

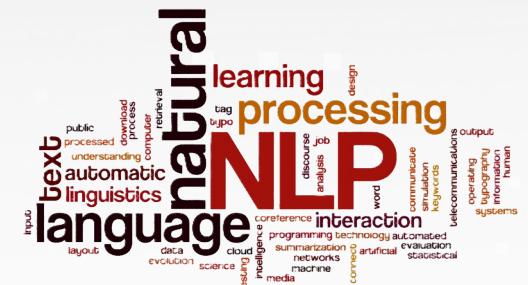


Proactive Customer Experience Management

Air India: A Data-Driven Strategy

Submitted By: Group-8

Name	Roll No



Business Understanding & Problem Statement

Why Customer Experience Matters

- In India's highly competitive aviation market, customer experience is a key differentiator.
- Passengers today rely heavily on social media, reviews, and digital platforms to shape perceptions.
- A single bad experience can amplify quickly online, damaging brand reputation.



Central Business Question

How can Air India leverage integrated text analytics and web analytics to proactively monitor, diagnose, and enhance customer experience?

Project Objectives

- Convert **real-time feedback** into actionable insights.
- Enhance **operational efficiency** and **service reliability**.
- Strengthen **brand reputation** by addressing customer pain points proactively.



Air India's Current Challenges

- **Safety & Maintenance Concerns** – long-standing issues raise trust deficits.
- **Service Delays** – frequent delays frustrate passengers and weaken loyalty.
- **Negative Digital Sentiment** – social media & review platforms highlight recurring dissatisfaction.



Data Understanding – Datasets Overview



Two Major Data Sources

1. Customer Reviews Dataset (Demand-Side Insights)

Data Source: www.airlinequality.com/airline-reviews (Web Scraping)

- Who:** Traveller profiles (Solo, Couple, Family)
- What:** Travel class (Economy, Premium, Business), Route details (Domestic & International)
- How They Felt:** Overall rating (1–10 scale) + detailed review text
- Why It Matters:** Provides **direct passenger sentiment**, highlighting satisfaction drivers & recurring pain points.

Feature	Description	Example
Keyword	The exact search query users type into search engines. Reflects customer interest areas.	"air india flight booking," "air india refund policy,"
Search Engine Position	Ranking of Air India's webpage for that keyword. Higher ranks bring higher visibility.	Position 1, Position 3, Position 9
URL	The Air India webpage that ranks for the keyword. Shows where users are directed.	https://www.airindia.com/in https://www.airindia.com/in/en/refund
Traffic	Shows estimated search traffic volume for each keyword. Indicates keyword popularity.	12,000 visits, 3,500 visits, 800 visits.

Feature	Description	Example
Type of Traveller	Shows the category of the traveller such as solo, couple, or family. Helps analyse experience differences across groups.	"Solo Leisure," "Couple Leisure," "Family Leisure."
Seat Type	Indicates the cabin booked by the traveller (Economy, Premium Economy, Business). Satisfaction often varies across classes.	"Economy Class," "Premium Economy," "Business Class."
Route	Lists the origin and destination of the journey. Different routes highlight variations between domestic and international services.	"Delhi to Toronto," "Kolkata to Mumbai," "Chicago to Delhi."
Date Flown	The actual flight date, useful for checking patterns during different seasons or periods.	"14 August 2023," "05 January 2024," "02 July 2025."
Overall Rating	A numerical score (1–10) summarizing satisfaction. Quick indicator of customer sentiment.	1, 4, 8
Review Title & Text	Title is a short summary, and text contains detailed opinions. Rich source for sentiment and complaint analysis.	Title: "Super unreliable airline" to Text: "Flight was delayed, missed connection."
Review Date	Date when the review was posted online. Useful for timeline analysis.	"15 September 2023," "02 May 2024," "31 August 2025."
Recommended	Shows whether the passenger recommends Air India (Yes/No). Direct indicator of loyalty.	"Yes," "No."

2. Web Analytics Dataset (Supply-Side Insights)

- What People Search:** Keywords, traffic volumes, engagement levels
- How Air India Performs:** Organic rankings, click-through potential, URL-level metrics
- Why It Matters:** Reflects **digital visibility & market demand**, showing how well Air India's online presence aligns with customer interest.

Data Preparation Process

Text Analytics

```
# Drop all rows that contain at least one null value
df = df.dropna(how='any')

# Reset index after dropping rows
df = df.reset_index(drop=True)
```

Scope: Unstructured customer reviews

Cleaning:

- Drop nulls → reset index

Transform:

- Lowercase → regex noise removal
- Tokenize → remove stopwords (+ custom: airindia, air, india, airline, airport)
- Lemmatize → normalize variants

Output:

- Clean_Review for sentiment and topics

Web Analytics

```
# ... Clean & Select Relevant Columns ...
page_df = df[["URL", "Keyword", "Search Volume", "Position", "Traffic", "CPC", "SERP Features by Keyword", "Keyword Intents"]].copy()

# Handle missing values
page_df = page_df.dropna(subset=["URL"])
```

Scope: SEO/traffic Excel export

Cleaning:

- Load Excel → select [URL, Keyword, Search Volume, Position, Traffic, CPC, SERP Features, Intent]
- Drop rows with missing URL

Transform(by URL):

- Keyword Count
- Total Search Volume
- Total Traffic
- Avg CPC
- Avg Position

Output:

- Page-level aggregated table for performance analysis

TEXT ANALYTICS

Text Analytics - Sentiment Analysis Models

Dual Sentiment Analysis Approach

TextBlob Sentiment Analysis:

- Polarity score: -1.0 (negative) to $+1.0$ (positive)
- Threshold: Positive (≥ 0.1), Negative (≤ -0.1), Neutral (-0.1 to 0.1)
- Average polarity: 0.134 (slightly positive)
- Average subjectivity: 0.551

Technique that computers use
to extract worthwhile information
from the human language
Text Analysis in a smart
and efficient manner.

```
# =====
# STEP 6: TextBlob Sentiment
# =====
def get_textblob_sentiment(text):
    analysis = TextBlob(str(text))
    if analysis.sentiment.polarity > 0.1:
        return "Positive"
    elif analysis.sentiment.polarity < -0.1:
        return "Negative"
    else:
        return "Neutral"

df['TextBlob_Sentiment'] = df['Clean_Review'].apply(get_textblob_sentiment)
```

VADER Sentiment Analysis:

- Compound score for overall sentiment
- Threshold: Positive (≥ 0.05), Negative (≤ -0.05)
- Average compound score: 0.311 (slightly positive)
- Better for social media and informal language

Why Dual Approach Matter

- Cross-validation: Using both ensures **robust results**.
- Different perspectives: One captures **grammatical nuance**, the other **informal tone**.

```
# =====
# STEP 7: VADER Sentiment
# =====
analyzer = SentimentIntensityAnalyzer()

def get_vader_sentiment(text):
    scores = analyzer.polarity_scores(str(text))
    if scores['compound'] >= 0.05:
        return "Positive"
    elif scores['compound'] <= -0.05:
        return "Negative"
    else:
        return "Neutral"

df['Vader_Sentiment'] = df['Clean_Review'].apply(get_vader_sentiment)
```

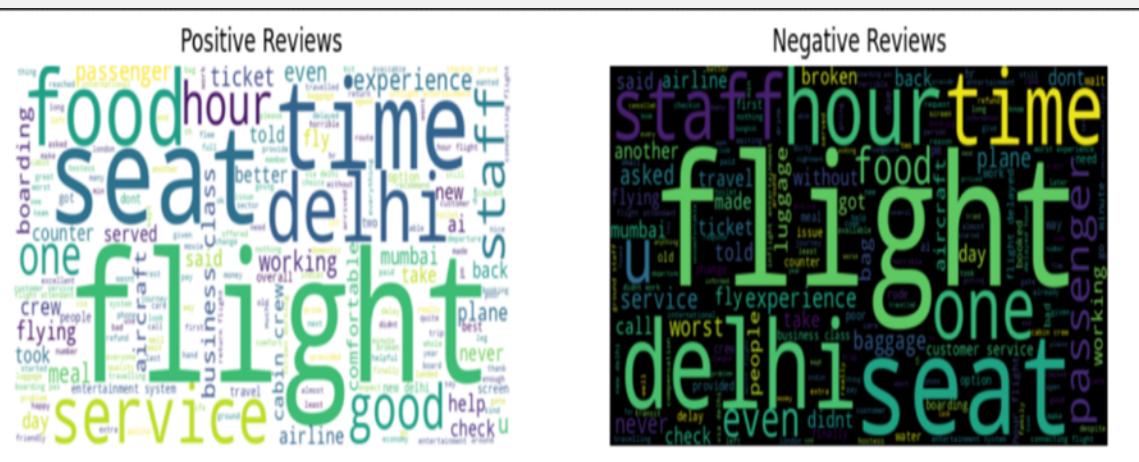
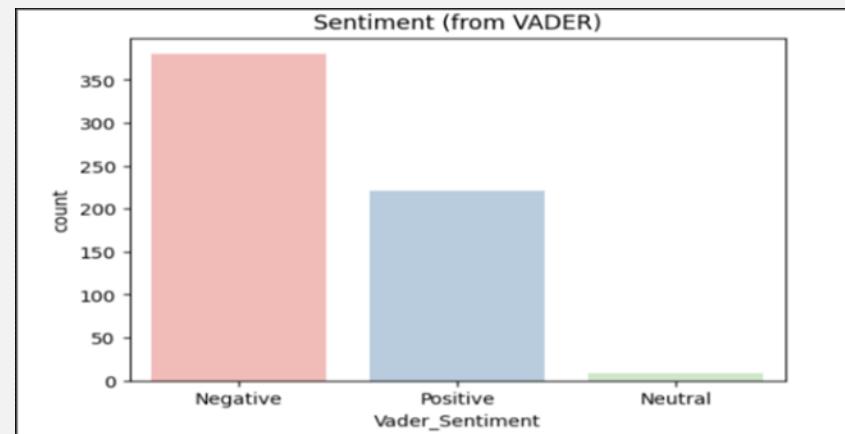


Text Analytics - Sentiment Analysis Models

Visualizing Customer Sentiment

Sentiment Distribution (VADER Analysis)

- Majority of reviews are **negative (~360)**, followed by **positive (~220)**, and very few neutral.
- Clear signal: Air India faces a **reputation challenge** with dissatisfaction dominating online feedback.



Word Cloud Insights

Positive Reviews:

- Highlight words: *flight, seat, food, service, staff, good*
- Strengths lie in **crew behavior, food quality, and select routes**.

Negative Reviews:

- Dominant words: *delay, seat, time, staff, flight, Delhi*
- Pain points are **delays, seat comfort, and ground/airport experience**.

Evidence of Negative Sentiment



YouTube
international ECONOMY CLASS ...

Mint



Travel influencer calls out Air India over 'broken seat, bad food, service' in business class | Today News

Visit >

K C Venugopal ✨ · Aug 11, 2025
@kcvenugopalmp · Follow

Air India flight AI 2455 from Trivandrum to Delhi - carrying myself, several MPs, and hundreds of passengers - came frighteningly close to tragedy today.

What began as a delayed departure turned into a harrowing journey. Shortly after take-off, we were hit by unprecedented

Air India ✨
@airindia · Follow

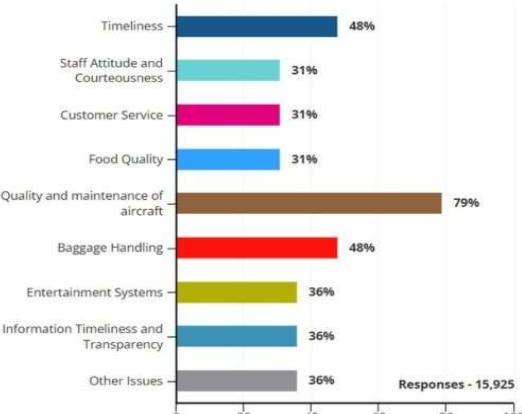
Dear Mr Venugopal, we would like to clarify that the diversion to Chennai was precautionary due to a suspected technical issue and poor weather conditions. A go-around was instructed by Chennai ATC during the first attempted landing at Chennai airport, not because of the presence [Show more](#)

1:14 AM · Aug 11, 2025

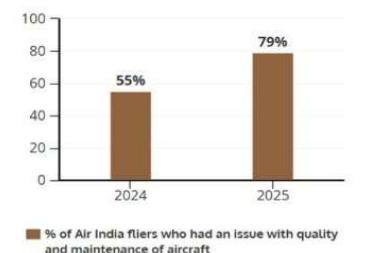
2.5K Copy link

79% Air India passengers surveyed confirmed experiencing quality and aircraft maintenance issues in the last 12 months, marking a jump from 55% who experienced such issues in 2024

What are some of the issues you have experienced when flying Air India in the last 12 months?



Sizable increase observed in Air India fliers who had an issue with airline's quality & maintenance of aircraft in 2025 as compared to 2024



**Some respondents have selected more than 1 options in the survey and hence the total does not equate to 100%

'8 Hours Of Absolute Chaos': Air India Traveller Describes Flight Experience Without Tray Table

Curated By : News Desk | Edited By : Qindra Mukherjee
Last Updated: March 26, 2025, 21:06 IST

The traveller, who was with two children on the Air India flight, detailed the chaotic experience on their social media and said they tried to balance meals and drinks on a standalone tray provided by the crew

Impact Shorts
Big news, made easy

Choose News18 on Google G



The traveller posted purported pictures of an Air India consent form and a standalone tray used by them during the journey. (Image: @mistercredible/X)

A seasoned traveller claimed to have had a chaotic journey on an Air India flight after they discovered that the tray table failed to open in the seat they were assigned.

The individual known as 'Mr. Credible' on their social media said they were travelling with two children and it was "8 hours of absolute chaos" as they tried to balance meals and drinks on a standalone tray provided by the crew.



From August 1 to September 30
Air India shifts A'bad-London operation to Heathrow from Gatwick under 'Safety Pause'

RECOMMENDED STORIES

1 Jammu Airport Chaos: Air India Express Passengers Stranded For Hours, Barred From Deboarding | Watch



Anmedabad-Gatwick route following the AI171 crash. From August 1 to September 30, a reduced thrice-weekly service to Heathrow will operate instead.

#airindia #ai171 #ahmedabadflights #heathrow #gatwick #aviationupdate #flightsuspension #dgca #boeing #787 #ahmedabadnews

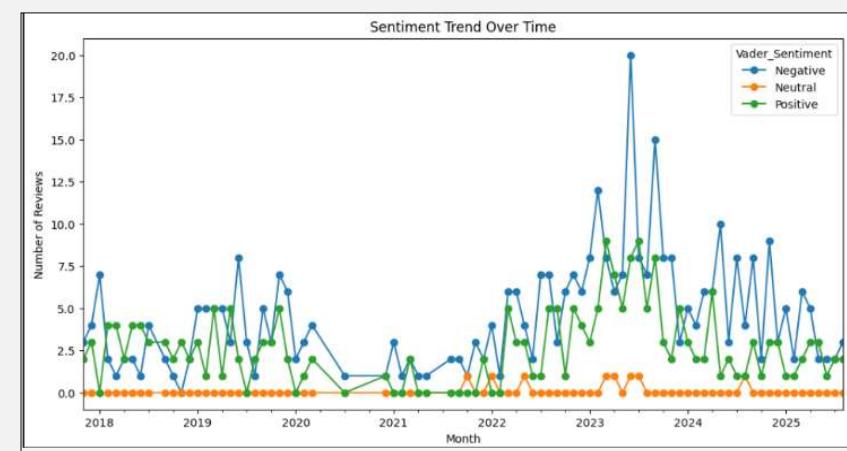
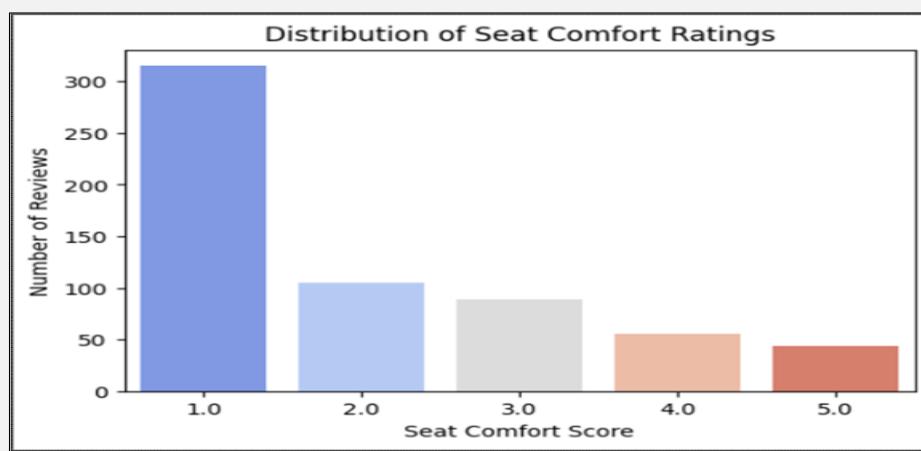
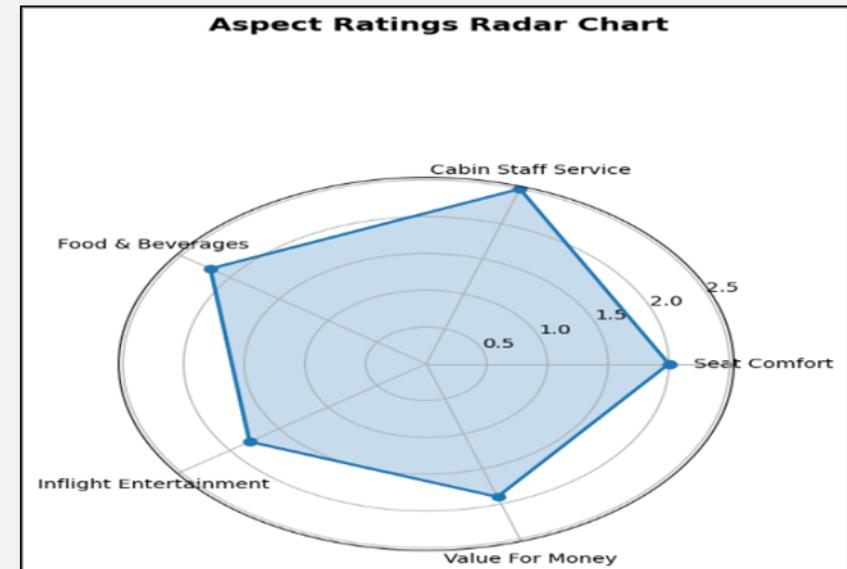
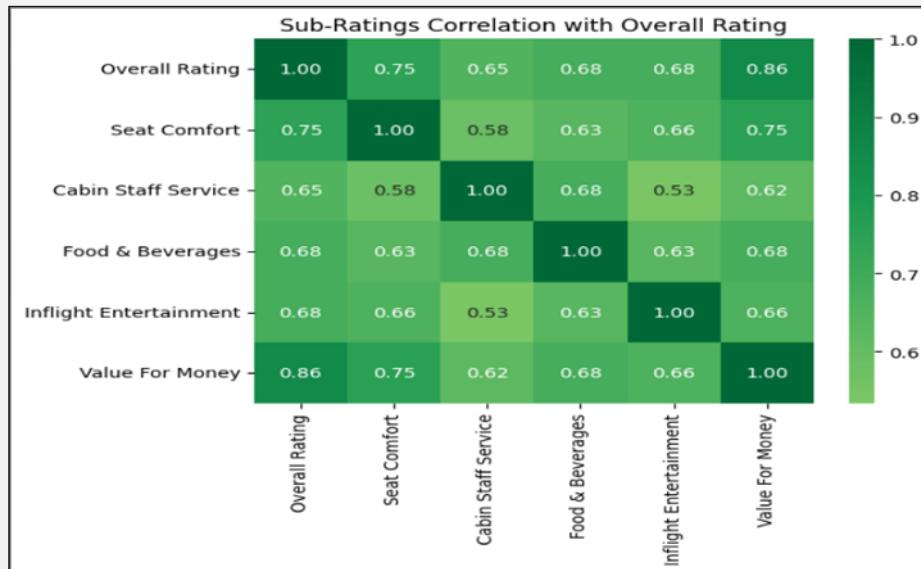


No comments yet.

Start the conversation.

10 likes
July 16

Text Analytics - Sentiment Analysis Models



Text Analytics - Topic Modeling Results



Topic 1 - Word Cloud



Topic 1: Food, Seats, Entertainment

Insight: In-flight quality and comfort drive satisfaction; staff links to overall experience.

Topic 2 - Word Cloud



Topic 2: Boarding & Seating Issues

Insight: Check-in/boarding inefficiencies, broken seats, delayed assistance.

Topic 3 - Word Cloud



Topic 3: Delays & Airport Experience

Insight: Long waits, baggage issues, weak ground support.

Topic 4 - Word Cloud



Topic 4: Crew & Systems Not Working

Insight: Malfunctioning IFE/seats; issues tied to crew response.

Topic 5 - Word Cloud



Topic 5: Business Class Service

Insight: Premium expectations vs reality; comparisons with economy

Text Analytics - Clustering and Classification Results

Document Clustering

Objective

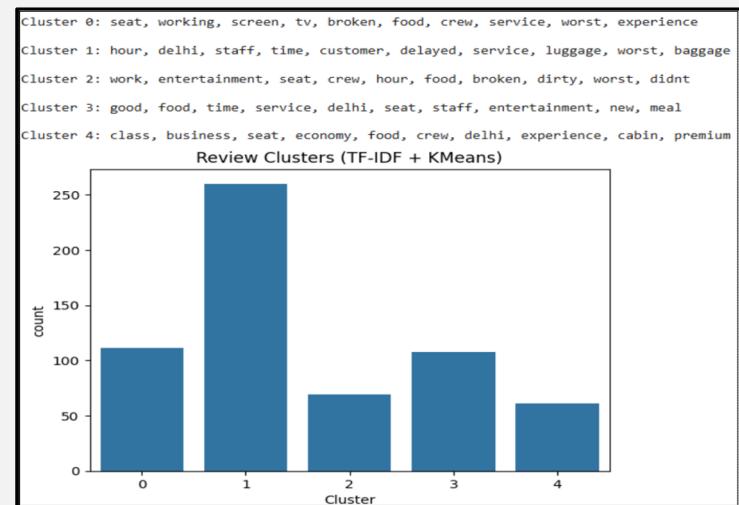
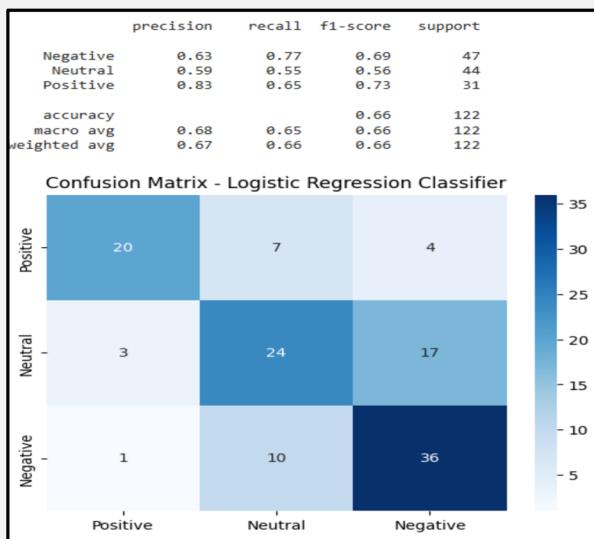
- Unsupervised grouping of reviews to surface major themes; no labels required.

Method

- TF-IDF vectorization to weight important terms across corpus.
- K-Means (k=5) to partition reviews into coherent clusters.

Output

- 5 clusters, each defined by dominant keywords indicating central topics.



Classifier Performance — Sentiment

Goal

- Auto-label reviews as Positive, Neutral, Negative for scalable monitoring.

Best model

- Linear SVM selected after model comparison; trained on labeled reviews.

Evaluation snapshot

- Overall accuracy: 0.66 on test set.
- Macro F1: 0.66; stable across classes.

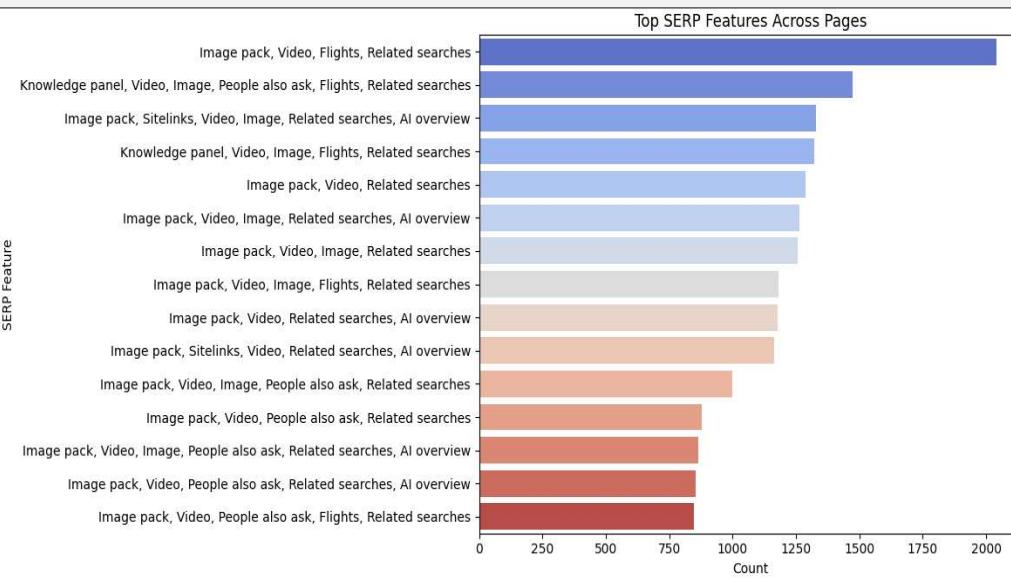
WEB ANALYTICS



WEB ANALYTICS - SERP & Keyword Intent Insights

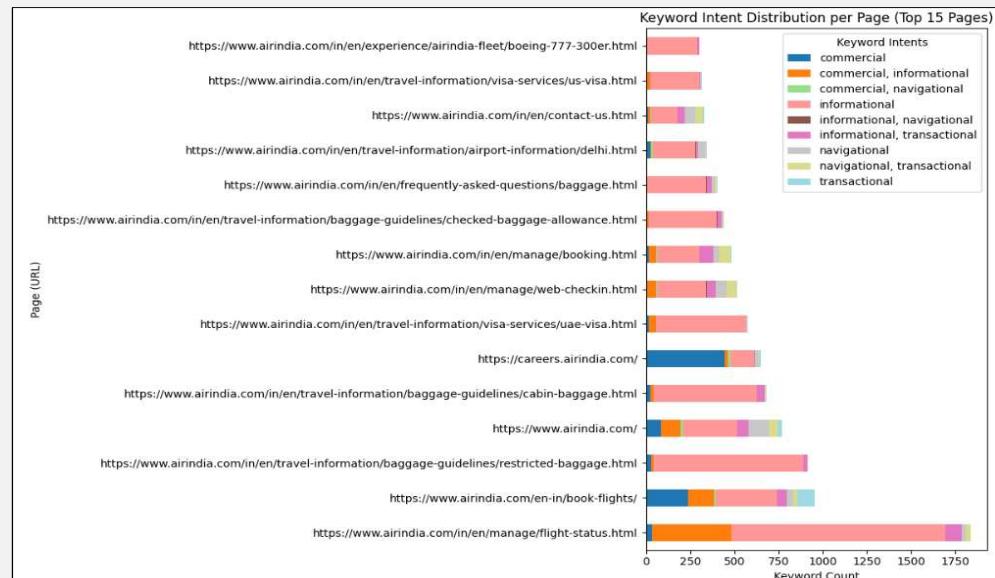
SERP Features

- Air India appears in **Image Pack, Video, Flights, Related Searches**, but ranking is **scattered**, not consistently in top 3 positions.
- **Missed opportunities** in high-value features like *People Also Ask* and *Knowledge Panel*, where competitors often dominate.
- Current visibility is **more visual (images, videos)** than **informational (FAQs, authority panels)**.
- **Implication:** Air India should strengthen structured content (FAQs, schema markup, SEO optimization) to capture Knowledge Panel & People Also Ask — boosting brand authority and click-through rates.



Keyword Intent Distribution

- Majority of keywords are **informational** (baggage rules, check-in, policies).
- **Transactional keywords** (flight booking, offers) are fewer → weaker monetization focus.
- Poor alignment between **user search intent** and **website content strategy**(e.g., high traffic to policy pages but low satisfaction).



Insight: SEO strategy should shift from only informational to transactional +service-oriented queries(refunds, cancellations, booking).

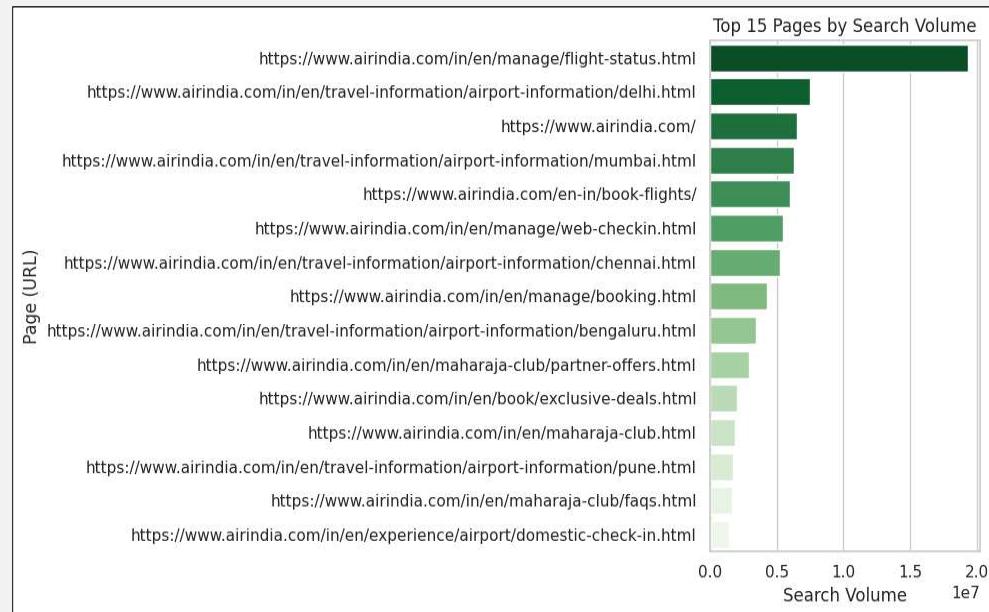
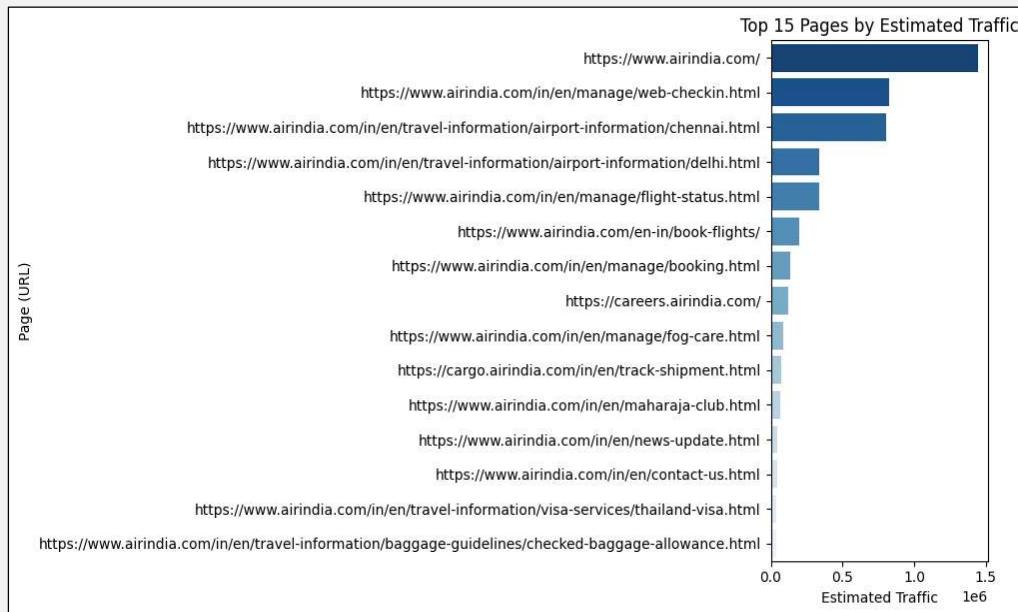
WEB ANALYTICS - Traffic & Search Volume Insights

Traffic Trends

- Homepage, Web Check-in, and Flight Status** dominate traffic → customers heavily rely on operational updates.
- Booking pages attract visitors but underperform** due to negative experiences (refund delays, booking failures).
- High-traffic service pages are **not fully optimized for usability**, creating drop-offs in customer journeys

Search Volume Trends

- Flight Status & Airport Information** pages generate the highest search demand → shows passengers prioritize reliability and real-time updates.
- Transactional searches (Book Flights, Manage Booking)** are lower vs service-related searches, signaling weak digital trust in booking flows.
- Customers are **more concerned with “post-purchase pain points” (status, cancellations, refunds)** than with new bookings.



Insight: Strengthen **service-related workflows (refunds, cancellations, tracking)** and **optimize booking journey** to convert high visibility into revenue.

Building the Recommendation System

01

Data Preprocessing

Filter essential columns: Author, Route, Overall Rating. Coerce ratings to numeric format and remove null values to ensure data quality.

02

Cold-Start Protection

Implement minimum threshold of ≥ 2 ratings per user and route. This ensures meaningful similarity calculations whilst balancing coverage and noise.

03

User-Route Matrix Construction

Pivot data structure with Author as rows, Route as columns, and rating values. Fill missing interactions with 0, indicating "unrated" rather than negative preference.

04

Similarity Computation

Calculate cosine similarity across user vectors. Output structured user_similarity_df table for efficient recommendation generation.

05

System Validation

Implement robust error handling: flag empty datasets, validate matrix dimensions, confirm successful engine build with appropriate logging.

```
# =====
# SECTION 3: BUILDING THE CORE ENGINE
#
print("\nStep 2: Building Collaborative Filtering Engine...")
def create_engine_matrices(df):
    df_filtered = df[['Author', 'Route', 'Overall Rating']].dropna()
    df_filtered['Overall Rating'] = pd.to_numeric(df_filtered['Overall Rating'], errors='coerce').dropna()
    user_counts = df_filtered['Author'].value_counts()
    route_counts = df_filtered['Route'].value_counts()
    df_final = df_filtered[df_filtered['Author'].isin(user_counts[user_counts >= 2].index) &
                           df_filtered['Route'].isin(route_counts[route_counts >= 2].index)]
    if df_final.empty: raise ValueError("Filtered dataset is empty.")
    user_route_matrix = df_final.pivot_table(index='Author', columns='Route', values='Overall Rating').fillna(0)
    user_similarity_df = pd.DataFrame(cosine_similarity(user_route_matrix),
                                       index=user_route_matrix.index, columns=user_route_matrix.index)
    return user_route_matrix, user_similarity_df

user_route_matrix, user_similarity_df = create_engine_matrices(main_df)
print("Engine built successfully!")
```

Key Assumptions

- Zero values represent unrated items
- Minimum 2 ratings ensure statistical validity

Technical Output

Success: "Engine built successfully!"

Error: Data validation flags raised

Recommendation Functions — User Scenarios

This recommendation system personalizes route suggestions by analyzing user ratings and travel patterns. For new users, the system highlights the highest-rated routes overall. For goal-driven travelers, destination-based and origin-based filtering helps identify the best possible options. As a result, every user scenario, whether frequent flyer or first-timer receives the most relevant, data-driven route recommendations.

New User

Provides route suggestions based on overall ratings when no prior history exists.



Destination Based

Recommends the best routes leading to a chosen destination city.

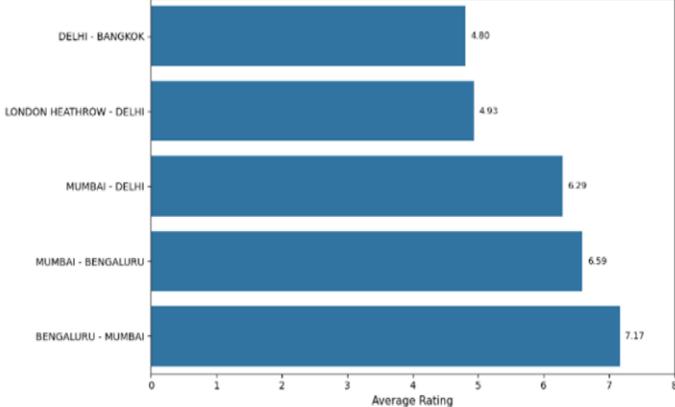


Origin Based

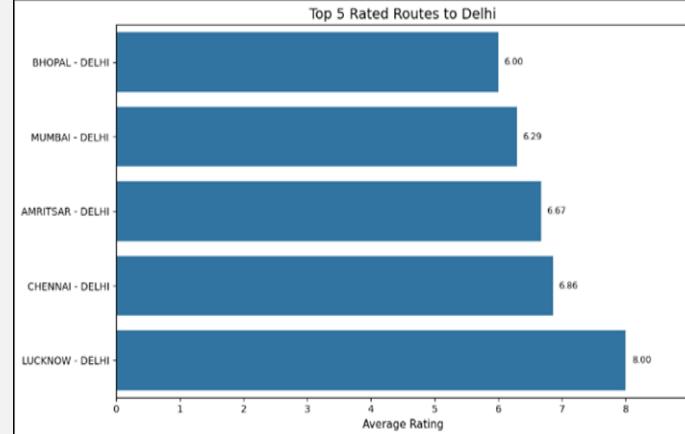
Recommends the best routes starting from a specified origin city.



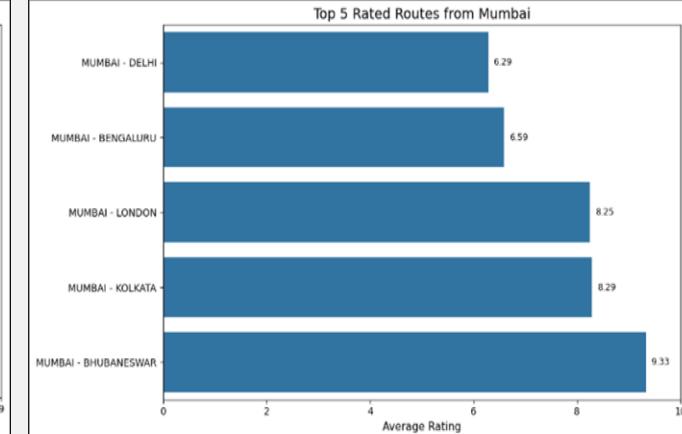
Top 5 Routes for a New User



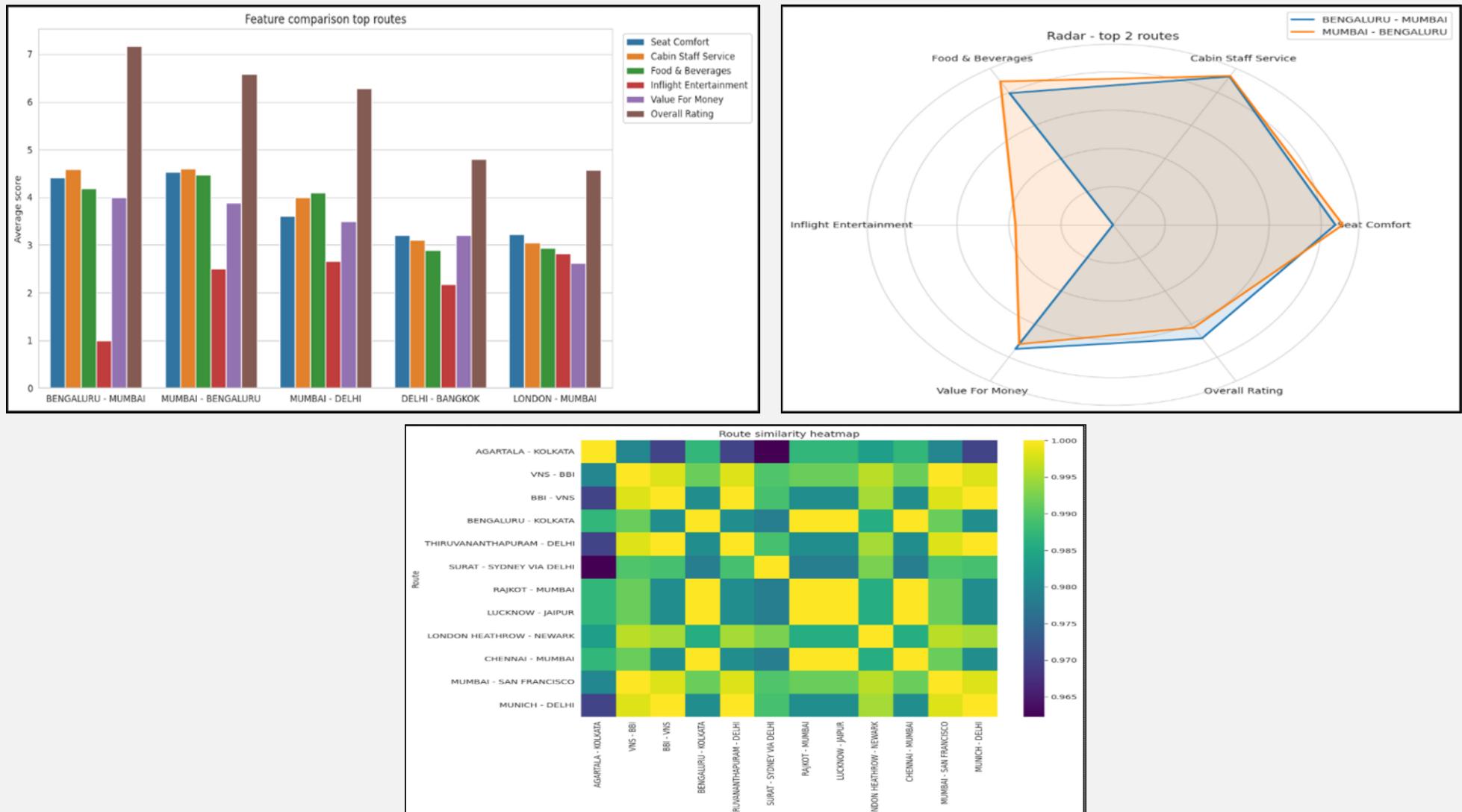
Top 5 Rated Routes to Delhi



Top 5 Rated Routes from Mumbai



Web Analytics



Business Implications and Recommendations



Customer Experience & Service Quality

Findings: Delays, poor professionalism, inconsistent service → 1/3rd reviews negative.

Actions:

- Fix top 3 complaints with ops & staff retraining.
- Service recovery: respond in 48 hrs, resolve 75% in 7 days.
- Reward & promote high-performing staff.
- **Impact:** Cut negative reviews by 40% in 6 months; +20% increase in 4–5★ ratings.



Marketing & Customer Acquisition

Findings: High-impression pages, low conversion; weak landing pages.

Actions:

- Launch bundled offers (Flight + Ancillaries).
- Leverage customer quotes/ratings in ads.
- Prioritize SEO for high-intent queries.
- **Impact:** Higher click-through & organic conversions.



Operations & Capacity Planning

Findings: Demand–capacity mismatch; weak routes drain resources.

Actions:

- Reallocate aircraft from weak to strong routes.
- Use route similarity model for expansion.
- Monitor cancellations, load factor & complaint rate.
- **Impact:** Higher yield & lower inefficiency.

Product & Strategic Growth

Findings: Untapped demand for premium/ancillary services.

Actions:

- Pilot priority boarding, express check-in, add-ons.
- Dynamic pricing experiments.
- Quarterly review of growth & pricing strategies.
- **Impact:** Ancillary adoption ≥6% in pilots; incremental revenue in 6 months.



Governance & Performance Measurement

Finding: Performance tracking is fragmented across functions, with limited use of structured data-driven decision-making.

Actions:

- Build integrated CX + Ops + Marketing dashboard.
- Institutionalize A/B testing for pricing & campaigns.
- Conduct quarterly reviews.
- **Impact:** Embed **data-driven decision-making** across Air India.

Recommendation - Prototype

The screenshot shows a web-based dashboard titled "Air India Analytics" with a red header bar. The header includes the title, a "Sign in" button, and navigation links for "Analytics", "Sentiment", and "Recommendations". Below the header, a section titled "Sentiment Analysis Overview" displays real-time analysis of Air India customer feedback and social media mentions. It features three main cards: "Positive Sentiment" (1,247, 45.2% of total feedback), "Neutral Sentiment" (892, 32.3% of total feedback), and "Negative Sentiment" (621, 22.5% of total feedback). A larger card below shows "Total Feedback Analyzed" (2,760) across various platforms: Social Media (1,820), Website Feedback (298), App Reviews (542), and Email Surveys (100). At the bottom, a section titled "Word Cloud Analysis" is mentioned as showing the most frequently mentioned words in customer feedback with sentiment analysis.

Air India Analytics
Sentiment Analysis Dashboard

Analytics Sentiment Recommendations

Sign in

Sentiment Analysis Overview

Real-time analysis of Air India customer feedback and social media mentions

Positive Sentiment: 1,247 (45.2% of total feedback)

Neutral Sentiment: 892 (32.3% of total feedback)

Negative Sentiment: 621 (22.5% of total feedback)

Total Feedback Analyzed: 2,760

Social Media: 1,820

Website Feedback: 298

App Reviews: 542

Email Surveys: 100

Word Cloud Analysis

Most frequently mentioned words in customer feedback with sentiment analysis

THANK YOU

