DESAUTELS FACULTY OF MANAGEMENT

INSY 669 - Text Analytics

# AIRLINES CUSTOMER EXPERIENCE ANALYSIS

Team 16

Liu, Alice    261007356
Zhu, Diwei    260761307
Zhang, Ziye   260766101
Liang, Jiahua 260711529
Wang, Kexin   260787422
Wang, Yichen  260761601


February 15, 2022

## 1. Introduction

Different from previous times that public opinions were created mainly by commentators and media, nowadays, through the internet, everyone's reviews and opinions can be heard. The "trails" and "voices" that users left on online platforms, known as user-generated content (UGC), are considered to be able to reflect a less distorted consumer attitude.

In our case, in order to compare airlines in terms of customer experience, we collect and analyze related UGCs rather than trusting polished news articles or advertisements produced by airlines themselves. With the UGC text analytical result in hand, we would be able to help airlines to engage in real-time interactions with potential customers and to generate insights for both airlines and consumers.

More specifically, we aim to 1) analyze popular airlines and understand whether people show positive or negative sentiments about each one of them as well as 2) providing customized recommendations of airlines to segmented customers based on their preferences by leveraging sentiment analysis and airlines segmentation, and 3) providing advice to airlines about which attributes to improve to effectively alleviate customer experience.

## 2. Methods and result interpretation

### 2.1. Scraping the airline review information

We perform data extraction using web scrapper, which is an excellent web extension for scraping website data via an easy-to-understand interface. The data is scraped from two famous airline review websites, namely Consumer Affairs and Skytrax. By scraping reviews from different sources, we reduce the bias on a specific platform and we are able to get the true consumer point of views and generalize our analysis. We focus our analysis on 10 different well-known airlines in the world, namely American Airlines, Delta Airlines, United Airlines, Southwest Airlines, Spirit Airlines, Frontier Airlines, Air Canada, Ryanair, Emirates Airlines, and British Airway. Among these, American Airlines, Delta Airlines, United

Airlines, Southwest Airlines, Spirit Airlines, Frontier Airlines are from the U.S.; Air Canada is from Canada; Ryanair is from the Republic of Ireland; Emirates is from the United Arab Emirates; and British Airway is from the U.K. Despite the fact that these airlines have different levels of popularity, and thus different amounts of reviews, we aim to scrape an equal number of reviews for these 10 airlines (2000 reviews per airline) so that our analysis is unbiased and balanced. The date ranges from 2022 February to 2010 January (Figure 1).

| COUNTRY | Canada | Ireland | United Arab Emirates | UK |
|---|---|---|---|---|
| AIRLINES | Air Canada | Ryanair | Emirates Airlines | British Airway |
| USA | USA | USA | USA | USA |
| American Airlines | Delta Airlines | United Airlines | Southwest Airlines | Spirit Airlines |

| | Airline | Review |
|---|---|---|
| 0 | Delta_Airline | No guarantee of seating if booked on Expedia. … |
| 1 | Delta_Airline | 3 days to get to London. Delta are still way b… |
| 2 | Delta_Airline | Our trip was from Phoenix to Reno with transit… |
| 3 | Delta_Airline | OUR FLIGHT FROM ST. LOUIS, MO TO ATLANTA, GA. … |
| 4 | Delta_Airline | My flight to JFK scheduled 2013 was canceled. … |

*Figure 1. scrapped comments from the source*

## 2.2. Pre-processing and word analysis

As the reviews are collected from themed forums for certain airlines, users who wrote reviews in the forum of airline A can be considered to show an intention of discussing the specific airline. Therefore, no matter the users did or did not mention the name of airline A in their reviews, we assume that one review contains one mention of the very airline. By attaching the corresponding airline names to the start of each review, we manage to capture the mentions even if the name of the airlines was not exactly mentioned in the reviews.

When it comes to tokenizing the reviews by the *nltk* package, we notice that there are spaces between the words of airline names and the word "Airlines" (we use "Spirit Airlines" and "Delta Airlines" as examples here), which would create obstacles for the word analysis as *nltk* tokenize sentences based on spaces. To deal with this problem, every time when we detect the mentions of airline names ("spirit" and "delta"), we replace the word with "spirit_airline" and "delta_airline". The same operation is done to "United Airlines", "Emirates Airlines", and "Southwest Airlines". However, as for "Air Canada", "British Airways" and "American Airlines", the airline names ("canada", "british", and "american") contain frequently used words that may cause a misunderstanding. We do not want to label "I flew to British Columbia in Canada in May." as one mention of "Air Canada" and one mention of "British Airways", so the above method does not apply. Alternatively, every time we detect "air", "british" or "american", we check if the following token is "canada", "airways", or "airlines" respectively. If the combinations are confirmed, we regard the two words as one mention of the very airline and replace the words with "air_canada", "british_airway" and "american_airline".

With the pre-processed word tokens, we removed duplicated mentions of each airline in each review. A word cloud is made to give an overview of frequently mentioned words across all reviews.

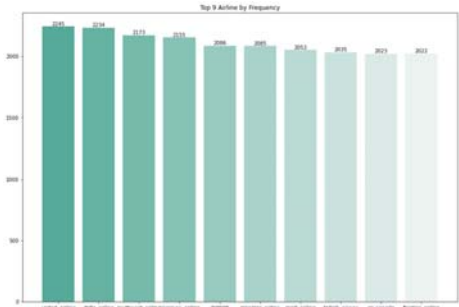*Figure 2. Word cloud showing frequently mentioned words by users*

*Figure. 3 Frequency of each airline mentioned in reviews*

We then count the number of mentions of airlines across all reviews. From Figure 3, we see that United Airlines gained the most mentions (2245) while Frontier Airline got mentioned the least (2022). The difference between airlines' frequencies of mentionings is not significant, as each airline has a base mentioning frequency of 2000. This is because we count each review in the airline forum as one mention. The exceeding counts are the mentions of an airline in other airlines' review sections, indicating the comparisons of two airlines made by the users.

### 2.3. Analyzing customer experience attributes

A list of airline customer experience attributes and corresponding keywords are needed to label and count the mentions of attributes in reviews. Based on the tokens and labeled part-of-speech by the *nltk* package, we filter out the 500 most frequently mentioned nouns, adjectives, and adverbs ('NN', 'JJ', 'JJR', 'JJS', 'NNS', 'NNP', 'NNPS', 'RB', 'RBR', 'RBS') across all reviews, as these words can contain the frequently mentioned keywords typed by users to describe their experiences with airlines. After eliminating some nouns and adjectives which cannot be categorized or too general (for example, "small", "good", and "London"), we manually grouped the remaining 177 words into seven major attribute categories: check-in & boarding, customer service, monetary values, comfort, food & beverage, in-flight entertainment and cleanliness. All the keywords mentioned in the reviews are substituted by the attribute category name (Figure 4). For example, the keyword "luggage" will be substituted into check-in & boarding in the original review post.

| Attribute category | Sample keywords |
|---|---|
| comfort | seat, noise, legroom |
| in-flight entertainment | screen, movies |
| customer service | attendant, staffs |
| cleanness | cleaned, dirty |
| monetary value | price, tax, discount |
| check-in & boarding | luggage, hour, delay |
| food & beverage | drink, cake, delicious |

*Figure 4. defined attribute categories and keywords examples*

Then, we counted the mentions of attributes and sorted the attributes by the mentioning frequency in the reviews. Figure 5 shows that check-in_boarding (16956 mentions among 20000 reviews) is the most popular attribute, followed by customer_service, and monetary values. In-flight experience attributes – comfort, foor_beverage, and in-flight_entertainment receive less attention. Only 1042 reviews mentioned cleanliness, which is the least mentioned attribute.
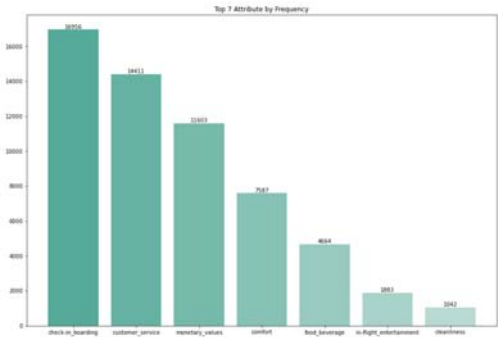


*Figure 5. attributes by mentions frequency*

### 2.4. Sentiment analysis

Aiming at creating a broader basis for a better understanding of airlines from customers' perspectives, we apply sentiment analysis to the collected reviews. With the *vader* package that examines both polarity and intensity of emotions reflected by sentences, we generate sentiment scores and calculate a compound score for each review that labels the overall polarity of the review. Then, we generate binary mentioning indicators of the airlines for each review. For example, as shown in Figure 6.1, any 1 in the "delta_airline" column means that the review mentioned the airline, and if it shows 0, it means the airline was not mentioned. Similar indicators are generated for attribute mentions (Figure 6.2).

| Airline | compound | Airline_Mention | Attribute_Menti | delta_airline | american_airlin | united_airline | spirit_airline | air_canada |
|---|---|---|---|---|---|---|---|---|
| Delta_Airline | 0.323 | ['delta_airline'] | ['comfort'] | 1 | 0 | 0 | 0 | 0 |
| Delta_Airline | -0.6641 | ['delta_airline'] | ['check-in_board] | 1 | 0 | 0 | 0 | 0 |
| Delta_Airline | -0.4767 | ['delta_airline'] | ['check-in_board] | 1 | 0 | 0 | 0 | 0 |
| Delta_Airline | 0.7804 | ['delta_airline'] | ['check-in_board] | 1 | 0 | 0 | 0 | 0 |
| Delta_Airline | -0.5186 | ['delta_airline'] | ['comfort', 'check] | 1 | 0 | 0 | 0 | 0 |
| Delta_Airline | -0.866 | ['delta_airline'] | ['cleanliness', 'cu] | 1 | 0 | 0 | 0 | 0 |
| Delta_Airline | -0.5789 | ['delta_airline'] | ['customer_servi] | 1 | 0 | 0 | 0 | 0 |
| Delta_Airline | 0 | ['delta_airline'] | ['check-in_board] | 1 | 0 | 0 | 0 | 0 |
| Delta_Airline | -0.7346 | ['delta_airline'] | ['check-in_board] | 1 | 0 | 0 | 0 | 0 |
| Delta_Airline | 0.9377 | ['delta_airline'] | ['monetary_value] | 1 | 0 | 0 | 0 | 0 |
| Delta_Airline | 0.9743 | ['delta_airline'] | ['check-in_board] | 1 | 0 | 0 | 0 | 0 |
| Delta_Airline | -0.6652 | ['delta_airline'] | ['monetary_value] | 1 | 0 | 0 | 0 | 0 |
| Delta_Airline | 0.3612 | ['delta_airline'] | ['check-in_board] | 1 | 0 | 0 | 0 | 0 |
| Delta_Airline | -0.5848 | ['delta_airline'] | ['customer_servi] | 1 | 0 | 0 | 0 | 0 |
| Delta_Airline | 0.7584 | ['delta_airline'] | ['customer_servi] | 1 | 0 | 0 | 0 | 0 |
| Delta_Airline | 0 | ['delta_airline'] | ['customer_servi] | 1 | 0 | 0 | 0 | 0 |
| Delta_Airline | -0.5122 | ['delta_airline'] | ['customer_servi] | 1 | 0 | 0 | 0 | 0 |
| Delta_Airline | -0.844 | ['delta_airline'] | ['monetary_value] | 1 | 0 | 0 | 0 | 0 |

*Figure 6.1. generated dataset of sentiment scores and airline mentioning indicators*

| compound | comfort | in-flight_enterta | customer_servi | monetary_value | cleanliness | check-in_board | food_beverage |
|---|---|---|---|---|---|---|---|
| 0.323 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| -0.6641 | 0 | 0 | 1 | 1 | 0 | 1 | 0 |
| -0.4767 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 0.7804 | 0 | 0 | 0 | 1 | 0 | 1 | 0 |
| -0.5186 | 1 | 0 | 0 | 1 | 0 | 1 | 0 |
| -0.866 | 1 | 0 | 1 | 0 | 1 | 1 | 0 |
| -0.5789 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 |
| -0.7346 | 0 | 0 | 1 | 1 | 0 | 1 | 0 |
| 0.9377 | 0 | 0 | 0 | 1 | 0 | 1 | 0 |
| 0.9743 | 1 | 1 | 0 | 1 | 0 | 1 | 1 |
| -0.6652 | 1 | 0 | 0 | 1 | 0 | 0 | 0 |
| 0.3612 | 0 | 0 | 1 | 0 | 0 | 1 | 0 |
| -0.5848 | 0 | 0 | 1 | 1 | 0 | 1 | 0 |
| 0.7584 | 1 | 0 | 1 | 1 | 0 | 1 | 0 |
| 0 | 0 | 0 | 1 | 1 | 0 | 1 | 0 |
| -0.5122 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| -0.844 | 0 | 0 | 1 | 1 | 0 | 0 | 0 |

*Figure 6.2. generated dataset of sentiment scores and attribute mentioning indicators*

With the dataset of compound sentiment scores, airlines mentioning indicators, and attributes mentioning indicators, we are able to compare average sentiment scores of airlines or reviews that mentioned certain attributes. Based on the dataset, we build an interactive dashboard (Figure 7) to help visualize the sentiments.



*Figure 7. Sentiment score visualization dashboard*

The left side of the dashboard showcases the average sentiment scores for each airline sorted from high to low. The airline with the most positive average sentiment score is Emirates Airlines while the one with the most negative average sentiment score is American Airlines. On the right side, a series of visualizations compare the average sentiment score of reviews that did and did not mention a certain attribute. For example, for all reviews that did mention "in-flight entertainment", their average sentiment score (0.345) is higher than all reviews that did not mention the attribute (-0.081), indicating that users usually discuss in-flight entertainment with a positive sentiment. The lower table lists the average score in terms of each attribute (Figure 8).

| Attribute | Avg score of mentions | Avg score of non-mentions |
|---|---|---|
| customer_service | -0.072 | -0.028 |
| in-flight entertainment | -0.081 | 0.345 |
| food_beverage | -0.113 | 0.198 |
| cleanliness | -0.059 | 0.302 |
| comfort | -0.135 | 0.115 |
| check-in_boarding | 0.082 | -0.062 |
| monetary_values | -0.003 | -0.068 |

*Figure 8. Average sentiment score comparison of reviews that did and did not mention certain attributes*

As allowed by the dashboard, we can zoom into one airline to compare the average sentiments of reviews in terms of different attributes. We can also compare the average sentiments of airlines among all reviews that mentioned "comfor

t". This interactive feature helps us generate recommendations for customers and airlines that will be discussed in the later section.

## 2.5 Regression

Additionally, also based on the sentiment score dataset, we are able to build a regression model that examines the relationship between airlines, attributes and user sentiments, shown in Figure 9 below. We use the 10 airlines and the 7 attributes as the features and set the sentiment score as the target variable. Among these airlines, United Airlines, Southwest Airlines, Ryanair, Emirates, and British Airway have positive relationships with sentiment scores; and Delta, American, Spirit Airlines, Air Canada, and Frontier Airlines have negative relationships with the sentiment score. Our results here are similar to the average sentiment scores we obtained above. The one major difference between the results is from the United Airline. The average sentiment score is -0.04 whereas the coefficient of the airline in regression is +0.11. We conclude this difference to the other controlled features such as the airline-related attributes. This implies that although United Airline has a decent reputation (because of its own positive coefficient), the services it provides frustrate the customers and cause its overall sentiment to decrease towards a negative value.

```
                              OLS Regression Results
===============================================================================
Dep. Variable:             compound   R-squared (uncentered):            0.105
Model:                          OLS   Adj. R-squared (uncentered):       0.104
Method:               Least Squares   F-statistic:                       138.0
Date:              Tue, 15 Feb 2022   Prob (F-statistic):                 0.00
Time:                      20:45:14   Log-Likelihood:                  -20842.
No. Observations:             20000   AIC:                            4.172e+04
Df Residuals:                 19983   BIC:                            4.185e+04
Df Model:                        17
Covariance Type:          nonrobust
===============================================================================
                          coef    std err          t      P>|t|      [0.025      0.975]
-------------------------------------------------------------------------------
delta_airline          -0.0268      0.018     -1.517      0.129      -0.061       0.008
american_airline       -0.1746      0.018     -9.652      0.000      -0.210      -0.139
united_airline          0.1108      0.018      6.281      0.000       0.076       0.145
spirit_airline         -0.1083      0.019     -5.627      0.000      -0.146      -0.071
air_canada             -0.0425      0.020     -2.134      0.033      -0.082      -0.003
ryanair                 0.0859      0.020      4.299      0.000       0.047       0.125
frontier_airline       -0.1350      0.019     -6.939      0.000      -0.173      -0.097
emirates_airline        0.1705      0.021      8.023      0.000       0.129       0.212
southwest_airline       0.2665      0.018     14.564      0.000       0.231       0.302
british_airway          0.1357      0.021      6.444      0.000       0.094       0.177
comfort                 0.1221      0.011     11.246      0.000       0.101       0.143
in-flight_entertainment 0.2243      0.018     12.220      0.000       0.188       0.260
customer_service       -0.0368      0.011     -3.459      0.001      -0.058      -0.016
monetary_values        -0.0327      0.010     -3.299      0.001      -0.052      -0.013
cleanliness             0.1777      0.022      7.916      0.000       0.134       0.222
check-in_boarding      -0.1513      0.012    -12.227      0.000      -0.176      -0.127
food_beverage           0.1218      0.014      8.741      0.000       0.095       0.149
```

*Figure 9. Sentiment score regression model result*

## 2.6. Classification Methodology

For the purpose of applying more text applications, we also attempt to build a classification model using our cleaned airlines' reviews data. Since we have the labeled reviews for all the airlines, we set the names of the airline as our target and use the reviews to predict which airline it belongs to. To compare the model accuracies across different tools, we try both TF-IDF vectorizer and count vectorizer in sklearn to preprocess the data and remove stop words. While removing stop words, we consider the case that the airline name may appear in the comment, and it would cause some information leakage. However, we think including the airline names in the reviews reflects the real situation when people comment. They write airline names in their review to express extra appreciation or intense loathing. Besides using the Naïve Bayes classifier, we also use SGD classifier to predict the classes. Based on our results, the SGD classifier with TF-IDF vectorizer fits best to our data, resulting in an accuracy of 68.75% with airline names removed and an accuracy of 86.9% without removing airline names.
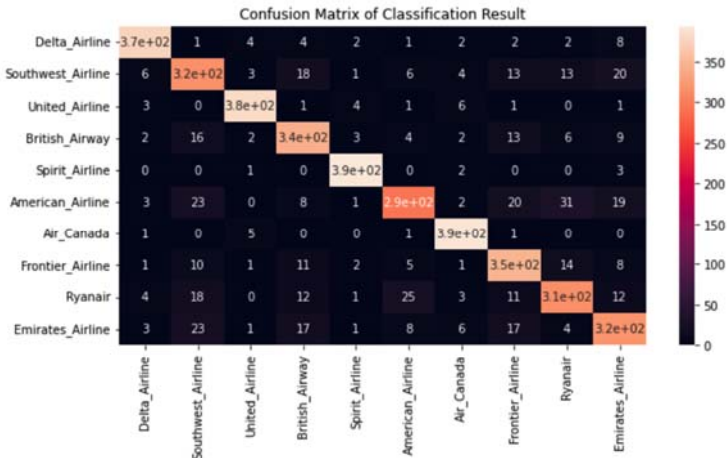
Figure 10. Confusion Matrix

## 2.7. Lift Calculation

We conduct lift analysis to study the association between airlines, between airlines and attributes, and between attributes. We start with calculating the lift between airlines.

### 2.7.1 Airline-Airline Lift

The lift of each pair of airlines (A&B) is calculated using the formula:

$$Lift(A,B) = N \times \frac{\#(A,B)}{\#(A) \times \#(B)}$$

where:

 N = the total number of reviews

#(A, B)  = the number of reviews mentioning both airline A and airline B

#(A)  = the number of reviews mentioning airline A

#(B)  = the number of reviews mentioning airline B

The values needed for this formula can be obtained from the airline mentioning count tables and the tables recording the airlines mentioned in each review. The lift table is shown in Figure 11. The heatmap of the lift table is shown in Figure 12. In order to properly plot the MDS map, the dissimilarity between each pair of airlines is calculated by 1/Lift (visualized in Figure 13 and Figure 14). From the result, we notice that all the lift values between airlines are low, which suggests that there are not many co-mentions of different airlines in the reviews. This is probably because we scrape the reviews for each airline separately and the majority of the reviews are talking about a specific airline. Regardless of that, we can still conclude that US airlines are relatively more associated with each other and the same for some non-US airlines like Ryanair, British Airway, and Emirates.  In our first iteration of calculation, we found that Frontier Airlines has no association with some other airlines (Lift=0) and this would be problematic for the MDS plot because the dissimilarity value will be infinity. Therefore, we exclude Frontier Airlines from lift analysis and MDS plot.

| Airline | united_airline | delta_airline | southwest_airline | american_airline | ryanair | emirates_airline | spirit_airline | british_airway | air_canada |
|---|---|---|---|---|---|---|---|---|---|
| **Airline** | | | | | | | | | |
| **united_airline** | NaN | 0.427 | 0.258 | 0.372 | 0.026 | 0.038 | 0.161 | 0.053 | 0.308 |
| **delta_airline** | NaN | NaN | 0.284 | 0.378 | 0.009 | 0.047 | 0.170 | 0.035 | 0.071 |
| **southwest_airline** | NaN | NaN | NaN | 0.248 | 0.022 | 0.004 | 0.184 | 0.005 | 0.005 |
| **american_airline** | NaN | NaN | NaN | NaN | 0.009 | 0.053 | 0.118 | 0.169 | 0.096 |
| **ryanair** | NaN | NaN | NaN | NaN | NaN | 0.046 | 0.009 | 0.405 | 0.019 |
| **emirates_airline** | NaN | NaN | NaN | NaN | NaN | NaN | 0.005 | 0.382 | 0.019 |
| **spirit_airline** | NaN | NaN | NaN | NaN | NaN | NaN | NaN | 0.014 | 0.005 |
| **british_airway** | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | 0.097 |
| **air_canada** | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN |

Figure 11. Lift table for airline-airline association
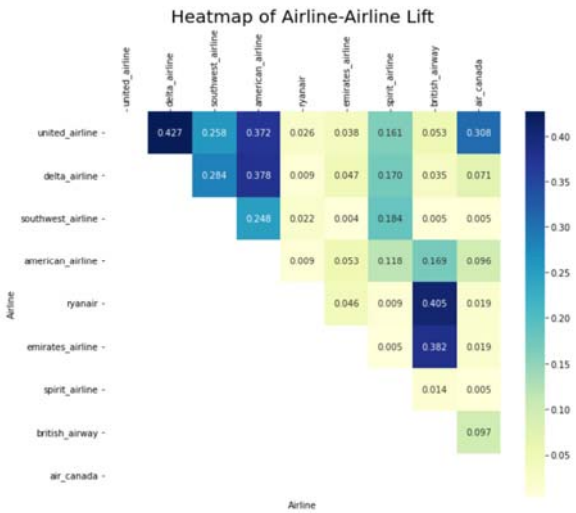
Figure 12. Heatmap for airline-airline Lift



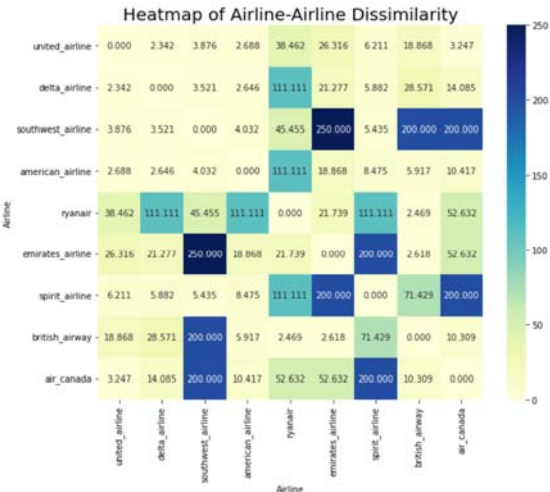Figure 13. Dissimilarity table for airline-airline association



Figure 14. Heatmap for airline-airline Dissimilarity

### 2.7.2 Airline-Attribute Lift

To study which attributes are strongly associated with each airline, the lift values between airlines and attributes are calculated following the same calculation process as for the lift between airlines except that the A and B in the formula become both the airline names and the attributes (generalized to "Word"). This method will also return the lift between airlines & airlines (redundant as it is repetitive) and airlines & attributes. We can extract the portion of the table for lifts between airline and attributes as shown in Figure 15. A heatmap for the airline-attribute lift table is shown in Figure 16.

| Word ⇕ | comfort ⇕ | in-flight_entertainment ⇕ | customer_service ⇕ | monetary_values ⇕ | cleanliness ⇕ | check-in_boarding ⇕ | food_beverage ⇕ |
|---|---|---|---|---|---|---|---|
| Word ⇕ | ⇕ | ⇕ | ⇕ | ⇕ | ⇕ | ⇕ | ⇕ |
| united_airline | 0.674 | 0.317 | 0.963 | 0.918 | 0.427 | 0.990 | 0.376 |
| delta_airline | 0.661 | 0.242 | 0.858 | 0.958 | 0.352 | 0.982 | 0.278 |
| southwest_airline | 0.809 | 0.337 | 0.970 | 1.019 | 0.883 | 0.975 | 0.478 |
| american_airline | 0.720 | 0.301 | 0.946 | 0.949 | 0.392 | 0.990 | 0.336 |
| ryanair | 1.204 | 0.494 | 1.008 | 1.187 | 1.454 | 1.079 | 0.886 |
| emirates_airline | 1.764 | 3.556 | 1.217 | 0.739 | 2.007 | 0.956 | 2.758 |
| spirit_airline | 0.633 | 0.181 | 0.854 | 1.290 | 0.374 | 1.012 | 0.384 |
| british_airway | 1.823 | 2.803 | 1.168 | 0.969 | 2.518 | 0.994 | 2.817 |
| air_canada | 1.200 | 1.764 | 1.105 | 0.872 | 1.338 | 1.003 | 1.528 |

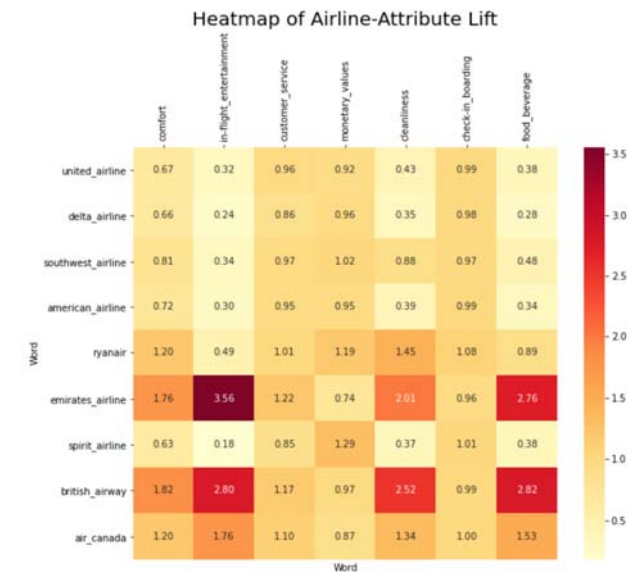Figure 15. Lift table for airline-attribute association

Figure 16. Heatmap for airline-attribute lift

### 2.7.3 Attribute-Attribute Lift

The lift between attributes can be extracted from the mixed table for airline and attribute lifts as shown in Figure 17. Dissimilarity is calculated as well (Figure 19). Heatmaps for attribute-attribute lift and dissimilarity are also provided (Figure 18, 20). From the heatmap, we can notice that the comfort, food & beverage, in-flight entertainment, and cleanliness attributes are highly associated with each other which suggests that customers tend to talk about them together in their reviews. This actually makes sense because these attributes are all related to in-flight experience.

| Word | check-in_boarding | customer_service | monetary_values | comfort | food_beverage | in-flight_entertainment | cleanliness |
|---|---|---|---|---|---|---|---|
| **Word** | | | | | | | |
| **check-in_boarding** | NaN | 1.011 | 1.014 | 1.006 | 1.011 | 0.983 | 0.992 |
| **customer_service** | NaN | NaN | 0.990 | 1.070 | 1.153 | 1.169 | 1.163 |
| **monetary_values** | NaN | NaN | NaN | 1.003 | 0.917 | 0.871 | 0.842 |
| **comfort** | NaN | NaN | NaN | NaN | 1.714 | 1.924 | 1.827 |
| **food_beverage** | NaN | NaN | NaN | NaN | NaN | 2.860 | 2.457 |
| **in-flight_entertainment** | NaN | NaN | NaN | NaN | NaN | NaN | 2.528 |
| **cleanliness** | NaN | NaN | NaN | NaN | NaN | NaN | NaN |

Figure 17. Lift table for attribute-attribute association
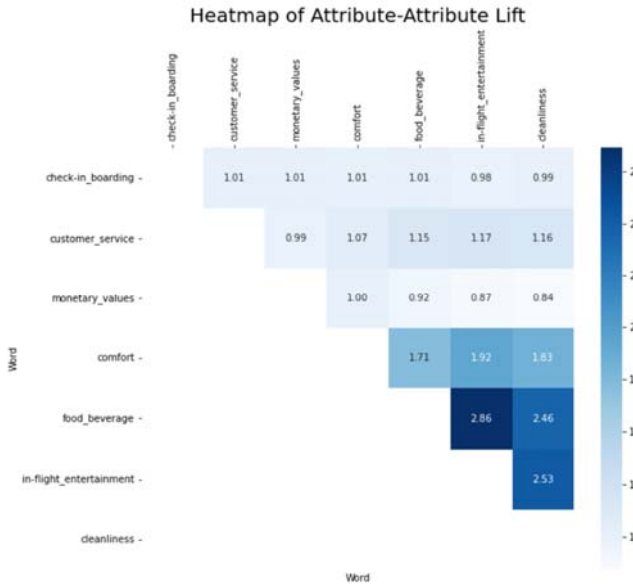


Figure 18. Heatmap for attribute-attribute Lift

| Word | check-in_boarding | customer_service | monetary_values | comfort | food_beverage | in-flight_entertainment | cleanliness |
|---|---|---|---|---|---|---|---|
| **Word** | | | | | | | |
| **check-in_boarding** | 0.000 | 0.989 | 0.986 | 0.994 | 0.989 | 1.017 | 1.008 |
| **customer_service** | 0.989 | 0.000 | 1.010 | 0.935 | 0.867 | 0.855 | 0.860 |
| **monetary_values** | 0.986 | 1.010 | 0.000 | 0.997 | 1.091 | 1.148 | 1.188 |
| **comfort** | 0.994 | 0.935 | 0.997 | 0.000 | 0.583 | 0.520 | 0.547 |
| **food_beverage** | 0.989 | 0.867 | 1.091 | 0.583 | 0.000 | 0.350 | 0.407 |
| **in-flight_entertainment** | 1.017 | 0.855 | 1.148 | 0.520 | 0.350 | 0.000 | 0.396 |
| **cleanliness** | 1.008 | 0.860 | 1.188 | 0.547 | 0.407 | 0.396 | 0.000 |

Figure 19. Dissimilarity table for attribute-attribute association
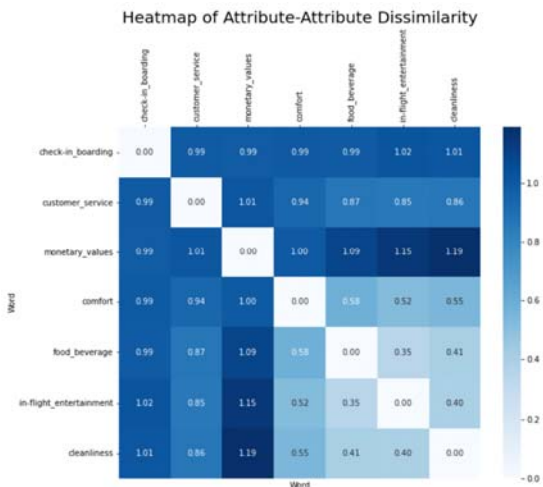
Figure 20. Heatmap of attribute-attribute dissimilarity matrix

# 3. MDS Visualization And Analysis

## 3.1 Airlines Clusters

The dissimilarity matrix displays the degree of variation between each airline. Therefore, smaller dissimilarity values indicate two airlines' affinity, while larger dissimilarity values indicate two airlines' distinction. From a machine learning perspective, we can understand values in the matrix as Euclidean distance between airlines, and with proper plotting, we are able to visualize patterns among all data points. Notice that the only available data here is airlines information, or predictors, we cannot construct a predictive model due to the lack of an outcome variable. Hence, unsupervised learning such as clustering is utilized to explore the relationships among all data points.

We choose K Means clustering to build our model. It's a popular unsupervised learning algorithm that is relatively simple to implement. guarantees convergence, can warm-start the positions of centroids and generalizes to clusters of different shapes and sizes. However, a disadvantage of K Means clustering is that the algorithm requires humans to manually indicate the number of clusters to generate. Since we lack an understanding of the data before machine learning, we can easily fall into the trap of choosing the wrong number of clusters and generate results with defected insights. A widely utilized methodology to resolve this issue is called the Elbow's method. It visualizes the performance of K Means clustering with different numbers of clusters with a line graph. The optimal number of clusters is located at the "elbow" position where model quality receives significant improvement. Here, with the airlines dissimilarity matrix calculated from lift ratio, we run a for loop of K Means clustering with the number of clusters ranging from 2 to 6. The Elbow Method plot is shown below.
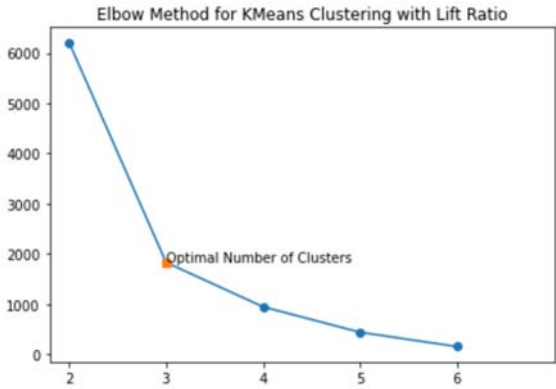


Figure 21. Elbow Method for Top 9 Airline Companies Clustering

It can be obviously observed that the last significant model improvement for airlines K Means clustering happens when the number of clusters equals three. Therefore, we choose 3 as the optimal number of clusters and re-run our algorithm. The algorithm outputs a specific cluster number (ranging from 1 to 3) for each airline. We then create a clustering label for each airline with its corresponding cluster number. Utilizing this label and dissimilarity matrix, along with the help of Python's *matplotlib* package, we are able to transform the clustering result into a multi-dimensional map as shown below.
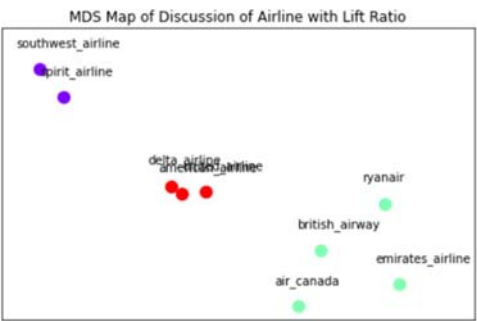
Figure 22. MDS Map of Discussion of Top 9 Airline Companies

## 3.2 Attributes Clusters

We also adopt similar methods for analyzing the 7 attributes extracted from the review forum. The Elbow Method suggests us split the 7 attributes into 3 groups. However, while we are running trials on the number of clusters, we find that the result is intuitively more explainable when the number of clusters equals 4. We then proceed to perform K Means clustering, as shown in Figure 14.
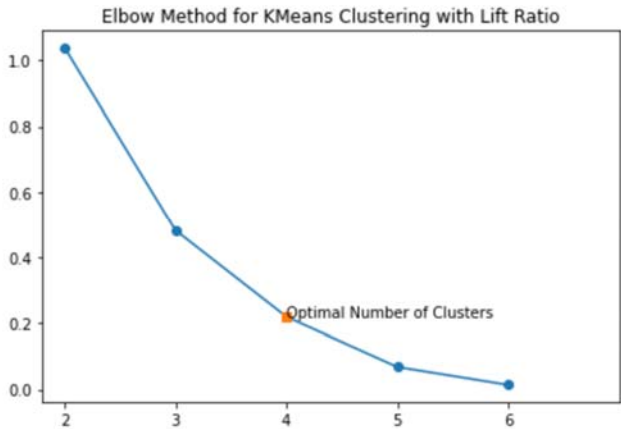


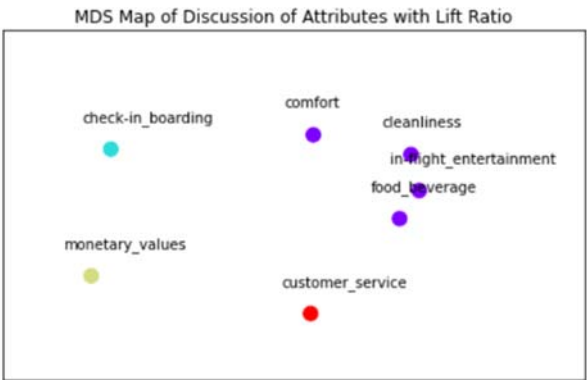Figure 23. Elbow Method for Top 7 Airlines Attributes



Figure 24. MDS Map of Top 7 Airlines Attributes

As we discussed above, k-means suggests that we use 3 clusters, but after implementing the MDS, we notice that 4 clusters can give a better explanation towards attributes. We have 3 clusters and each of them contains only one attribute, and 1 cluster contains 4 attributes. It's interesting to see from the attribute ranking (figure 24), the top 3 attributes belong to the 3 clusters separately, and the last four attributes are grouped together in the one cluster.

## 3.3 Clusters Analysis

From the result of airlines K Means clustering (see Figure. 22) we see that the Top 9 airlines can be split into three groups, with group 1 consisted of Southwest Airlines and Spirit Airlines, group 2 consisted of Delta Airlines, United Airlines and American Airlines, as well as group 3 consisted of British Airway, Air Canada, Ryanair and Emirates Airline.

We first take a look at airlines from group 1. Southwest Airlines is one of the major airlines in the United States and the largest low-cost carrier in the world. Headquartered in Dallas, Texas, it operates between 121 destinations in the United States and 10 in other surrounding countries. Spirit Airlines, the eighth largest passenger carrier in North America as of 2020, is an ultra-low-cost airline in the United States. Headquartered in Miramar, Florida, it operates between the United States, Caribbean and Latin America. These two airlines are both United States based and low cost, thri

ves on large scale operation with low marginal profit and marginal cost. Customers who want to look for a cheap airline ticket or a quick business trip in North America can choose any airlines within this cluster. However, low-cost airlines also operate with low marginal cost, which means any extra services will cost a significant amount of money. For example, several extra luggages might end up costing more than the original ticket price.

We then take a look at airlines from group 2. Delta Airlines, American Airlines and United Airlines are three major airlines of the United States. They are all internationally operated airline companies with services around the world, operating thousands of flights from and to hundreds of locations everyday. The on-board service is also divided into different classes: economic class, business class and first class, ranking from the cheapest to most expensive. However, since Delta, American and United are not low-cost airlines, even the lowest ticket price for economy class costs more than a regular ticket price from Southwest or Spirit Airlines. This also means the services offered from these airlines are with higher quality than cheaper airlines. For example, check-in luggages is free, and meals/drinks are offered during flight.

Last but not least, we take a look at group 3. Group 3 consists of airlines outside of the United States, including airlines from the UK, Ireland, Canada and United Arab, ranging from ultra low-cost Ryanair to luxury Emirates. We also see in the MDS map that the distance between each point represented by airlines in Group 3 is farther than the ones in Group 1 and Group 2. This means differences between each airline in Group 3 is quite large, which is intuitively understandable, since Ryanair, British Airway, Emirates and Air Canada are from three different scountries.

Based on business analysis, we believe the reason behind that interesting cluster distribution is, the top 3 attributes: check-in boarding, customer service and monetary values are essential in determining customer preference when they choose an airline, so each of them can be clustered into 1 group, and still has significant means. When we take a look at the last four attributes, we realize they are all the factors within the in-flight experience, so it's reasonable to group them into one cluster, and that compound cluster can represent those customers who care about the in-flight experience.

### 3.4 Topic Modelling

For topic modelling, we apply regular expression tokenizer and WordNet lemmatizer in nltk to process all words in the dataset. Then we use count vectorizer in the sklearn package to vectorize and count the words in the text. Next, we import *lda* package, which is a package for Latent Dirichlet Allocation, a model frequently used in topic modelling. We fit the model to our airline reviews by pre-specifying the number of topics we would like to discover. By trial and error, 3 topics suits best our data and the resulting topics are aligned with our results in MDS and Sentiment Analysis. Lastly, we extract the topic-word distribution and airline-topic distribution from the LDA model's output.

By combining our results from sentiment analysis and MDS clustering, we are able to interpret the results of the LDA topic modelling. As shown in the table, the resulting 3 topics have different association scores with the 9 airlines. We assign each airline to its most relevant topic that has the highest scores among all three topics. In this way, we end up with the same 3 clusters obtained from our MDS plot. From the topic-word distribution, we summarize and come up with 8 words that are most likely associated with each topic. From the table, Topic 1 contains Air Canada, British Airway, Emirates, and Ryanair, which reside in one cluster in the MDS plot. These airlines have positive sentiment scores except for Air Canada that is slightly negative. This implies that people are content with these airlines and the words appearing in this topic are positive and optimistic. These airlines provide good service, comfortable seats, and flavorful food. Furthermore, people mention them as good business class airlines. In fact, Emirates and British Airways are in the list of World's Best Business Class Airlines 2021 by World Airline Rewards. ([https://www.worldairlineawards.com/worlds-best-business-class-airlines-2021/](https://www.worldairlineawards.com/worlds-best-business-class-airlines-2021/)). Looking into the second topic where we have American Airlines, Delta Airlines, and United Airlines, we could see these three airlines are US based full-service airlines. From our online research, we found many articles complaining that US airlines are the worst airlines in the world. This finding is in line with what we obtained using text mining. Besides they have very low averaged airline sentiment scores, they also perform poorly on many areas of services such as check-in & boarding and in-flight services. The topic-word distribution table shows that the issues that are frequently discussed are canceled flights, delayed flights, changed gates at last minute. It appears these airlines have notorious reputations in terms of time, punctuality, and reliability. Our last topic consists of two US low-cost carriers: Southwest Airlines, the world's largest low-cost carrier, and Spirit Airlines, a US ultra-low-cost airline. Although these two airlines have opposite sentiment scores (positive sentiment for Southwest and negative for Spi

rit), being low-cost carriers, they share similar characteristics and thus are grouped together into another topic by the LDA. When talking about these two airlines, customers often mention services or areas related to money, price, payment, and refund. This makes sense since people who buy low-cost carriers' tickets are more sensitive to price. In conclusion, the three topics have discovered what are being discussed for the three airlines groups, and have informatively supplemented                    our               analysis               of               airlines               comparison.

| Airline | Topic 1 | Topic 2 | Topic 3 |
|---|---|---|---|
| Air_Canada | 0.4168 | 0.3863 | 0.1970 |
| American_Airline | 0.0906 | 0.5081 | 0.4014 |
| British_Airway | 0.7789 | 0.1057 | 0.1154 |
| Delta_Airline | 0.1000 | 0.4802 | 0.4199 |
| Emirates_Airline | 0.7821 | 0.1127 | 0.1051 |
| Ryanair | 0.4698 | 0.1603 | 0.3699 |
| Southwest_Airline | 0.2217 | 0.3423 | 0.4360 |
| Spirit_Airline | 0.0589 | 0.3928 | 0.5483 |
| United_Airline | 0.1005 | 0.4946 | 0.4049 |

*Figure 25. Airline - Topic Distribution*

| Topic 1 | Topic 2 | Topic 3 |
|---|---|---|
| Service | U.S. | Customer |
| Seat | Time | Ticket |
| Food | Delayed | Pay |
| Time | Canceled | Never |
| Business | Gate | Money |
| Good | Hour | booked |
| Crew | Day | refund |
| friendly | Minutes | Price |

*Figure 26. Summary of Topic - Word Distribution*

# 4. Managerial Insights & Recommendations

## 4.1 For Customers

All the reviews collected from the forum are written by customers who had past experience with certain airlines or heard how it went from their friends or relatives. Therefore, the reviews extracted could reflect customer preference towards certain airlines and attributes. We could implement sentiment scores on the 7 attributes to do the customer segmentation analysis and give our recommendation. Going through each attribute, we find that Emirates Airline achieves the highest sentiment score among almost all the attributes, so we would recommend it to customers who prefer excellent customer service experience and don't care too much about cost-effectiveness. On the contrary, for customers who are price-sensitive, we would recommend Southwest Airline, which is the world's largest low-cost carrier, it achieves the highest sentiment score in monetary values. For customer care about their time, and try to avoid any accident in delay, we would recommend British Airline, it does a great time managed work at on-time checking-in & boarding. After we sum all attributes related to in-flight experience (Comfort, Cleanliness, In-flight Entertainment, and Food & Beverage), Ryanair gets the highest score among low-cost airlines while Emirates gets the highest score among full-service airlines. We would recommend Ryanair to customers for domestic or short-range flight. For cost-effective purposes, we would recommend British Airway to customers of mid-to-long-range flights who pay attention to in-flight experience. In summary, our recommendation to customers only considers sentiment score from text analytics for this report, but we know there are other useful metrics and techniques used by the airline industry to analyze customer preference and segmentation by demographic, psychographic or behavioral data to give proper recommendation. We may further extract customers information like username, location, data and time and the others into our analysis.

## 4.2 For Airlines

We can also recommend the airlines to improve certain aspects of their services based preliminarily on the sentiment analysis results with lift analysis results as complement. The lift analysis between airline and attributes provides information about what services of this airline customers more frequently talked about. With the help of the sentiment analysis dashboard we can easily investigate the impact of certain attributes on the average sentiment score of the revie

ws of an airline. This impact can be determined by observing the overall average sentiment score for all the reviews of an airline, the average sentiment scores for reviews that mention or do not mention this particular attribute.

Figure 27 shows the result of sentiment analysis for Air Canada as an example. The left panel shows the average compound sentiment score of -0.066 for all reviews for Air Canada. The panel on the right shows the average sentiment score of all the reviews that mention or do not mention certain attributes. As an example, the average sentiment score of all the reviews that mention customer service attribute (the first one in the panel) is -0.047, while it is -0.138 for all the reviews that do not mention customer service attribute. Since the -0.138 < -0.066, we can imply that the average sentiment score would be even lower without those reviews that mention customer service attributes. In another word, the reviews for customer service would bring the overall average sentiment score up from -0.138 to -0.066. Therefore, we can conclude that the customer service attribute has a positive impact on the customer review for Air Canada which in turn suggests that Air Canada is doing a good job in their customer service. We would recommend Air Canada to keep it up.

On the other hand, the average sentiment score of all the reviews that mention check-in & boarding attribute is -0.099, while it is +0.119 for all the reviews that do not mention check-in & boarding attribute. Since the +0.119 > -0.066, we can imply that the average sentiment score would be even higher without those reviews that mention check-in & boarding. In another word, the reviews for check-in & boarding would bring the overall average sentiment score down from +0.119 to -0.066. Therefore, we can conclude that the check-in & boarding attribute has a negative impact on the customer review for Air Canada which in turn suggests that Air Canada is not that good at providing customers with good check-in & boarding experience. We would recommend Air Canada to improve their check-in & boarding service efficiency and quality.

The same analysis can be applied to all of the attributes for Air Canada and also to all the airlines as well. In this way, we can provide recommendations targeted to each airline.
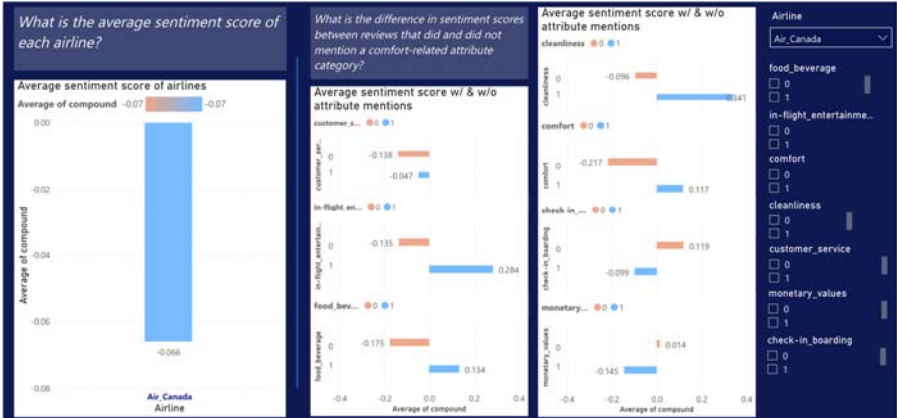


Figure 27. Sentiment analysis dashboard display for Air Canada