**Flight Delays and Passenger Satisfaction: Uncovering Key Drivers of Airline Experience**

Overview:

The project aimed to analyze the factors influencing passenger satisfaction in the airline industry, particularly focusing on ho flight disruptions (e.g., delays, cancellations) impact customer perceptions. By understanding these relationships, airlines can improve operational processes and enhance customer experiences.

Problem Statement:

Airline passenger satisfaction is a critical measure for any airline's success. Delays, cancellations, and poor service can negatively impact a customer's experience, leading to lower loyalty and satisfaction. The project’s key objective was to explore:

1. How do flight disruptions (like departure and arrival delays) affect passenger satisfaction?

2. What operational and service-related factors most strongly correlate with passenger satisfaction?

This study was intended to help airlines make data-driven improvements to service quality and operational efficiency.

Requirement Scoping:

To address the problem, the following requirements were identified:

1. Target Variable: Passenger Satisfaction (categorical or scaled metric).

2. Features of Interest:

- Flight-related metrics: Departure Delay, Arrival Delay, Flight Distance.

- Service-related metrics: In-flight Wi-Fi Service, Food and Drink, Entertainment, Baggage Handling.

- Demographic Data: Age, Gender, Class, Customer Type (First-time or Returning), Travel Type (Business or Personal).

- Booking-related metrics: Ease of Online Booking, Check-in Service, Boarding Service.

3. Analysis Goals:

- Identify the top factors influencing passenger satisfaction.

- Quantify the impact of flight disruptions on customer satisfaction.

4. Methods: Feature selection using statistical methods like Mutual Information and Chi-Squared Test to determine which factors have the most significant impact on satisfaction.

Data Collection using Data Scraping:

For this project, publicly available airline satisfaction datasets were sourced. The process involved:

1. Data Sources: Data from platforms like Kaggle or public airline datasets was collected. These datasets often include passenger feedback, demographic details, and operational metrics.

2. Data Scraping Tools:

- Web scraping tools such as BeautifulSoup or Scrapy (in Python) could be used to gather data if needed.

- APIs (if available) from airlines or travel platforms to get real-time operational data, e.g., delays or flight cancellations.

3. Challenges: Data scraping from airline websites or third-party review sites might require handling CAPTCHA or dynamically loaded content, which can complicate the process. Ethical concerns like terms of service for data use were also considered.

---

Data Cleaning and Transformation:

Once the data was collected, extensive cleaning and preprocessing were required to ensure its quality for analysis:

1. Handling Missing Data:

- Imputation techniques (mean, median, or mode) were applied for missing numerical values.

- For categorical values, missing entries were filled using the most frequent value or dropped based on relevance.

2. Encoding Categorical Variables:

- Categorical variables such as Class, Customer Type, and Travel Type were encoded using methods like One-Hot Encoding to convert them into numerical form for analysis.

3. Outlier Detection and Removal:

- Extreme values in metrics like departure delay and arrival delay were detected and addressed (either capped or removed) to prevent them from skewing the analysis.

4. Feature Scaling:

- Continuous variables like Flight Distanceand Age were scaled (e.g., using StandardScaler or MinMaxScaler) to bring them into comparable ranges for model training and analysis.

5. Feature Engineering:

- Additional features such as delay ratios (e.g., proportion of flights delayed beyond a threshold) or satisfaction categories(binary classification: satisfied vs. unsatisfied) were created to enhance model performance.

6. Data Validation:

- Ensured that the data after transformation adhered to the correct data types, ranges, and distributions for modeling.

Analysis Approach:

After the data was cleaned, the feature selection process began:

1. Mutual Information:

- This method measured the dependency between the target variable (satisfaction) and other features. It highlighted the most relevant features (e.g., in-flight entertainment, delays) by quantifying their predictive power.

2. Chi-Squared Test:

- For categorical features like class, customer type, and travel type, a Chi-Squared test was applied to assess the independence of satisfaction from these factors.

From this analysis, you identified the top 5 features impacting passenger satisfaction, such as:

- Departure Delay.

- Class (Business vs. Economy).

- In-flight Wi-Fi Service.

- Baggage Handling.

- Customer Type(First-time vs. Returning).

Conclusion:

The analysis revealed key insights into the factors driving passenger satisfaction:

- Flight delays had a significant negative impact on satisfaction, particularly for business travelers who prioritize punctuality.

- Passengers in premium classes (Business, First Class) reported higher satisfaction levels, especially when in-flight services like Wi-Fi and entertainment were rated well.

- Service quality, particularly check-in service and baggage handling, were crucial in shaping customer satisfaction, underscoring the importance of smooth ground operations.

Recommendations:

- Improving On-Time Performance: Airlines should focus on minimizing delays and providing real-time updates to passengers to enhance satisfaction.

- Upgrading In-flight Services: Enhancing Wi-Fi and entertainment systems can significantly boost passenger experience, especially on long-haul flights.

- Ground Service Enhancements: Streamlining check-in, boarding, and baggage handling can further improve customer satisfaction, especially for economy passengers.