**PUBLIC TRANSPORT ANALYSIS**

**OBJECTIVE**

To transform the design and ideas from the Problem Definition and Design Thinking phase into practical solutions for analyzing public transportation data using Jupyter notebooks and JupyterBook.

**Step 1: Data Preprocessing**

* **Data Cleaning**: Begin by cleaning the public transportation dataset using Python libraries like Pandas. Handle missing values, correct inconsistencies, and ensure data quality.
* **Data Integration**: If multiple data sources are involved, integrate them into a unified dataset for analysis.
* **Data Transformation**: Transform the data into a format suitable for analysis, such as aggregating data to calculate performance metrics and efficiency.

**Step 2: Jupyter Environment Setup**

* **Environment Setup**: Ensure that your Jupyter environment is properly configured and ready for data analysis.
* **Library Installation**: Install the necessary Python libraries like Pandas, Matplotlib, Seaborn, and NumPy for data analysis and visualization.

**Step 3: Data Analysis and Visualization**

* **Chart Creation**: Design and create charts and graphs using Matplotlib and Seaborn to visually represent performance metrics and efficiency of public transportation services.
* **Statistical Analysis**: Utilize Python's capabilities for statistical analysis, such as calculating mean values and standard deviations.

**Step 4: Insights Generation**

* **Pattern Identification**: Analyze the visualizations and statistical results to identify patterns, trends, or areas for improvement in public transportation services.
* **Insight Formulation**: Translate these patterns into actionable insights, such as suggesting improvements for specific routes or services.

**Step 5: JupyterBook Creation**

* **Installation**: Install JupyterBook, a tool to create interactive, shareable, and documentation-ready books from Jupyter notebooks.
* **Book Structure**: Define the structure of your JupyterBook, including chapters and sections.
* **Notebook Integration**: Include your Jupyter notebooks, code, visualizations, and insights into the JupyterBook

**JUPYTERBOOK**

**VISUALYISING TOP 10 STOPS BY BOARDING POINTS USING JUPYTERNOTE BOOK**

import pandas as pd

import matplotlib.pyplot as plt

data = pd.read\_excel('/content/sample\_data/Public Transportation Data Set V2.xlsx')

print(data.head())

missing\_values = data.isnull().sum()

duplicates = data.duplicated().sum()

data = data.dropna()

data = data.drop\_duplicates()

plt.figure(figsize=(12, 6))

data.groupby('StopName')['NumberOfBoardings'].sum().sort\_values(ascending=False).head(10).plot(kind='bar', color='skyblue')

plt.title('Top 10 Stops by Total Number of Boardings')

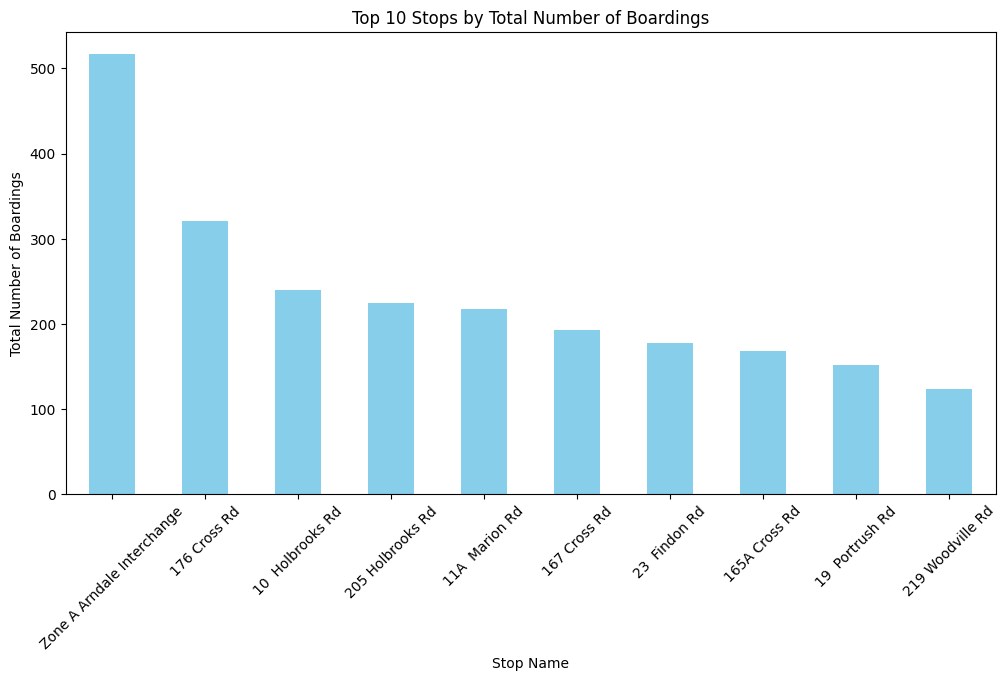
plt.xlabel('Stop Name')

plt.ylabel('Total Number of Boardings')

plt.xticks(rotation=45)

plt.show()

**OUTPUT**

****