PUBLIC TRANSPORTATION ANALYSIS

Phase 1: problem definition and Design thinking

Documentation:

Problem Background:

Public transportation plays a critical role in many cities, providing a convenient and cost-effective mode of transportation for individuals to move around. However, in recent times, there have been complaints about the inefficiency of the system, leading to delays, over-crowding, and overall poor passenger experiences. Addressing this problem requires a comprehensive understanding of data and trends in public transportation to identify areas that need improvement.

Objectives:

To tackle this problem, our objective is to analyze public transportation data to understand on-time performance, passenger satisfaction, and service efficiency. We aim to provide insights that support transportation improvement initiatives and enhance the public transportation experience.

Data Collection:

We will collect data from various sources, including schedules, real-time updates, and passenger feedback. The data collected will be in the form of a structured database, csv files, and unstructured feedback data. Also, different APIs will be used to collect data on real-time updates and station-wise/bus-wise details.

Visualization Strategy:

The insights identified will be visualized using IBM Cognos, creating informative dashboards and reports. The data visualizations will be highly interactive and easy to understand, providing valuable insights to stakeholders.

Code Integration:

To improve the accuracy and efficiency of the analysis, we will integrate code into various aspects of the analysis, such as data cleaning, transformation, and statistical analysis. We will use python tools to perform various open-source data analytical libraries to interactively explore the data and derive new insights for meaningful storytelling.

Python code for analyzing public transportation

```python
Import pandas as pd

# Read the public transportation data into a DataFrame
Public_transportation_data = pd.read_csv('public_transportation.csv')

# Explore the data
Print(public_transportation_data.head())

# Calculate the total number of trips
Total_trips = public_transportation_data['TripID'].nunique()
Print('Total number of trips:', total_trips)

# Calculate the average trip duration
Average_duration = public_transportation_data['Duration'].mean()
Print('Average trip duration:', average_duration, 'minutes')

# Calculate the most popular transport mode
Popular_mode = public_transportation_data['TransportMode'].mode()[0]
Print('Most popular transport mode:', popular_mode)

# Calculate the total revenue earned
Total_revenue = public_transportation_data['Revenue'].sum()
```

```
Print('Total revenue earned:', total_revenue)


# Group the data by transport mode and calculate the average revenue per mode

Average_revenue_per_mode =
public_transportation_data.groupby('TransportMode')['Revenue'].mean()

Print('Average revenue per mode:\n', average_revenue_per_mode)


# Plot a bar chart of revenue per mode

Import matplotlib.pyplot as plt


Average_revenue_per_mode.plot(kind='bar')

Plt.title('Average Revenue per Transport Mode')

Plt.xlabel('Transport Mode')

Plt.ylabel('Average Revenue')

Plt.show()
```

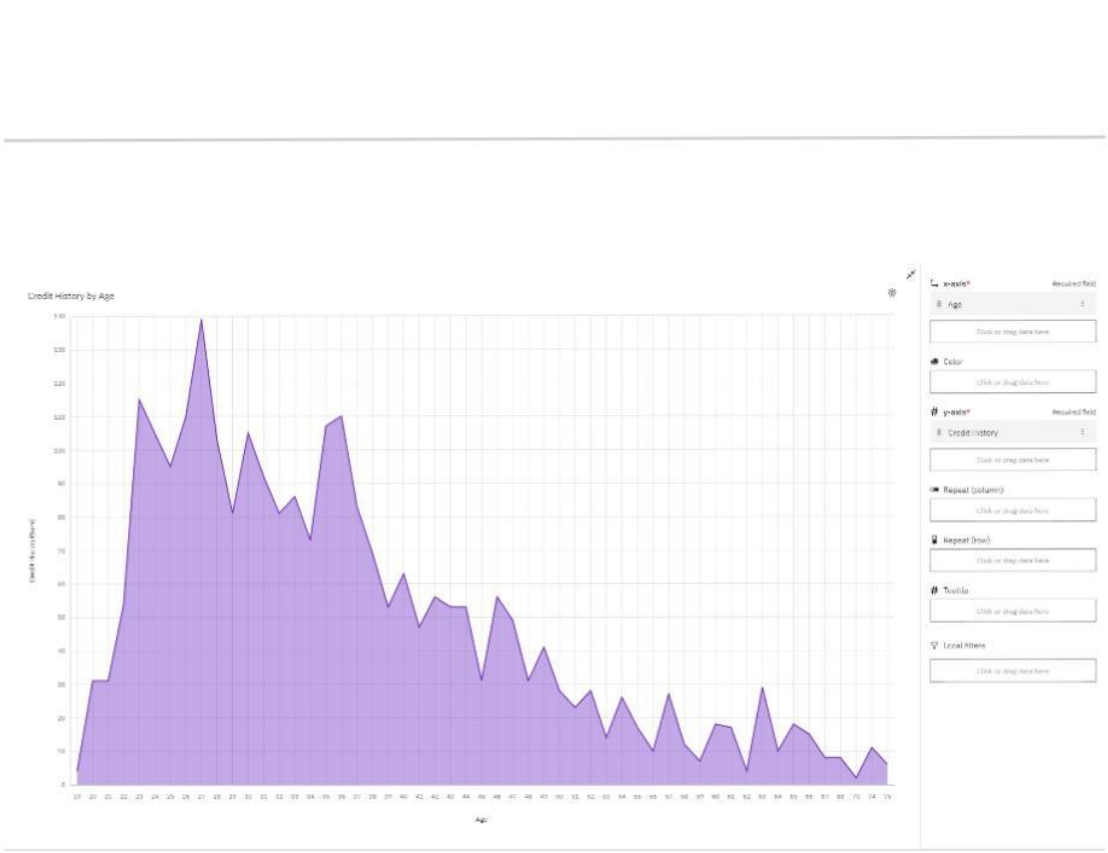
Please note that this code assumes you have a CSV file named `public_transportation.csv` containing the relevant public transportation data. You may need to modify the code accordingly based on your dataset structure.

This code will help you analyze various aspects of the public transportation data, such as the total number of trips, the average trip duration, the most popular transport mode, the total revenue earned, and the average revenue per mode.

Conclusion:

Through this design thinking approach, we aim to create a comprehensive data analysis plan for the project, emphasizing the importance of collecting, analyzing, and visualizing data to enhance public transportation efficiency and the passenger experience.

BAR CHART


Credit History by Age
x-axis*
Required field
Age
Color
y-axis*
Required field
Credit history
Repeat (column)
Repeat (row)
Tooltip
Local filters