Forage

BRITISH AIRWAYS

Task-1

Web scraping to gain company insights.

By

Miraj Ahmed

Data Science

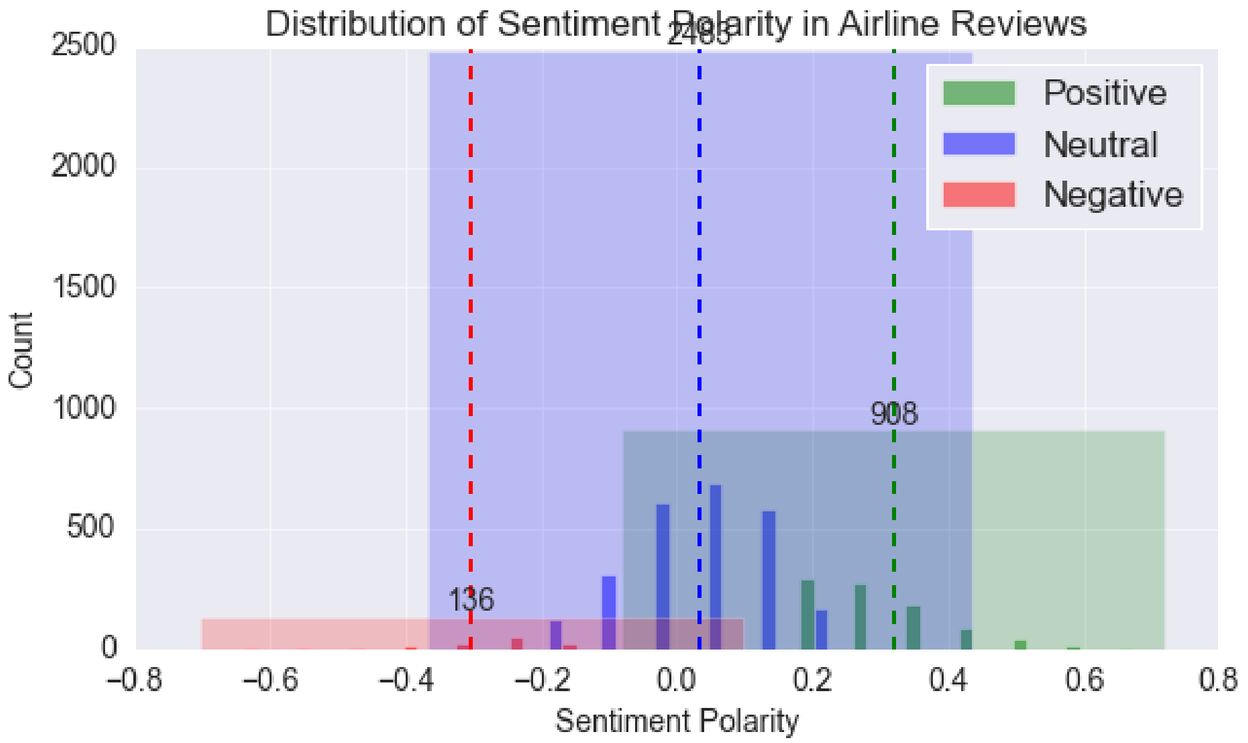
Virtual Internship

Summary

The project involves three main tasks. Firstly, scraping review data from the website called Skytrax. Secondly, analysing the collected data by performing data cleaning and various analyses like topic modelling, sentiment analysis, and word clouds. Finally, summarizing the findings in a single PowerPoint slide, which includes visualizations, metrics, and clear explanations. The team can use Python or any other tool of their choice for analysis, and they can access resources in the "Resources" section below to help them complete the tasks..

WordCloud

Customers are interested in airport and airline-related aspects of their travel experience such as flight delays, cabin crew service quality, food quality, airline staff efficiency, helpfulness and overall airline experience.



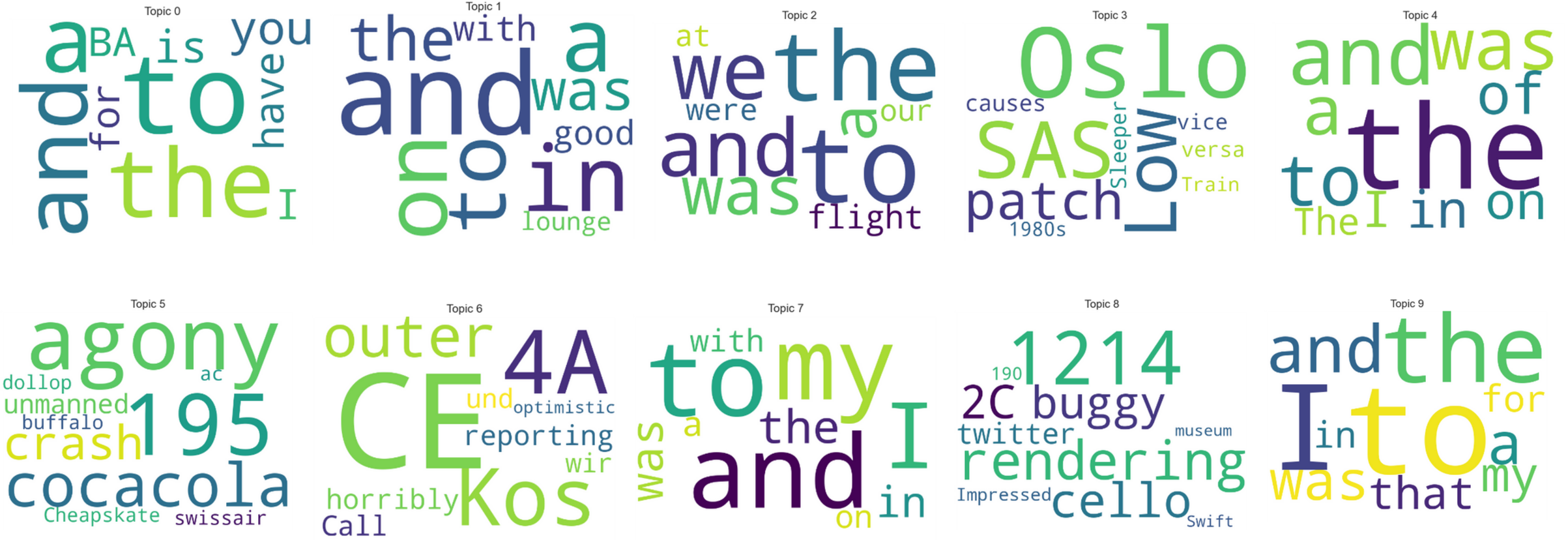
Sentiment Analysis

The image shows that the mean polarity is on 0.1 and the numbers are more into neutral segment given that Comments with sentiment polarity values close to 0 can be interpreted as either slightly positive or neutral, causing some comments to fall into both the neutral and positive sentiment categories when using a 0.2 threshold.

The threshold is subjective and it can be set differently based on the needs and interpretation of the data.

Topic Modeling with Wordcloud

Insights are shared in the following page



Insight

TOPIC 0

a common topic in all datasets where the most common words such as 'the', 'to', 'and', 'I', and 'was' are being used. This topic does not provide any useful information.

TOPIC 1

not giving any useful information, as it is just a collection of random words.

TOPIC 2

suggests that the dataset may contain reviews about British Airways' flights to Abu Dhabi, Amman, and Hamburg.

Insight(Continues)

TOPIC 3

not providing any information about the dataset, as it is just a collection of common words such as 'video', 'to', 'the', 'flight', and 'with'.

TOPIC 4

not providing any meaningful information either.

TOPIC 5

like Topic 0, is a common topic where the most common words are 'the', 'and', 'was', 'a', and 'to'. This topic does not provide any useful information.

Insight(Continues)

TOPIC 6

appears to be a random set of words and doesn't provide any insights into the dataset.

TOPIC 8

suggests that the dataset may contain reviews about British Airways' flights.

TOPIC 7

suggests that the dataset may contain information about British Airways as a company, its products, and services.

TOPIC 9

similar to Topic 5 and Topic 0 where it is a common topic where the most common words

Future Work

Future work could include exploring different sources of data, experimenting with machine learning algorithms and hyperparameter tuning, performing comparative analysis with other airlines, and incorporating natural language generation techniques to automate reports and summaries of findings.