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.

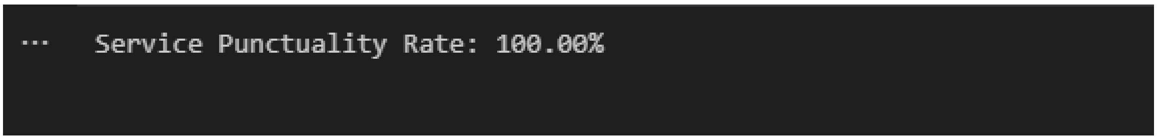
Service punctuality rates

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:

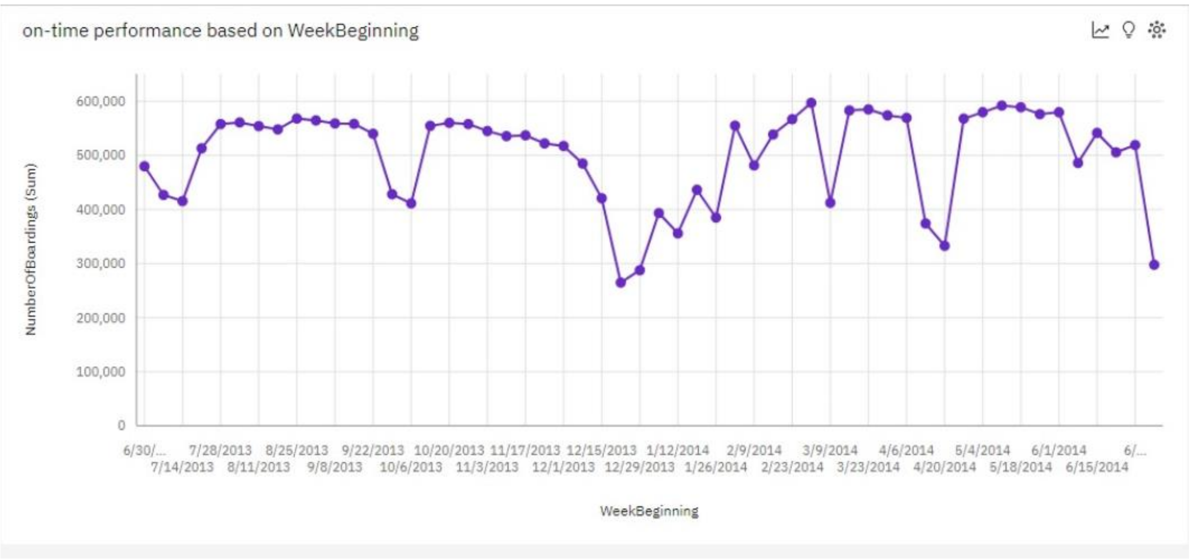
```
# Assuming a tolerance of 5 minutes for punctuality
tolerance = pd.Timedelta(minutes=5)


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}%")
```



```
···   Service Punctuality Rate: 100.00%
```

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.

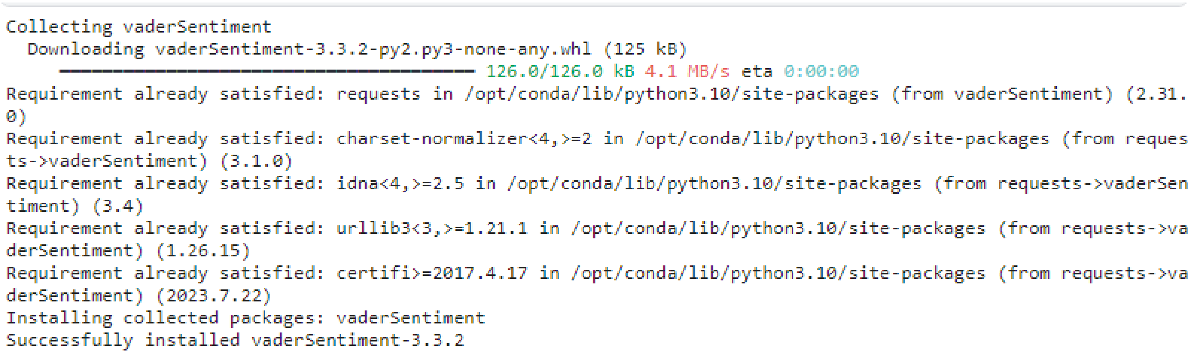
Load the dataset

```
# 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

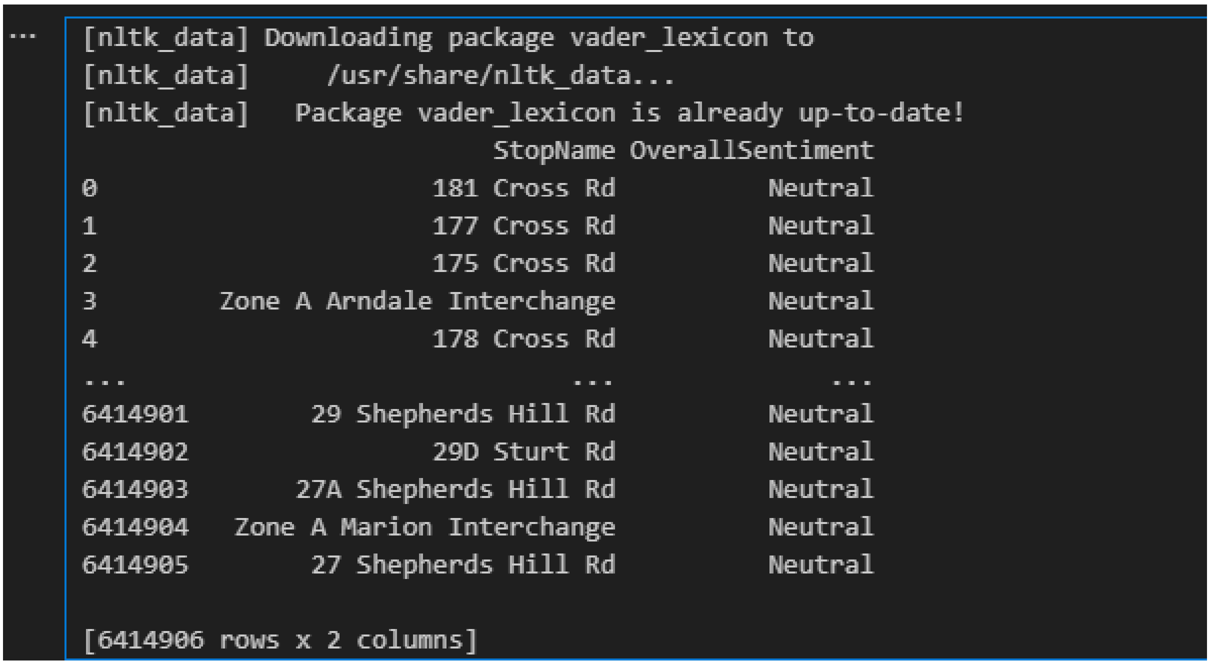
The VADER (Valence Aware Dictionary and sEntiment Reasoner) model is a lexicon and rule-based sentiment analysis tool that is specifically designed to understand and analyze sentiments in text data.The analysis is based on a lexicon of words, emojis, and sentiment intensity scores. pip command is used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

```
!pip install vaderSentiment
```



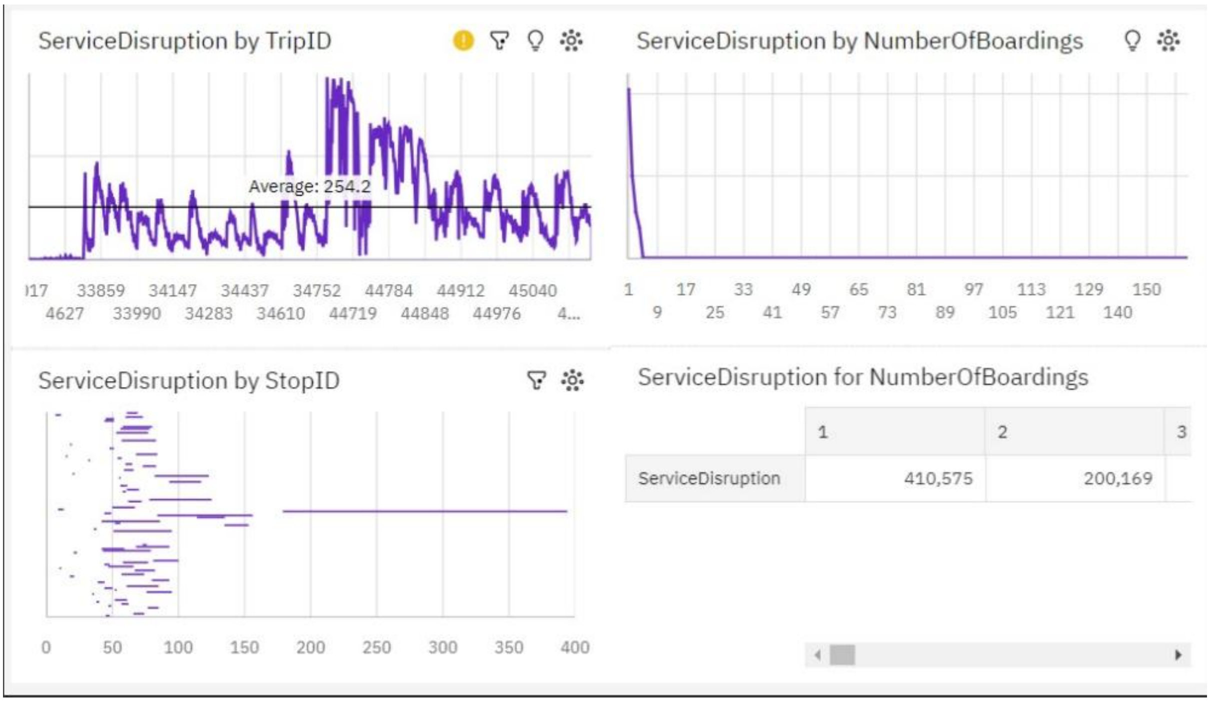
```
Collecting vaderSentiment
  Downloading vaderSentiment-3.3.2-py2.py3-none-any.whl (125 kB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 126.0/126.0 kB 4.1 MB/s eta 0:00:00
Requirement already satisfied: requests in /opt/conda/lib/python3.10/site-packages (from vaderSentiment) (2.31.
0)
Requirement already satisfied: charset-normalizer<4,>=2 in /opt/conda/lib/python3.10/site-packages (from reques
ts->vaderSentiment) (3.1.0)
Requirement already satisfied: idna<4,>=2.5 in /opt/conda/lib/python3.10/site-packages (from requests->vaderSen
timent) (3.4)
Requirement already satisfied: urllib3<3,>=1.21.1 in /opt/conda/lib/python3.10/site-packages (from requests->va
derSentiment) (1.26.15)
Requirement already satisfied: certifi>=2017.4.17 in /opt/conda/lib/python3.10/site-packages (from requests->va
derSentiment) (2023.7.22)
Installing collected packages: vaderSentiment
Successfully installed vaderSentiment-3.3.2
```

```
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']])
```



```
[nltk_data] Downloading package vader_lexicon to
[nltk_data]     /usr/share/nltk_data...
[nltk_data]   Package vader_lexicon is already up-to-date!
                           StopName OverallSentiment
0                      181 Cross Rd          Neutral
1                      177 Cross Rd          Neutral
2                      175 Cross Rd          Neutral
3         Zone A Arndale Interchange          Neutral
4                      178 Cross Rd          Neutral
...                             ...              ...
6414901          29 Shepherds Hill Rd          Neutral
6414902                 29D Sturt Rd          Neutral
6414903         27A Shepherds Hill Rd          Neutral
6414904    Zone A Marion Interchange          Neutral
6414905          27 Shepherds Hill Rd          Neutral

[6414906 rows x 2 columns]
```

Visualisation in service metrics:



ServiceDisruption for NumberOfBoardings

	1	2	3
ServiceDisruption	410,575	200,169	

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