6%.

Since SVM model outperforms VADER Lexicon, we choose the SVM model to predict reviews sentiment for each airline company. The disadvantage of SVM model is that if the human-created labels are not accurate, our model accuracy could be affected. Its advantage is that SVM allows us to train our own sentimental classifier for future airline reviews sentiment prediction, which is more accurate than the fixed dictionary of VADER Lexicon.

Topic Modeling

The second is topic modeling. We split the dataset into six subsets based on the airline that the tweet comments on because we want to perform topic modeling for each airline. Then we repeated the following process of topic modeling to each subset: we normalized the data first, used the bag-of-words method to vectorize the data and then used Latent Dirichlet Allocation to form four topics in this subset. Next, we displayed the 10 most frequent words in each topic and visualized the result containing the intertopic distance map and the chart of top 30 most exclusive terms for each topic at lambda = 0.2.

In the analysis of United, the topics are about the situation and service at the airport gate, flight reservation, airport customer service, and customer service by email/phone. It is worth mentioning that we can see the negative words such as “bad”, “terrible”, “poor”, and “worst” in the top 30 most relevant terms, indicating that people might be not satisfied with the customer service in the airport. (See Appendix 2,3,4,5)

In the analysis of US Airways, the first topic is also about the situation and service at the airport gate, the second topic is about the customer service by phone, the third topic is unclear since we cannot tell the correlation among the words in this group (i.e. the group contains both positive and negative words such as “like”, “fail”, “good”, “bad”, “love” in the top 30 most relevant terms), and the fourth topic is about the reservation of the flights. (See Appendix 6,7,8,9)

The words in each topic of the analysis of American Airlines are quite informative.The first topic is also about the situation at the airport. We can see locations such as “jfk”, “Miami " and “Chicago " are in the list, so we might conclude that the company should keep an eye on the situation in these locations, either good or bad. Frequent words such as “cancel”, “rebook”, “hotel”, “reschedule” in the second topic indicate that people might suffer a lot from the cancellation of the flight, since we didn't

see other companies have a topic talking about the cancellation specifically. The third topic is about customer service in general. We can see positive words such as "thank", "love","care", "handle", "appreciate", and negative words like "terrible" and "rude" in the list. After all, the positive words dominate the list, so we might conclude that American Airlines's customer service meets the requirement in general, but still has room for improvement. The fourth topic is about customer service calls specifically. Words such as "hold", "hang", "busy", "wait", "horrible" and "frustrate" indicate that people might wait for a very long time to get an agent to pick up their calls. The frequency of words such as "hour" and "min" might also indicate that the wait time is abnormally long.
(See Appendix 10,11,12,13)

In the analysis of Southwest, the topics are the status of the flights, the customer service, and the campaign "Destination Dragon Tour" partnered with the music band Imagine Dragons. The fourth one is unclear. We didn't see many emotional words in the first and second topic, so it might indicate that people just care about the status and customer service in general. Moreover, from the words in the third topic, we learn that this campaign attracted lots of attention online and achieved great success. (See Appendix 14,15,16,17)

In the analysis of Jetblue, the topics are the flight conditions in general, the complement of customer service, and contents are not related to the flights or customer service in the third and fourth topics. In the second topic, we can see words like "awesome" ,"appreciate", "thanks" and "quick" dominate the list, so we might say that Jetblue performs really well in customer service. And the content in the last two topics might be left to the PR team to conduct deeper analysis. On the other hand, it could be because the sample size of Tweets about Jetblue is relatively small so that we cannot get insight here. (See Appendix 18,19,20,21)

The last analysis is for Virgin. The topics are ticket reservations, customer service, flight status, and lady gaga. There are no explicitly positive or negative words in each topic. And the reason why lady gaga is one of the popular topics about Virgin might be that lady gaga claimed that she would sing in the rocket made by Virgin Galactic, which is a subsidiary of Virgin Atlantic Airways. This news might gain lots of public attention on Twitter. (See Appendix 22,23,24,25)

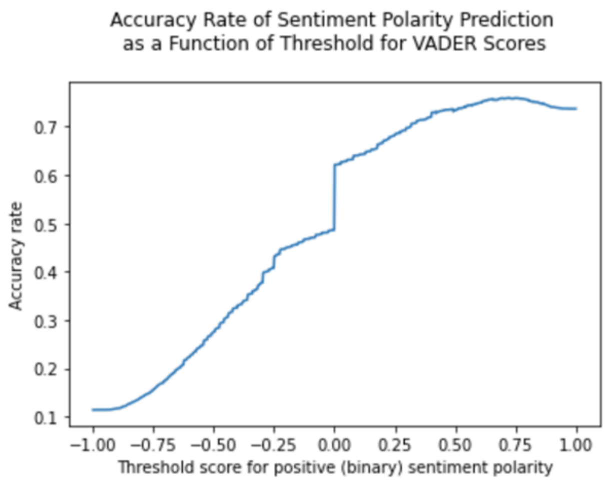
Conclusion and Recommendation

From sentiment analysis, negative review rate for VirginAmerica, United, Southwest, JetBlue, US Airways and American Airlines are respectively 36.5%, 71.4%, 49.9%, 43.1%, 84.9%, 81.9%. Therefore, the top three negative review rate companies are US Airways, American Airlines and United Airline. The top three positive review rate companies are VirginAmerica, JetBlue and Southwest Airline. (See Appendix 26)

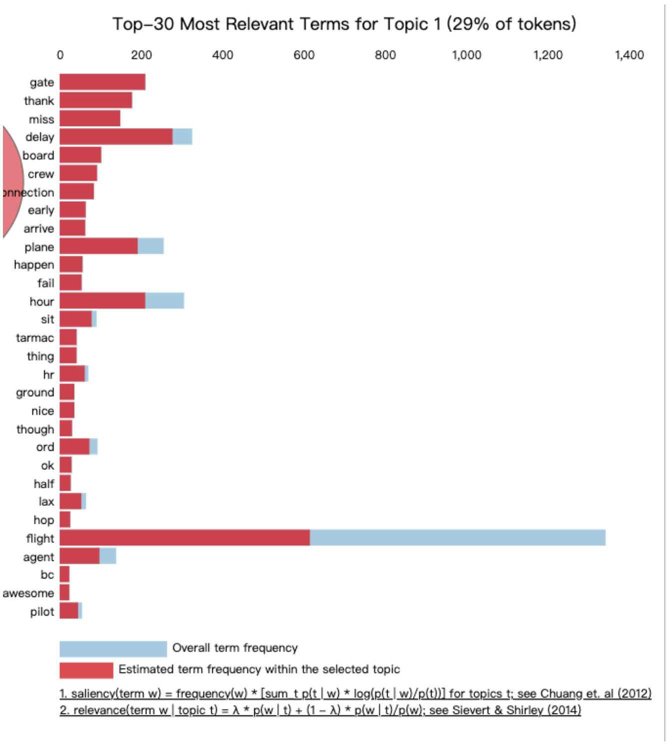
From topic modeling, we can learn that all three companies with the most negative reviews didn't do well in the customer service part, no matter the airport customer services or the customer services calls. It makes sense that people complain a lot on social media when customer service cannot solve their problems. However, the companies definitely should look into the issues that customers complain about and make improvements. For example, United should improve their customer services at the airport, giving more employee training or making mitigation plans in advance; And American Airlines should focus on improving the customer services call specifically. Since people complained a lot about the length of wait time, this issue might result from the lack of agents, employee training or solutions, or the complexity of internal communication between different departments. On the other hand, we also learn that Jetblue is the only one having the topic of compliment. Therefore, other companies can earn some experience of good customer services from Jetblue if possible.

Appendix

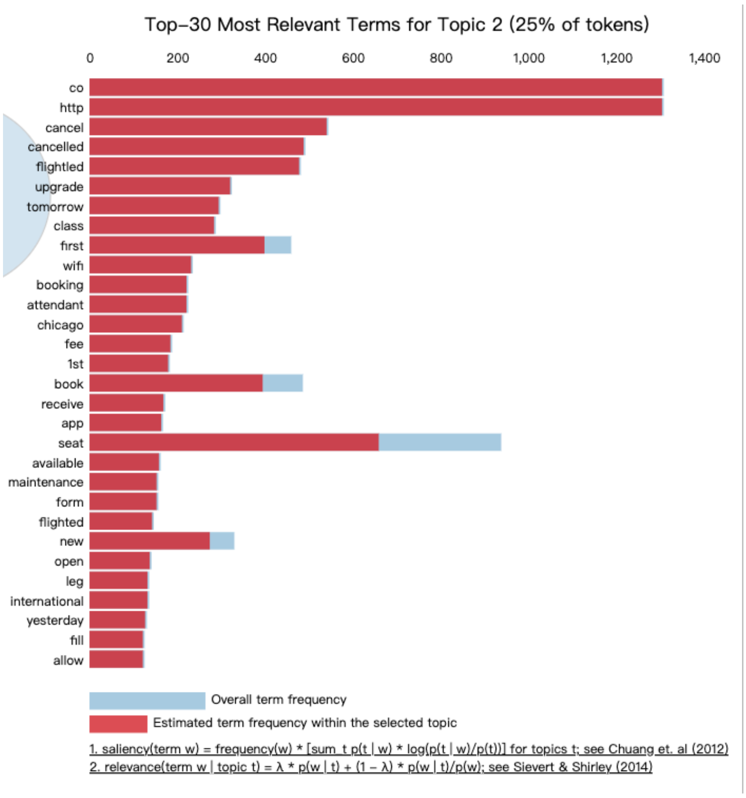
Appendix 1



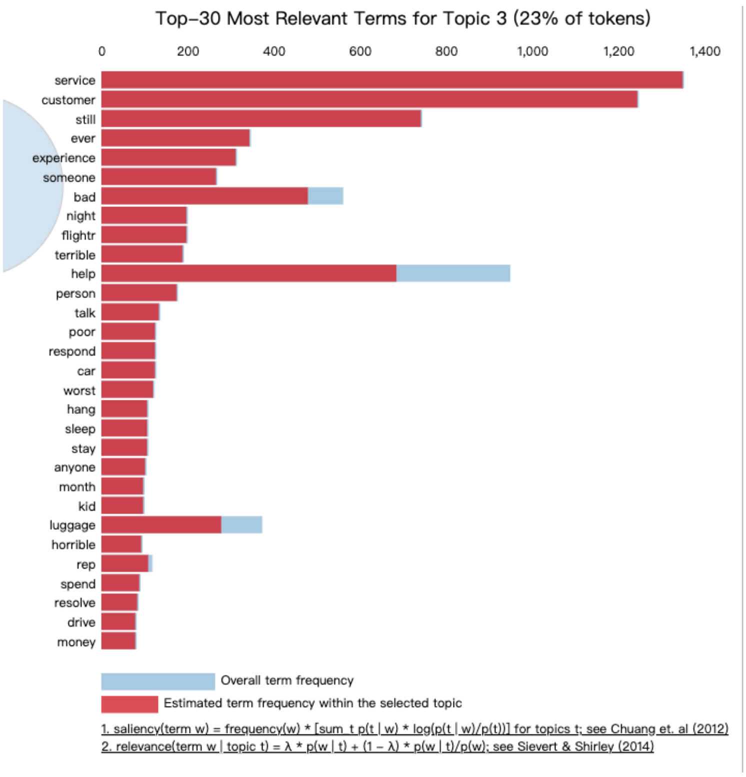
Appendix 2 (Topic 1 for United)



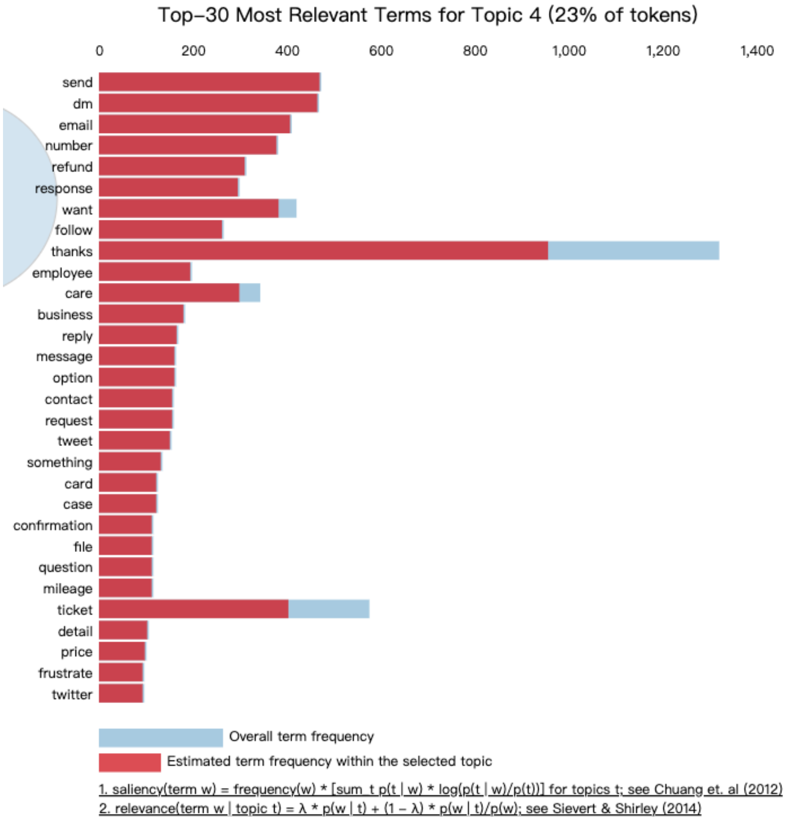
Appendix 3 (Topic 2 for United)



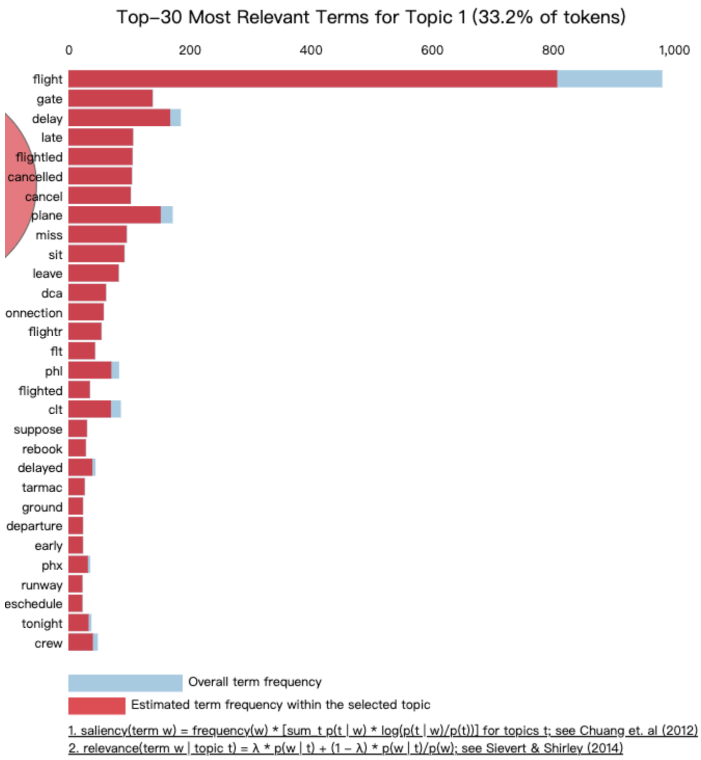
Appendix 4 (Topic 3 for United)



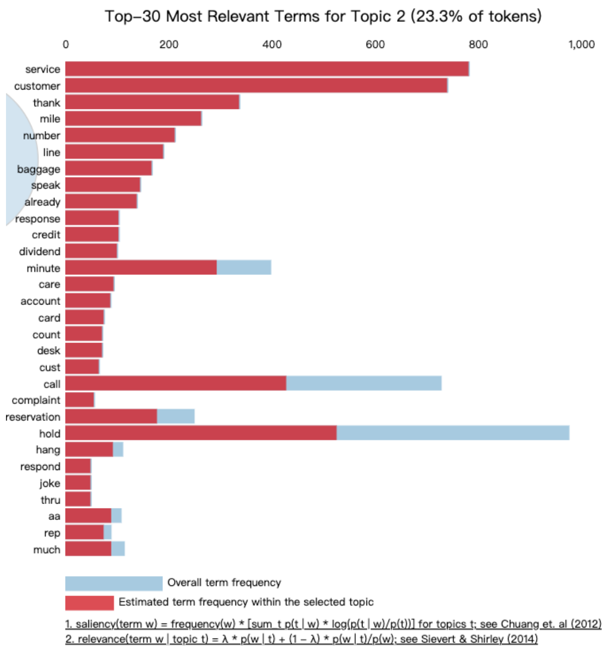
Appendix 5 (Topic 4 for United)



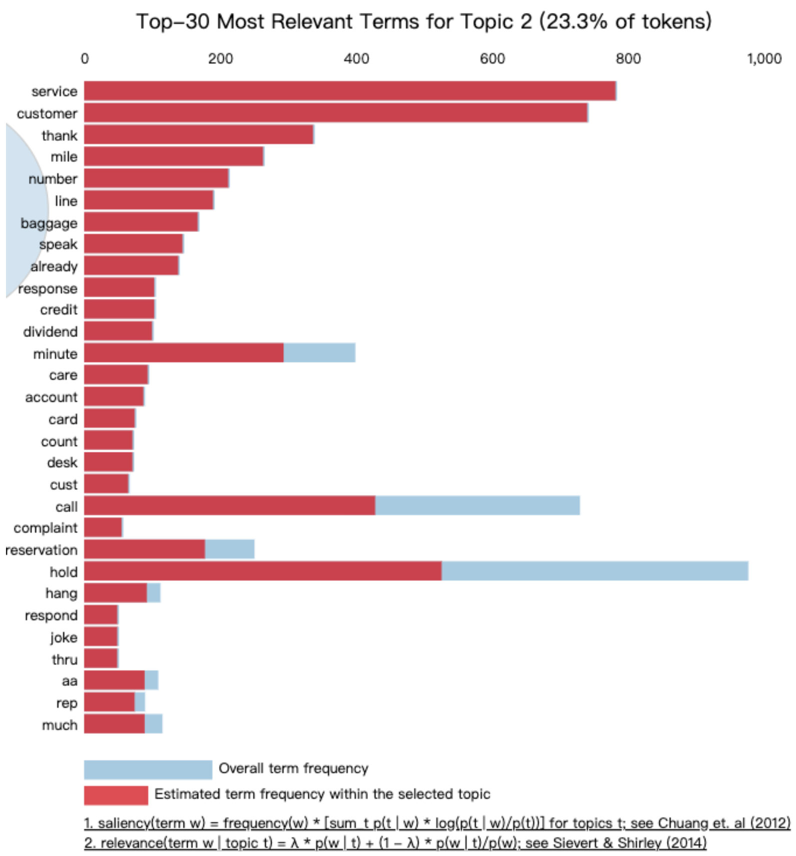
Appendix 6 (Topic 1 for US Airways)



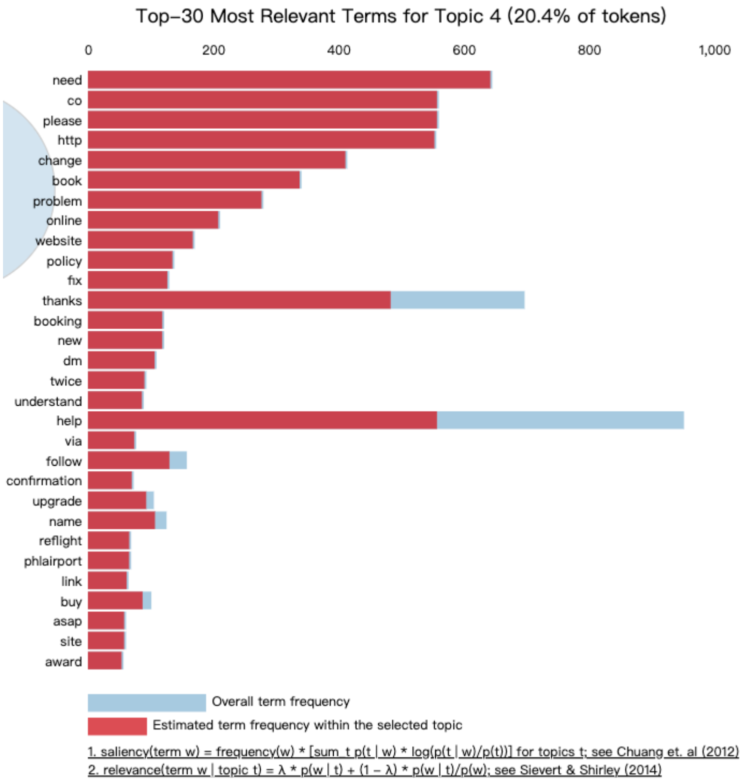
Appendix 7 (Topic 2 for US Airways)



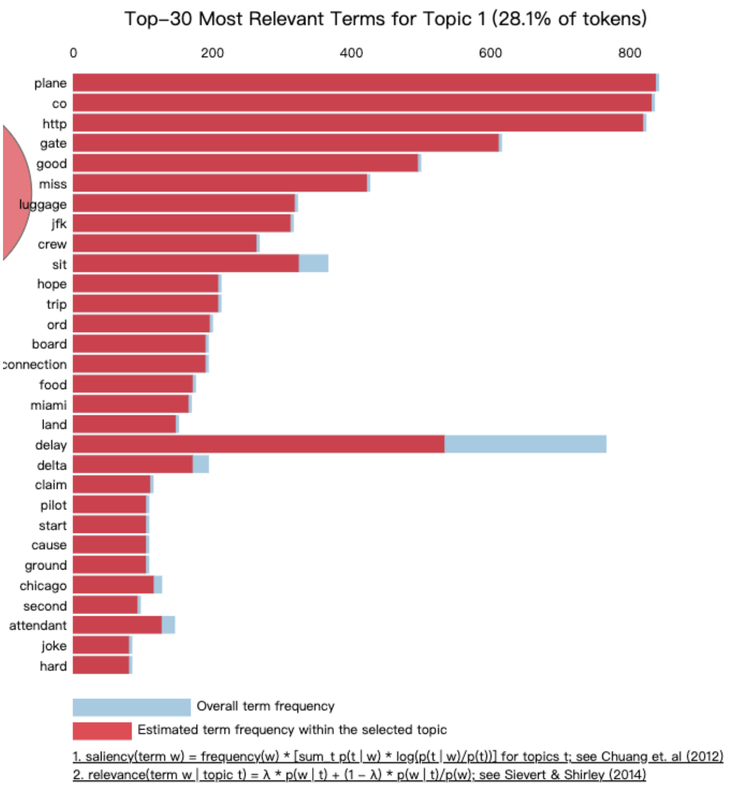
Appendix 8 (Topic 3 for US Airways)



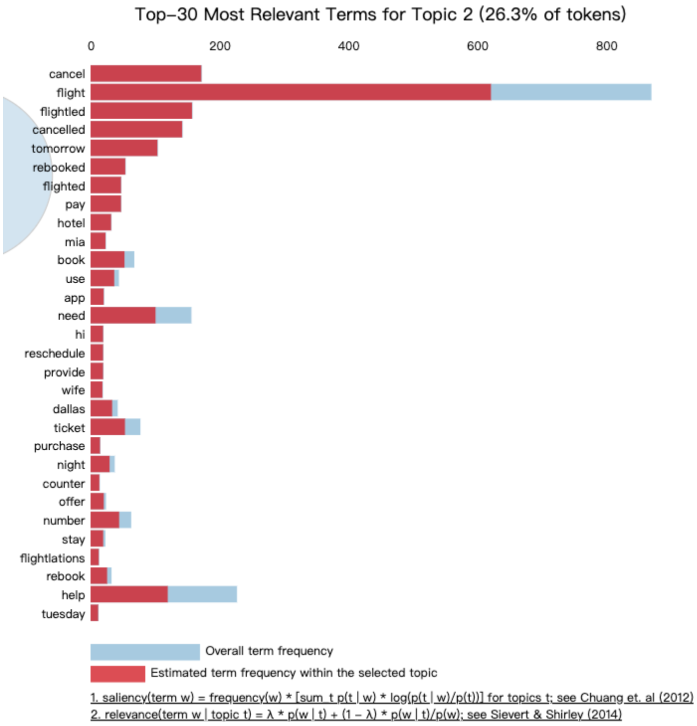
Appendix 9 (Topic 4 for US Airways)



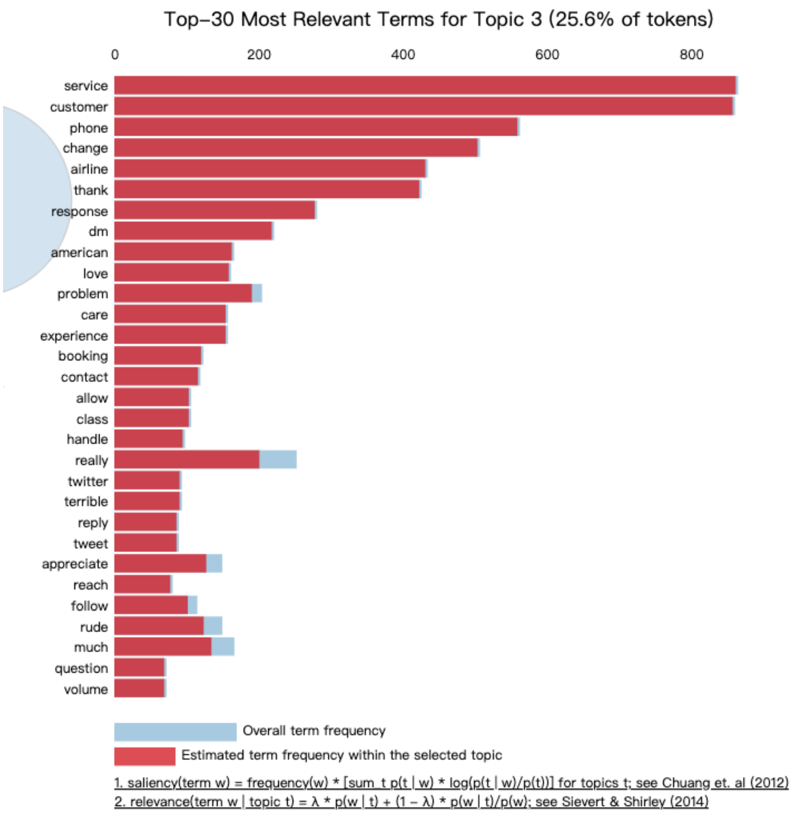
Appendix 10 (Topic 1 for American Airlines)



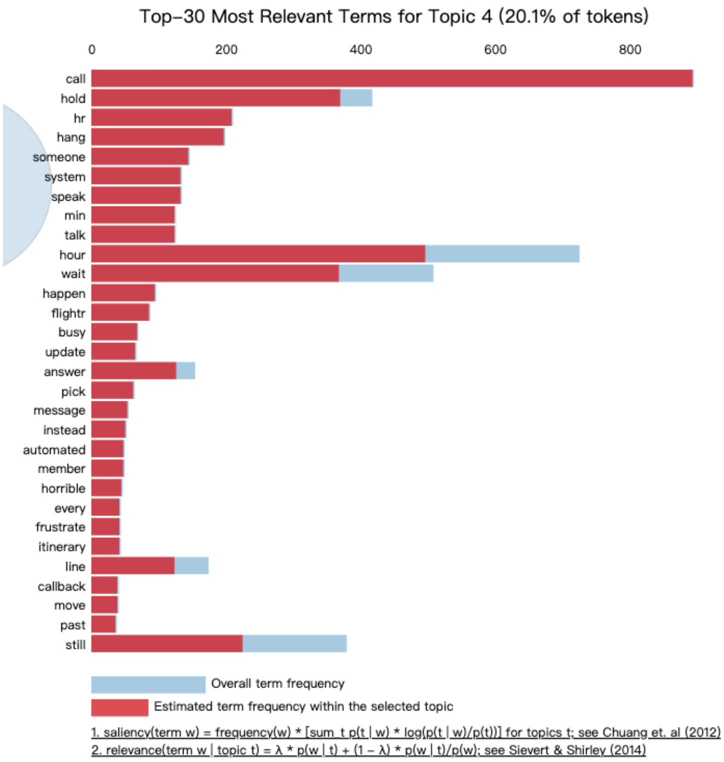
Appendix 11 (Topic 2 for American Airlines)



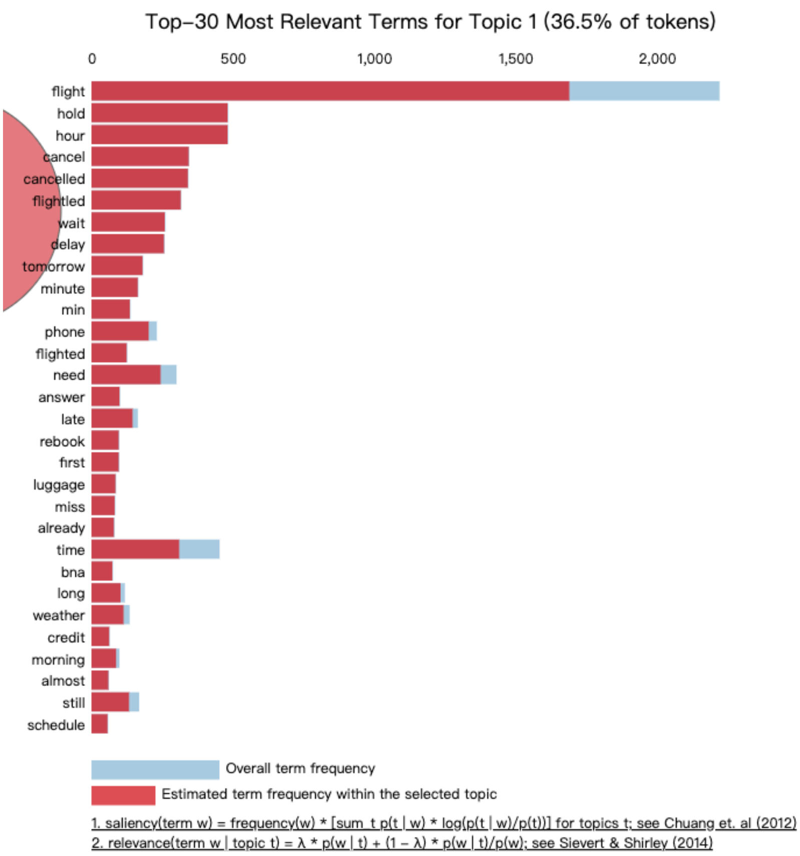
Appendix 12 (Topic 3 for American Airlines)



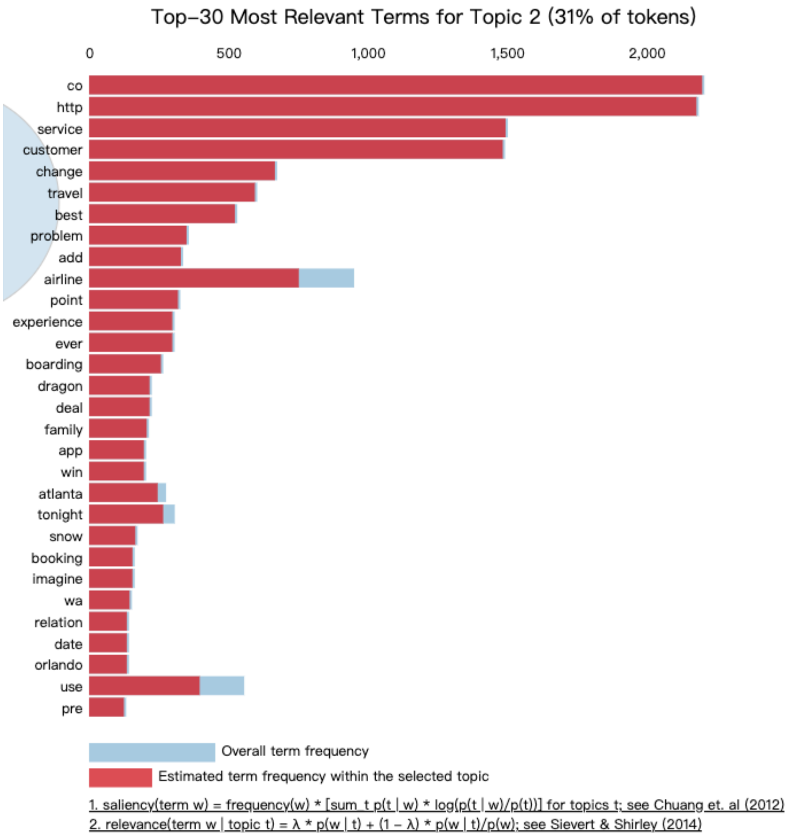
Appendix 13 (Topic 4 for American Airlines)



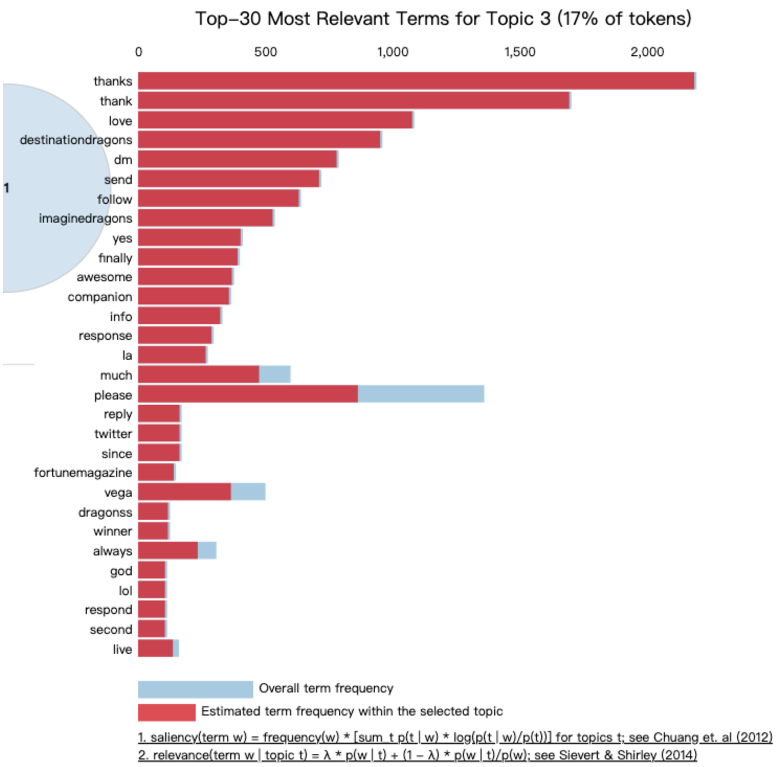
Appendix 14 (Topic 1 for Southwest)



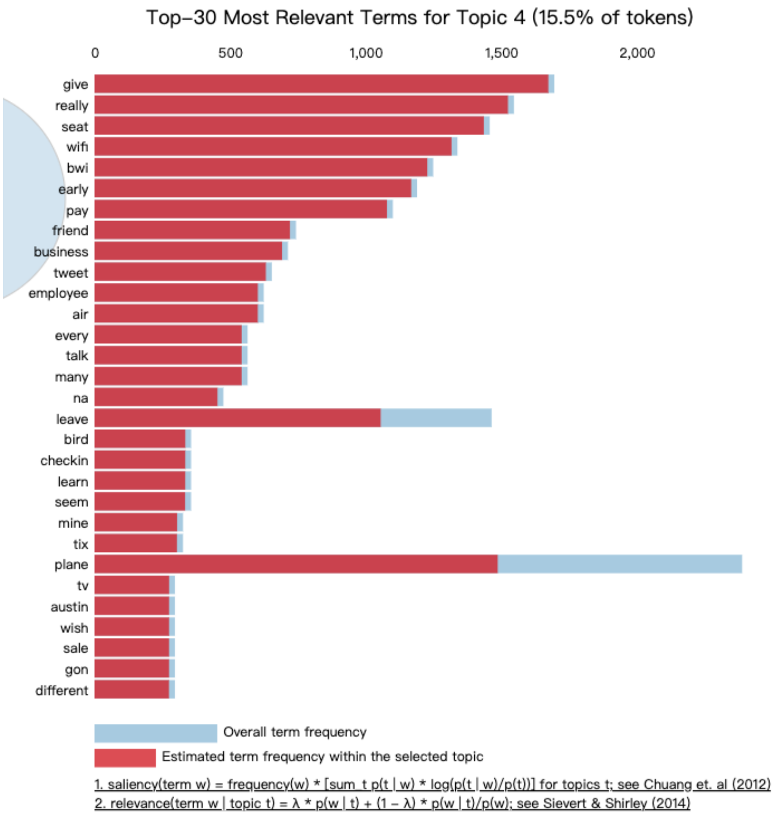
Appendix 15 (Topic 2 for Southwest)



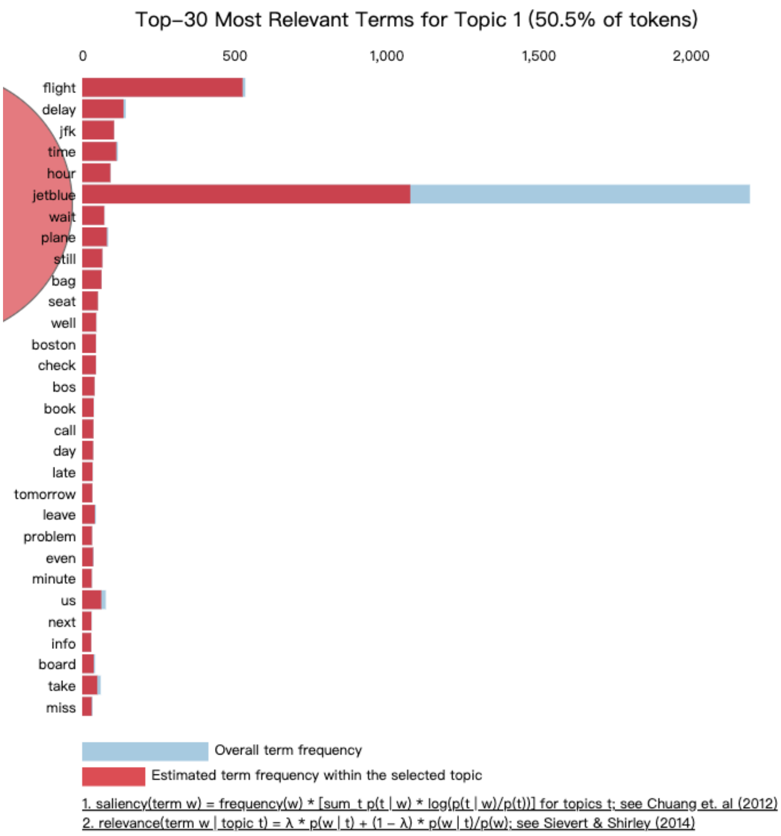
Appendix 16 (Topic 3 for Southwest)



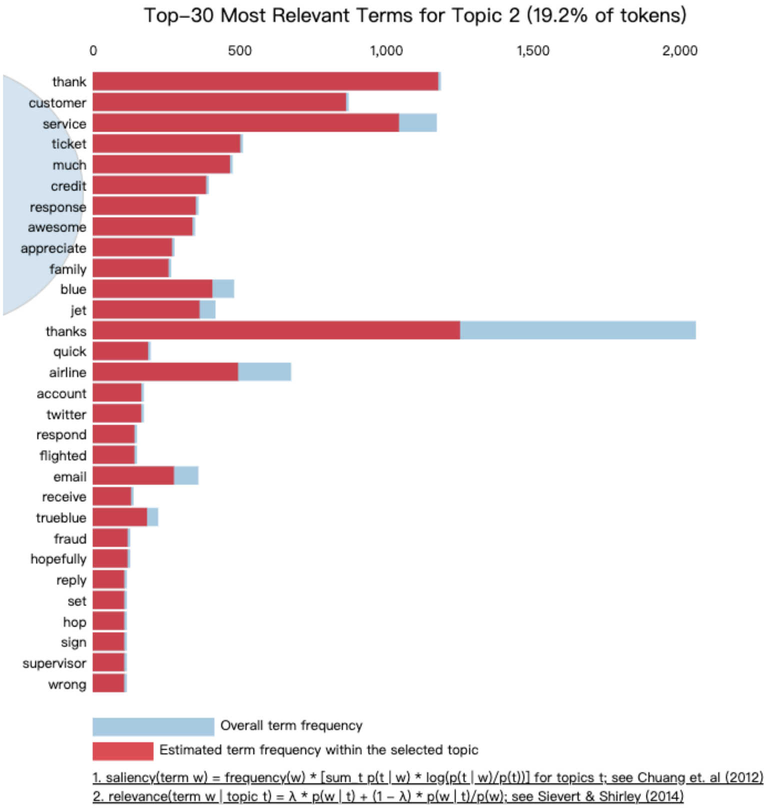
Appendix 17 (Topic 4 for Southwest)



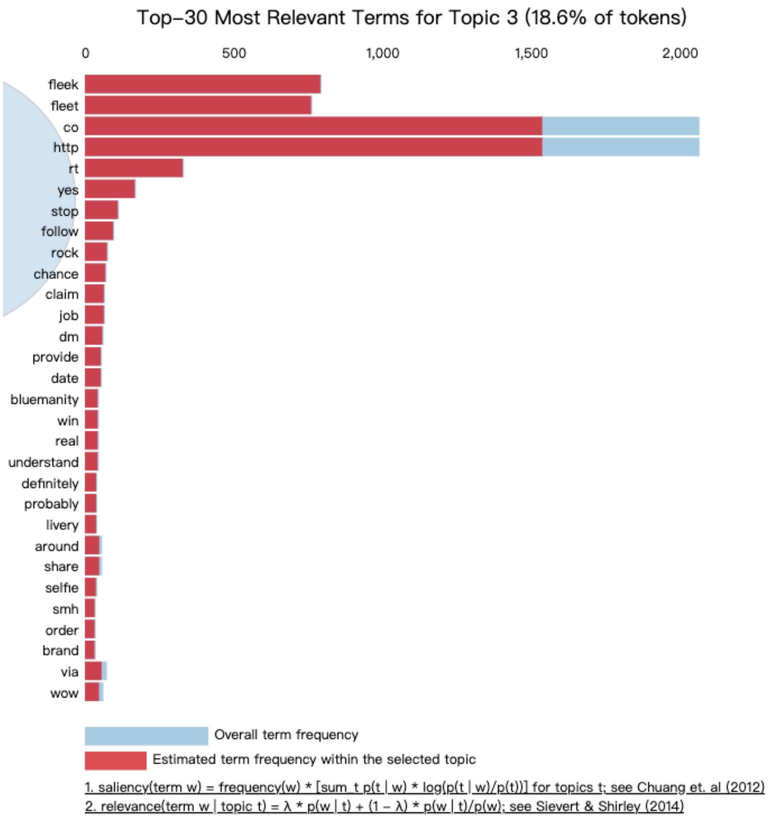
Appendix 18 (Topic 1 for Jetblue)



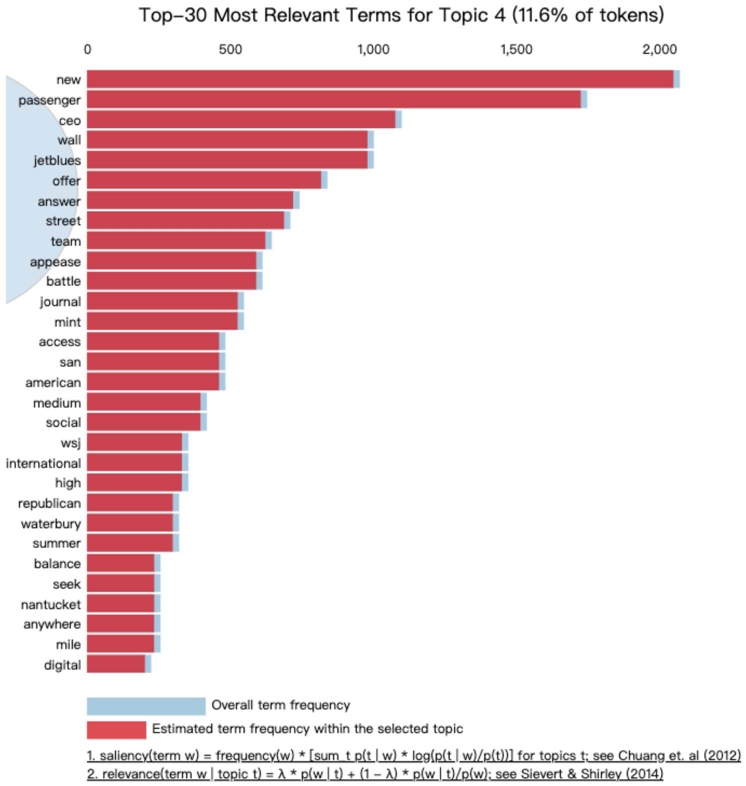
Appendix 19 (Topic 2 for Jetblue)



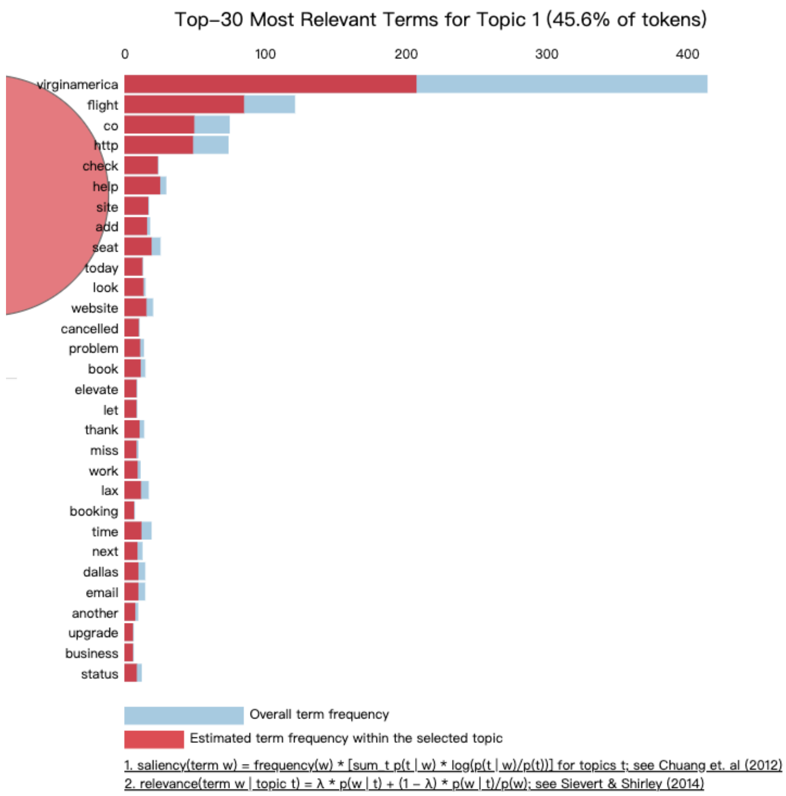
Appendix 20 (Topic 3 for Jetblue)



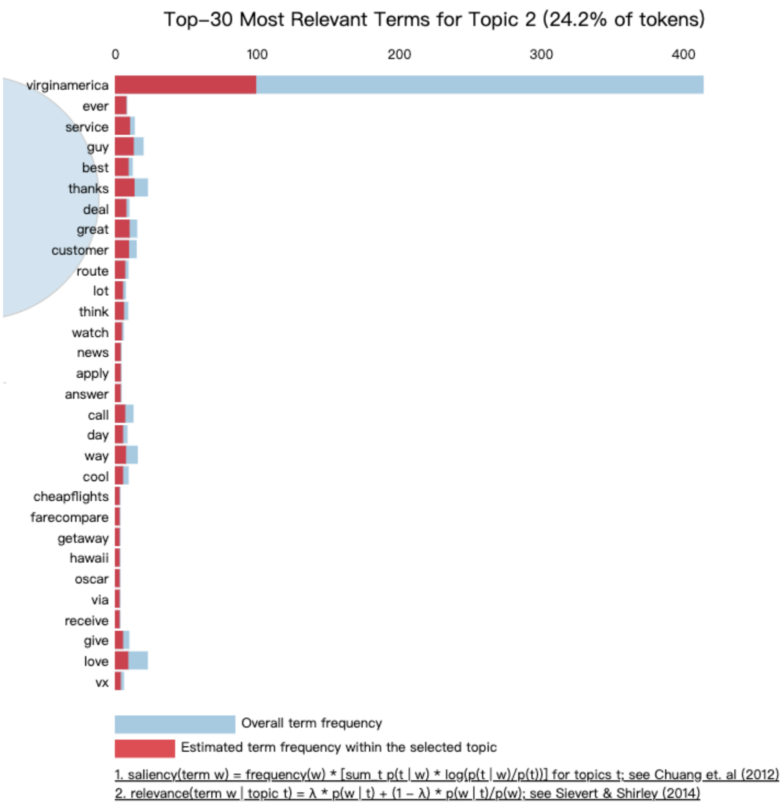
Appendix 21 (Topic 4 for Jetblue)



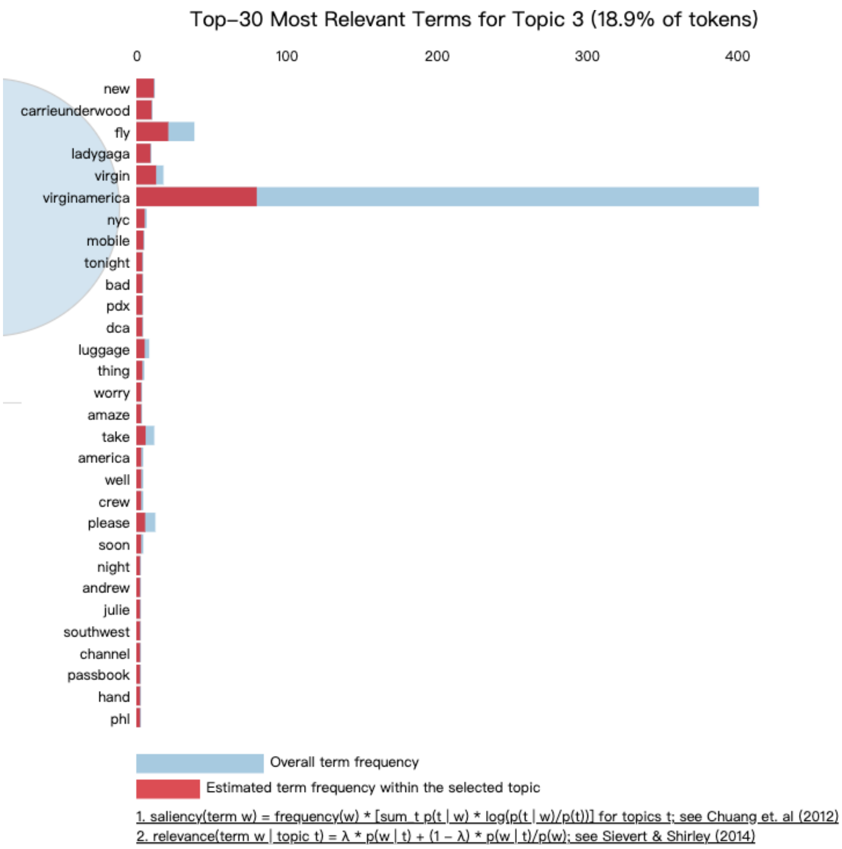
Appendix 22 (Topic 1 for Virgin)



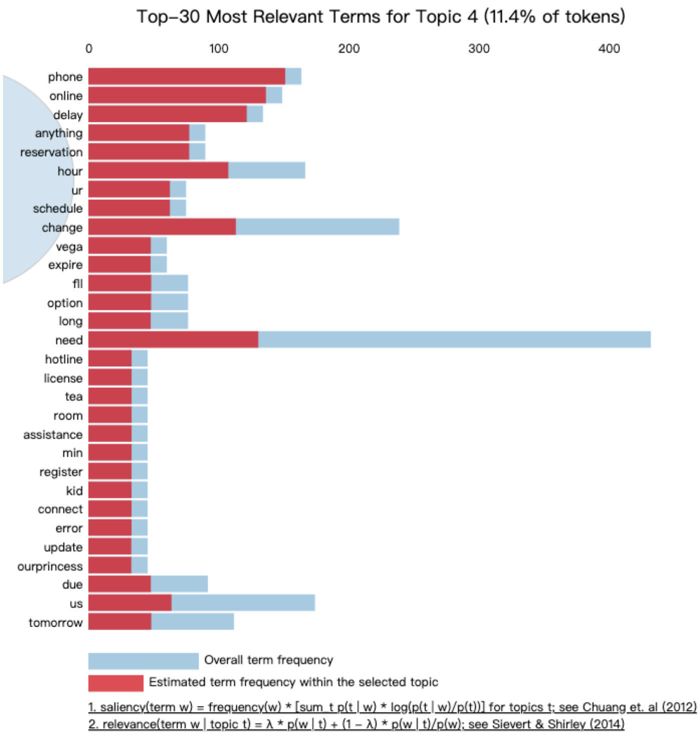
Appendix 23 (Topic 2 for Virgin)



Appendix 24 (Topic 3 for Virgin)



Appendix 25 (Topic 4 for Virgin)



Appendix 26