

Topic Classification and Modelling

Ambika Prasad Swain FT251013

Aniket Singh FT251019

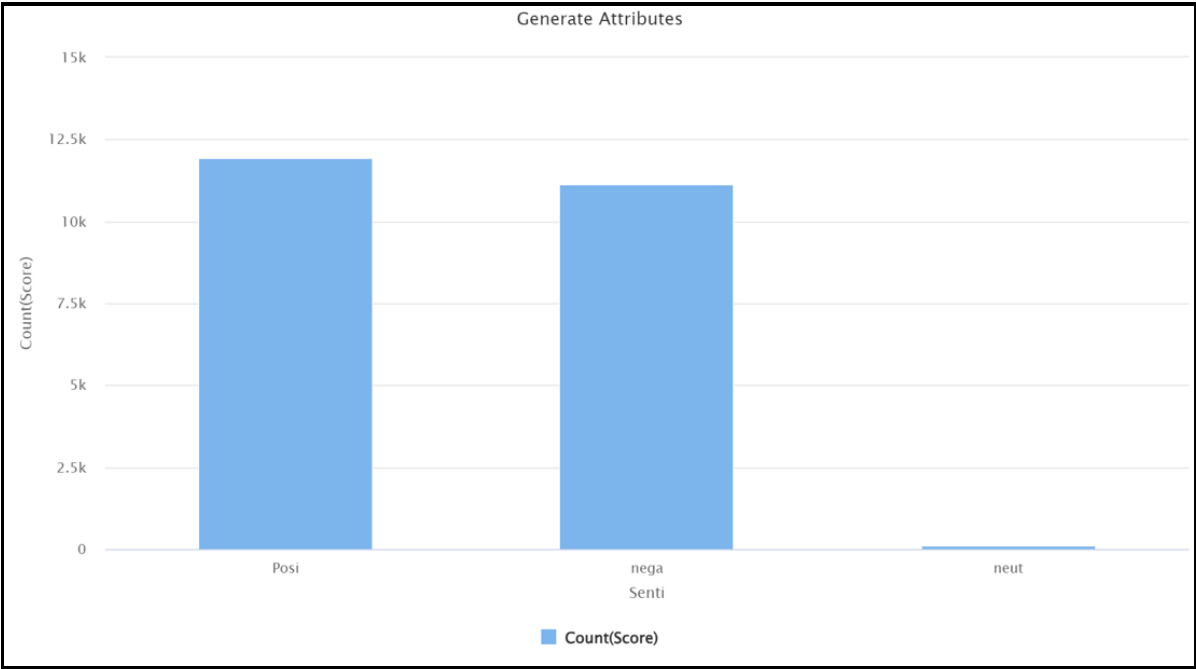
Ankit Mamgai FT251020

Anubhav Verma FT251021

Mayank Agrawal FT251042

Rohit Behara FT251066

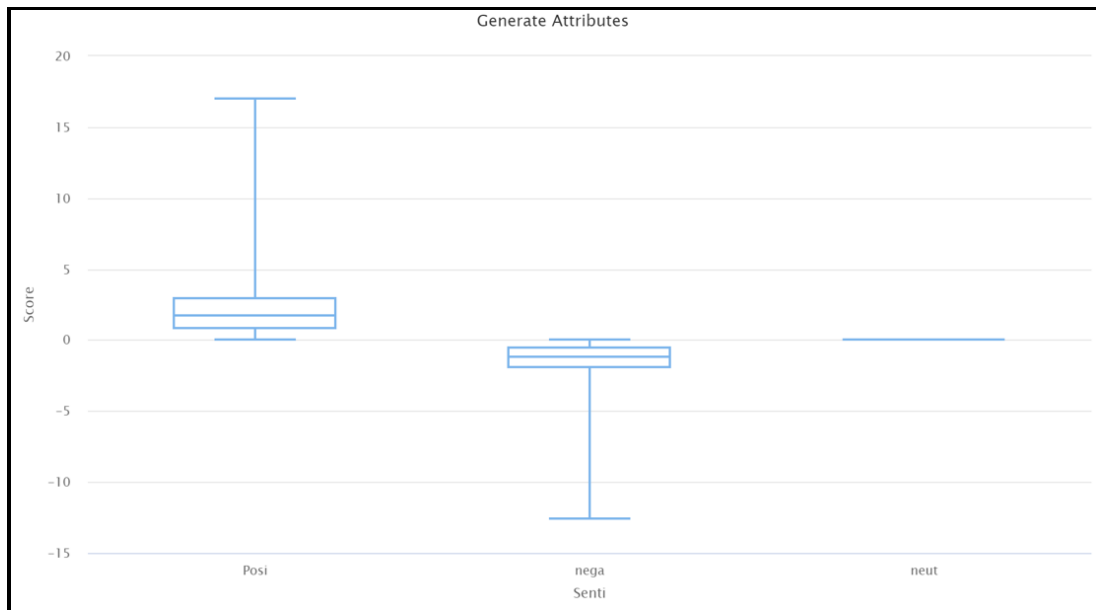
SENTIMENT DISTRIBUTION ANALYSIS:



Key Business Insights:

- 1. **Leverage Strengths:** Positive reviews highlight customer satisfaction; use them in marketing to boost loyalty and attract new customers.
- 2. **Address Weaknesses:** A high volume of negative reviews signals dissatisfaction; focus on resolving common issues like service quality or delays.
- 3. **Polarized Opinions:** Low neutral sentiment indicates strong customer opinions; prioritize handling negative feedback to protect brand reputation.
- 4. **Segment Customers:** Target positive customers with loyalty rewards and address negative ones with recovery initiatives to improve retention.
- 5. **Explore Neutral Sentiment:** Neutral reviews represent an untapped opportunity; targeted actions can convert them into positive advocates.

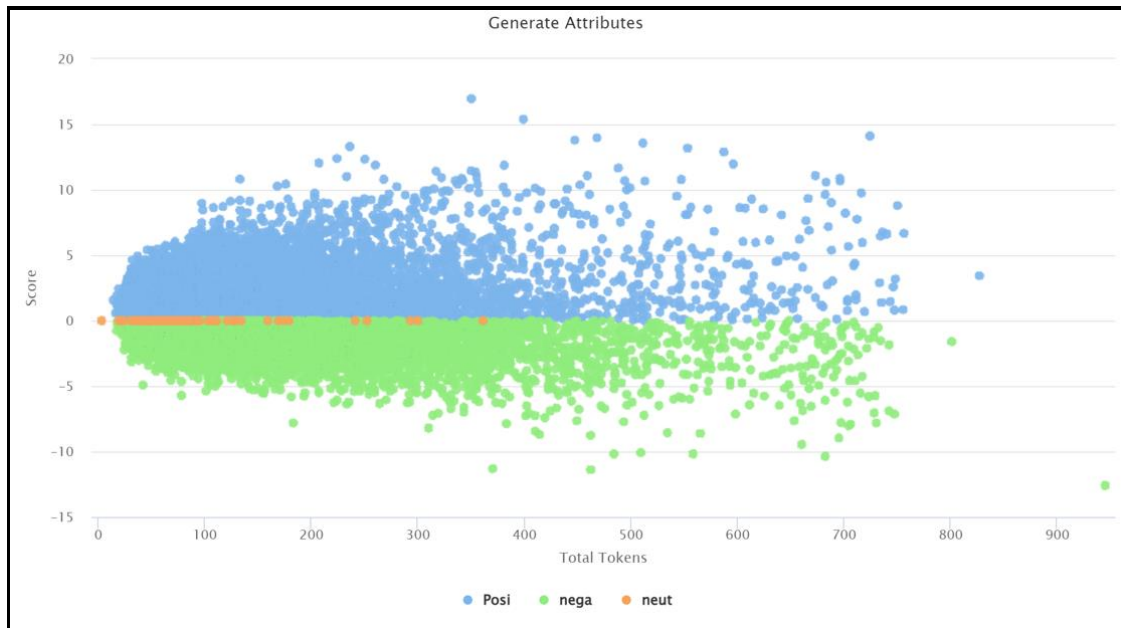
SENTIMENT INTENSITY ANALYSIS:



Key Business Insights from Sentiment Intensity Analysis (Box Plot)

1. **Wide Positive Range:** Positive reviews vary widely with high outliers. Action: Use consistent positive feedback for marketing and replicate praised experiences.
2. **Tight Negative Range:** Negative scores show extreme low outliers. Action: Address severe issues like delays and poor service to reduce dissatisfaction.
3. **Minimal Neutral Reviews:** Neutral feedback is rare. Action: Engage neutral customers to convert them into positive advocates.
4. **Polarized Sentiments:** Significant differences between positive and negative scores highlight polarization. Action: Build loyalty for positive customers; recover negative ones with tailored efforts.
5. **Focus on Outliers:** Extreme negative outliers indicate urgent improvement areas. Action: Perform root cause analysis and monitor sentiment trends.

SENTIMENT TREND OVER TOKEN:

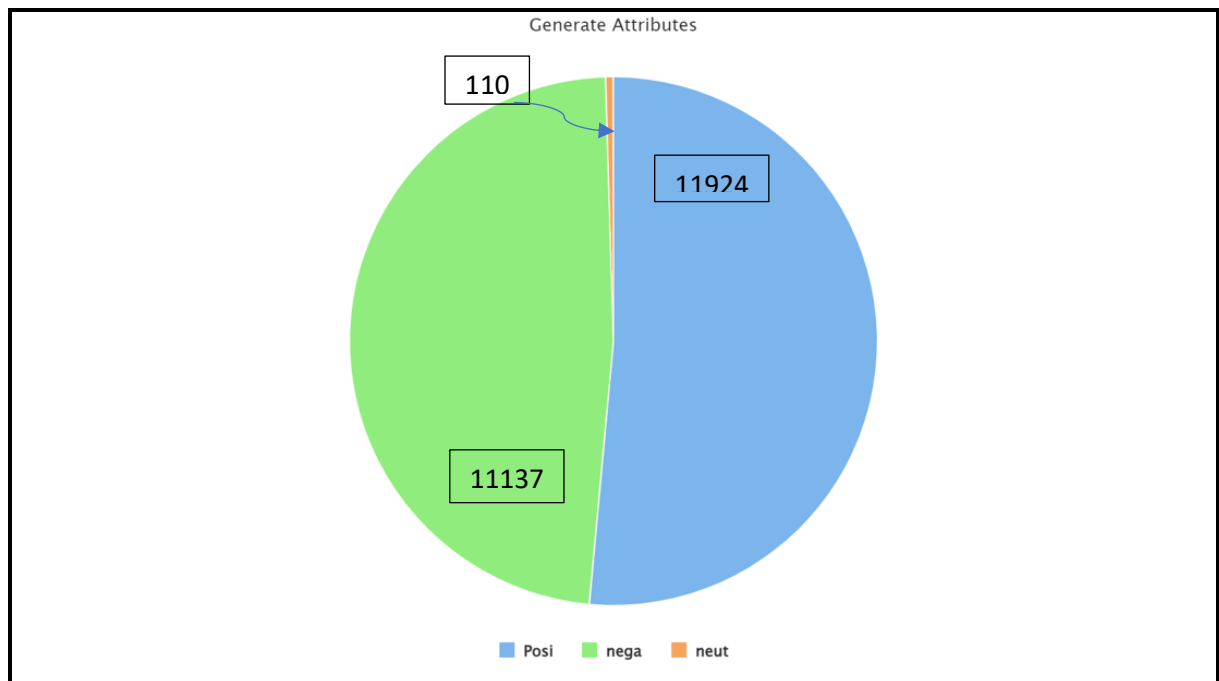


KEY BUSINESS INSIGHTS:

Key Insights:

1. **Long Reviews are Polarized:** Longer reviews tend to have more extreme sentiments (highly positive or highly negative). Focus on these for detailed feedback.
2. **Neutral Reviews are Short:** Neutral reviews are generally shorter, indicating less detailed feedback. Engage these customers to gather actionable insights.
3. **Positive Sentiment Increases with Length:** Highlight detailed positive reviews in marketing to showcase customer satisfaction.
4. **Negative Sentiment Trends with Length:** Longer negative reviews often highlight critical issues; prioritize resolving these to improve satisfaction.

SENTIMENT PROPORTION ANALYSIS:



KEY BUSINESS INSIGHTS:

Key Insights:

1. **Balanced Sentiments:** Positive (11,924) and Negative (11,137) reviews are nearly equal, reflecting a polarized customer base. Focus on reducing negatives to shift the balance.
2. **Low Neutral Feedback:** Neutral reviews (110) are minimal, indicating strong customer opinions. Encourage feedback from less vocal customers.
3. **Leverage Positives:** Highlight positive reviews in marketing to reinforce brand trust.
4. **Address Negatives:** Analyse negative reviews to identify and fix recurring issues, improving customer satisfaction.

Classification Matrix

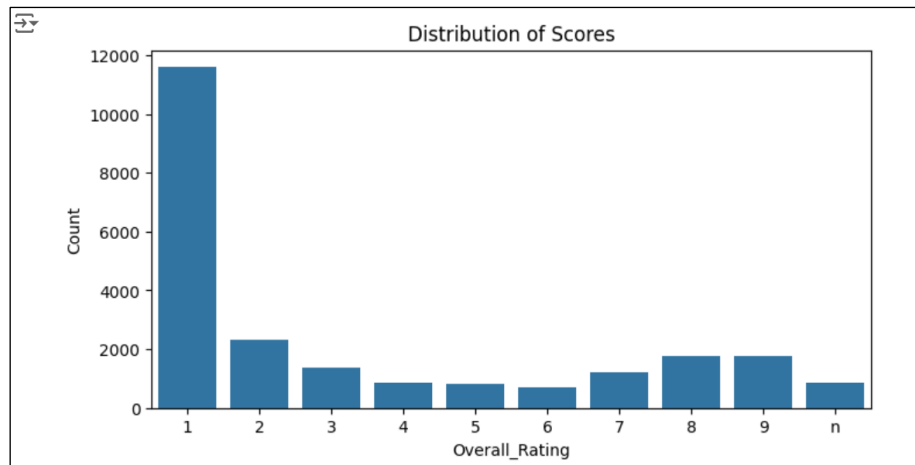
	precision	recall	f1-score	support
HighRating	0.58	0.63	0.61	1310
LowRating	0.86	0.93	0.90	5081
MediumRating	0.95	0.04	0.07	561
accuracy			0.80	6952
macro avg	0.80	0.53	0.52	6952
weighted avg	0.82	0.80	0.78	6952
0.8049482163406214				

1) Recall = $TP / (TP + FN)$	2) Precision = $TP / (TP + FP)$
3) False Negative (FN) = Support-TP	4) True Negative(TN) = Total Samples - (TP+FP+FN)

From the classification matrix we can say:

- 1) **High Customer dissatisfaction** - TP for low Rating = 4725.33, means majority of the reviews are negative.
- 2) **Neutral feedback** - Low recall for medium rating = 4%, means airline is not properly addressing moderately satisfied customers sentiments.
- 3) **Limited positive sentiments** - FN for High Rating = 484.7, means less number of delighted customers.

Analyzing Score Distribution (Plot between Overall Rating and Count)



Interpretation of the score distribution

- 1) Majority of the reviews have a rating of 1, which shows many of the customers are dissatisfied with the airlines services.
- 2) Very few customers have given a rating of 8,9,n , which shows only a small segment of the customers are satisfied with the airline service.
- 3) Rating of the customers between 2-7 shows that customers experiences are very polarized, either it is very negative or very positive.

Inferences

- 1) Majority poor score of 1 shows that many customers are unhappy may be due to repeated operational failures, poor service quality.
- 2) The polarized feedback shows inconsistencies in service provided across different flights.

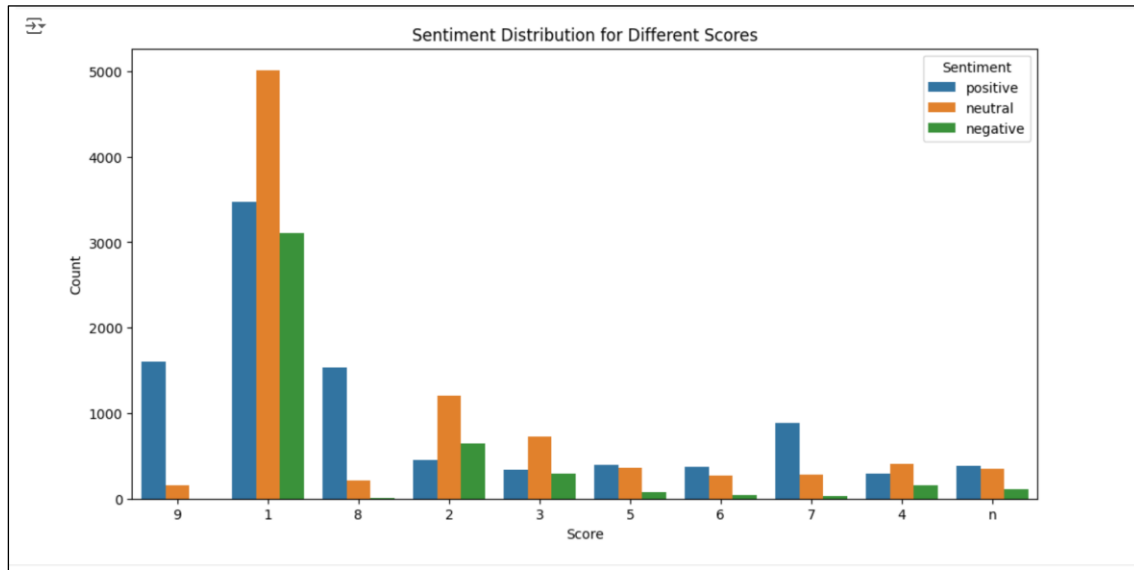
Suggestions to Airlines

- 1) Do the root cause analysis for this low rating. Analyze text reviews of the customer highlighting delays and cancellation.
- 2) Make a team to solve these customer queries.
- 3) Standardize the customer experiences across all the flights so that there is no polarization of the reviews, and train the staff accordingly.

Business Insight

- 1) Airlines seems to be at the risk of losing many customers unless solving the customer complaints.
- 2) Train staffers to solve the customer issues.

Sentiment Distribution for Different Scores



Interpretation

- 1) Score =1, categorized as neutral and has a negative sentiments. This means customers may have used neutral language although they were dissatisfied.
- 2) Score = 8-9, categorized as positive, shows complete customer satisfaction.
- 3) Score between 3-7, has equal number of positive, negative and neutral sentiments, which shows inconsistency in customer insights for these scores.

Business Insights

- 1) The large number of low score and mixed sentiments shows that a large number of customers may switch to competitors.
- 2) Address the issues of the neutral sentiments for low score customers and solve them.

Negative Reviews Word Cloud

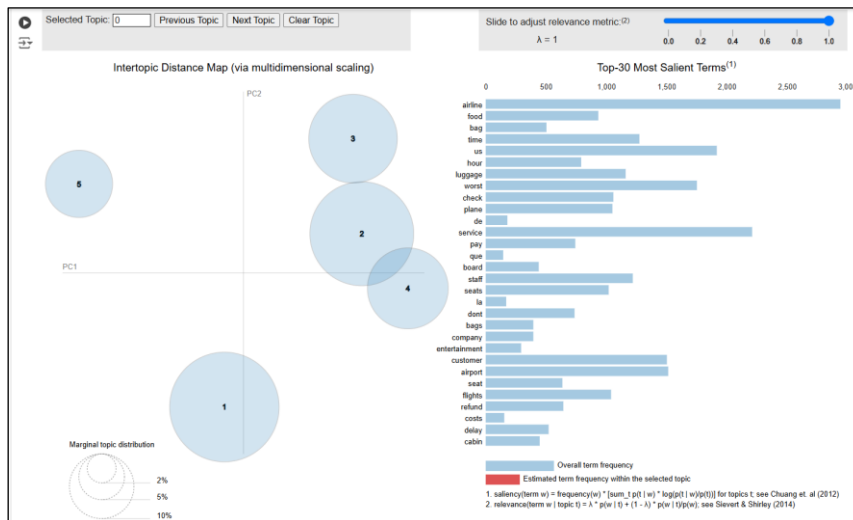
The word cloud displays a variety of terms related to negative travel experiences. The most frequent and largest words include 'airline', 'flight', 'service', 'customer', 'seat', 'delay', 'baggage', 'time', 'passenger', 'ticket', 'refund', 'food', 'rude', 'lost', 'worst', 'horrible', 'disappointed', 'late', 'missed', 'boarding', 'waiting', 'pay', 'take', 'way', 'now', 'said', 'back', 'boarding', 'worst', 'hotel', 'phone', 'got', 'travel', 'even', 'experience', 'told', 'minute', 'asked', 'called', 'never', 'flew', 'bag', 'paid', 'money', 'extra', 'line', 'another', 'terrible', 'will', 'never', 'flew', 'bag', 'paid', 'money', 'extra', 'line'. Other visible words include 'plane', 'luggage', 'information', 'put', 'small', 'check', 'flights', 'departure', 'due', 'company', 'said', 'boarding', 'reason', 'worst', 'finally', 'hotel', 'book', 'offer', 'already', 'travel', 'response', 'even', 'experience', 'told', 'minute', 'asked', 'called', 'never', 'flew', 'bag', 'paid', 'money', 'extra', 'line', 'another', 'terrible', 'will', 'never', 'flew', 'bag', 'paid', 'money', 'extra', 'line'.

- 1) Word like (airline, airport, hour, time, seat) are mostly mentioned in the negative reviews.
- 2) Clustering the issues -
 - a. Words like (delay, hour, airport) suggest operational delays
 - b. Word like (staff, rude, customer) suggest service related issues.
- 3) Words like (Worst, horrible, rude) shows intensity of customer dissatisfaction.

- 1) Bold mentioning of the words like (airport, hour, time) - shows the dissatisfaction of the customers with the punctuality and time management.
- 2) Words like (staff, rude, customer) shows customer service not upto the mark of the expectation.
- 3) Words like (refund, ticket, compensation) shows airline's policies on cancellation or refunds which dissatisfied the customers.
- 4) Words like (baggage, lost) shows luggage management fails.

- 1) Reduce the operational delay and optimize the schedule and pre inform any delays or cancellation to the customers.
- 2) Refund policy should be more streamline and a proper system to be in place to handle that.
- 3) Upgrade the baggage tracking management to reduce the baggage missing complaints.
- 4) High focus on comfortable seats.
- 5) Staffers should be train properly, so they talk politely with customers and not being rude.

Visual extract topics from LDA model



Interpretation

- 1) Map shows 5 distinct topics extracted by the LDA model
- 2) Topic 5 and Topic 1 have less overlap and are far apart, seems to be distinct
- 3) Topic 2,3 and 4 share more common words and are related more.
- 4) Common Terms are = airline, food, bag, time, service, customer, airport, staff, plane, refund, delay, bags.
- 5) Relevance Metric ($\lambda=1$) - shows the frequency and contribution for each topics, like - (airline, service) are very commonly repeated.

Suggestions to Airline

- 1) Airlines focus should be on reducing delays and timely performance, as words like (time, hour, delay) are there.
- 2) High mention of 'food', shows dissatisfaction of the customers with the food service. So improve the menu and offer good food.
- 3) Focus on staff training, as customer frequently mention (staff, service)
- 4) Simplify the refund policy and improve the baggage management.

Business Insights

- 1) Words like (refund, costs, baggage) shows high dissatisfaction with the airlines refund policies and baggage handline, this could be very bad for brand image.
- 2) By solving operational and service related complaints customers retention and satisfaction will increase.
- 3) Utilize the data drive insights and work on the high priority areas.