Assignment 2

Ehsan Keikhavani (105665705)

Analysis and Visualization of Flight Delays at Atlanta Airport

Introduction

Hartsfield-Jackson Atlanta International Airport (ATL) serves as one of the busiest airports globally, handling a vast number of passengers and flights daily. Despite its efficiency, flight delays remain a significant challenge that impacts airline operations, passenger satisfaction, and airport management. The goal of this analysis is to explore the underlying causes of flight delays, provide actionable insights, and demonstrate the steps to create a visualization dashboard that can aid decision-makers in improving operational efficiency.

Methodology

Data Preparation

The dataset for this analysis includes flight information from Atlanta Airport, featuring attributes such as flight schedules, delay times, and reasons for delays. Extract connection been used to data reduces the load on the source database and faster processing, this helps improve performance and minimizes strain on the database, especially when dealing with large datasets (and other benefits such as Offline access etc.).

Step 1: Data Cleaning

1. **Handling Missing Values:** Columns with missing or null values, such as delay reasons (AIR_SYSTEM_DELAY, SECURITY_DELAY, etc.), were reviewed. Missing values were replaced with zeros if the absence indicated no delay.

2. **Outliers:** Delays over 400 min has been excluded, due to their small counts and heavy weight, for more data accuracy.

Step 2: Aggregation

- Delay reasons were Averaged to assess their contributions to overall delays.
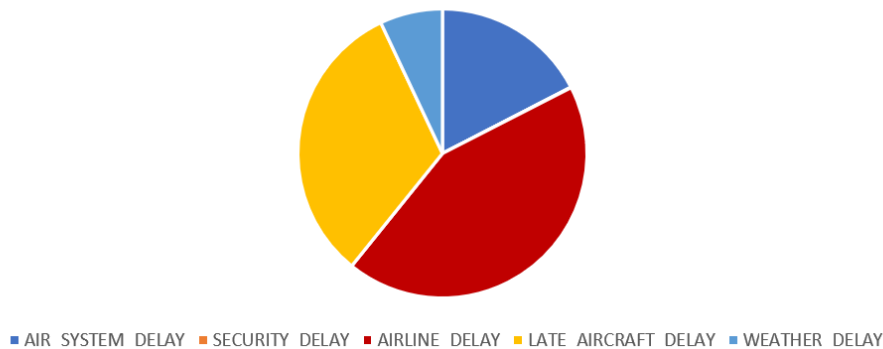
Step 3: Validation

- The cleaned data was validated for consistency, ensuring the absence of anomalies such as negative delay values (those indicating early arrivals), according to the business need which is analysing the delays and their reasons. Decision been made to use only above zero Arrival

Delays. Used SQLite to create a new table for by Arrival Delay > 0 condition, the result dropped from 107947 to 32578 rows.

Exploratory Data Analysis

The exploratory phase aimed to uncover patterns and insights:

1. **Descriptive Statistics:** Calculated the percentage of delayed flights (30.23%), average delay duration (29 min), and the contribution of each delay reason.

■ AIR_SYSTEM_DELAY ■ SECURITY_DELAY ■ AIRLINE_DELAY ■ LATE_AIRCRAFT_DELAY ■ WEATHER_DELAY

2. **Trend Analysis:** Examined delays over time to identify seasonal patterns or anomalies.

3. **Correlation Analysis:** Explored relationships between delay types and overall delays.

AVERAGE of the delays been used instead of the SUM in most cases (not all) because it reduces the impact of outliers, providing a more representative measure of typical delay durations.

Visualization Report:

Step 1: Connecting Data to Tableau

- The processed dataset was saved in a .csv format and imported into Tableau.
- Fields were cleaned using Data Interpreter in Tableau.

Step 2: Creating Individual Visualizations and Designing the Dashboard

1. **KPIs:**
   o Total delayed flights and average delay time, average Airline delays were prominently displayed.

2. **Daily/Monthly Trends of Delays:**

   o A Heatmap allows to observe **trends** and **patterns** across different days and months. For example, you can quickly identify:

   Months or days with the **highest delays** (darker or brighter colours depending on the colour scale). Periods with fewer delays (lighter colours).

   o Filters were added for Day and Month ranges to allow dynamic exploration.

3. **Delay Reasons Breakdown:**

   o A stacked bar chart was used to show the total delay minutes by reason (Air System Delay, Security Delay etc.). Stacked bar been used for better comparison of different data.

   o A colour scheme was applied to differentiate delay types.

4. **Airline and Airport Delay Proportions:**

   o A Map (Geospatial) chart displayed the proportion of Airline and Airport Delays using Colour mark and Arrival Delay for Airports using Size mark vs. on-time flights.

   o Filters were added for Airline and Airport to allow dynamic exploration. Ability to check their values and have an understanding of which Airline or Airport has the most delays.

5. **Forecasting Delays:**

   o Scatter plot been used to show whether certain airlines are more prone to delays and how severe those delays are. By adding and extending the trend line, you can forecast the expected arrival delay for each airline in the future based on historical data and communicate the delay patterns visually to stakeholders.

6. **Flight Delay Proportions:**

   o A pie chart using SUM of Arrival Delay, displayed to provide a concise overview of the reasons behind the delays.

7. **Stacked Airline/Type Bar Chart:**

   o This visualization effectively compares delay types across airlines, providing a clear breakdown of the contributions of each delay type.
   Highlights which airlines or delay types contribute most to total delays, allowing for targeted improvements.

The **layout** flows naturally from high-level overviews (bar chart breakdowns) to more detailed analysis (scatter chart forecasting), provides a logical progression for deeper analysis, starting with comparisons and moving to actionable forecasts.

KPIs are positioned for immediate visibility, reflecting the most critical metrics at a glance. Allows stakeholders to quickly assess performance without searching through detailed visuals.

Interactive elements, such as filters for date, airline and airport, were added allow users to drill down into specific airlines, delay types, or time periods.

Hoover Tooltips provide additional context, such as exact delay numbers, without cluttering the visualization, enables data exploration without overwhelming the user.

Accessibility considerations include clear, high-contrast visuals, intuitive layout for easy navigation, and interactive elements like tooltips and filters for enhanced usability by diverse audiences.

Choice of Metrics:

1. **Average Delay Time**:

   o Focuses on typical delay durations to reduce the impact of outliers and provide a more representative measure.

2. **Delay Causes by Minutes**:

   o Quantifies the total delay time attributed to each cause (e.g., airline, weather, security) to identify key contributors.

3. **KPI Metrics**:

   o Total delayed flights, average delay time, and total delay minutes by cause are prominently displayed to provide a quick snapshot of performance.

Forecasting Model:

   o A scatter plot with a trend line is used in forecasting models to visualize the relationship between variables, identify patterns, and predict future outcomes based on historical trends.

Intended Audience

The dashboard is tailored specifically for Atlanta Airport's **Customer Service** and **Customer Success Teams**:

1. **Customer Service Team**:

   Why used Departure Delays? Departure delays reflect service efficiency at the origin, helping identify bottlenecks in airport operations or turnaround times.

Why used Late Aircraft Delays? These delays often trigger a chain reaction throughout the schedule, highlighting areas where service teams can improve aircraft readiness or scheduling to reduce downstream impacts.

Reason: Departure delays are operational challenges that directly impact the service team's efficiency in managing flights and schedules. Late aircraft delays (caused by previous flights) are a primary cause of departure delays, so analysing both together is essential.

Forecasting Goal: To help the service team improve operational planning, reduce delays at the source, and optimize scheduling to prevent compounding issues.

2. **Customer Success Team**:

- **Why used Airlines?** Different airlines have varying performances. Analysing arrival delays by airline helps forecast which airlines are likely to provide better or worse experiences for passengers.

- **Reason:** Arrival delays directly affect the customer's experience at the destination. They measure how well airlines meet their schedules, which is critical for customer satisfaction.

- **Forecasting Goal:** To predict and improve customer satisfaction by identifying patterns in arrival delays and helping mitigate them for airlines with recurring issues.

Step 3: Customization and Formatting

- Titles, labels, and tooltips were customized to ensure clarity.

- A consistent colour palette and font style were applied for a professional appearance.

- Calculated fields were created using the Analysis tab to enhance data clarity and usability. For example:

  " Day Name": Converts numerical representations of days into their corresponding names (e.g., 1 to "Monday") for better readability.

  "Total Delay": Aggregates various delay metrics into a single KPI, providing a concise summary for key performance indicators. These fields simplify the data and improve user understanding for more effective analysis.

- The dashboard is designed with a layout that flows logically from left to right and top to bottom, prioritizing the presentation of critical data in a clear and accessible manner.

Findings and Insights

1. **Delay Patterns**

   o Approximately 30% of flights experienced delays, with an average delay duration of 29 minutes.

   o Delays were more common during peak travel seasons, indicating seasonal variation.

2. **Primary Causes**

   o Airline-related issues accounted for the highest delay minutes, followed by late aircraft connections.

   o Weather delays, though less frequent, caused significant disruptions during specific periods.

3. **Forecast**

   o Forecasting customer success: (**0.696**) shows a strong positive dependency, meaning arrival delays are heavily influenced by airline delays.

   o Forecasting customer service: (**0.707**) indicates a significant positive influence of late aircraft delays on departure delays.

   o Future delays showed a consistent trend, suggesting that operational improvements could stabilize the pattern.

Recommendations

1. **Operational Improvements:**

   o Airlines should focus on reducing turnaround times and ensuring better schedule adherence to address airline-related delays.

2. **Proactive Weather Management:**

   o Invest in advanced weather monitoring systems and flexible scheduling during adverse conditions.

3. **Resource Optimization:**

   o Enhance resource allocation during peak travel periods to reduce air system delays.

4. **Passenger Communication:**

   o Improve communication regarding delays, especially during seasonal peaks, to maintain passenger satisfaction.

My Ethical Stance as an Analyst

- **Ensure Transparency:** Data has presented truthfully without selectively omitting information that could mislead stakeholders.

- **Maintain Integrity:** Avoided biases that serve specific interests at the expense of accuracy or ethical standards.

- **Promote Responsible Decision-Making:** Provided a detailed breakdown of all delay causes, including airline-related delays, in reports intended for internal decision-makers. This ensures that internal stakeholders have a complete understanding of the factors affecting delays, enabling

  Selective reporting of delay causes, such as omitting airline-related delays, can lead to biased conclusions. The UK Government's Data Ethics Framework (2020) stresses the importance of transparency in data reporting to build trust and support informed decision-making. (UK Government. (2020). Data Ethics Framework)

- **Check for PII and PHI:** The dataset has been reviewed for sensitive information and must be handled with strict care to ensure compliance with privacy regulations such as GDPR.

While business needs often focus on preserving reputation and profitability, these objectives must be pursued without compromising ethical values. Misrepresenting data could harm the company's credibility and lead to misguided strategies.

Conclusion

The analysis and visualization of flight delays at Atlanta Airport highlight the complexity of managing a high-volume aviation hub. By leveraging Tableau to present key insights, decision-makers can identify actionable areas for improvement and mitigate delays effectively. The procedures detailed in this report serve as a robust framework for similar analyses at other airports or transportation hubs.

By combining rigorous data preprocessing, exploratory analysis and interactive visualizations using Tableau, the report provides a comprehensive view of the challenges and opportunities in optimizing airport operations. This approach ensures informed decision-making, improving both efficiency and customer experience.

References

- Airline Delay Statistics: Bureau of Transportation Statistics

- Tableau Documentation: Tableau Help

- UK Government. (2020). Data Ethics Framework. Retrieved from https://www.gov.uk/government/publications/data-ethics-framework