Public Transportation Analysis

2023 Naan Mudhalvan - IBM Data Analytics with Cognos - Group 1 - Project 8
College : NM001 - College of Engineering Guindy
Proj_200340_Team_2
Members: Abinithi R, Abirami S V, Adithya R U, Akshaya G R, Sai Rishi L
Faculty Advisor : Dr. Bama S

PHASE 1

Problem Definition and Design Thinking

Problem Definition :

Analyse public transportation data to assess **service efficiency**, **on time performance**, and **passenger feedback.**

Provide insights that **support transportation improvement initiatives** and enhance the overall public transportation experience.

Design Thinking :

1. **Analysis Objectives**: Define specific objectives for analysing 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. **Visualisation Strategy**: Plan how to visualise 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 :

1. **On-Time Performance :**
   - *Punctuality Rate*, *Average Delay* and *Peak Hour Delays*

2. **Reliability :**
   - *Service Reliability* and *Headway Adherence*

3. **Ridership :**
   - *Ridership Volume, Ridership Growth* , and *Occupancy Rate*

4. **Service Coverage :**
   - *Service Area Coverage* and *Accessibility*

5. **Service Efficiency :**
   - *Average Speed, Dwell Time* and *Vehicle Utilisation*

6. **Customer Satisfaction :**
   - *Customer Surveys* and *Complaints and Feedback Analysis*

7. **Safety :**
   - *Accident Rate* and *Emergency Response Time*

8. **Environmental Impact :**
   - *Emissions Reduction* and *Vehicle Efficiency*

9. **Financial :**
   - *Cost per Rider* and *Revenue Generation*

10. **Infrastructure :**
    - *Maintenance and Repairs* and *Infrastructure Investment*

11.. **Data Visualization :**
    - *Dashboard Usage*

12. **Operational Efficiency :**
    - *Operational Costs* and *Route Optimization*

With the provided dataset, metrics **3**, **4** and **12** can be applied to provide relevant and consistent results.

Dataset Summary :

Obtained from **https://www.kaggle.com/datasets/rednivrug/unisys?select=20140711.CSV** as per the project instructions.

TripID : an **integer** attribute which uniquely identifies trips
RouteID : a **string** attribute which uniquely specifies a transport route
StopID : an **integer** attribute which uniquely identifies stops in a transport network
StopName : a **string** attribute which gives a name to the corresponding StopID
WeekBeginnning : a **date/time** attribute which represents the first day of any given week
NumberOfBoardings : an **integer** attribute which keeps count of all the boardings at that specific stop for that specific trip over the week.

For example, a record in the dataset can be read as :

```
23631, 100, 14156, 181 Cross Rd, 2013-06-30 00:00:00, 1
```

> " For **StopName** *181 Cross Road* (corresponding **StopID** of *14156*) in the week which began at *2013-06-30*, had *1* boarding(s) during the trip identified by **TripID** *23631* which followed the route identified by **RouteID** *100*. "