Airline Flight Dataset Analysis – Internship Task 1

1. Introduction

This project was carried out as part of the Cloud Tech Internship program. The primary objective was to perform exploratory data analysis (EDA) on an airline flight dataset to uncover trends, generate insights, and prepare visualizations for business decision-making. The dataset contained 300,153 rows and 11 columns (12 if the index column is included).

The dataset contains the following columns:

- Index – Unique row number
- Airline – Flight carrier
- Flight – Flight code
- Source_City – City of departure
- Departure_Time – Time of departure (Morning/Afternoon/Evening/Night)
- Stops – Number of stops (Non-stop, 1-stop, 2-stops, etc.)
- Arrival_Time – Time of arrival
- Destination_City – City of arrival
- Class – Ticket class (Economy, Business)
- Duration – Flight duration (in hours)
- Days_Left – Days left before departure when the ticket was booked
- Price – Ticket price (in USD or NGN)

2. Methodology (Steps Taken)

- Imported required Python libraries (Pandas, NumPy, Matplotlib, Seaborn, etc.).
- Loaded the dataset and inspected its dimensions, column names, and data types.
- Performed data cleaning: handled missing values, duplicates, and ensured correct data types.
- Conducted exploratory data analysis (EDA): explored distributions of booking windows, ticket prices, and routes.
- Aggregated data by airline, booking window, and route for comparative insights.
- Applied descriptive statistics (mean, sum, percentages) to identify performance indicators.
- Created visualizations (bar charts, histograms, scatter plots, etc.) to support findings.
- Derived KPIs to assess airline performance and customer booking behaviors.

3. Insights Generated

- Early booking (0–3 days) is associated with higher average prices compared to later booking windows.

- Airline performance varies significantly across routes, with some routes contributing disproportionately to revenue.
- Customer booking behavior shows clear peaks around short booking windows (last-minute bookings).
- Correlation observed between delay frequency and certain routes, suggesting scheduling inefficiencies.
- Revenue optimization can be achieved by adjusting dynamic pricing strategies across booking windows.

4. Conclusion

The analysis revealed distinct booking patterns, price differences, and performance variations across airlines and routes. The findings suggest opportunities for revenue optimization through dynamic pricing, better scheduling, and targeted promotions. Future work may include predictive modeling to forecast demand.

5. Appendices

Key code snippets and visualizations are provided in the accompanying Jupyter Notebook and Tableau snapshots

Expanded Details

Detailed Methodology

1. Data Import & Inspection:

 - Used Pandas to load CSV file into a DataFrame.

 - Inspected dataset shape (300,153 rows × 11 columns).

 - Reviewed column names, data types, and first few records.

2. Data Cleaning:

 - Removed duplicate records.

 - Checked and handled missing values.

 - Converted data types for numerical analysis (e.g., price to float, dates to datetime).

3. Exploratory Data Analysis (EDA):

 - Distribution analysis of booking windows and ticket prices.

 - Route-level analysis to identify high-traffic and low-performing routes.

 - Airline-level aggregation to compare KPIs across competitors.

4. Visualizations:

 - Booking window vs. average ticket price (line/bar charts).

 - Route performance (bar plots by revenue contribution).

 - Delay frequency vs. routes (heatmaps/scatter plots).

 - Passenger booking trends (histograms).

5. KPI Calculation:

 - Average Revenue per Booking Window.

 - Route Revenue Contribution (% of total revenue).

 - Airline Market Share (based on bookings).

 - Customer Booking Behavior Trends.

Expanded Insights

-Booking Window Behavior: Last-minute bookings (0–3 days) showed significantly higher average ticket prices, indicating an opportunity for airlines to maximize revenue from urgent travelers.

-Revenue Distribution: A small set of routes accounted for a disproportionate share of revenue, suggesting that airlines should prioritize service quality and marketing on these profitable routes.

-Delay Patterns: Specific routes experienced recurrent delays, which can negatively impact customer satisfaction. This insight highlights the need for better operational efficiency.

-Airline Comparison: While some airlines dominated in volume of passengers, others achieved higher revenue per passenger, pointing to different pricing strategies and customer bases.

-Seasonality & Demand Peaks: Booking volumes fluctuated with time, suggesting potential for predictive forecasting and seasonal promotions.

Recommendations

1. Implement dynamic pricing models to optimize revenue across booking windows.
2. Enhance customer experience on high-revenue routes to maintain competitive advantage.
3. Address operational inefficiencies on routes with recurrent delays.
4. Introduce targeted promotions for routes with lower occupancy.
5. Develop predictive models for seasonal demand to support strategic planning.