# Airline Customer Satisfaction & Flight Delay Analysis

By: Bhoomika c

## 1.Project Overview

Airlines face frequent challenges with flight delays, which directly impact customer satisfaction. This project analyzes airline operational data and customer feedback to identify:

- Major causes of flight delays
- Impact of delays on customer satisfaction
- Routes, airlines, and time periods with poor performance

## 2.Business Objective

- Reduce flight delays
- Improve customer satisfaction scores
- Identify high-risk routes and airlines
- Support data-driven operational improvements

# 3.Dataset Description

### 1. Flights Table

| Column | Description |
|---|---|
| flight_id | Unique flight identifier |
| airline | Airline name |
| origin | Departure airport |
| destination | Arrival airport |
| scheduled_dep_time | Scheduled departure |
| actual_dep_time | Actual departure |

| Column | Description |
|---|---|
| departure_delay | Delay in minutes |
| arrival_delay | Arrival delay in minutes |
| day_of_week | Day of flight |
| weather_delay | Weather delay (min) |
| technical_delay | Technical delay (min) |

### 2. Customers Table

| Column | Description |
|---|---|
| customer_id | Unique customer ID |
| flight_id | Flight taken |
| age | Passenger age |
| class | Economy / Business |
| satisfaction_score | 1–5 |
| feedback_category | Service / Delay / Comfort |

## 4.Dataset Overview

**1.flight data** C:\Users\Administrator\Downloads\Flights_Cleaned.xlsx

**2.customer data** C:\Users\Administrator\Downloads\Customers_Cleaned.xlsx

## 5.SQL Core

1. **Average Delay by Airline**

```
from google.colab import files
uploaded = files.upload()

!pip install duckdb
import duckdb
con = duckdb.connect()

query = """
```

```
SELECT airline,
    AVG(departure_delay) AS avg_dep_delay,
    AVG(arrival_delay) AS avg_arr_delay
FROM 'Flights_Cleaned.xlsx'
GROUP BY airline
ORDER BY avg_dep_delay DESC
"""
con.execute(query).df()
```

**OUTPUT :**

Requirement already satisfied: duckdb in /

| | airline | avg_dep_delay | avg_arr_delay |
|---|---|---|---|
| 0 | SpiceJet | 75.700000 | 65.200000 |
| 1 | Vistara | 69.000000 | 81.333333 |
| 2 | IndiGo | 65.166667 | 70.833333 |
| 3 | Air India | 57.375000 | 82.750000 |
| 4 | Akasa Air | 52.333333 | 60.777778 |

## 2. Most Delayed Routes

```
import duckdb
con = duckdb.connect() # Ensure the connection is established if not already

query_sql = """
SELECT origin, destination,
    AVG(departure_delay) AS avg_delay
FROM 'Flights_Cleaned.xlsx'
GROUP BY origin, destination
HAVING AVG(departure_delay) > 30;
"""

# Execute the SQL query using duckdb and display the result
```

```
result_df = con.execute(query_sql).df()
print(result_df)
```

**OUTPUT :**

```
    origin destination   avg_delay
0      MAA         BLR   96.000000
1      BLR         BOM   58.000000
2      HYD         MAA   66.333333
3      MAA         HYD   72.000000
4      DEL         BOM   54.000000
5      BOM         HYD   75.000000
6      HYD         DEL   99.000000
7      BOM         DEL   56.000000
8      DEL         MAA   67.250000
9      HYD         BOM   79.666667
10     BLR         HYD  116.000000
11     DEL         BLR   71.500000
12     MAA         DEL   58.000000
13     MAA         BOM   51.000000
14     DEL         HYD   60.666667
```

## 3.Satisfaction vs Delay

```
import duckdb
con = duckdb.connect()

query_sql_join = """
SELECT f.airline,
     AVG(c.satisfaction_score) AS avg_satisfaction
FROM 'Flights_Cleaned.xlsx' AS f
JOIN 'Customers_Cleaned.xlsx' AS c ON f.flight_id = c.flight_id
GROUP BY f.airline;
"""
result_satisfaction_df = con.execute(query_sql_join).df()
print(result_satisfaction_df)
```

**OUTPUT:**

```
      airline  avg_satisfaction
0   SpiceJet          2.700000
1     IndiGo          3.000000
2  Akasa Air          3.555556
3  Air India          3.375000
4    Vistara          2.333333
```

### 4. Peak Delay Days

```
import duckdb
con = duckdb.connect()

query_sql_day_of_week = """
SELECT day_of_week,
     AVG(departure_delay) AS avg_delay
FROM 'Flights_Cleaned.xlsx'
GROUP BY day_of_week
ORDER BY avg_delay DESC;
"""

result_day_of_week_df = con.execute(query_sql_day_of_week).df()
print(result_day_of_week_df)
```

**OUTPUT:**

```
   day_of_week   avg_delay
0     Thursday  104.000000
1      Tuesday   73.000000
2    Wednesday   69.875000
3     Saturday   65.454545
4       Sunday   63.666667
5       Friday   61.333333
6       Monday   45.125000
```

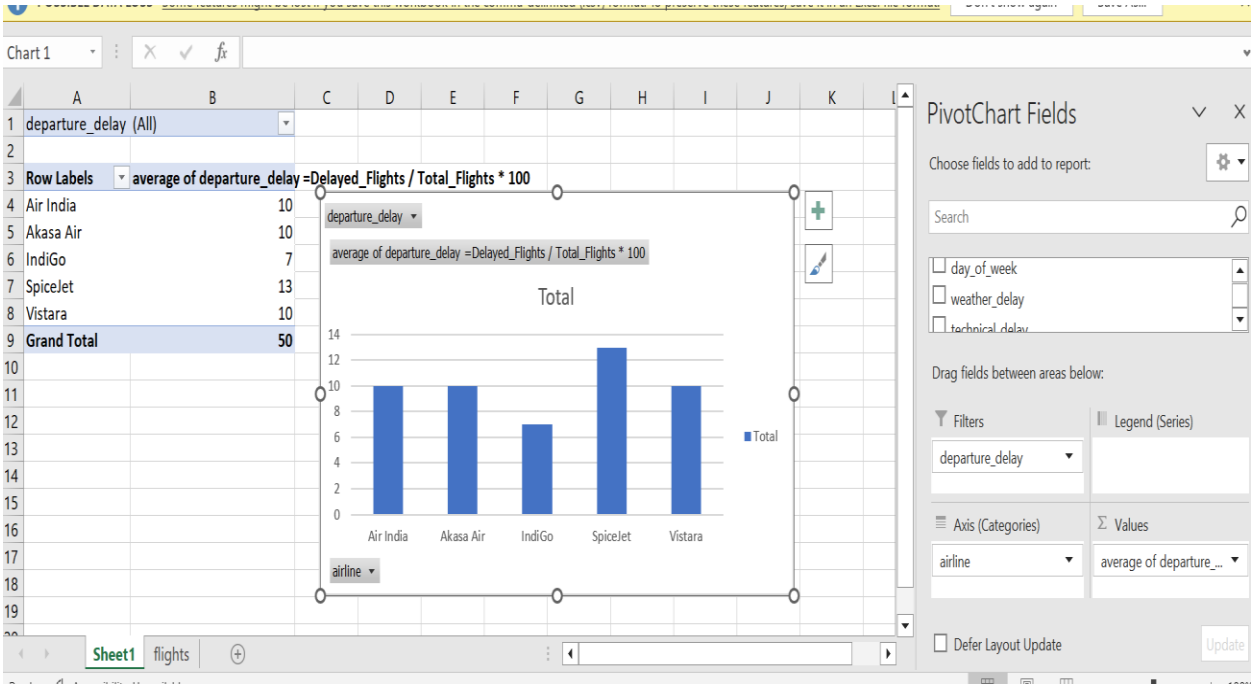
## 6.Excel Analysis & Dashboard

### Excel Tasks

- Clean missing delay values
- Create **Pivot Tables**
- Calculate:
    - % delayed flights
    - Avg satisfaction per airline
    - Delay by day & route

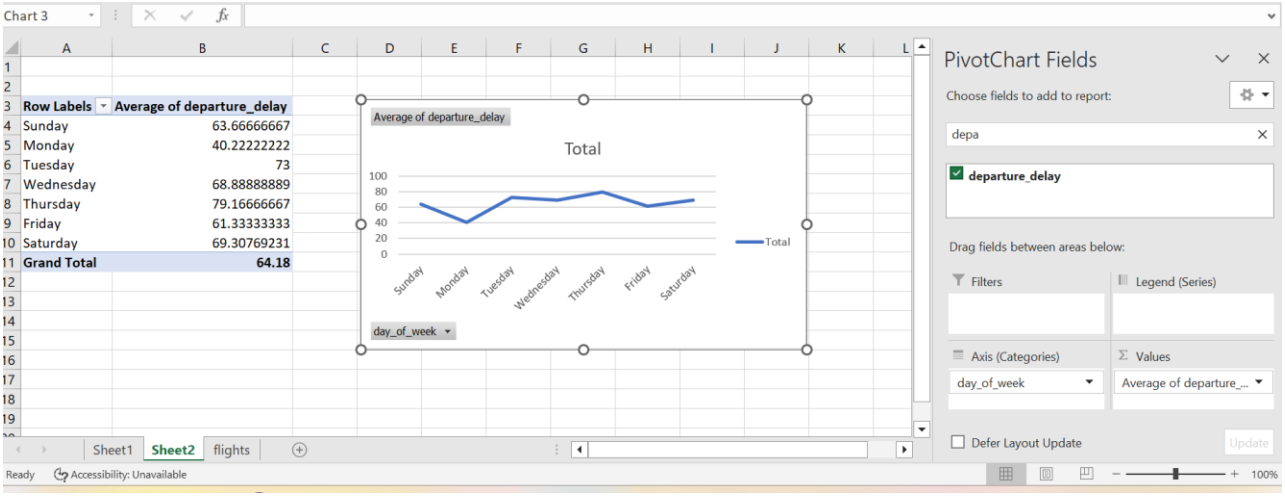
## Excel Dashboard Includes

### 1.Bar chart: Avg delay by airline




"This bar chart compares the average departure delay across airlines, helping identify which airlines experience higher operational delays."

### 2.Average Flight Delay by Day



"This line chart shows the average flight departure delay by day of the week. Monday has the lowest average delay, while Thursday experiences the highest delays. This helps identify peak delay days and supports better flight scheduling and resource planning."

## 3.Route vs delay

| Average of departure_de | Column Labels | | | | | |
|---|---|---|---|---|---|---|
| **Row Labels** | **BLR** | **BOM** | **DEL** | **HYD** | **MAA** | **Grand Total** |
| BLR | 62 | 290 | 28 | 232 | | 612 |
| BOM | | | 112 | 225 | 2 | 339 |
| DEL | 143 | 108 | | 182 | 269 | 702 |
| HYD | 51 | 239 | 198 | 244 | 199 | 931 |
| MAA | 96 | 51 | 58 | 216 | 204 | 625 |
| **Grand Total** | **352** | **688** | **396** | **1099** | **674** | **3209** |

- Routes involving **HYD and DEL** show **higher delays**, especially:
  - **HYD → DEL**
  - **DEL → HYD**
- Routes like **BLR → DEL** and **MAA → BOM** have **lower delays**, indicating better performance.
- **Grand Total row/column** shows overall delay contribution, with some airports contributing more to delays.

# 6.Python Analysis & Visualization

## Load Data

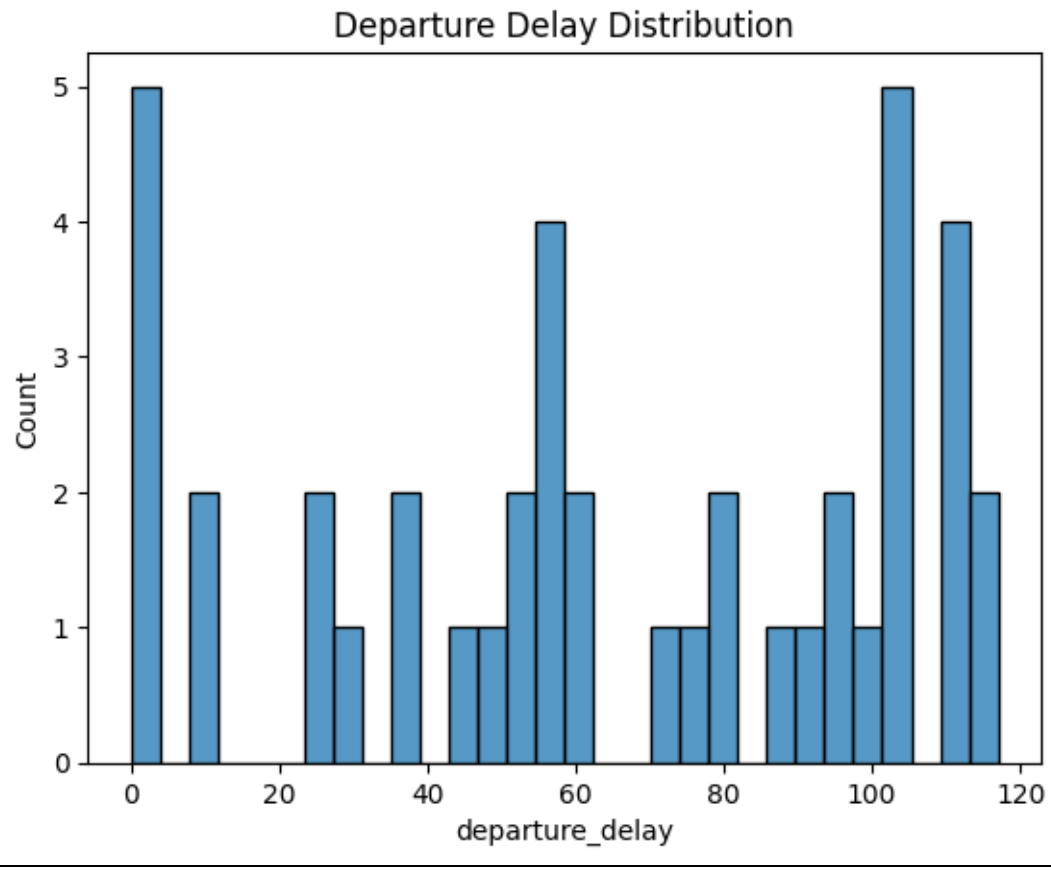
```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
flights = pd.read_csv("flights.csv")
customers = pd.read_csv("customers.csv")

df = flights.merge(customers, on="flight_id")
```

```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import io

# Corrected to read the uploaded Excel files
flights = pd.read_excel(io.BytesIO(uploaded['Flights_Cleaned.xlsx']))
customers = pd.read_excel(io.BytesIO(uploaded['Customers_Cleaned.xlsx']))

df = flights.merge(customers, on="flight_id")
print("DataFrames 'flights' and 'customers' loaded and merged into 'df' successfully.")
```

```
DataFrames 'flights' and 'customers' loaded and merged into 'df' successfully.
```

## Delay Distribution

```
sns.histplot(df['departure_delay'], bins=30)
plt.title("Departure Delay Distribution")
plt.show()
```

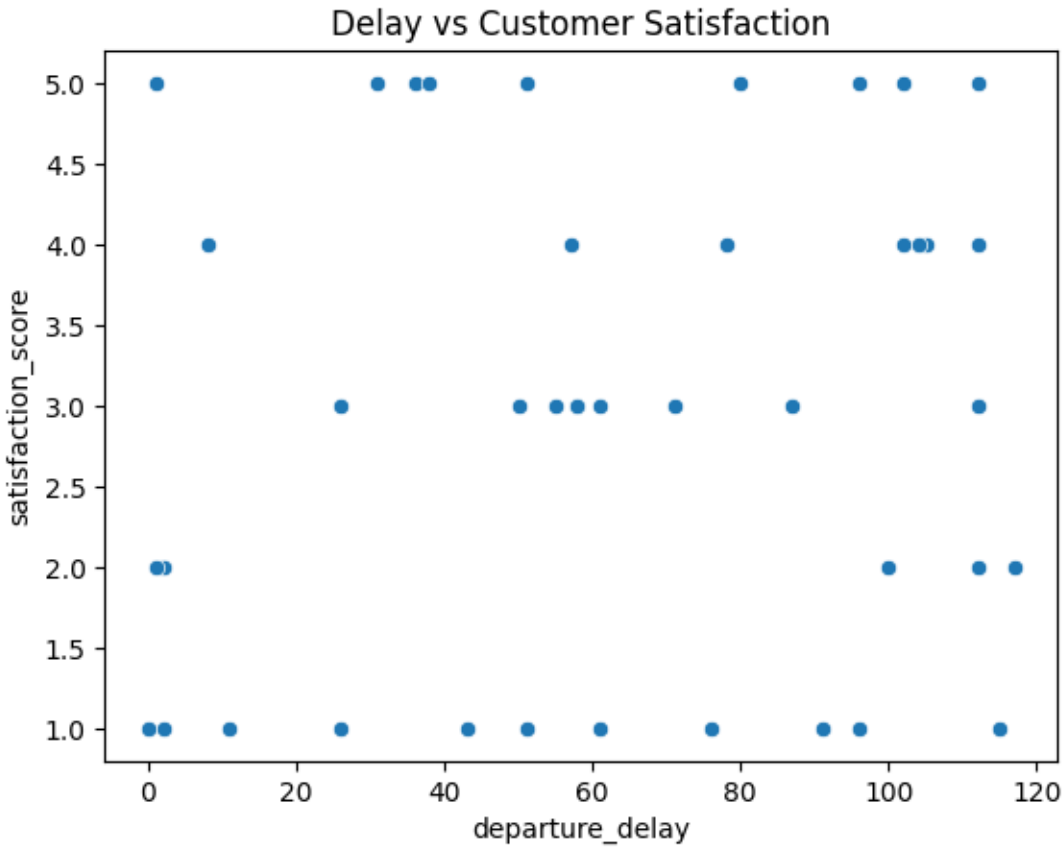


“This chart shows how flight departure delays are distributed, with most delays falling between 40 and 110 minutes.

“The distribution shows that while some flights depart on time or with low delay, a large number experience moderate to high delays, indicating frequent operational issues.”

## Delay vs Satisfaction

```
sns.scatterplot(x='departure_delay', y='satisfaction_score',
data=df)
plt.title("Delay vs Customer Satisfaction")
plt.show()
```

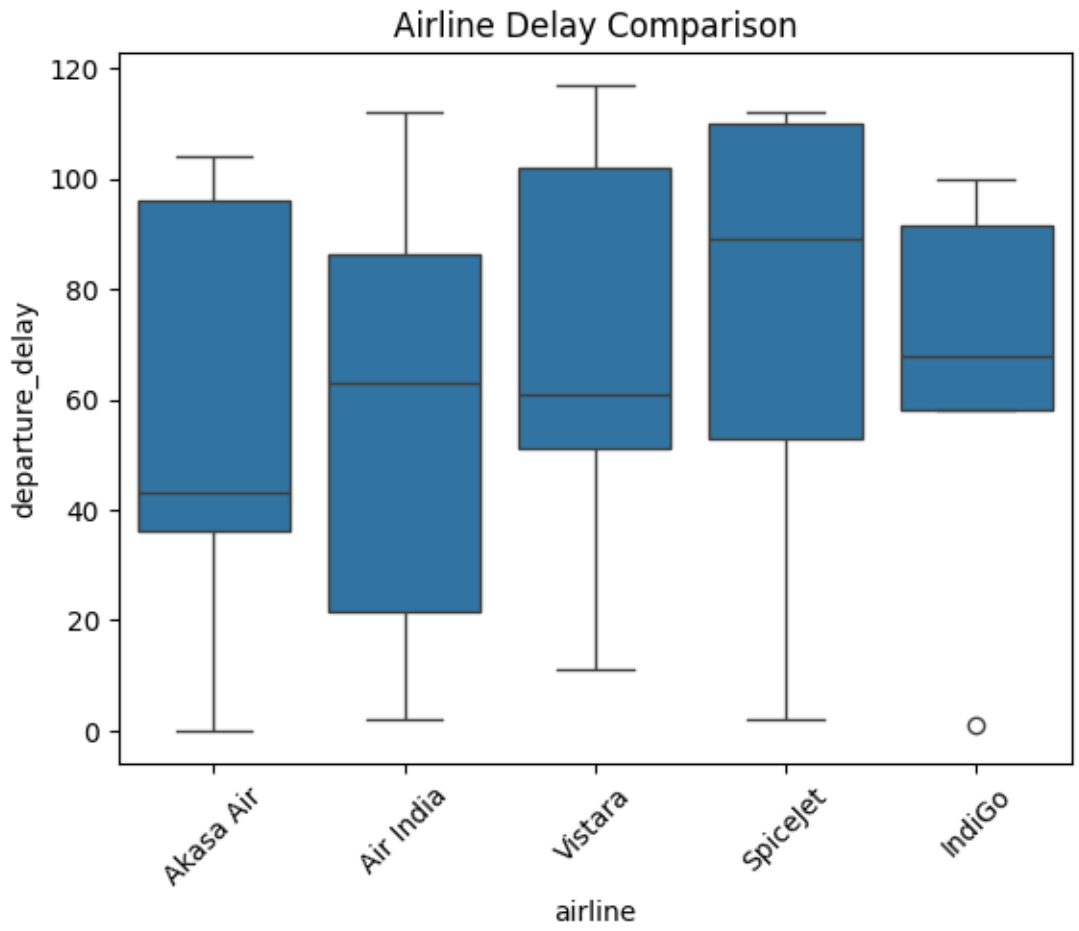



This scatter plot shows the relationship between departure delay and customer satisfaction, where higher delays generally correspond to lower satisfaction scores.”

“The chart indicates a negative relationship between departure delay and customer satisfaction, meaning longer delays tend to reduce passenger satisfaction.”

## Airline Performance Comparison

```
sns.boxplot(x='airline', y='departure_delay', data=df)
plt.xticks(rotation=45)
plt.title("Airline Delay Comparison")
plt.show()
```




“This box plot compares departure delay distributions across airlines, showing that some airlines (like SpiceJet and Vistara) have higher median delays and greater variability, while others (like IndiGo) show more consistent performance with fewer extreme delays.”

“The chart shows the spread, median, and outliers of departure delays by airline, helping identify which airlines experience higher and more inconsistent delays.”

## 7.Key Insights

- Flights delayed **>30 minutes** have satisfaction scores **below 3**
- Certain airlines consistently perform worse during **weekends**
- Weather-related delays have a stronger impact than technical delays
- **Flight delays are common and often significant**
  Most flights experience moderate to high departure delays, indicating operational inefficiencies rather than occasional issues.
- **Delays vary noticeably by day of the week**
  Mid-week and weekend flights show higher average delays compared to Mondays, suggesting peak-traffic and resource-constraint effects.
- **Certain routes consistently face higher delays**
  The route-level heatmap highlights specific origin–destination pairs as high-delay routes, indicating location-specific operational bottlenecks.
- **Airline performance differs significantly**
  Some airlines show higher median delays and greater variability, while others demonstrate more consistent on-time performance.
- **Higher delays lead to lower customer satisfaction**
  There is a clear negative relationship between departure delay and customer satisfaction—longer delays reduce passenger satisfaction scores.
- **Delay consistency matters, not just average delay**
  Airlines with more predictable delays (lower variability) tend to deliver a better customer experience than those with frequent extreme delays.
- **Operational delays directly impact customer experience**
  Delays are one of the strongest drivers of poor feedback, making them a critical focus area for improving satisfaction.

## 8.Business Recommendations

- Add buffer time for high-risk routes
- Improve communication during delays
- Prioritize preventive maintenance
- Offer compensation for delays >45 minutes

## 9.Conclusion

This project analyzed airline flight delays and customer satisfaction using SQL, Excel, and Python. The analysis revealed that flight delays vary significantly by airline, route, and day of the week, with certain routes and mid-week or weekend flights experiencing higher delays. A strong negative relationship was observed between departure delays and customer satisfaction, highlighting the need for airlines to focus on reducing delays to improve passenger experience and overall service quality.
Flight delays differ by airline, route, and day, and higher delays lead to lower customer satisfaction. Targeted operational improvements on high-delay routes can significantly enhance passenger experience.


**By: Bhoomika C**