**Part 4: Project Recap & Lessons Learned**

**Topic: Airlines Dataset Analysis**

**Introduction and Initial Hypothesis:**

We initiated our project hypothesizing that passenger age and gender significantly influence flight punctuality. We aimed to analyze a Kaggle Airlines dataset to uncover patterns and correlations that would either support or contradict this premise. Our focus was on the impact of these factors on market trends and environmental effects in the airline industry. (Kaggle 2023)

**Data Acquisition:**

We acquired our data from the Kaggle repository to secure a robust dataset capable of providing the necessary detail for analyzing the impact of gender and age on flight punctuality.

**Data Preparation:**

During data preparation, we cleaned the dataset, encoded binary attributes for diagnoses, and addressed missing values to maintain high data quality for precise analysis.

**Information Modeling:**

We adopted a two-pronged modeling approach: first, to determine the underlying relationship between passenger demographics and flight punctuality, and second, to leverage this understanding to predict potential impacts on the airline industry.

**Visualizations:**

We applied various visualization techniques and got the below visualizations:

We used bar charts to compare categorical data frequencies, such as age, gender.

A graph with a pink and blue bar chart

Description automatically generated (Notebook 2023)

A graph of people with different colored bars

Description automatically generated with medium confidence (Notebook 2023)

A graph with different colored bars

Description automatically generated (Notebook 2023)

We utilized box plots to depict age distribution and identify outliers.

A diagram of a flight status

Description automatically generated

(Notebook 2023)

**Effectiveness and Result of Analysis:**

The analysis delineated demographic trends within the dataset effectively but yielded inconclusive results regarding the initial hypothesis about flight punctuality.

* The frequency of the variables of flight status is almost equal. In addition, the number of men and women in flight status is almost the same, so it can be said that gender is not effective in-flight status.
* The airline's primary customers are working adults, it can be said that age is effective in-flight punctuality status.

**Demonstration of your Hypothesis result:**

Based on the considered visualizations, you interpret each one to demonstrate or refute your hypothesis—that passenger age and gender significantly influence flight punctuality—as follows:

**Travelers by Gender:**

The bar chart reveals a slight predominance of female passengers over male passengers. Although this chart alone doesn't directly correlate with flight punctuality, it provides a demographic baseline for further analysis. (Notebook 2023)

**Travelers by Age:**

The second bar chart categorizes passenger counts by age groups. It shows that the majority of passengers belong to the 36-60 age group, with the 19-35 age group following. This indicates that the airline's primary customers are working adults, a factor that could be pertinent to punctuality if different age groups exhibit distinct time sensitivities or travel behaviors. (Notebook 2023)

**Flight Punctuality by Gender:**

This bar chart directly compares the counts of on-time, delayed, and canceled flights between genders. Observing a significant punctuality difference between male and female passengers would lend support to the hypothesis. However, since the chart reveals no substantial difference in delay or cancellation rates between genders, it suggests that gender may not significantly impact punctuality, which does not support the hypothesis. (Notebook 2023)

**Flight Punctuality by Age:**

The box plot details the age distribution for on-time, delayed, and canceled flights. This visualization can identify if certain age groups tend to experience more punctuality issues. For instance, if delays or cancellations are more prevalent in older age groups, this could suggest a pattern that supports the hypothesis. On the other hand, if the age distribution is similar across all flight statuses, this would imply that age is not a significant factor in flight punctuality, which would support the null hypothesis. (Notebook 2023)

This demonstrates that the hypothesis we considered tends to be null hypothesis through visualizations analysis of results above.

**Challenges faced and lessons learned:**

**Challenges:**

We faced the complexity of high-dimensional data and the curse of dimensionality, which obscured potential correlations.

Our visualization techniques succeeded in depicting clear demographic distributions, crucial for stakeholder understanding.

We did not observe the hypothesized direct correlation between demographics and flight punctuality as clearly as expected.

**Lessons Learned:**

The project highlighted the critical nature of thorough data preparation and the power of visualizations to reveal trends. We also learned about the difficulties of handling high-dimensional data and the necessity of producing clear, interpretable results in data analytics.

**References:**

(Kaggle, Airlines Data 2023) <https://www.kaggle.com/datasets/iamsouravbanerjee/airline-dataset/data>

(Notebook 2023)

Sanisha-- AIT-664-DL3-M4 Project Deliverable.ipnyb