ISSS609 Text Analytics & Applications

Sentiment & Topic Analysis in Aviation Industry During Covid-19

Project Report

Group 10

Andre Justin Lee Sheng Wei

Choong Shi Lian, Selene

Jiang Weiling, Angeline

Son Yejin

Wong Jia Wei

Date of Submission: 8 August 2020

1. Introduction

The COVID-19 pandemic has resulted in a significant impact on the commercial aviation industry globally due to severe travel restrictions imposed by numerous countries as a measure to combat the virus. No one in the industry has been spared, with many suspending their flights and scaling back operations (Orion, 2020). This has prompted bankruptcies and massive layoffs in many aviation companies, with many stakeholders in the aviation industry seeking financial support from governments. Given the constant changes and uncertainties in the market outlook, it is difficult for stakeholders in the aviation industry to continuously monitor the market changes and plan for long term measures to tide through the pandemic. As such, it is crucial for the aviation players to understand how the industry outlook has evolved over the pandemic period and identify areas of concern or potential growth in the industry.

The dataset used for this project consists of global daily news articles related to aviation and aerospace industries, which were scrapped from FlightGlobal[1] – a media platform specialising in aerospace. As the scope of this project aims to focus on the COVID-19 period, we only looked at the articles published between 1st January to 30th June 2020 (which consists of a total of 2389 articles). Two analytical tasks – Sentiment Analysis (Section 3.1) and Topic Modeling (Section 3.2) – were performed to derive useful insights for stakeholders in the aviation industry from the extracted news articles.

Sentiment analysis aims to provide the sentiment change in the aviation industry over time to allow stakeholders in the industry to evaluate the existing market situation and make long term decisions. For example, airlines and airports can assess the potential for recovery in commercial travel demand by analyzing the sentiment trends of the articles over time and plan for long-term business strategies, such as the gradual resumption of flights. Apart from airlines, aviation authorities can leverage on findings of sentiment analysis to review regulatory policy based on the sentiments of the global aviation industry.

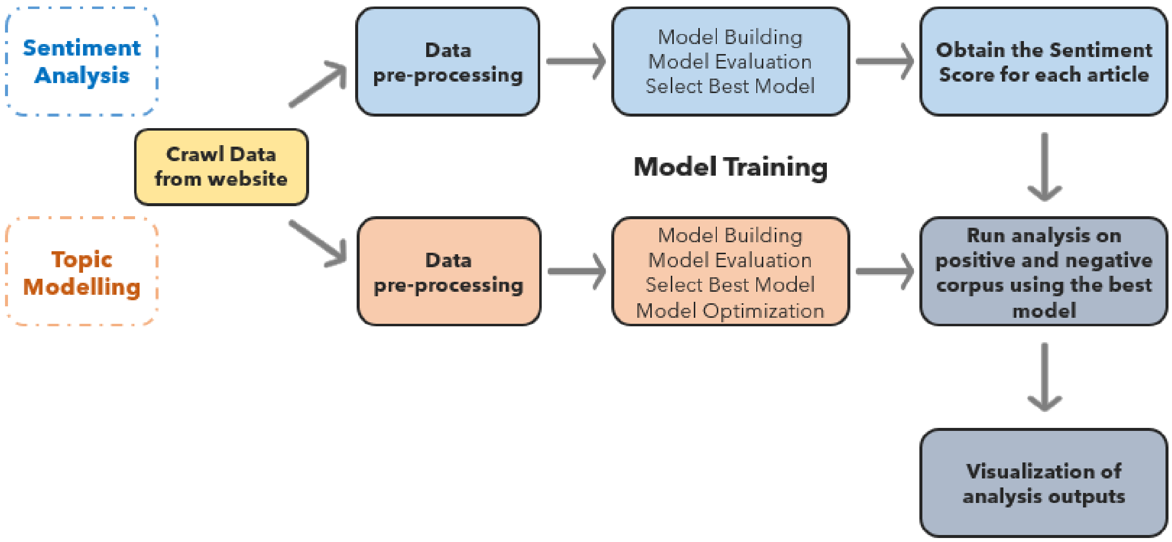
On top of understanding the sentiments, topic modelling will be applied to the outputs of sentiment analysis to delve deeper and understand the trending keywords among the positive and negative sentiments. This allows aviation authorities to prioritize support for parts of the aviation sectors depending on their current outlook and look out for any unexpected or newly emerged topics under each sentiment. Businesses in the sector are also able to gather pointed insights from the plethora of news articles on the industry based on their specific topics of interest and strategically plan business focus.

2. Solution Overview

The tasks dealt with two forms of challenges – input data and analysis challenges. As the dataset was crawled directly from the FlightGlobal website, the input corpus might not be in the most appropriate form for applying the respective analyses (e.g. lexicons with varied inflectional endings or "noisy" aviation-specific terms used across the articles). The sentiment analysis then presented the challenge of determining the optimal model to generate appropriate sentiment scores without any pre-labels available on the dataset, while the topic modelling posed the challenge of appropriately determining the topics of each cluster, while ensuring that we extract topics of acceptable quality that are segregated and meaningful.

We first performed data pre-processing and model evaluation for sentiment analysis and topic modeling separately to determine the most optimal model for each analysis. Upon determining the most suitable sentiment model, sentiment analysis will first be performed on the full dataset using that model to obtain the sentiment score for each news article. The sentiment scores will subsequently be used to segregate the dataset into the positive and negative corpuses, before the optimal topic modeling technique will be applied on both corpuses to obtain the topic clusters within each corpus. A brief overview of the solution process flow is as shown below:

[1] https://www.flightglobal.com/news



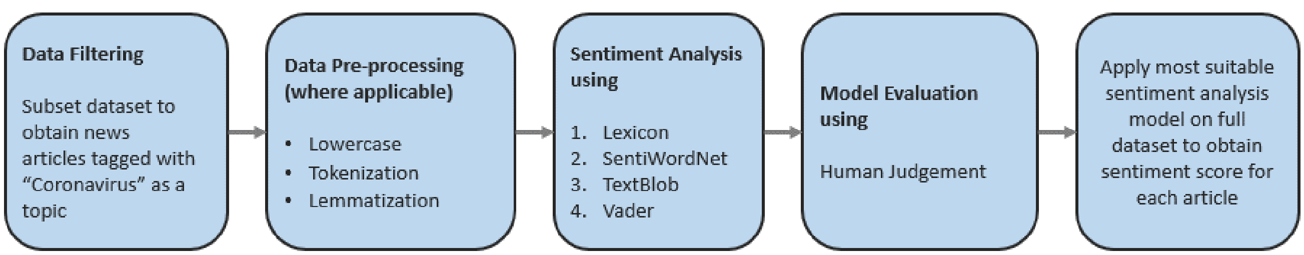
3. Solution Details

3.1 Sentiment Analysis

Sentiment analysis is the interpretation and classification of text data into polarity. For our project, we are interested in classifying global news articles in the aviation and aerospace industries into 3 broad sentiments, namely (i) positive, (ii) negative, and (iii) neutral. As pre-labels are not available, unsupervised and lexicon-based approach will be required to classify the articles into their respective polarities. We will be considering four sentiment analysis techniques, namely (i) Lexicon, (ii) SentiWordNet, (iii) TextBlob and (iv) Vader, with each to their benefits and limitations.

To determine the most suitable sentiment analysis technique, the dataset was subset to obtain the news articles tagged with “Coronavirus” under the topic in FlightGlobal website (comprised of 708 articles i.e. ~30% of the full dataset). By analyzing aviation news related to “Coronavirus”, human evaluation will most likely be able to identify any misclassification more accurately as intuitively, the airline industry was badly affected by travel restrictions etc.

Thereafter, the most suitable sentiment analysis technique will be applied on the full dataset to obtain the overall sentiment of the airline industry in the first half of 2020. A brief process flow is as shown below:



<u>Data Pre-processing and Model Training</u>

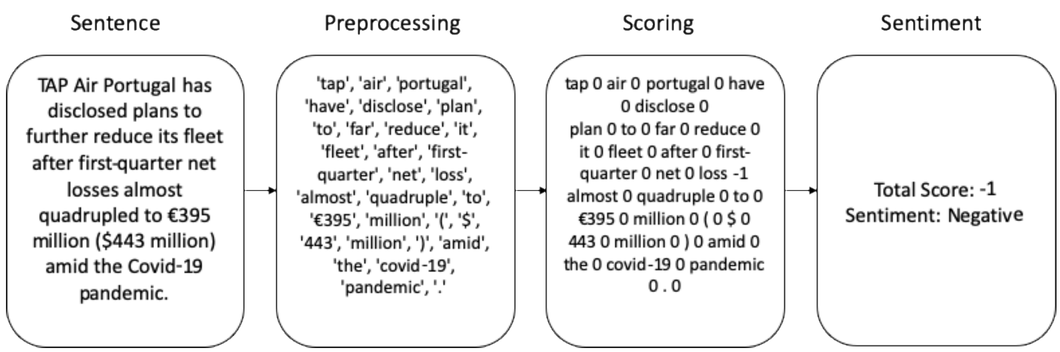
(a) Lexicon

The sentiment lexicon used here has been built and compiled by Liu and Hu since 2004. It consists of around 6800 positive and negative sentiment words in the English language.

Tokenization and subsequently lowercasing of the tokenized words were done for all the articles as part of pre-processing. Lemmatization was then carried out to return the words to their root dictionary form. This was because inspection of the lexicon revealed that for a single root word, the lexicon did not always contain all the variations with inflectional endings. For example, the word ‘approve’ appears as a positive word, but its past tense ‘approved’ was not in the list of positive words, although ‘approved’ should arguably receive a positive score as well. In this case, lemmatization can help to minimize underscoring.

We scored the pre-processed form of each article by going through all the words and counting how many positive and negative words it contains from the lexicon. Each positive word received a score of 1 while each negative word received a score of -1, and words not found in the lexicon did not receive any score. Eventually, an article’s score is the sum of the scores of all its words. The article is judged to have Positive sentiment if the overall score is greater than 0, Negative sentiment if the overall score is less than 0, and Neutral sentiment if the overall score is equal to 0.

A more detailed look at how the sentiment of a single sentence in an article is scored using the Lexicon model is shown in the flowchart below:



(b) SentiWordNet

SentiWordNet is a lexical resource that contains a list of English terms which have been assigned a score of positivity, negativity and objectivity (or neutrality). Each synset of WordNet gets a score between 0 and 1 for each of the three categories mentioned above, and the total score of the three categories will sum up to 1.

Tokenization and lowercasing of the tokenized words were done for all the articles as part of pre-processing. After each article has been pre-processed, we assign a score to each word in the article using the word’s first synset. Each word’s score is derived by subtracting the negative score from the positive score. After we get the score for each word, we then take the article score to be an average of all its words’ scores. For our context, if the article score is ≥0.01, the article is judged to have Positive sentiment. If the article score is ≤-0.01, the article is judged to have Negative sentiment. Otherwise, it is judged to have Neutral sentiment.

(c) TextBlob

TextBlob sentiment module contains two implementations, namely (i) PatternAnalyzer based on the pattern library – default implementation and (ii) NaiveBayesAnalyzer – an NLTK classifier trained on movie reviews. Using the default implementation, TextBlob returns 2 values, ‘Polarity’ and “Subjectivity”. The polarity score, a value between -1 to 1, is assigned to the text based on the most commonly occurring positive and negative adjectives. While, the subjectivity score will be assigned to the text with a value between 1 (opinion, emotion or judgement) to 0 (factual information). For the purpose of our study, we will only consider the polarity score.

As TextBlob can handle modifier words, different inflected forms of a word and is case insensitive, no pre-processing such as lowercase conversion and lemmatization were done on the dataset prior to implementation.

(d) Vader (Valence Aware Dictionary for Sentiment Reasoning)

Vader is a rule-based and lexicon sentiment analysis tool which was built to detect sentiments expressed in microblog-like contexts. In addition, Vader is sensitive to both the polarity (positive, negative, neutral) and the intensity of the sentiments expressed in social media contexts such as Twitter, Facebook, Instagram etc. Hutto and Gilbert (2014) study indicated that Vader also works well on texts from other domains.

Unlike typical bag-of-words model, Vader also implements the grammatical and syntactical rules, and hence is sensitive to word-order between terms in sentence-level text. For instance, degree modifiers such as “extremely good” will get higher sentiment intensity compared to “good”. Vader can also handle punctuations (“It’s great!!!), negations (“not great”), slangs (“kinda”, “hella”), emoji etc. Hence, the only pre-processing done was to convert all the words to lowercase though the polarity score obtained was similar with or without lowercase conversion.

Using the lexicon, each word will be tagged to a valence score, and thereafter summed and normalized to obtain a compound score (also called ‘normalized, weighted composite score’) between -1 (most extreme negative) and +1 (most extreme positive).

Model Evaluation

Applying the four algorithms on the articles which were tagged by FlightGlobal with “Coronavirus” as a topic, a summary of the polarity scores is shown in Table 1. Among the four techniques, SentiWordNet and TextBlob classified 69% to 87% of the news articles as Positive sentiments. However, based on human evaluation, majority of these articles were incorrectly classified as positive sentiment (Table 2). As the performance of these two techniques were poor, they were deemed as not suitable for our data domain.

Performance of Lexicon and Vader were similar in terms of proportion of articles labelled as Negative sentiment. However, compared to Lexicon, Vader had only classified 1 article as having neutral sentiment and 10% more articles with Positive sentiment than Lexicon (Table 1). Given that Vader was built specifically on social media such as Twitter which has a maximum of 140 characters, the technique may not perform as well as expected on news articles which has more than 140 characters. As a result, Lexicon was determined to be the most suitable technique for our context.

Table 1: Results from Sentiment Analysis using articles with “Coronavirus” tag

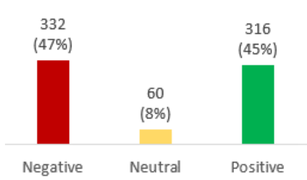
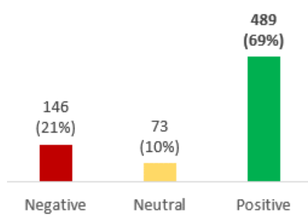
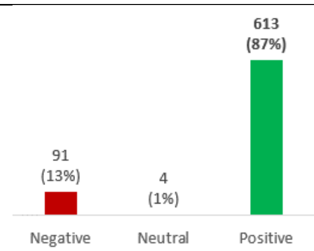
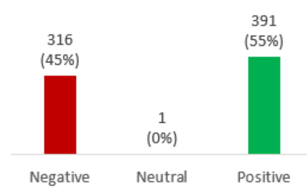
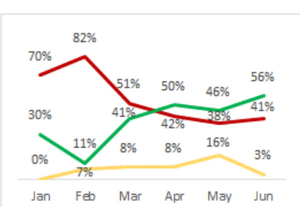
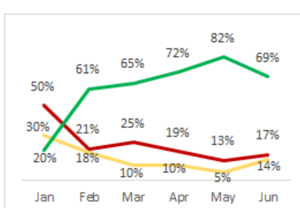
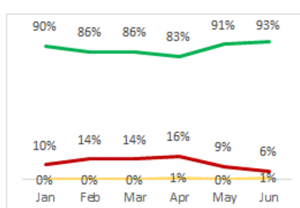
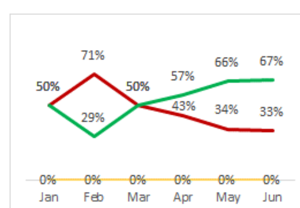
Lexicon	SentiWordNet	TextBlob	Vader
Pre-processing			
Tokenization Lowercasing Lemmatization	Tokenization Lowercasing	None *TextBlob is not case sensitive	Changed words to lowercase
Overall Results			
			
Results by Months Neg Neu Pos			
			

Table 2: Examples of incorrectly labelled news articles from SentiWordNet and TextBlob

Article Title	Article Overview	Scoring Details
Norwegian cancels 97 Boeing aircraft orders[2]	The article is about how Norwegian, a low-cost carrier from Norway, has cancelled orders for 97 Boeing aircraft, some of which included the now-grounded 737 Max. The article also detailed how the airline had suffered a long hiatus of flights following the coronavirus pandemic.	The article got a net Positive sentiment score of 0.053 (SentiWordNet) and 0.024 (TextBlob). Human evaluation of the article however leads one to conclude that it should unanimously be labelled as a Negative article, as did the Lexicon model. A detailed look at how the first synset of the individual words were scored by SentiWordNet sheds light on how the article got a Positive score. For example, the adjective 'financial' has a first synset of 'fiscal', for which it has a 0.25 Positive score. However, in the context of the article, the word was just part of the term 'financial restructuring', and thus should not have a polarity score at all. Another example would be the verb 'has', for which it has the first synset of 'have'. 'Have' receives a 0.25 Positive score as well, but one would argue that it should be a neutral term. The aggregation of multiple incorrectly labelled words thus leads to the article being wrongly classified as a Positive article.
Singapore Airlines and SilkAir reduce capacity to China over coronavirus outbreak[3]	The article is about how Singapore Airlines and SilkAir suspended flights to multiple destinations in China as the coronavirus outbreak caused "weak demand and operational constraints". The article also highlighted examples of other airlines which had also cut flights to China.	The article got a net Positive sentiment score of 0.048 (SentiWordNet) and 0.045 (TextBlob). Human evaluation of the article however leads one to conclude that it should unanimously be labelled as a Negative article, as did the Lexicon model. A detailed look at how the first synset of the individual words were scored by SentiWordNet sheds light on how the article got a Positive score. For example, the verb 'cut' has a same first synset of 'cut', for which it mysteriously has a 0.125 Positive score. The word cut was used in the context of cutting flights, which obviously should receive a Negative score if any. Also, the adverb 'well' had a 0.375 Positive score. This word was used in the sentence 'The carrier has also reduced frequencies to eight other Chinese cities, including Guangzhou, Kunming, Nanjing, Tianjin as well as Macau.' 'As well as' is a conjunction that should not assigned any polarity score. The aggregation of multiple incorrectly labelled words thus leads to the article being wrongly classified as a Positive article.

[2] https://www.flightglobal.com/fleets/norwegian-cancels-97-boeing-aircraft-orders/139050.article
[3] https://www.flightglobal.com/strategy/sia-group-slashes-capacity-to-china-over-coronavirus-outbreak/136466.article

Model Implementation

With the most suitable model i.e. the Lexicon model, we applied the model on the universe of news articles (i.e. full corpus) along with the same pre-processing rules applied to the training model for Lexicon to obtain the sentiment score of each article. A screenshot of the output is as shown below, where the sentiment labels ("sentiment_lexicon") will be used as input for topic modelling.

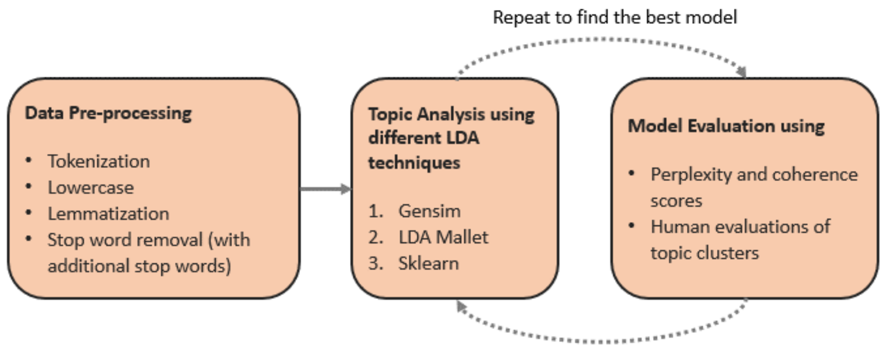
	PageLink	Title	Article	Month	Topic	sent	lem	score	sentiment_lexicon
0	https://www.flightglobal.com/fleets/tap-to-cut...	TAP to cut more aircraft as losses mount	TAP Air Portugal has disclosed plans to furthe...	6	Coronavirus, Europe, Fleets, Networks	[tap, air, portugal, has, disclosed, plans, to...	[tap, air, portugal, have, disclose, plan, to,...	4	positive
1	https://www.flightglobal.com/aerospace/europea...	European Aviation responds to pandemic with pl...	For most of the aviation sector, the coronavir...	6	Aerospace	[for, most, of, the, aviation, sector, ,, the,...	[for, most, of, the, aviation, sector, ,, the,...	4	positive
2	https://www.flightglobal.com/airlines/easyjet-...	EasyJet proposes UK base closures including tw...	UK budget carrier EasyJet is proposing to clos...	6	Air Transport, EasyJet, Europe	[uk, budget, carrier, easyjet, is, proposing, ...	[uk, budget, carrier, easyjet, be, propose, to...	-1	negative
3	https://www.flightglobal.com/airlines/easyjets...	EasyJet's agreement with Stelios ends as his s...	EasyJet's relationship agreement with founder ...	6	Airlines, Europe	[easyjet, 's, relationship, agreement, with, f...	[easyjet, 's, relationship, agreement, with, f...	-2	negative
4	https://www.flightglobal.com/business-aviation...	Gulfstream G280 hits 200th-delivery milestone	Gulfstream delivered the 200th G280 in late Ju...	6	Business Jets	[gulfstream, delivered, the, 200th, g280, in, ...	[gulfstream, deliver, the, 200th, g280, in, la...	1	positive
...	...	...	...	...	...	...	...	...	...
2384	https://www.flightglobal.com/airlines/iata-see...	IATA sees strong response to airline gender di...	IATA is pointing to a strong response from air...	1	Airlines	[iata, is, pointing, to, a, strong, response, ...	[iata, be, point, to, a, strong, response, fro...	16	positive
2385	https://www.flightglobal.com/business-aviation...	Airbus Corporate Helicopters and Aston Martin ...	Airbus Corporate Helicopters (ACH) is partneri...	1	Aerospace, Business & General Aviation, Europe...	[airbus, corporate, helicopters, (, ach,), is...	[airbus, corporate, helicopter, (, ach,), be,...	18	positive
2386	https://www.flightglobal.com/news/wake-separat...	Wake separation adequate before Bek Fokker 100...	Kazakhstan's air navigation service is rejecti...	1	Air Transport, Asia Pacific, Safety	[kazakhstan, 's, air, navigation, service, is,...	[kazakhstan, 's, air, navigation, service, be,...	-6	negative
2387	https://www.flightglobal.com/military-uavs/ind...	Indonesia rolls out indigenous MALE UAV	Indonesian Aerospace has rolled out the first ...	1	Asia Pacific, Military UAVs	[indonesian, aerospace, has, rolled, out, the,...	[indonesian, aerospace, have, roll, out, the, ...	-1	negative
2388	https://www.flightglobal.com/helicopters/taiwa...	Taiwan to investigate UH-60M crash that killed...	Taiwan will probe the fatal crash of a Sikorsk...	1	Asia Pacific, Defence, Helicopters, Safety	[taiwan, will, probe, the, fatal, crash, of, a...	[taiwan, will, probe, the, fatal, crash, of, a...	-7	negative

2389 rows × 9 columns

3.2 Topic Modelling

Topic modelling, also known as topic analysis, is the discovery of topics within the documents where each topic is a set of words frequently co-occurring together. For our project, we are interested to cluster the articles within the positive and negative corpuses classified using sentiment scores from sentiment analysis to discover the broad topics of the articles in each corpus and understand the aspects of aviation industry which were performing better or suffering over the pandemic period. As each article might contains multiple topics, the unsupervised Latent Dirichlet Allocation (LDA) method was used. We will be evaluating three topic modelling techniques, namely (i) Gensim, (ii) LDA Mallet and (iii) sklearn, to determine the most optimal model for our context.

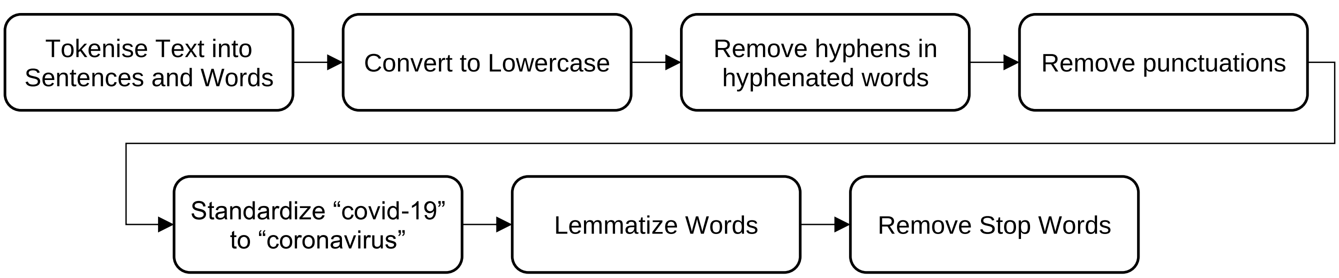
Pre-processing of the corpus must first be done before performing LDA. Performances of the LDA models are largely dependent on the corpus analysed. As such, model evaluation was conducted through assessing the perplexity and coherence scores of the three LDA models (where applicable). While these scores are convenient measures to judge the models, human evaluations were additionally performed to ensure that the generated topic clusters could be summarized into meaningful topics. Based on the best performing model, an interactive topic model was visualised for the business users to better understand the technical outputs. A brief process flow is as shown below:



Data Pre-processing

During the pre-processing stage for topic analysis, the corpus of news article was first tokenized to split the text into sentences and the sentences into words, before converting the words to lowercase. To avoid removing words with hyphens during punctuation removal through regular expression method, hyphens were first excluded from those words before removing punctuations. Articles interchanges between the terms “coronavirus” and “covid-19” which could affect effectiveness of clustering, hence both terms were standardised to “coronavirus”. Stemming was not performed to keep the words in dictionary form to allow for easier inspection of the learned topics. Instead, we performed lemmatization using NLTK WordNet Lemmatizer to further facilitate effective clustering of non-repetitive words. Following initial preliminary analysis of the topic clusters, we removed stop words with an additional custom list of stop words on top of the NLTK stop word English list to further reduce the “noise”.

A summary of the text pre-processing steps is shown below:



Apart from the corpus, a dictionary containing all the words with numerical identifiers was also created to be used as an input to the LDA model.

Model Training

LDA is a generative probabilistic model of a corpus (Blei, Ng, & Jordan, 2003) which is commonly used to perform topic analysis. Prior to running LDA on the corpuses of positive and negative sentiments to obtain a summary of the major themes for each sentiment corpus, we first used the overall corpus to build the optimal model. In this paper, we tested LDA algorithms from three different implementations – namely Gensim, Mallet and sklearn.

(a) Gensim (“Generate Similar”)

Gensim is a technique to perform unsupervised semantic modelling from plain text. The LDA implementation transforms documents from bag-of-words counts into a latent topic space of lower dimensionality. Topics generated are probability distributions over words, and these distributions are inferred from the training corpus using the Variational Bayes sampling method. The pre-processed corpus was first converted into vectors for implementing the Gensim model. Parameter tuning was performed in Gensim, where “alpha” was set to ‘auto’ for the model to learn an asymmetric prior from the corpus. As the model also required input of the number of topics (k), we iterated the model between 4 to 20 topics. The outputs generated for each k value is k number of lists of topics each comprising of the probability distributions of words.

(b) Mallet (MAchine Learning for LanguagE Toolkit)

LDA Mallet is a Java-based package developed by University of Massachusetts at Amherst. The implementation learns the probability distributions based on Gibbs Sampling, which often offers a better quality of topics than the Gensim's standard LDA albeit running slightly slower. To perform LDA using the Mallet implementation, the pre-processed corpus needs to be converted into vectors as well. Parameter finetuning was performed where "iterations" was set at 1000 and "optimize_interval" was set at 100, meaning that the model will optimize hyperparameters every 100 iterations. Number of topics (k) was also set to iterate and build models between 4 to 20 topics. K number of topic lists containing the probability distributions of words were generated as the output for the Mallet implementation.

(c) sklearn (Sci-kit Learn)

sklearn is an LDA technique that learns the probability distributions using an online variational Bayes algorithm. For sklearn implementation, the pre-processed corpus is converted into a document-term matrix with term frequencies by first initialising the CountVectorizer class and applying fit_transform to create the matrix. Similar to the previous two implementations, models between 4 to 20 topics were built for sklearn as well. The outputs generated by sklearn is a list of topics with each topic represented as a list of terms but without any probability distributions shown.

Model Evaluation

To determine the best performing implementation for our FlightGlobal corpus, we first used perplexity score – a common supported intrinsic evaluation metric across all three implementations. Perplexity is a measure of the log-likelihood on the test data (exp(-1. * log-likelihood per word)), where the lower the perplexity, the better the model. To compare the perplexity scores, the evaluation was performed on a test data set containing 108 FlightGlobal articles from 1 to 8 July 2020. Due to time constraints, we were not able to evaluate perplexity using a larger test dataset. Based on the perplexity scores, sklearn implementation did not perform as well as the other two models with high perplexity and increasing perplexity with number of topics.

Table 3: Perplexity Scores of Topic Models for 4 to 20 topics

# Topics	**Gensim**	**LDA Mallet**	**sklearn**
4	-8.90874	-8.43984	60089.54801
5	-9.02440	-8.43337	97095.76973
6	-9.14986	-8.46574	107194.4752
7	-9.23564	-8.46437	168567.0805
8	-9.39326	-8.47162	182254.7746
9	-9.51098	-8.4826	159106.8298
10	-9.60777	-8.50025	211537.5585
11	-9.69490	-8.51341	197137.075
12	-9.82617	-8.52613	210325.733
13	-9.91552	-8.52091	217480.5644
14	-10.04324	-8.53493	210941.998
15	-10.04905	-8.53873	240629.9993
16	-10.15314	-8.55143	248649.9244
17	-10.18972	-8.5564	248270.7704
18	-10.33980	-8.5782	313380.8702
19	-10.46335	-8.57417	289488.0072
20	-10.47458	-8.61087	294843.1227

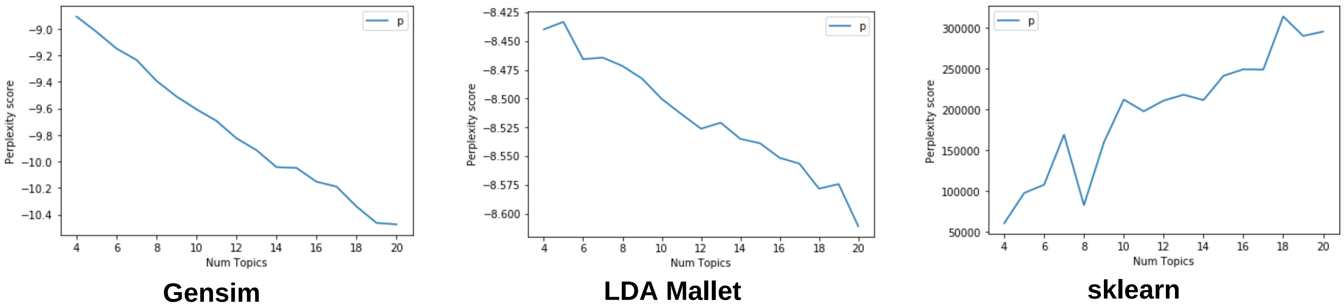


Figure 1: Graph for perplexity scores for overall corpus with each model

As perplexity scores were difficult to be interpreted by business users, we adopted the coherence scoring to further evaluate the Gensim and LDA Mallet implementations. Coherence is a measure of the degree of semantic similarity between the top words in the topic, where the higher the coherence, the better the model. Based on results in Table 3 and 4, LDA Mallet gave the best performance on both perplexity and coherence scoring (lowest perplexity and highest coherence) for our corpus.

Table 4: Coherence Scores of Topic Models for 4 to 20 topics

# Topics	**Gensim**	**LDA Mallet**
4	0.30297	0.52523
5	0.30232	0.53941
6	0.30689	0.54654
7	0.31037	0.54654
8	0.30427	0.52366
9	0.30377	0.54283
10	0.32230	0.53044
11	0.32220	0.49723
12	0.33176	0.53220
13	0.32668	0.51470
14	0.32825	0.53508
15	0.31636	0.52841
16	0.32066	0.54985
17	0.33201	0.52296
18	0.34998	0.52621
19	0.34346	0.54258
20	0.33801	0.51811

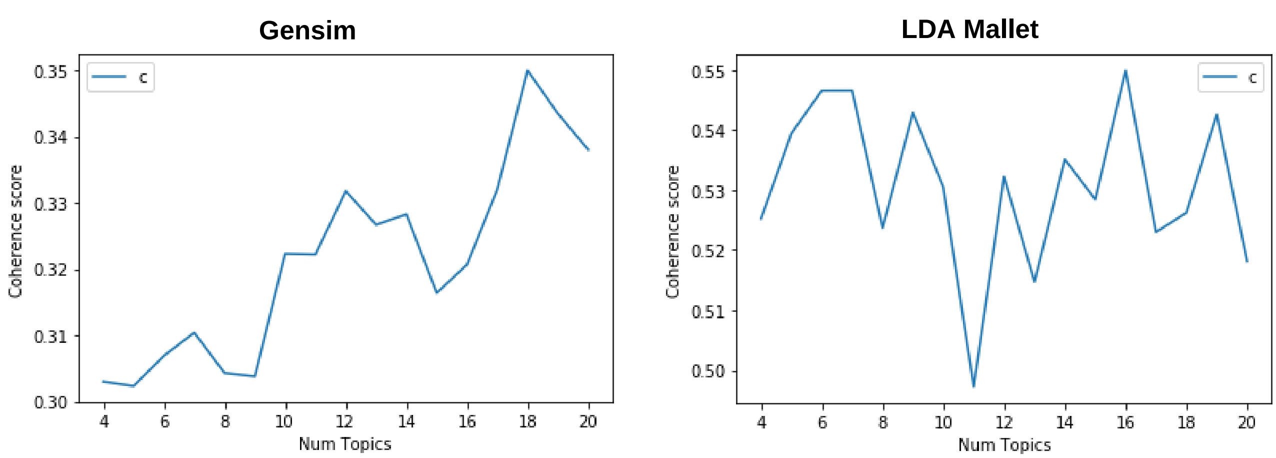


Figure 2: Graph for coherence scores for overall corpus with Gensim and Mallet Models

Visualisation of topics using pyLDAviz was also done to ensure minimal overlaps between the topic clusters. As each of the implementations have an optimal number of topics, we created the visualizations below based on their respective optimal k values. The visualizations in Figure 3 showed that the LDA Mallet has the least overlaps, whereas Gensim had some overlaps between multiple clusters and sklearn has major overlap between two of its clusters.

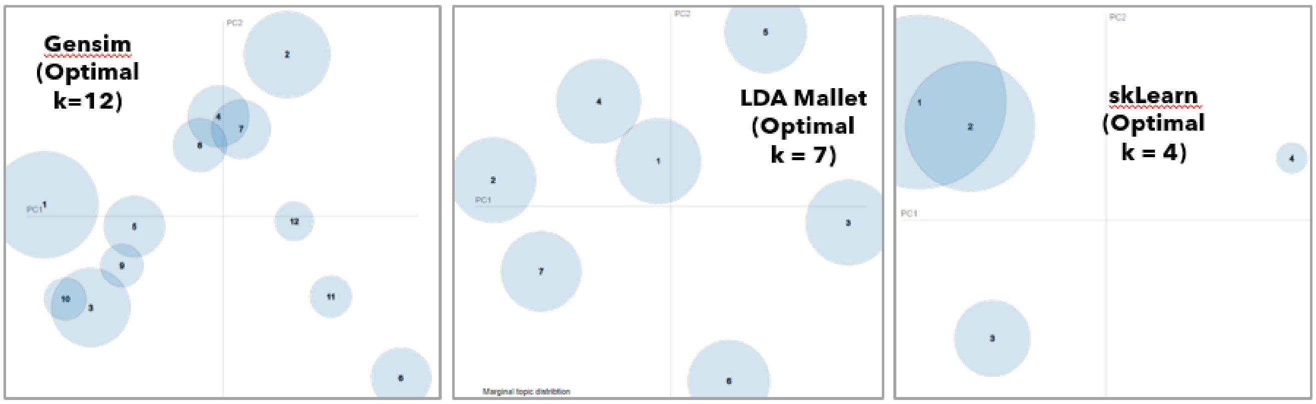


Figure 3: pyLDAviz visualization of topic models from Gensim, LDA Mallet and sklearn

While these metrics were useful for evaluating the model performances, they do not address the goal of topic modeling which is to create interpretable topic clusters (Chang, Boyd-Graber, Gerrish, Wang, & Blei, 2009). As such, human evaluation was performed on the top 10 most frequent words generated for each topic for LDA Mallet with the most optimal k to assess the quality of the topics and ensure the interpretability. As the corpus revolves around aviation industry, we chose the smaller k (i.e. 7) with the highest coherence score. Using these ten words, we manually assessed each topic to ensure that these topics are relevant to the research question and similar in context.

```
(0, '0.014*"government" + 0.010*"plan" + 0.010*"carrier" + 0.009*"business" + 0.009*"state" + 0.008*"agreement" + 0.007*"suppor
t" + 0.007*"uk" + 0.007*"share" + 0.007*"add"')
(1, '0.017*"crew" + 0.013*"runway" + 0.012*"pilot" + 0.010*"faa" + 0.008*"safety" + 0.008*"approach" + 0.007*"land" + 0.007*"en
gine" + 0.007*"investigation" + 0.007*"inquiry"')
(2, '0.026*"carrier" + 0.022*"service" + 0.016*"international" + 0.015*"airport" + 0.015*"route" + 0.013*"passenger" + 0.012*"o
perate" + 0.012*"schedule" + 0.011*"march" + 0.010*"operation"')
(3, '0.016*"helicopter" + 0.014*"force" + 0.013*"system" + 0.009*"service" + 0.009*"test" + 0.007*"capability" + 0.007*"militar
y" + 0.007*"usaf" + 0.007*"programme" + 0.007*"lockheed"')
(4, '0.013*"coronavirus" + 0.012*"government" + 0.011*"industry" + 0.011*"passenger" + 0.010*"travel" + 0.008*"measure" + 0.007
*"include" + 0.007*"support" + 0.007*"work" + 0.007*"country"')
(5, '0.013*"production" + 0.012*"engine" + 0.010*"jet" + 0.009*"programme" + 0.009*"business" + 0.008*"test" + 0.007*"service"
+ 0.007*"work" + 0.007*"include" + 0.007*"commercial"')
(6, '0.012*"coronavirus" + 0.011*"month" + 0.011*"fleet" + 0.011*"expect" + 0.010*"demand" + 0.010*"impact" + 0.010*"cost" + 0.
009*"capacity" + 0.009*"revenue" + 0.009*"cut"')
```

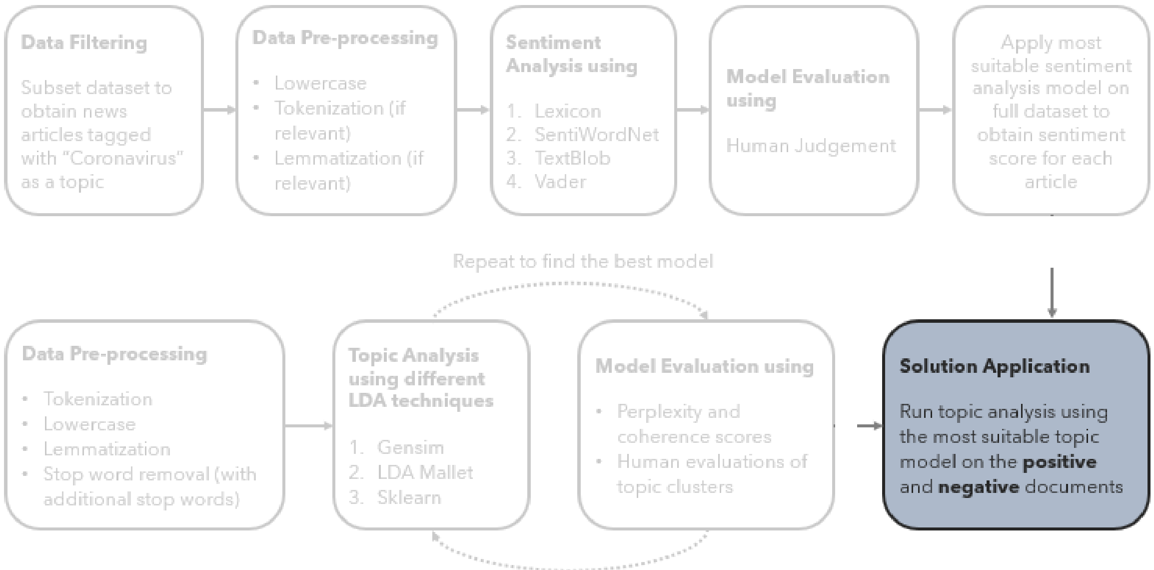
Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7
Business agreement	Aviation safety	Flight plans	Military planes	Government support	Plane production	Coronavirus impact

Figure 4: Top 10 keywords that form the 7 topics for overall corpus using LDA Mallet Model and their corresponding topics

3.3 Solution Implementation

<u>Choosing the right number of topics</u>

To analyse the topics for each sentiment, we separated the overall corpus of news articles into 2 different corpuses based on their sentiment scores generated from the sentiment analysis process described above.



Articles assigned with a negative sentiment score were allocated to the negative sentiment corpus whereas articles assigned with a positive sentiment score were added to the positive sentiment corpus. The most optimal topic modelling implementation (LDA Mallet) was then applied to both corpuses separately, using differing number of topics. This was necessary as we do not have prior knowledge of the optimal number of topics in each of the corpuses.

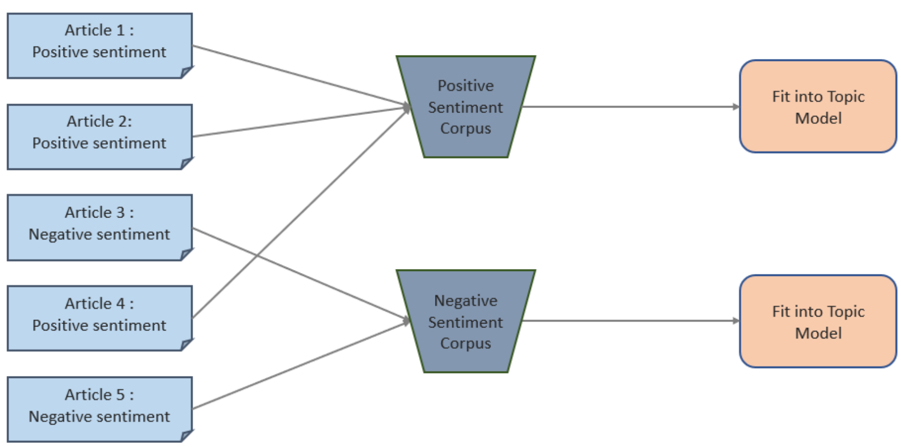


Figure 5: Process flow for separating the overall corpus into Positive Sentiment and Negative Sentiment Corpuses

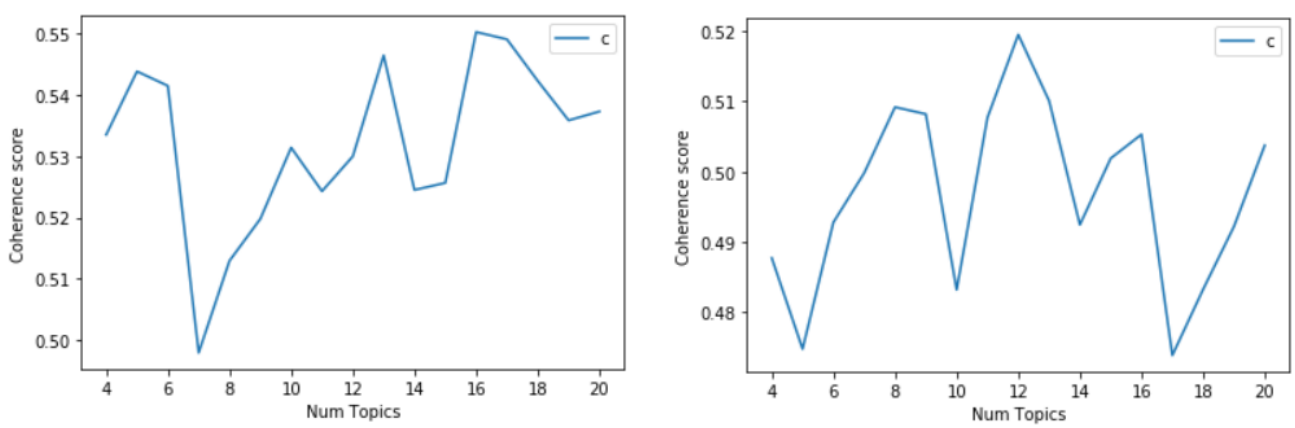


Figure 6: Graph for coherence scores for negative corpus (left) and positive corpus (right) with LDA Mallet model

The graph in Figure 6 shows the coherence scores for both corpuses between 4 to 20 topics. Optimal number of topics is chosen based on 2 criteria: 1) lowest number of topics with the highest coherence score before the curve flattens out and 2) each topic cluster can be generalized well from a business perspective. For each corpus, the peaks were first identified. After which, we manually assessed the keywords in the topic clusters by identifying "general areas of interest" which might be applicable to the business users.

Negative Corpus

For the negative corpus, the coherence scores peaked at 5, 13 and 16 topics. The keywords for each model are shown in the figures below, with a short summary of the topics' validity in a business sense.

```
Negative Model
Topic 0:['passenger', 'coronavirus', 'month', 'impact', 'capacity', 'demand', 'loss', 'revenue', 'expect', 'operating']
Topic 1:['crew', 'runway', 'pilot', 'inquiry', 'land', 'approach', 'state', 'investigation', 'incident', 'engine']
Topic 2:['order', 'delivery', 'service', 'faa', 'test', 'production', 'include', 'work', 'fleet', 'issue']
Topic 3:['government', 'business', 'state', 'executive', 'uk', 'plan', 'add', 'airway', 'chief', 'carrier']
Topic 4:['coronavirus', 'international', 'march', 'carrier', 'airport', 'service', 'travel', 'country', 'schedule', 'april']
```

Figure 7: Top 10 keywords that form the 5 topics for the negative corpus

1. Relatively straightforward but vague topics which are easily generalizable.
2. General areas of interest include coronavirus impact on profit and revenue, incidents and aviation policies and regulations.

```
Negative Model
Topic 0:['order', 'business', 'jet', 'delivery', 'production', 'customer', 'deliver', 'commercial', 'quarter', 'include']
Topic 1:['test', 'force', 'helicopter', 'fire', 'system', 'service', 'attack', 'design', 'usaf', 'missile']
Topic 2:['executive', 'chief', 'cut', 'add', 'situation', 'staff', 'crisis', 'make', 'coronavirus', 'employee']
Topic 3:['passenger', 'loss', 'operating', 'revenue', 'traffic', 'month', 'increase', 'capacity', 'decline', 'result']
Topic 4:['airport', 'fleet', 'service', 'data', 'carrier', 'operation', 'passenger', 'number', 'operate', 'operator']
Topic 5:['carrier', 'march', 'international', 'outbreak', 'suspend', 'capacity', 'schedule', 'april', 'united', 'coronavirus']
Topic 6:['engine', 'maintenance', 'fuel', 'takeoff', 'result', 'failure', 'system', 'problem', 'power', 'issue']
Topic 7:['crew', 'runway', 'inquiry', 'land', 'approach', 'accident', 'investigator', 'captain', 'incident', 'pilot']
Topic 8:['government', 'uk', 'carrier', 'business', 'state', 'share', 'south', 'airway', 'financial', 'plan']
Topic 9:['industry', 'crisis', 'demand', 'coronavirus', 'global', 'iata', 'market', 'level', 'cargo', 'expect']
Topic 10:['international', 'authority', 'civil', 'state', 'investigation', 'case', 'add', 'january', 'airasia', 'crash']
Topic 11:['pilot', 'faa', 'issue', 'safety', 'include', 'report', 'require', 'training', 'ground', 'order']
Topic 12:['travel', 'government', 'country', 'coronavirus', 'measure', 'restriction', 'passenger', 'state', 'march', 'canada']
```

Figure 8: Top 10 keywords that form the 13 topics for the negative corpus

1. Additional general areas of interest, such as employment-related issues, travel restrictions due to the pandemic and military, can be derived from these topics.
2. Topics are more well-defined. For example, the topics expanded more on the reason behind revenue impacts. Topic 3 is relating to revenue loss due to decline in passengers, while topic 2 is talking about cost cut relating to employment issues.
3. Overlap between topic 6, 7 and 11 which are talking about incidents and safety-related issues.

```
Negative Model
Topic 0:['fleet', 'ground', 'european', 'europe', 'return', 'lufthansa', 'crisis', 'carrier', 'price', 'part']
Topic 1:['engine', 'safety', 'faa', 'issue', 'operator', 'system', 'work', 'maintenance', 'failure', 'fuel']
Topic 2:['loss', 'revenue', 'operating', 'cost', 'quarter', 'result', 'profit', 'expect', 'financial', 'end']
Topic 3:['business', 'state', 'union', 'south', 'rescue', 'government', 'carrier', 'airway', 'add', 'saa']
Topic 4:['pilot', 'crew', 'officer', 'captain', 'control', 'fly', 'approach', 'inquiry', 'ground', 'thrust']
Topic 5:['add', 'state', 'make', 'provide', 'additional', 'call', 'return', 'united', 'trade', 'change']
Topic 6:['cut', 'crisis', 'coronavirus', 'canada', 'staff', 'executive', 'operation', 'pandemic', 'week', 'reduce']
Topic 7:['coronavirus', 'outbreak', 'suspend', 'international', 'country', 'carrier', 'february', 'united', 'american', 'day']
Topic 8:['delivery', 'order', 'production', 'customer', 'business', 'jet', 'deliver', 'commercial', 'programme', 'include']
Topic 9:['service', 'airport', 'march', 'carrier', 'route', 'operate', 'schedule', 'april', 'international', 'number']
Topic 10:['add', 'include', 'investigation', 'state', 'airasia', 'commission', 'office', 'al', 'authority', 'court']
Topic 11:['runway', 'crew', 'land', 'incident', 'investigator', 'accident', 'data', 'takeoff', 'state', 'inquiry']
Topic 12:['passenger', 'capacity', 'traffic', 'demand', 'month', 'cargo', 'impact', 'outbreak', 'carrier', 'fall']
Topic 13:['test', 'helicopter', 'force', 'missile', 'system', 'service', 'attack', 'usaf', 'design', 'fire']
Topic 14:['travel', 'government', 'restriction', 'country', 'measure', 'coronavirus', 'passenger', 'include', 'risk', 'industr
y']
Topic 15:['uk', 'lease', 'carrier', 'share', 'government', 'capital', 'airway', 'loan', 'flybe', 'plan']
```

Figure 9: Top 10 keywords that form the 16 topics for the negative corpus

1. Topics are too granular and sparse i.e. most of the new topics are country specific.
2. There are no new general areas of interest.

In summary, the model with 16 topics is not considered because the topics were too drilled down and focused. The model with 5 topics, although has generalizable and straightforward topics, tend to be a little vague and broad. The model with 13 topics is the best performing model based on manual human assessment because it has a wide spread of topics which are not too specific to a certain context.

Positive Corpus

For the positive corpus, the coherence scores peaked at 8, 12 and 16 topics. The keywords for each model are shown in the figures below, with a short summary of the topics' validity in a business sense.

```
Positive Model
Topic 0:['force', 'helicopter', 'system', 'capability', 'service', 'programme', 'contract', 'mission', 'lockheed', 'usaf']
Topic 1:['chief', 'executive', 'market', 'uk', 'change', 'industry', 'work', 'focus', 'singapore', 'add']
Topic 2:['coronavirus', 'pilot', 'industry', 'include', 'passenger', 'government', 'support', 'united', 'work', 'provide']
Topic 3:['carrier', 'service', 'passenger', 'operation', 'international', 'airport', 'route', 'operate', 'schedule', 'cargo']
Topic 4:['expect', 'coronavirus', 'end', 'cost', 'reduce', 'demand', 'month', 'march', 'impact', 'production']
Topic 5:['fleet', 'order', 'jet', 'business', 'delivery', 'customer', 'service', 'include', 'deliver', 'deal']
Topic 6:['engine', 'test', 'design', 'technology', 'system', 'certification', 'part', 'power', 'cabin', 'development']
Topic 7:['government', 'plan', 'state', 'support', 'carrier', 'share', 'business', 'australia', 'financial', 'agreement']
```

Figure 10: Top 10 keywords that form the 8 topics for the positive corpus

1. Slight overlap between topic 2 and 4 with the keyword "coronavirus". However, the other keywords in these topics suggest that the context beyond said keyword is different - topic 2 suggest that it is relating to some form of support that the government is providing for the industry, while topic 4 is relating to the expected end of the pandemic, as well as the impact and reduce in demand and production in March.
2. Slight overlap between topic 2 and topic 7 as well, which are both relating to support which is or planned to be provided by the government.
3. Keywords in topic 4 looks to be of a negative connotation because of words like 'reduce', 'cost' and 'impact', which suggests the impact of the pandemic on the industry, such as decline in production and demand.
4. General areas of interest include military demand and contracts, government support, aircraft manufacturing, cargo operations and industry strategy.

```
Positive Model
Topic 0:['passenger', 'coronavirus', 'travel', 'airport', 'cargo', 'transport', 'measure', 'country', 'operation', 'pandemic']
Topic 1:['government', 'state', 'share', 'agreement', 'financial', 'carrier', 'lease', 'support', 'capital', 'lufthansa']
Topic 2:['fleet', 'order', 'jet', 'delivery', 'customer', 'service', 'business', 'include', 'deliver', 'embraer']
Topic 3:['carrier', 'service', 'international', 'route', 'operate', 'domestic', 'schedule', 'australia', 'city', 'operation']
Topic 4:['pilot', 'work', 'report', 'safety', 'change', 'director', 'faa', 'training', 'agency', 'board']
Topic 5:['plan', 'business', 'operation', 'support', 'include', 'process', 'offer', 'rescue', 'interest', 'june']
Topic 6:['expect', 'cost', 'demand', 'end', 'reduce', 'month', 'fleet', 'capacity', 'increase', 'result']
Topic 7:['market', 'chief', 'executive', 'industry', 'make', 'commercial', 'global', 'add', 'focus', 'programme']
Topic 8:['march', 'employee', 'april', 'support', 'industry', 'coronavirus', 'government', 'include', 'request', 'day']
Topic 9:['force', 'helicopter', 'system', 'capability', 'mission', 'contract', 'lockheed', 'service', 'usaf', 'fighter']
Topic 10:['test', 'design', 'technology', 'system', 'land', 'launch', 'development', 'fly', 'project', 'speed']
Topic 11:['production', 'engine', 'work', 'part', 'facility', 'maintenance', 'programme', 'line', 'cabin', 'site']
```

Figure 11: Top 10 keywords that form the 12 topics for the positive corpus

1. General areas of interest remain the same.
2. There are more topics within some general areas of interest. For example, for aircraft manufacturing related topics, there is now a separate topic related to engine production and maintenance.
3. A new topic which focuses more on news in March and April. As a general guideline, topics which are too specific to a certain time period are not useful. However, March and April are the prime periods when the industry was hit hard during the pandemic. It would be interesting to see the positive articles which fall under this topic.

```
Positive Model
Topic 0:['uk', 'european', 'operator', 'state', 'include', 'transport', 'operation', 'base', 'airport', 'traffic']
Topic 1:['industry', 'government', 'support', 'provide', 'march', 'coronavirus', 'request', 'carrier', 'april', 'loan']
Topic 2:['government', 'share', 'support', 'state', 'financial', 'lufthansa', 'shareholder', 'capital', 'package', 'secure']
Topic 3:['pilot', 'safety', 'united', 'faa', 'agency', 'training', 'process', 'change', 'report', 'approval']
Topic 4:['chief', 'executive', 'cargo', 'director', 'world', 'industry', 'global', 'work', 'supply', 'transport']
Topic 5:['carrier', 'service', 'route', 'international', 'operate', 'network', 'schedule', 'city', 'domestic', 'airport']
Topic 6:['coronavirus', 'passenger', 'travel', 'country', 'measure', 'airport', 'pandemic', 'april', 'march', 'june']
Topic 7:['production', 'work', 'facility', 'line', 'employee', 'part', 'programme', 'operation', 'include', 'spirit']
Topic 8:['engine', 'cabin', 'power', 'certification', 'feature', 'offer', 'deliver', 'maintenance', 'jet', 'mro']
Topic 9:['expect', 'cost', 'demand', 'end', 'month', 'reduce', 'increase', 'quarter', 'result', 'revenue']
Topic 10:['plan', 'business', 'australia', 'virgin', 'government', 'carrier', 'process', 'rescue', 'june', 'interest']
Topic 11:['technology', 'development', 'design', 'programme', 'project', 'effort', 'develop', 'future', 'create', 'industry']
Topic 12:['fleet', 'order', 'delivery', 'customer', 'lease', 'jet', 'agreement', 'deal', 'data', 'lessor']
Topic 13:['force', 'helicopter', 'capability', 'service', 'contract', 'lockheed', 'usaf', 'fighter', 'mission', 'martin']
Topic 14:['test', 'system', 'range', 'launch', 'design', 'carry', 'control', 'land', 'uav', 'unmanned']
Topic 15:['business', 'service', 'canada', 'jet', 'embraer', 'bombardier', 'market', 'include', 'regional', 'commercial']
```

Figure 12: Top 10 keywords that form the 16 topics for the positive corpus

1. Topics are too granular and sparse i.e. most of the new topics are country specific.
2. There are no new general areas of interest.

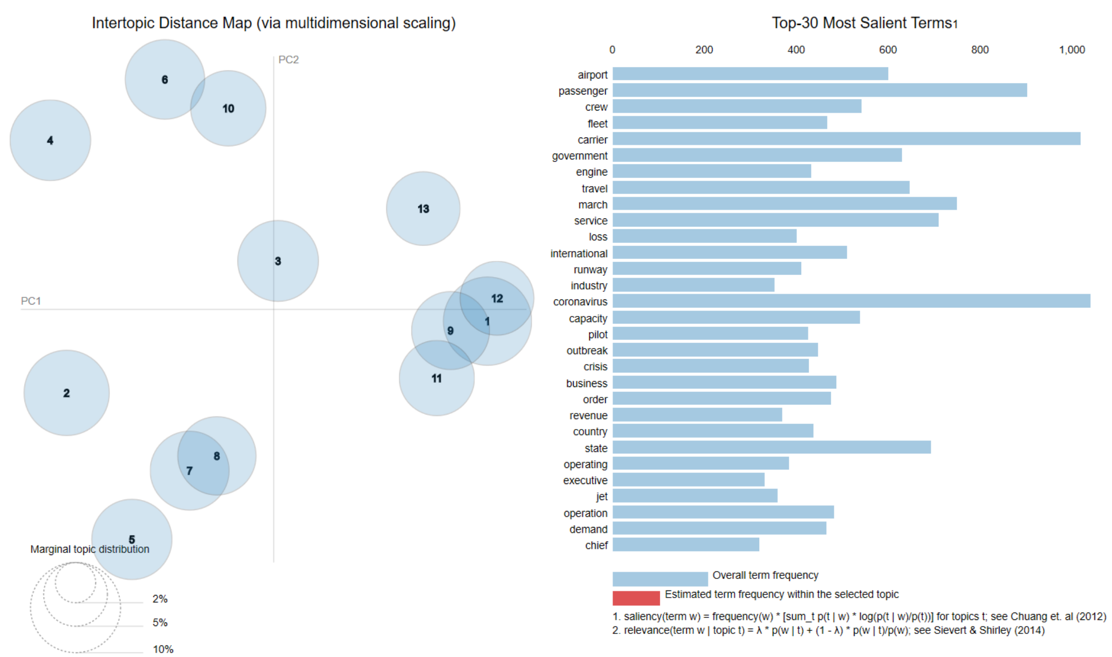
In summary, the model with 16 topics is not considered because the topics were too drilled down and focused. The model with 12 topics is the best performing model based on manual assessment as it covers more areas of interest than the model with 8 topics.

<u>pyLDAvis</u>

Upon deciding the optimal number of topics based on coherence score and manual assessment, we used pyLDAvis to visualize the optimal topics and associated keywords. This is to ensure the model that we have built do not have excessive overlapping topic clusters and covers a wide spectrum of topics. In the Intertopic Distance Map, a good model will have bubbles which is spread out across the graph and not have too many overlapping bubbles. Additionally, the bubble size indicates how prevalent the topic is. If there are too many small bubbles, it means that the topics are too sparse and granular.

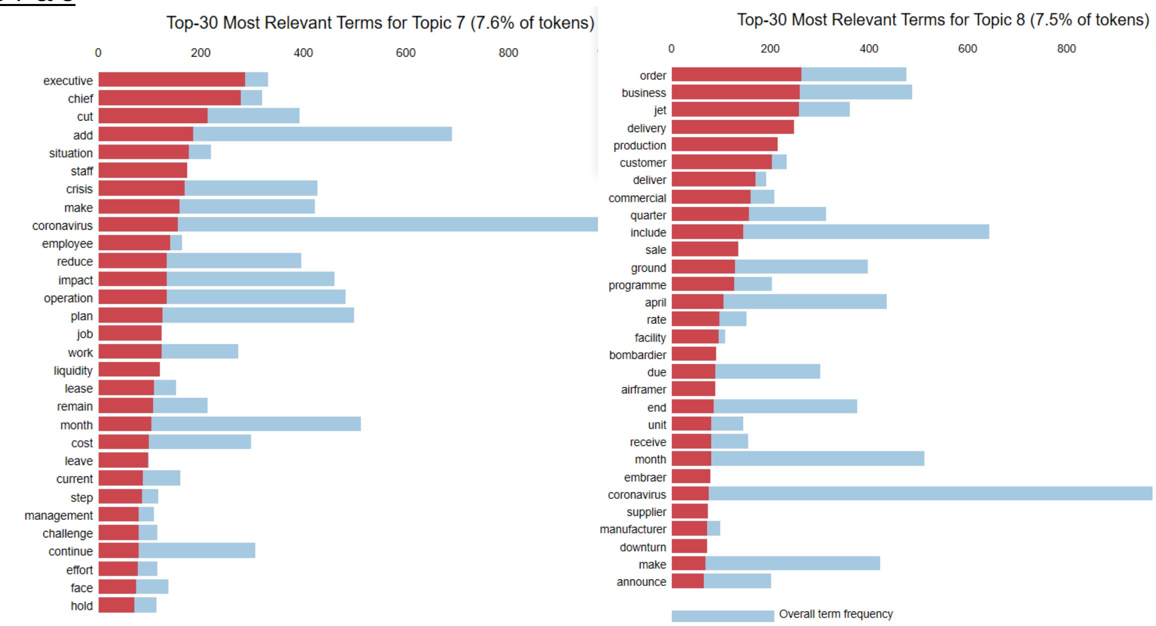
Another point to note is that the topic numbers in pyLDAvis do not coincide with that of the LDA Mallet model above. Hence, topic 1 stated in the LDA Mallet model above might not be the same as the topic 1 in the pyLDAvis output. However, the keywords used are exactly the same.

Negative Corpus



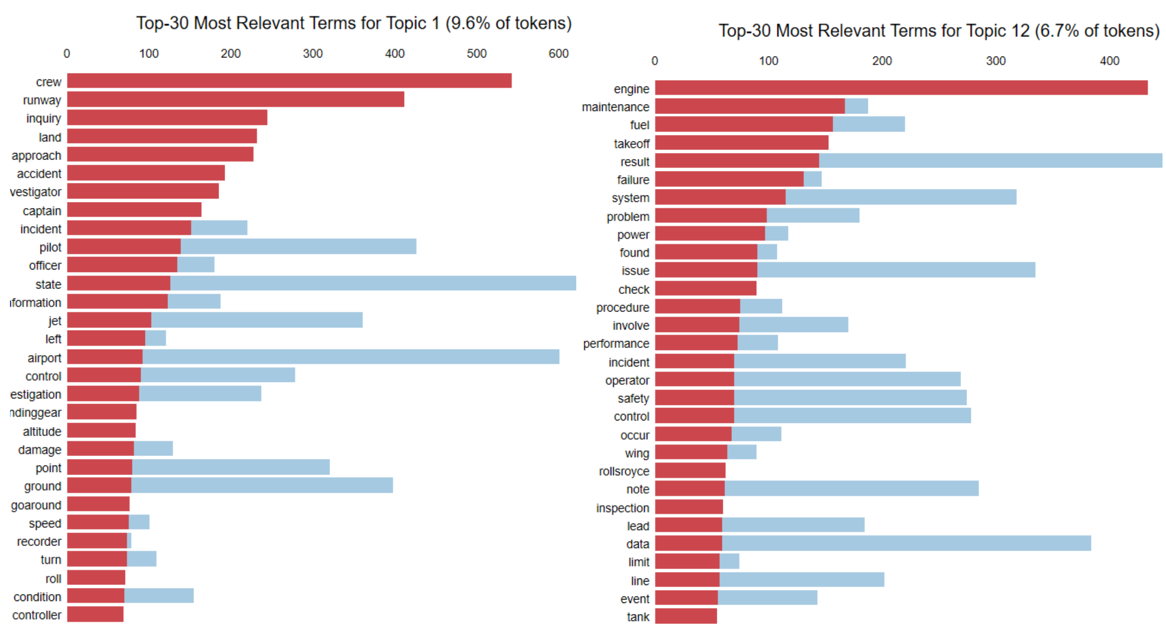
The topics in the negative corpus are well spaced out across the graph and the topics are relatively prevalent as well. However, we can see 2 major overlapping clusters – 7 & 8 and 1, 12 & 9. We further analysed these 2 overlapping clusters by looking at the keywords associated with the topics and their term frequency.

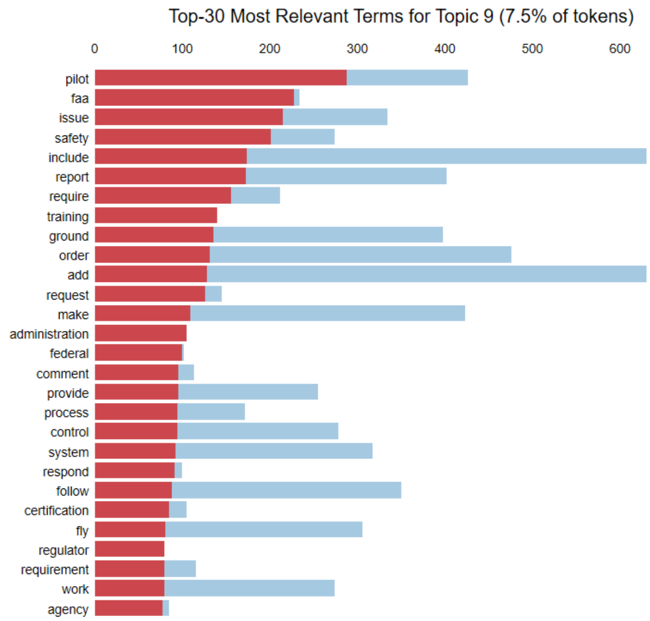
<u>Topics 7 & 8</u>



Topic 7 is related to employment issues during the pandemic, while topic 8 is related to the commercial business for aircraft delivery and production. On first look, these two topics might not be closely related to each other. However, the topics might be overlapping because some of these keywords from the two topics often appear in the same article.

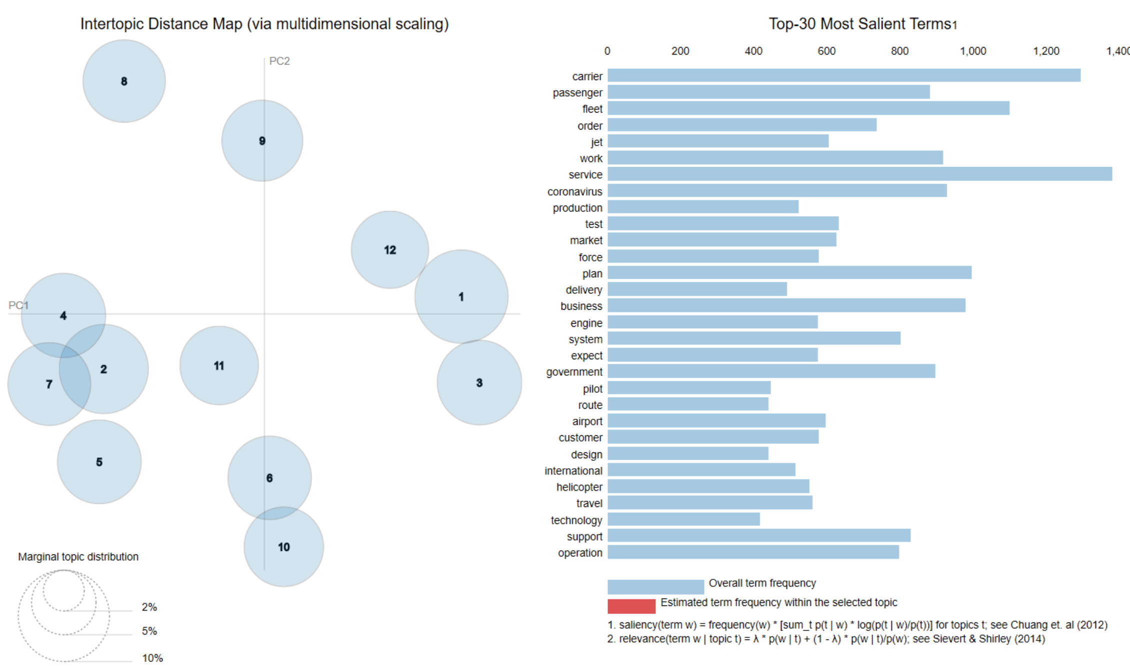
<u>Topics 1, 12 & 9</u>





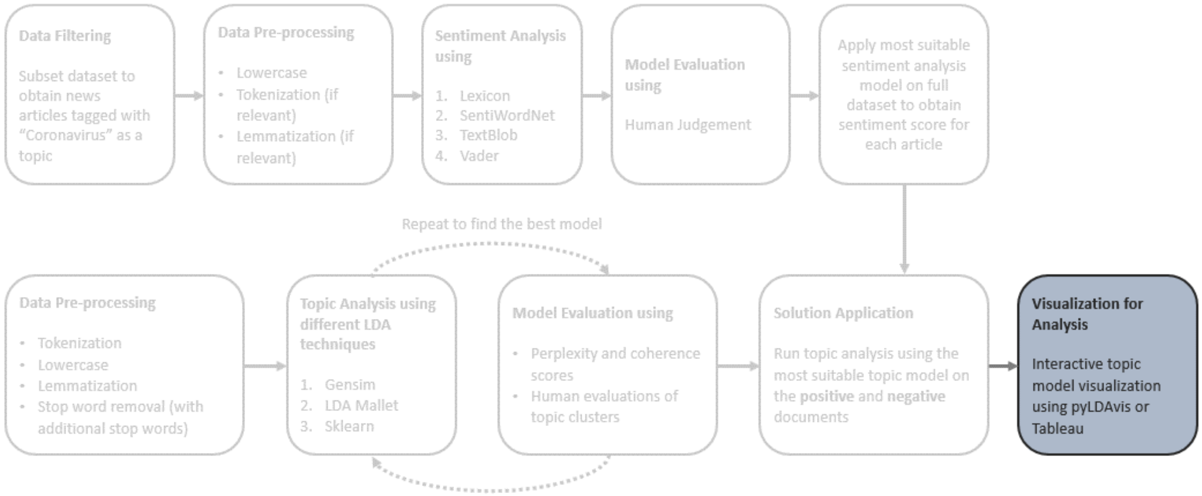
All topics look at air-related accidents. Topic 12 looks at some form of engine failure or problems with the aircraft which might have caused the accidents. Topic 9 looks at the safety measures from the FAA. The overlap is justified in this case.

Positive Corpus



The topics are well spaced out across the graphs. There is a minor overlap between topics 2, 4 & 7. However, the overlaps seem relatively small and hence is not an issue.

4. Results and Analyses



Results showed that there were significantly more news articles labelled as Positive sentiments during the first half of 2020 i.e. Covid-19 period, despite subdued travel demand, tepid outlook for new aircraft orders etc. Compared to the sentiment analysis done purely on articles tagged with 'Coronavirus' as a Topic by FlightGlobal, this shows that the articles not tagged with 'Coronavirus' as a Topic tend to be mostly Positive in nature.

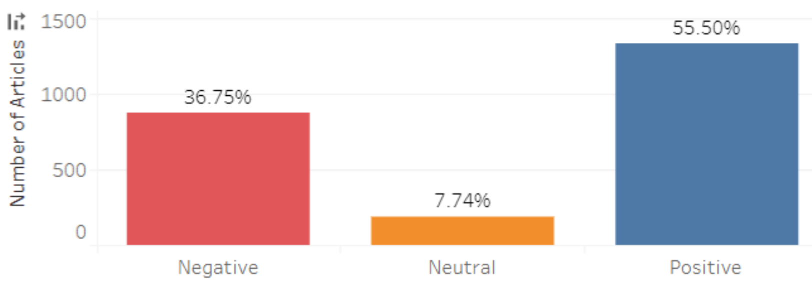


Figure 13: Sentiment Distribution for the Entire FlightGlobal Corpus

Mth	Neg	Neu	Pos
Jan	127 (36%)	21 (6%)	204 (58%)
Feb	145 (35%)	29 (7%)	242 (58%)
Mar	219 (44%)	39 (8%)	236 (48%)
Apr	151 (39%)	31 (8%)	236 (53%)
May	117 (32%)	39 (11%)	204 (57%)
Jun	119 (31%)	26 (7%)	234 (62%)

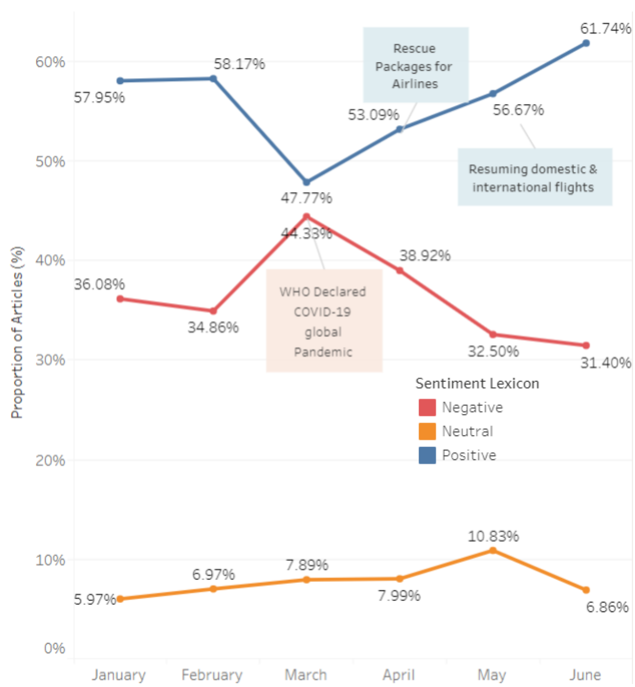


Figure 14: Sentiment Distribution for the Corpus across the months

However, for the purpose of this study, we are not interested in the proportion of Positive vs Negative articles for the overall corpus. As mentioned above, it is observed that the number of articles with Positive sentiments still outnumbers that of Negative sentiments for every month, so there is no pattern of interest just by looking at the counts themselves. Rather, we are interested in how the proportion of sentiments changes over time to see how the impact of the Covid-19 pandemic played out over the months in relation to actual developments in the real world.

The proportion of negative sentiments in January and February were similar at ~36%. During these two months, the brunt of Covid-19 cases was still confined within China, and while countries around the world were slowly experiencing the emergence of local clusters, the aviation industry still did not have a clear picture as to the global economic damage that the virus would eventually cause.

However, moving from February to March, we see that the proportion of negative sentiments has increased from 35% to 44%. On 11 March 2020, the World Health Organisation officially labelled Covid-19 as a global pandemic (World Health Organisation, 2020), as it became obvious that the virus had spread across continents at a rapid pace. March was also the month which had the greatest number of countries imposing their first national or state-wide lockdowns, such as United States of America, United Kingdom, Italy, France etc. This directly translated to disaster for the aviation industry as international bans on non-essential travel were imposed and airlines began to ground most of their fleet.

Moving from March to April, the downtick in Negative sentiments from 44% to 39% could be explained by the emergence of state rescue packages for airlines across the globe. For example, the US Treasury Department agreed a ~$25bn rescue package for 10 of the country's biggest airlines to support their payrolls (BBC News, 2020). Air France-KLM also secured €7bn in French government aid to prop up the airline (Reuters, 2020). The slew of bailouts for airlines helped to dampen the negative sentiments from the previous month.

The even bigger dip in Negative sentiments from 39% in April to 32% in May and 31% in June was largely due to countries beginning to lift their lockdowns gradually. As May went by, many countries began to announce their own plans to begin reopening their borders, and intra-regional travel in Europe was expected to start in June (International Air Transport Association, 2020). This likely helped to lift the gloomy outlook for air travel.

In summary, the shift in proportion of Negative sentiments against Positive sentiments over the months as COVID-19 panned out can indeed be explained by the development of real-world events. Thus, sentiment analysis is a potentially useful quick tool for industry participants to assess the nature of developments in the industry.

Extending the results further, we reviewed the topic clusters generated by the topic modelling implementation. We analysed the document topic probabilities (theta) for each article in the overall corpus to determine the number of dominant topics, which we defined as topics that had at least 2/k of the probabilities, where k = optimal number of topics (i.e. 2/7 = 28.6%). Analysis showed that majority of the articles are dominated by one topic (i.e. only one topic taking up ≥28.6% of the probability). As such, each article was represented as a single topic for subsequent analysis of topic modelling results.

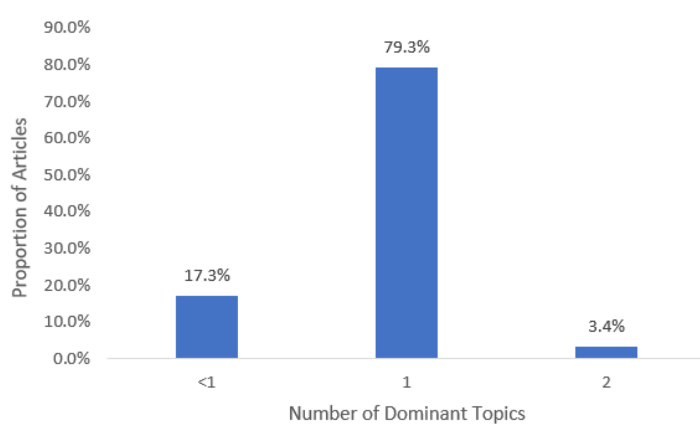


Figure 15: Distribution of Articles by Number of Dominant Topics

While the number of topics in the positive sentiment corpus remained relatively similar across the first 6 months of 2020, there was a spike in number of topics in March for the negative sentiment corpus. The spike was contributed by the increase in proportion of Topics 6, 13, 3 and 5. Further analysis of these 4 topics showed that:

- Topic 6 focused on the suspension of flight schedules with the coronavirus outbreak, which coincided with the declaration of the virus as a pandemic.
- Topic 13 covered the travel restrictions imposed by the governments.
- Topic 3 discussed the cut of employees due to the crisis.
- Topic 5 talked briefly about the airport, plane fleets and operators.

Results showed that applying topic modelling on the positive and negative sentiment corpuses allowed business users to delve deeper into the topics that have contributed to the sentiments for each month, which will be useful for government to plan for support measures and adjust policies based on the existing situation.

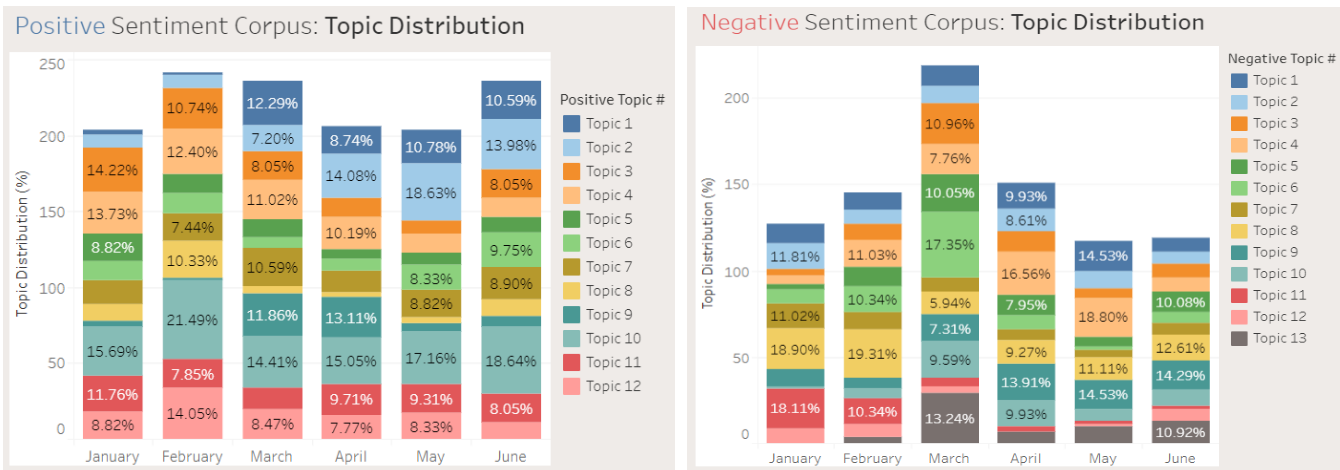


Figure 16: Topic Distribution for the Corpus across the months

5. Discussions and Gap Analysis

Traditional sentiment analysis stops short at polarity classification. This project, however, went further by generating the sentiment scores for each news articles and broke down the evolution of sentiments across months. Our group successfully evaluated and applied the most suitable sentiment analysis model that could produce a sentiment scoring that made the most logical sense. Thereafter, we trained the most suitable topic model that produces topic clusters of satisfactory quality and applied it on the positive and negative sentiment corpuses to derive topics in each corpus. The results derived were easy to be interpreted and provided us more insights to the corpus. Our domain knowledge in the airline industry has also provided advantages in the pre-processing and analysis of the news articles.

Nonetheless, there were some challenges faced in the project. During the sentiment classification process, some articles tend to provide a balanced view resulting in neutral sentiments. Although, it is not surprising since the data scrapped was on news articles, having too many neutral labelled news articles may not be useful in giving an accurate outlook of the airline industry during the Covid-19 period. In addition, LDA models require large amount of data to obtain optimal results. However, due to time constraints for this project, our group was not able to scrap more articles to obtain a larger corpus to facilitate the LDA modelling. During the topic modeling process, it was flagged that there was a lack of standardization across articles in the use of terms with different variations (e.g. COVID-19, Coronavirus, Corona / United States, US, USA). Terms which were identified were manually standardized, however, there may be terms which were not flagged out due to the large vocabulary of words.

6. Future Work and Conclusion

In this project, we analysed news articles scrapped from FlightGlobal during the first half of 2020 to gain insights of the sentiments in the aviation industry across the months of the coronavirus pandemic. We observed that variation in the sentiment trends aligned with the real-world developments of the pandemic and the sentiment analysis could potentially offer industry players a quick sense of nature of developments in the industry. Topic modelling adds further value by summarizing the common key topics within the positive and negative corpuses, allowing stakeholders to gain more insights on areas of concerns or aspects that were gradually recovering from the impact of the pandemic.

For future work, some possible extension of this project will be to explore readily available pre-labelled training data for sentiment analysis on airline related news, which could potentially increase the classification accuracy. In addition, users may wish to expand the corpus by crawling airline news articles from other platforms which could help in identifying more granular topic by each sentiment corpus. In our project, we manually created a long list of stop words for topic modelling, users may try different pre-processing method to compare the results or evaluate the topic models using domain-specific dictionary.

Other relevant analytical tasks which might be useful development for this project would be information extraction to extract companies or entities of interest from the news article to keep track on their updates regarding the entities, as well as summarization which allow aviation stakeholders to get a quick sense of the happenings in the aviation industries without having to read through individual articles.

7. Project Experiences/Reflections

Members	Reflections
Andre Justin Lee Sheng Wei	This project allowed me to explore different sentiment analysis models and have a deeper understanding of how each model carried out their scoring across similar articles. Understanding the scoring methodology of each model guided us to fine-tune our pre-processing of the articles. Evaluating each model's scoring methods also gave me a feel of how human evaluation of models is done.
Choong Shi Lian, Selene	This project allowed me to explore the various techniques for topic modelling, understand the impact of different pre-processing steps on the model performance, try different evaluation metrics and extend evaluation beyond metrics to human assessment. While Python offers existing packages that allow users to easily perform LDA topic modelling, effort is required to prepare the corpus for analysis with relevant text pre-processing steps and perform human assessment on outputs to ensure its interpretability. The collaboration with team members allowed us to brush ideas off each other on ways to improve the analytical models built.
Jiang Weiling, Angeline	Through this project, I explored various sentiment analysis techniques, understanding how it works, advantaged and limitations of each model. In addition, during the group discussion, we found a more efficient way of evaluating the sentiment models. Instead of evaluating on whole dataset, we found that it may be more reasonable to evaluate on partial of the dataset i.e. Coronavirus tagged articles to determine which technique is more suitable base on domain prior to running on the whole dataset as it will be tedious if we were to do human labelling on all the 2389 articles.
Son Yejin	This project gave me an opportunity to apply topic modelling to the actual industry example, going beyond the theoretical learning from the class. While building the solution, I learned that training the right model for a specific business model can have a lot of variables to perform in the best way. Also, it allowed me to think again to focus on how users would benefit from this study.
Wong Jia Wei	Prior to this course, text analytics has always been a mystery to me. Work-related projects had always used traditional machine learning techniques with numbers for analysis or even prediction. This project opened up my eyes to the world of text analytics. Using textual data, we were able to perform analysis which are not possible through quantifiable measures. I am also very grateful to my groupmates for coming up with this use case as it shows how we can apply what we have learnt in the course for real-world scenarios.

8. References

Air France-KLM wins aid deal, flags likely share issue. (2020, April 25). Retrieved July 24 2020, from https://www.reuters.com/article/us-health-coronavirus-airfranceklm-renau/air-france-klm-wins-aid-deal-flags-likely-share-issue-idUSKCN22631H

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research 3*, 993-1022.

Chang, J., Boyd-Graber , J., Gerrish, S., Wang, C., & Blei, D. M. (2009). Reading Tea Leaves: How Humans Interpret Topic Models. *Neural Information Processing Systems*, 288-296.

Hutto, C. J., & Gilbert, E. (2014, May). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Eighth international AAAI conference on weblogs and social media*.

International Air Transport Association. (2020). *Economic Performance of the Airline Industry.* Retrieved July 24 2020, from https://www.iata.org/en/iata-repository/publications/economic-reports/airline-industry-economic-performance-june-2020-report/

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

WHO Director-General's opening remarks at the media briefing on COVID-19—11 March 2020. (n.d.). Retrieved July 26, 2020, from https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020

US airlines to receive $25bn rescue package. (2020, April 15). Retrieved July 24 2020, from https://www.bbc.com/news/business-52288860