**Data analytics with Cognos**

**Topic:** Building the public transportation efficiency analysis using IBM Cognos for visualization



**Introduction:**

Public transportation analysis involves evaluating the efficiency, accessibility, and sustainability of urban transit systems to enhance mobility and reduce environmental impacts, aiming to improve overall urban livability.

**Challenges involved in loading a public transportation analysis dataset:**

* **Data Integration:**

Gathering, integrating, and managing data from various sources, including sensors, schedules, and user feedback, to make informed decisions.

* **Route Optimization:**

Balancing route efficiency and coverage to meet diverse commuter needs while minimizing costs and environmental impact.

* **Funding Constraints:**

Securing adequate funding for infrastructure development, maintenance, and service improvements in the face of budget limitations.

* **User Behaviour Understanding:**

Predicting and adapting to changing travel patterns and preferences of commuters.

* **Technological Advances:**

Keeping pace with advancements in transportation technology, such as autonomous vehicles and smart city solutions.

* **Environmental Sustainability:**

Reducing the environmental impact of public transportation, including emissions and energy consumption.

**How to load the dataset:**

**Step 1:** The first step is to identify the dataset that you want to load. This dataset may be stored in a local file, in a database, or in a cloud storage service.

**Step 2**: Once you have identified the dataset, you need to load it into the IBM cognos.

**Step 3:** Once the dataset is loaded into the IBM cognos, you may need to preprocess it before you can start training and evaluating your model. This may involve cleaning the data, transforming the data into a suitable format, and splitting the data into training and test sets.

**Data cleaning:**

This involves identifying and correcting errors and inconsistencies in the data. For example, this may involve removing duplicate records, correcting typos, and filling in missing values.

**Code:**

import pandas as pd

dataset\_path = 'path/to/your/content/publictransportationanalysis.csv'

transportation\_data = pd.read\_csv('publictransportationanalysis.csv')

transportation\_data.drop\_duplicates(inplace=True)

columns\_with\_missing\_values = ['TripID', 'RouteID', 'StopID']

transportation\_data.dropna(subset=columns\_with\_missing\_values, inplace=True)

z\_scores = (transportation\_data['TripID'] - transportation\_data['RouteID'].mean()) / transportation\_data['StopID'].std()

transportation\_data = transportation\_data[(z\_scores < 3)]

transportation\_data['column\_name'] = transportation\_data['column\_name'].astype('new\_data\_type')

transportation\_data['new\_column'] = transportation\_data['column1'] + transportation\_data['column2'

transportation\_data['column1\_normalized'] = (transportation\_data['column1'] - transportation\_data['column1'].min()) / (transportation\_data['column1'].max() - transportation\_data['column1'].min())

Visualization using IBM Cognos:

