**PUBLIC TRANSPORT ANALYSIS**

**PHASE 3 : DEVELOPMENT PART 1**

**Abstract:**

The rapid urbanization in modern cities has led to an increased demand for efficient and sustainable public transport systems. This project, "Public Transport Analysis," aims to conduct a thorough analysis of existing public transport networks to enhance their efficiency, accessibility, and sustainability. The study focuses on employing advanced data analytics and simulation techniques to evaluate the current state of public transport, identify bottlenecks, and propose data-driven strategies for improvement. The project will begin with comprehensive data collection, including passenger usage patterns, traffic flows, peak hours, and geographic data. Advanced analytical tools, such as machine learning algorithms and geographic information systems (GIS), will be applied to process and interpret this data. The research will delve into various aspects, including route optimization, scheduling, fleet management, and fare structures.

Code:

import pandas as pd

import numpy as np

import seaborn as sns

import matplotlib.pyplot as plt

import plotly as pt

from scipy import stats

import sklearn

from sklearn import preprocessing

from sklearn.impute import SimpleImputer

from math import floor,ceil

import tabulate as tb

from sklearn.preprocessing import StandardScaler

from sklearn.model\_selection import train\_test\_split

from imblearn.over\_sampling import SMOTE

from sklearn.linear\_model import LogisticRegression,RidgeClassifier,SGDClassifier,PassiveAggressiveClassifier

from sklearn.svm import SVC,LinearSVC,NuSVC

from sklearn.neighbors import KNeighborsClassifier

from sklearn.tree import DecisionTreeClassifier

from sklearn.ensemble import RandomForestClassifier,AdaBoostClassifier,GradientBoostingClassifier

from xgboost import XGBClassifier

from sklearn.naive\_bayes import GaussianNB,BernoulliNB

from sklearn.model\_selection import GridSearchCV

from sklearn.metrics import precision\_score,accuracy\_score

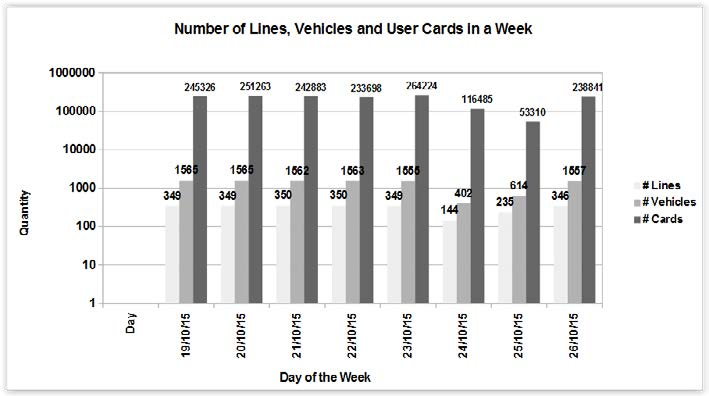
from sklearn.metrics import classification\_report

import warnings

warnings.filterwarnings('ignore')

Data set :

1. **Geographic Data:**
   * **Bus Stops/Stations:** Name, location coordinates (latitude, longitude), capacity.
   * **Routes:** Route number, stops along the route, distance, travel time estimates.
   * **City Layout:** Maps, road networks, and geographic obstacles affecting routes.
2. **Demographic Data:**
   * **Population Data:** Total population, age distribution, income levels.
   * **Employment Data:** Job locations, working hours, employment sectors.
   * **Special Needs Population:** Data on individuals with disabilities, elderly citizens.
3. **Usage Patterns:**
   * **Ridership Data:** Daily/weekly/monthly ridership numbers per route and station.
   * **Peak Hours:** Time of the day with maximum passenger load.
   * **Ticket Sales Data:** Revenue generated, types of tickets (single ride, monthly pass, etc.).
4. **Operational Data:**
   * **Bus/Train Fleet:** Number of vehicles, types of vehicles, fuel type (diesel, electric, hybrid).
   * **Maintenance Records:** Vehicle repair and maintenance schedules, downtime data.
   * **Driver Information:** Number of drivers, shifts, experience levels.
5. **Environmental Data:**
   * **Emission Data:** Greenhouse gas emissions from vehicles.
   * **Fuel Consumption:** Fuel efficiency data for different vehicle types.
   * **Air Quality:** Pollution levels at different times and locations.
6. **Customer Feedback:**
   * **Surveys:** Passenger satisfaction surveys, feedback on specific routes or services.
   * **Complaints Data:** Types of complaints, frequency, resolution times.
7. **Technological Data:**
   * **GPS Data:** Real-time vehicle location data for tracking routes.
   * **Mobile App Data:** Usage statistics for transport-related apps.
   * **Digital Payment Data:** Number of digital transactions, types of payment methods.
8. **External Factors:**
   * **Weather Data:** Impact of weather conditions on ridership and vehicle efficiency.
   * **Events Data:** Impact of local events (sports, concerts) on public transport demand.
9. **Financial Data:**
   * **Budget Allocation:** Funds allocated for public transport projects.
   * **Expenditure:** Operational costs, maintenance expenses, staff salaries.
10. **Regulatory and Policy Data:**
    * **Laws and Regulations:** Pertinent laws governing public transport operations.
    * **Policy Changes:** Historical data on policy changes and their impact on public transport.



In this article I will show the Exploratory Data Analysis (EDA) process with visualization techniques to analyze how many vehicles used in a week.

**Data Content :**

**1.** . **Geopgraphic data :** Geographic data refers to information that is associated with a specific location on the Earth's surface. This data is typically represented using geographical coordinates (latitude and longitude) and can include various attributes related to that location.

**2.** **Demographic data :** Demographics comprise an array of socioeconomic information, including the breakdown of a population by gender, age, ethnicity, income, employment status, home ownership, and even internet access**.**

**3. Usage patterns :** Usage patterns show you when a particular app is being used during the day, highlighting the hours when users are most active. The Usage Patterns tab displays data based on the local time zone of each user and is calculated by the monthly average sessions for each hour of the day.

**4. Operational data :** Operational data is a form of strategic data that captures information on the internal functions and processes of a business. Different types of operational data include business operational data, IT operational data and Integrated business-IT operational data**.**

**5.Customer feedback :** Customer feedback is the information, insights, issues, and input shared by your community about their experiences with your company, product, or services. This feedback guides improvements of the customer experience and can empower positive change in any business — even (and especially) when it's negative.

**6. Technological data** : Data technology sector includes solutions for data management, and products or services that are based on data generated by both human and machines**.**

**Conclusion :**

While the number of vehicles operating in a city increases rapidly, the change in infrastructure is a rather slow process. In order to accommodate the users and their transportation needs, a city must carefully analyze data sources to determine the user needs and possible changes in transportation to support those needs. This paper presented an initial investigation in order to identify scenarios and implications from the public transportation data from the city of Curitiba, integrating several sources, including stationary and dynamic open data. Detected characteristics can be used for further analysis in order to optimize the transportation system. This work will be extended to provide recommendation and optimization for alternative and existing bus routes that minimizes various parameters including travel time and number of transfers. Lastly, we will investigate various sapproaches for integrating other data sources.