**Project Title: Public Transportation Analysis**

**Phase 1: Project Definition and Design Thinking**

**Project Definition:** The project involves analyzing public transportation data to assess service efficiency, on time performance, and passenger feedback. The objective is to provide insights that support transportation improvement initiatives and enhance the overall public transportation experience. This project includes defining analysis objectives, collecting transportation data, designing relevant visualizations in IBM Cognos, and using code for data analysis.

**Design Thinking:**

1. Analysis Objectives: Define specific objectives for analyzing public transportation data, such as assessing on-time performance, passenger satisfaction, and service efficiency.
2. Data Collection: Identify the sources and methods for collecting transportation data, including schedules, real-time updates, and passenger feedback.
3. Visualization Strategy: Plan how to visualize the insights using IBM Cognos to create informative dashboards and reports.
4. Code Integration: Decide which aspects of the analysis can be enhanced using code, such as data cleaning, transformation, and statistical analysis.

**Analysis Objectives:**

* **On-Time Performance:** Measure and improve the punctuality of public transportation by analyzing data to determine the percentage of buses/trains arriving on time and identifying factors contributing to delays.
* **Passenger Satisfaction:** Assess passenger satisfaction by collecting and analyzing feedback through surveys or social media sentiment analysis, focusing on aspects like cleanliness, safety, and overall experience.
* **Service Efficiency:** Evaluate service efficiency by analyzing operational data, such as the cost per passenger mile, frequency of breakdowns, and utilization of resources.

**Data Collection:**

* **On-Time Performance:**

Data Sources: Real-time GPS tracking systems, historical schedule data, traffic data, incident reports.

Data Collection Methods: Integration with GPS APIs, sensors on vehicles, and automated schedule tracking.

* **Passenger Satisfaction:**

Data Sources: Passenger surveys, social media mentions, customer service logs.

Data Collection Methods: Online surveys, sentiment analysis tools, and manual entry of feedback.

* **Service Efficiency:**

Data Sources: Financial records, maintenance logs, fleet management systems.

Data Collection Methods: Data extraction from financial systems, monitoring of maintenance logs, and integration with fleet management software.

**Visualization Strategy (Using IBM Cognos):**

* **On-Time Performance:**

Create a real-time dashboard displaying the current status of buses/trains.

Generate line charts showing historical on-time performance trends.

* **Passenger Satisfaction:**

Develop a sentiment analysis dashboard summarizing positive and negative feedback.

Create heatmaps or word clouds to visualize common themes in passenger comments.

* **Service Efficiency:**

Design a cost analysis dashboard with bar charts or pie charts showing cost breakdowns.

Create a maintenance performance dashboard to track vehicle breakdowns and repairs.

**Code Integration:**

* **Data Cleaning:**

Use Python or R scripts to automate data cleaning processes, removing duplicates, handling missing values, and standardizing data formats.

* **Data Transformation:**

Write code to transform raw data into a format suitable for analysis, e.g., aggregating GPS data by time intervals.

* **Statistical Analysis:**

Utilize statistical libraries in Python or R to perform in-depth analysis, such as regression analysis to identify factors affecting on-time performance.

**Next Step:**

* incorporating machine learning algorithms to predict service disruptions or analyze passenger sentiment from feedback.
* building the public transportation efficiency analysis using IBM Cognos for visualization.
* Continue building the analysis by creating visualizations using IBM Cognos and integrating code for data analysis.
* Document the public transportation efficiency analysis project and prepare it for submission.

**Overview View of Our Final Programming Code in Python**

**#Importing necessary libraries**

import pandas as pd

import matplotlib.pyplot as plt

import networkx as nx

**# Load the public transportation data from a CSV file**

public\_transportation\_df = pd.read\_csv('publictransportationanalysis.csv')

**# Print the first 5 rows of the DataFrame**

print(public\_transportation\_df.head())

**# Print the DataFrame information**

print(public\_transportation\_df.info())

**# Create a network graph of the public transportation system**

public\_transportation\_graph = nx.Graph()

**# Add the transit stops to the graph as nodes**

for stop in public\_transportation\_df['stop\_name']:

public\_transportation\_graph.add\_node(stop)

**# Add the transit routes to the graph as edges**

for route in public\_transportation\_df['route\_id']:

for stop\_1,stop\_2inzip

(public\_transportation\_df['stop\_sequence'][public\_transportation\_df['route\_id']==route].values[0],public\_transportation\_df['stop\_sequence'][public\_transportation\_df['route\_id'] == route].values[1:]):

public\_transportation\_graph.add\_edge(stop\_1, stop\_2)

**# Identify the shortest path between two transit stops**

shortest\_path = nx.shortest\_path(public\_transportation\_graph, 'origin\_stop', 'destination\_stop')

**# Visualize the public transportation network graph**

nx.draw(public\_transportation\_graph, with\_labels=True)

plt.show()

**Output:**

TripID RouteID StopID StopName WeekBeginning \

0 23631 100 14156 181 Cross Rd 30-06-2013 00:00

1 23631 100 14144 177 Cross Rd 30-06-2013 00:00

2 23632 100 14132 175 Cross Rd 30-06-2013 00:00

3 23633 100 12266 Zone A Arndale Interchange 30-06-2013 00:00

4 23633 100 14147 178 Cross Rd 30-06-2013 00:00

NumberOfBoardings

0 1

1 1

2 1

3 2

4 1

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 1048575 entries, 0 to 1048574

Data columns (total 6 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 TripID 1048575 non-null int64

1 RouteID 1048575 non-null object

2 StopID 1048575 non-null int64

3 StopName 1048575 non-null object

4 WeekBeginning 1048575 non-null object

5 NumberOfBoardings 1048575 non-null int64

dtypes: int64(3), object(3)

memory usage: 48.0+ MB

None

**Demonstration:**



