



Analyzing Top Airlines Reviews

Presented by:
Prashansa Thapa
Bikin Ghimire

Improving People's experience



Objectives

To help airlines understand their customer's needs better

- Identify customer's rating based on class
- Identify customer's rating based on airlines on lowest rated class
- Identify areas of improvement for the lowest rated class
- Find the most popular words used in negative reviews
- Calculate the likelihood of people recommending the airlines to others

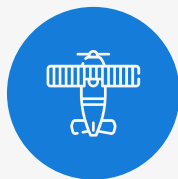


Data Identification and Acquisition



Data Source

Airlinequality.com



Dataset Source

Thesujalsuthar's
GitHub profile



Dataset Scope

Top 10 rated airlines
in 2024



Review Count

8100

Title	Name	Review Date	Airline	Verified	Reviews	Type of Traveller	Month Flown	Class	Seat Comfort	Staff Service	Food & Bever	Inflight Entertainment	Value For Money	Overall Rating
Flight was an	Alison Soetar	01/03/2024	Singapore Air	TRUE	Flight was a	Solo Leisure	Dec-23	Business Cla	4	4	4	4	4	9
seats on this	Robert Wats	21/02/2024	Singapore Air	TRUE	Booking	Solo Leisure	Feb-24	Economy Cla	5	3	4	4	1	3
Food was ple	S Han	20/02/2024	Singapore Air	TRUE	Excellent pe	Family Leisure	Feb-24	Economy Cla	1	5	2	1	5	10
,Áúhow much	D Laynes	19/02/2024	Singapore Air	TRUE	Pretty comfo	Solo Leisure	Feb-24	Economy Cla	5	5	5	5	5	10
,Áúservice w	A Othman	19/02/2024	Singapore Air	TRUE	The service v	Family Leisure	Feb-24	Economy Cla	5	5	5	5	5	10
,Áúseat is ab	Robert Wats	19/02/2024	Singapore Air	TRUE	This flight wa	Solo Leisure	Feb-24	Economy Cla	5	1	5	5	1	3
Very ordinary	S Holger	19/02/2024	Singapore Air	TRUE	Boarding pro	Solo Leisure	Feb-24	Economy Cla	1	3	1	2	2	5
I like flying th	A Jabil	19/02/2024	Singapore Air	TRUE	Pleasant flig	Family Leisure	Feb-24	Economy Cla	2	5	3	4	5	10
left me deepl	A Kalarsan	18/02/2024	Singapore Air	TRUE	I embarked c	Business	Feb-24	Economy Cla	5	4	5	5	3	4
can,Áôt fault	S Dayle	17/02/2024	Singapore Air	TRUE	No queu	Family Leisure	Feb-24	Economy Cla	4	5	4	3	5	10
No complain	J Hassin	17/02/2024	Singapore Air	TRUE	Competitive	Solo Leisure	Feb-24	Economy Cla	5	5	5	5	5	10
,ÁúI was impr	N Allen	16/02/2024	Singapore Air	TRUE	This was my f	Family Leisure	Feb-24	Economy Cla	5	5	5	5	4	10
,ÁúSuperb se	N Allen	16/02/2024	Singapore Air	TRUE	Superb servi	Solo Leisure	Feb-24	Economy Cla	4	5	5	5	5	10
regional seat	L Hee	16/02/2024	Singapore Air	TRUE	I was one of	Solo Leisure	Jan-24	Business Cla	5	1	5	5	3	2
,Áúno proble	Jason Keh	09/02/2024	Singapore Air	FALSE	The flight is v	Family Leisure	Feb-24	Business Cla	1	5	4	5	5	10

Data Extraction and Transformation

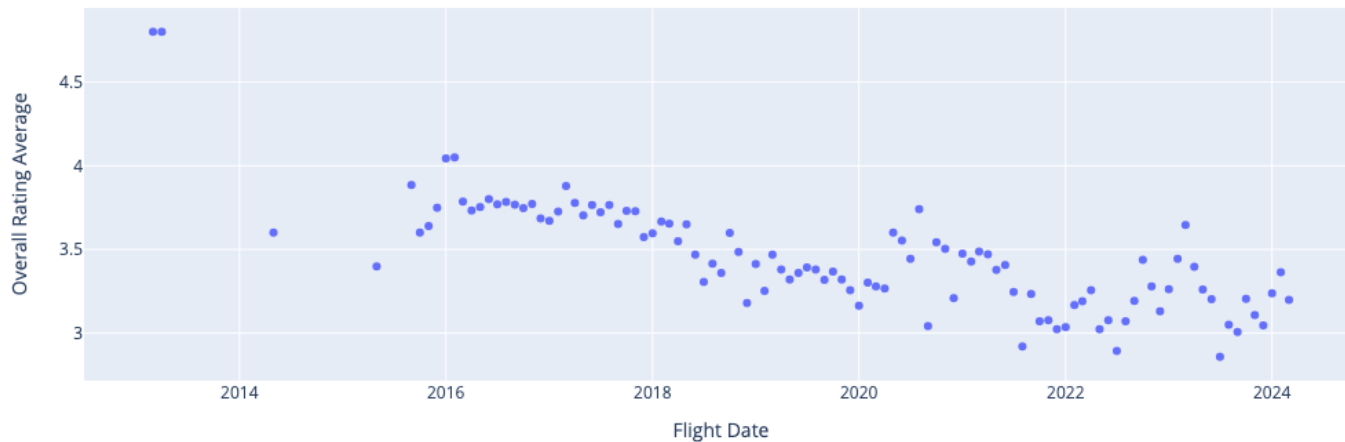


	What	How
1	Data Schema	Pyspark datetime
2	Removed Special Characters	Regex_replace
3	Converted overall rating to 5	withColumn
4	Added Review ID	monotonically_increasing_id()

Data Analysis



Customer Satisfaction Over Time



Data Analysis



Customer Satisfaction in Airline's services has (almost) linearly decreased over time.

Data Analysis

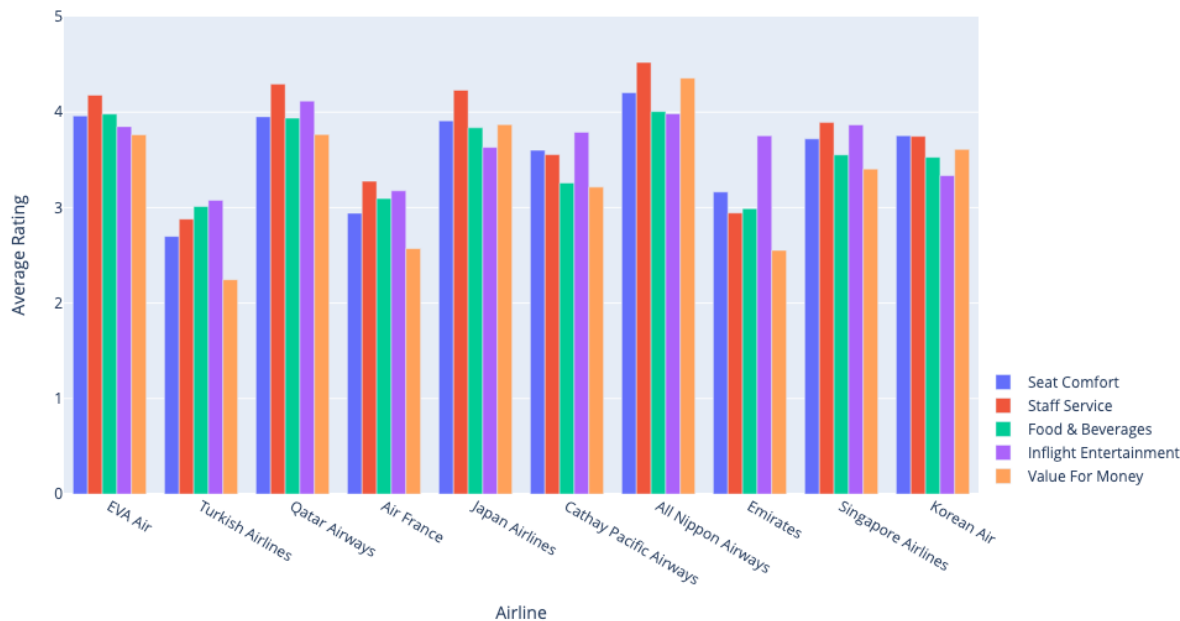


Passengers flying in economy class are the ones that give the poorest review for their flights

Data Analysis



Average Ratings by Category for Each Airline (Economy Class)

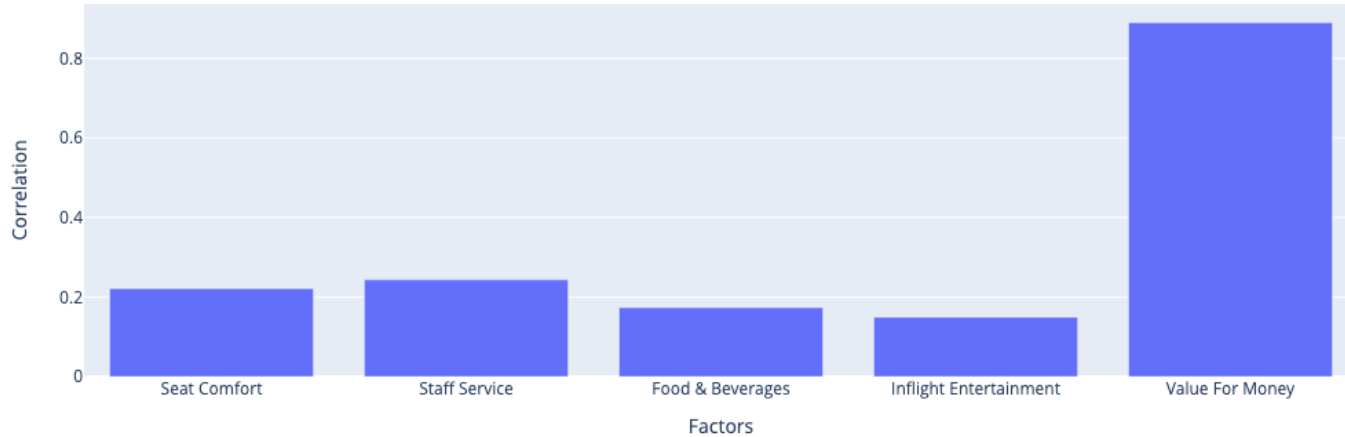


For all the lower rated airlines, Value for Money is the least rated category.

Data Analysis



Correlation between Factors and Overall Rating (Economy Class)



People's satisfaction in the services received compared to the price of the flights is the factor that impact's their overall rating of the flights the most.

Sentiment Analysis



Using RoBERTa model

- Step 1: Load the tokenizer from the pre-trained model
- Step 2: Load the model
- Step 3: Pass the input and get the scores (negative, neutral, positive)
- Step 4: Apply softmax to convert the scores to probability
- Step 5: Store the scores in the dataframe

	Review ID	negative	neutral	positive	Review Title	Reviewer	Review Date	Airline	Review Content	Flight Date	Class
0	0	0.004829	0.033120	0.962051	Flight was amazing	Alison Soetantyo	2024-03-01	Singapore Airlines	Flight was amazing	2023-12-01	Business Class
1	1	0.937025	0.055582	0.007393	seats on this aircraft are dreadful	Robert Watson	2024-02-21	Singapore Airlines	seats on this aircraft are dreadful	2024-02-01	Economy Class
2	2	0.003312	0.043574	0.953113	Food was plentiful and tasty	S Han	2024-02-20	Singapore Airlines	Food was plentiful and tasty	2024-02-01	Economy Class

Sentiment Analysis



Using RoBERTa model

- Step 1: Load the tokenizer from the pre-trained model
- Step 2: Load the model
- Step 3: Pass the input and get the scores (negative, neutral, positive)
- Step 4: Apply softmax to convert the scores to probability
- Step 5: Store the scores in the dataframe

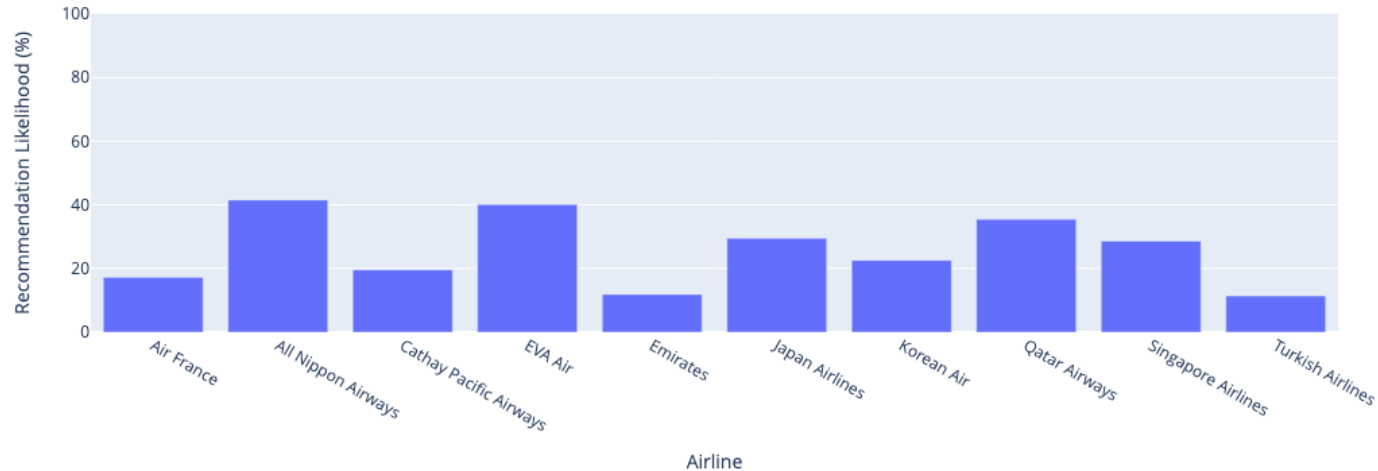
Review with the highest positive sentiment score:
one of the few airlines offering a truly distinctive premium service

Review with the highest negative sentiment score:
the food was utterly inedible and honestly hideous to look at

Sentiment Analysis



Likelihood of reviewers recommending the airlines to others (Economy Class)



The reviewers of Turkish Airlines, Emirates, Air France, and Korean Air are very unlikely to recommend the airlines to others.

[illegible]

The words represent the areas where reviewers are most disappointed in and suggests areas for improvement for the airlines.

Conclusion

Customer satisfaction in airline services have steadily declined over the years, and the airlines need to improve their services.

- | | |
|---|---|
| 1 | Airlines need to focus on improving services for Economy Passengers to improve their overall rating. |
| 2 | Turkish Airlines, Air France, and Emirates need to up their game the most and improve their services for customers. |
| 3 | Airlines should balance the services provided compared to the price of the flights, because that is the factor that impact's their overall rating the most (Economy Class). |
| 4 | The reviewers of Turkish Airlines, Emirates, Air France, and Korean Air are very unlikely to recommend the airlines to others (Economy Class). |
| 5 | Airlines need to improve their quality of food, luggage, seat, staff, comfort, and helpfulness among others in their negative reviews (Wordcloud result). |

Tools and Technologies Used

Code Editor and IDE	VSCode, Jupyter Notebook, Python
Data Extraction and Transformation	PySpark
Data Analysis	Numpy, PySpark, Pandas
Sentiment Analysis	AutoTokenizer, Softmax, RoBERTa model
Visualization	Plotly, Matplotlib, Wordcloud

Thanks!

Do you have any questions?

Credits: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, infographics & images by **Freepik**

