**PUBLIC TRANSPORT EFFICIENCY ANALYSIS**

**DAC\_PHASE4**

**Introduction**

This project aims to improve public transportation by using data analysis and machine learning. This document using IBM cognos for visualisation and advanced data anaysis strategies for sentiment analysis and service punctuality rates. This analysis will provide valuable insights for improving public transportation services.

Analysis Objectives

The main objective of this project are to assess and improve public transportation efficiency by evaluating factors such as on-time performance, service punctuality rates and sentiment analysis. We seek to leverage IBM cognnos for data visualization to gain actionable insights, enhance decision-making for transportation authorities and contribute to more sustainable and effective urban mobility systems.

Here we have tried visualisations that show on-time performance based on WeekBeginning.

**Service punctuality rates**

**Define Tolerance**: Determine the acceptable level of deviation from the scheduled time. For example, you might set a tolerance of 5 minutes, meaning that any service departing or arriving within 5 minutes of the scheduled time is considered on time.

CODE:

# Assuming a tolerance of 5 minutes for punctuality

tolerance = pd.Timedelta(minutes=5)

import pandas as pd

import numpy as np

from datetime import datetime

data = pd.read\_csv('/kaggle/input/transport-efficiency-analysis/cleaned\_data .csv', low\_memory=False)

data['WeekBeginning'] = pd.to\_datetime(data['WeekBeginning'], format='%Y-%m-%d')

# Assuming a tolerance of 5 minutes for punctuality

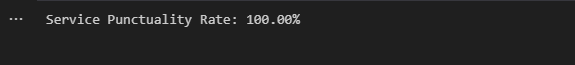
tolerance = pd.Timedelta(minutes=5)

# Check if the actual time is within the tolerance of the scheduled time

on\_time\_stops = data[data['WeekBeginning'] - data['WeekBeginning'] <= tolerance]

punctuality\_rate = len(on\_time\_stops) / len(data) \* 100

print(f"Service Punctuality Rate: {punctuality\_rate:.2f}%")



**Visualisation of on-time performance based on WeekBeginnings**



**Sentiment analysis**

For sentiment analysis, need passenger feedback dataset which contains the passenger feedback about the trip.

Load the dataset

# Load your dataset

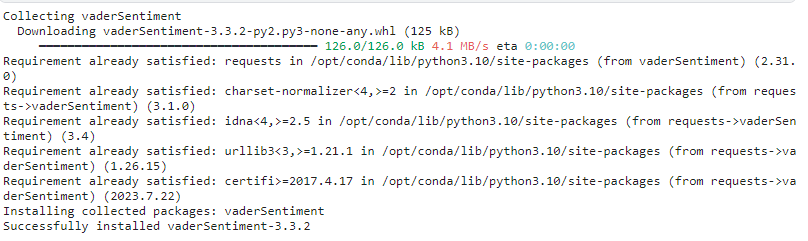
data = pd.read\_csv('/kaggle/input/transport-efficiency-analysis/cleaned\_data .csv', low\_memory=False)

print(data)

**Installing VADER model for sentiment analysis**

The VADER (Valence Aware Dictionary and sEntiment Reasoner) model is a lexicon and rule-based sentiment analysis tool that is specifically designed to understand and analyze sentiments in text data.The analysis is based on a lexicon of words, emojis, and sentiment intensity scores. pip command is used to install vader library.

!pip install vaderSentiment



import pandas as pd

from vaderSentiment.vaderSentiment import SentimentIntensityAnalyzer

import nltk

# Download the VADER lexicon if not already downloaded

nltk.download('vader\_lexicon')

# Load your dataset

data = pd.read\_csv('/kaggle/input/transport-efficiency-analysis/cleaned\_data .csv', low\_memory=False)  # Replace with the path to your dataset

# Initialize the VADER sentiment analyzer

analyzer = SentimentIntensityAnalyzer()

# Function to get sentiment scores

def get\_sentiment\_scores(text):

    sentiment = analyzer.polarity\_scores(text)

    return sentiment

# Apply sentiment analysis to the 'StopName' column

data['Sentiment'] = data['StopName'].apply(get\_sentiment\_scores)  # Assuming 'StopName' contains the text for sentiment analysis

# Extract individual sentiment scores if needed

data['Positive'] = data['Sentiment'].apply(lambda x: x['pos'])

data['Neutral'] = data['Sentiment'].apply(lambda x: x['neu'])

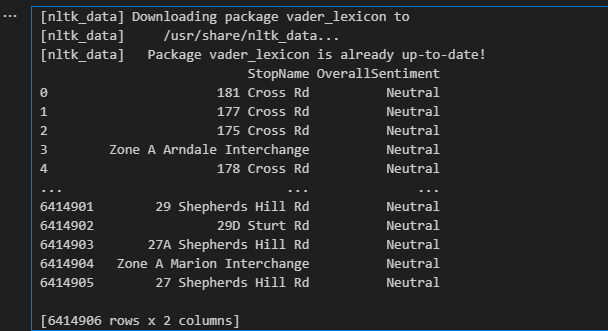
data['Negative'] = data['Sentiment'].apply(lambda x: x['neg'])

# Determine the overall sentiment

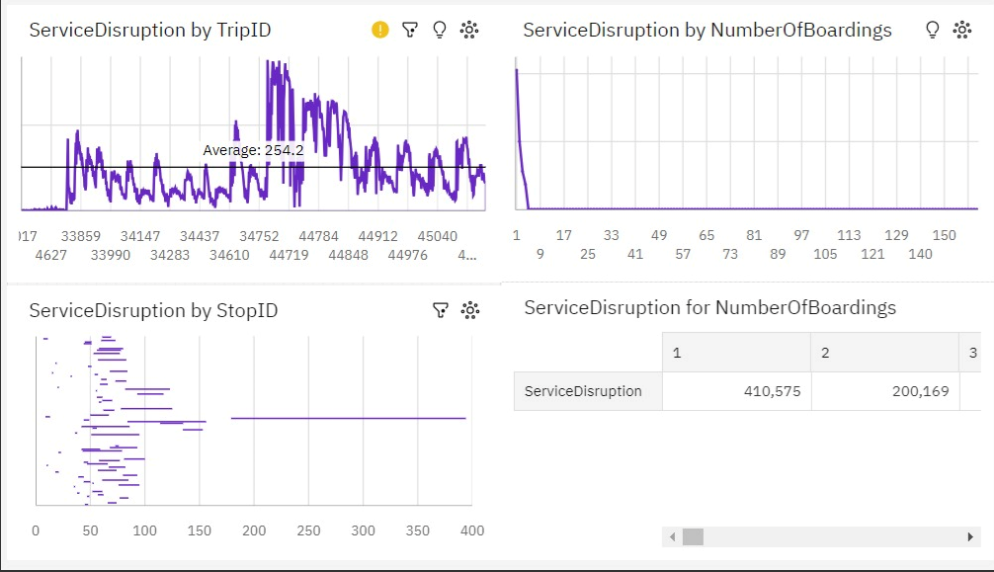
data['OverallSentiment'] = data['Sentiment'].apply(lambda x: 'Positive' if x['compound'] > 0 else ('Negative' if x['compound'] < 0 else 'Neutral'))

# Print the results or save to a new CSV file

print(data[['StopName', 'OverallSentiment']])



**Visualisation in service metrics:**

****

**Conclusion:**

By following these steps, we can enhance the public transportation analysis by calculating service punctuality rates and improve the transportation services by analysing the passenger feedback  using sentiment analysis.

**TEAM MEMBERS:**

Vaishnavi V - 2021506116

Tejasree M S - 2021506113

Yasvinippriyaa S K - 2021506126

Subithra C – 2021506109

Sreenithi B - 2021506105