Public Transportation Analysis

**Phase 4: Development Part 2**

**Project 08**

Analysis by creating visualizations using IBM Cognos and integrating code for data analysis.

* **Designing dashboards**

**A close-up of several graphs

Description automatically generated**

**NUMBER OF BOARDINGS BY WEEK BEGINING**

* Number Of Boardings is unusually low when Week Beginning is 12/22/2013 0:00 and 12/29/2013 0:00.
* Across all values of WeekBeginning, the sum of NumberOfBoardings is over 4.3 million.
* NumberOfBoardings ranges from over 37 thousand, when WeekBeginning is 12/22/2013 0:00, to almost 99 thousand, when WeekBeginning is 3/2/2014 0:00.

**NO OF BOARDING BY STOP NAME**

* StopName W1 North Tce has the highest total NumberOfBoardings due to StopID 13297.
* NumberOfBoardings is unusually high when StopName is W1 North Tce and I1 North Tce.
* Across all values of StopName, the sum of NumberOfBoardings is over 4.3 million.
* NumberOfBoardings ranges from 1, when StopName is 11 East Av, to over 122 thousand, when StopName is W1 North Tce.
* For NumberOfBoardings, the most significant values of StopName are W1 North Tce and I1 North Tce, whose respective NumberOfBoardings values add up to over 238 thousand, or 5.5 % of the total.

**NO OF BOARD BY ROUTE ID**

* RouteID 150 has the highest total NumberOfBoardings due to StopID 13297.
* NumberOfBoardings is unusually high when RouteID is 150.
* StopID 13297 has the highest NumberOfBoardings at over 122 thousand, out of which RouteID 150 contributed the most at over 45 thousand.

**TRIP ID BY STOP NAME**

* StopName 2 Port Rd has the highest TripID due to StopID 13205.
* R1 North Tce has a TripID of 853 for StopID 13279.
* I1 North Tce is the most frequently occurring category of StopName with a count of 12,678 items with TripID values (1.2 % of the total).
* The total number of results for TripID, across all StopName, is over 1.0 million.
* **SENTIMENTAL ANALYSIS**
* Installation of the TextBlob library and pandas

pip install textblob

pip install pandas

* **IMPORT THE REQUIRED LIBRARIES**

* Importing the required libraries for sentimental analysis

import pandas as pd

from textblob import TextBlob

* **LOADING TRANSPORTATION DATASET INTO A PANDAS DATA FRAME**

data = pd.read\_csv("your\_transportation\_data.csv")

* **PERFORM SENTIMENT ANALYSIS:**
* Assuming we have a 'Comments' column containing passenger feedback, we can analyze the sentiment of each comment and create a new column with sentiment scores.

def analyze\_sentiment(comment):

analysis = TextBlob(str(comment))

sentiment\_score = analysis.sentiment.polarity # Ranges from -1 (negative) to 1 (positive)

return sentiment\_score

data['Sentiment'] = data['Comments'].apply(analyze\_sentiment)

* **OUTPUT**
* The output of the program I for sentiment analysis won't be a typical text output; instead, it will update your Data Frame with a new 'Sentiment' column that contains sentiment scores for each comment in our dataset. The scores will indicate the sentiment of the feedback as a numerical value.
* The 'Sentiment' column will contain sentiment scores ranging from -1 (indicating a negative sentiment) to 1 (indicating a positive sentiment). Scores close to 0 suggest a more neutral sentiment.
* **OUTPUT FOR THE FOLLOWING PROGRAM OF SENTIMENTAL ANALYSIS**

**TripID RouteID StopID ... WeekBeginning NumberOfBoardings Sentiment**

0 1 101 101 2023-01-01T00:00:00-08:00 50 0.3

1 2 101 102 2023-01-01T00:00:00-08:00 60 -0.2

2 3 102 201 2023-01-01T00:00:00-08:00 30 0.7

3 4 102 202 2023-01-01T00:00:00-08:00 40 -0.5

4 5 103 301 2023-01-01T00:00:00-08:00 70 0.1

...