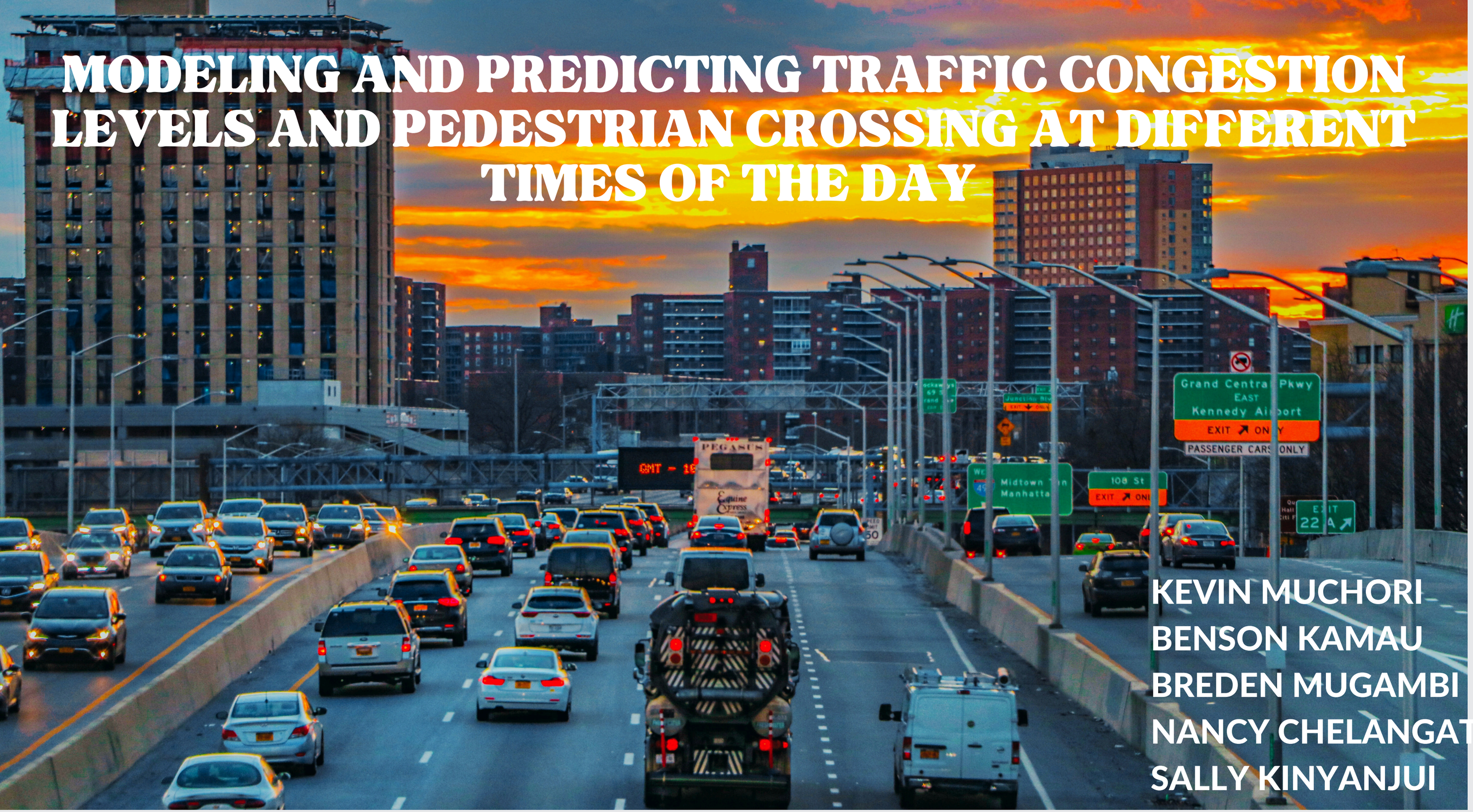
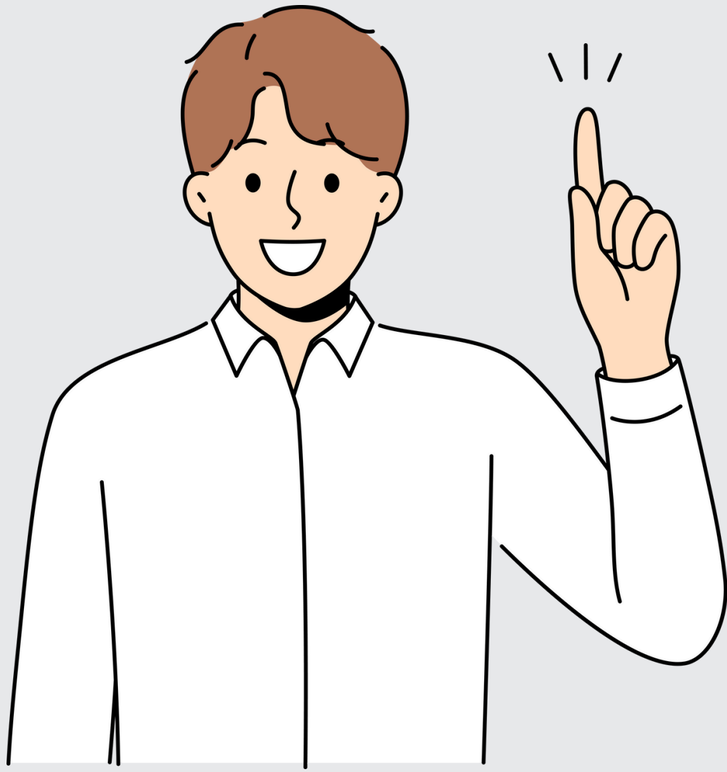
MODELING AND PREDICTING TRAFFIC CONGESTION LEVELS AND PEDESTRIAN CROSSING AT DIFFERENT TIMES OF THE DAY



KEVIN MUCHORI
BENSON KAMAU
BREDEN MUGAMBI
NANCY CHELANGAT
SALLY KINYANJUI

Overview

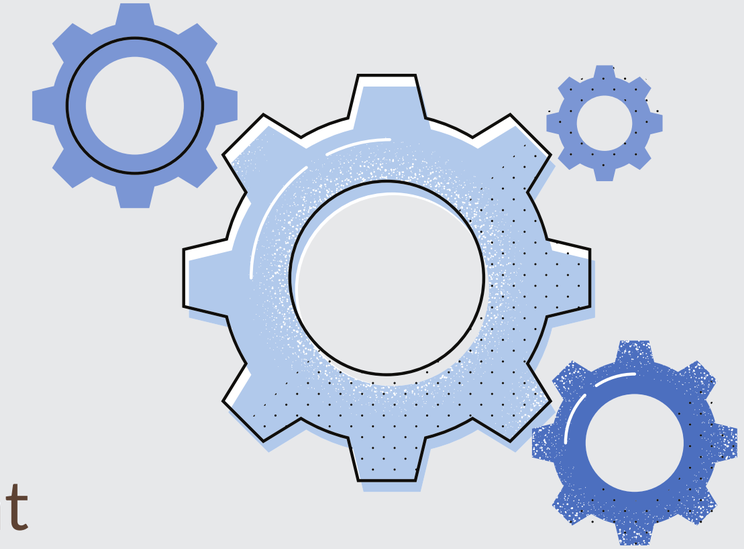
The urban mobility and transportation sector are vital for the functioning of modern cities, enabling the movement of people and goods efficiently. Within this industry, traffic management and pedestrian safety are crucial components that directly impact the quality of life in urban areas. Effective traffic pattern analysis and prediction can help mitigate congestion, enhance safety, and improve overall urban mobility. Well managed traffic leads to minimized economic losses, improved quality of life especially on the side of pedestrians.



PROBLEM STATEMENT.

Urban areas continue to face significant challenges in managing their traffic congestion and ensuring pedestrian safety. The changing nature of these areas together with the increasing volume of both vehicle and pedestrian traffic, makes it hard for one to predict traffic patterns affectively.

Objectives



Primary objective is to create an accurate time series model(s) that can model, analyze and predict traffic congestion levels and pedestrian crossings at different times of the day.

Specific objectives:

To identify key factors that influence traffic and pedestrian movement

2. To develop predictive models for forecasting future traffic congestion and pedestrian crossing patterns.
3. To provide recommendations for urban planners and traffic management authorities to improve traffic flow and pedestrian safety.

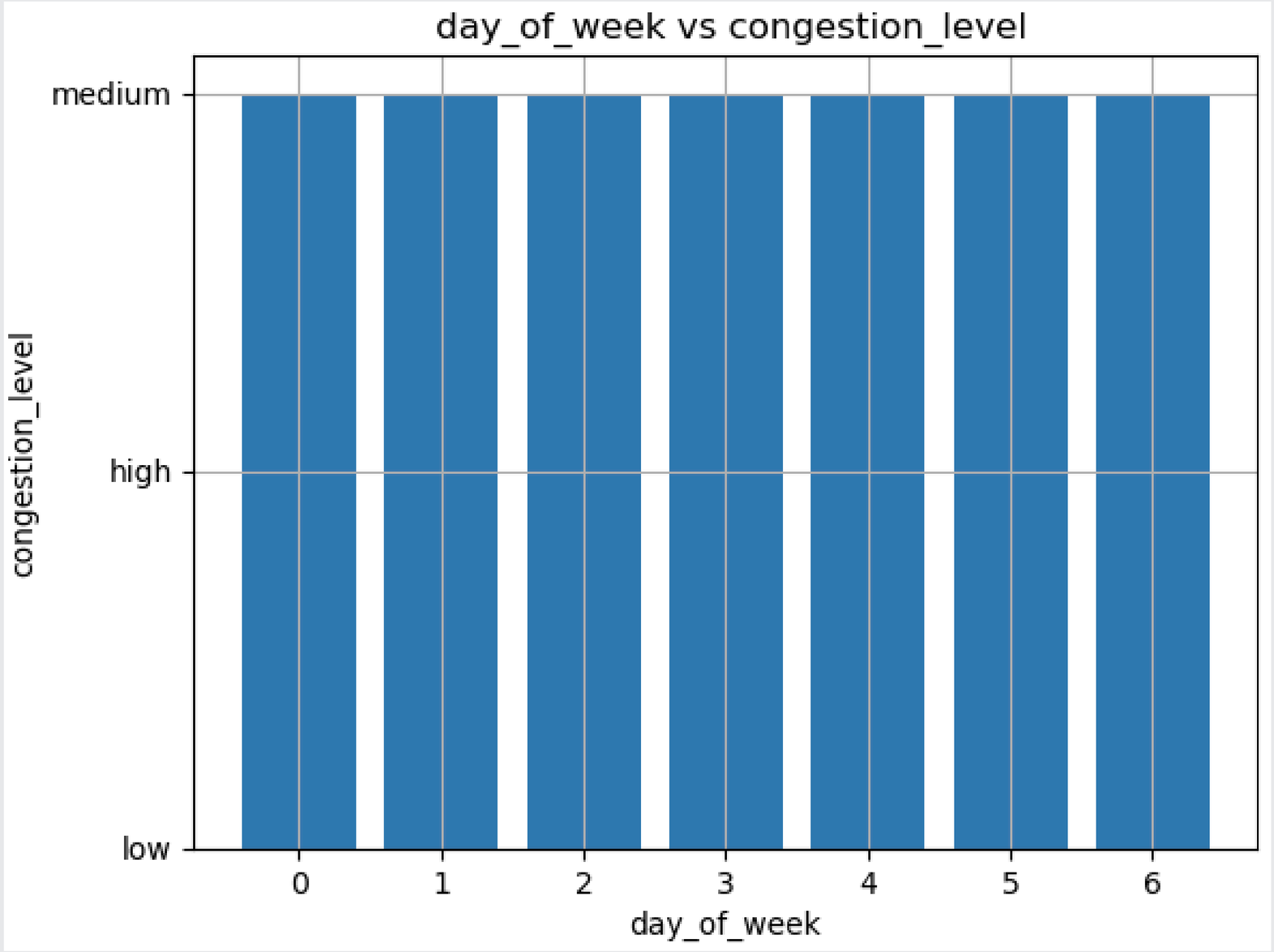
Data Understanding

The data to use in this study is sourced from the UC Irvine Machine Learning Repository. It has 4760 rows and 22 data features.

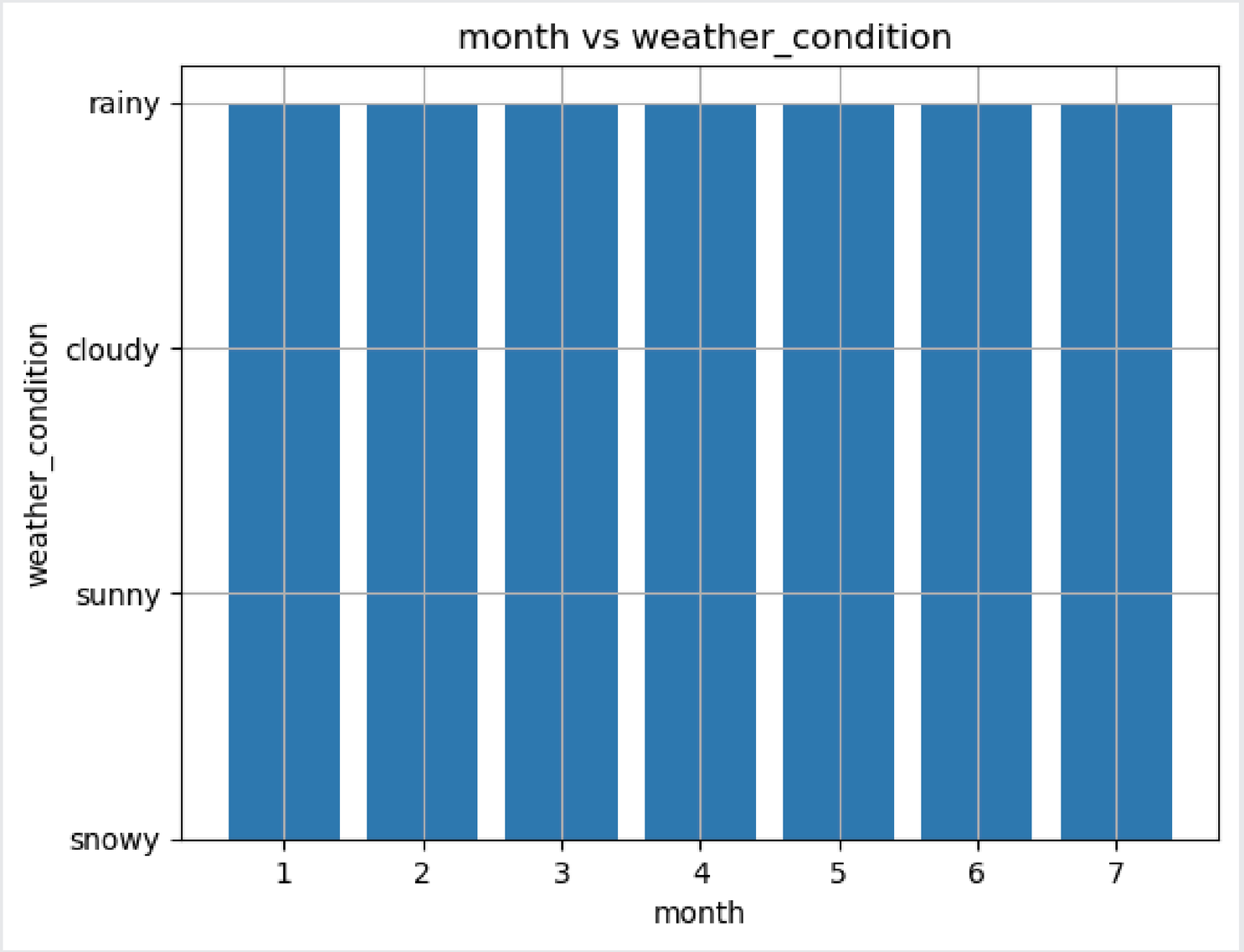
Data Preparation

- Dropped the empty columns
- Checked for duplicates
- Dropped duplicate rows
- Converted timestamp columnn to datetime format
- Set timestamp as the index
- Checked for outliers
- Carried out Exploratory Data Analysis(EDA)

EDA



The following visualization shows how there are different levels of congestion are in different days of the week.



The illusration shows the different weather conditions in different months of the year.

Modeling

- We carried out data preprocessing first.For our modeling bit we used different models to check for vehicle count,pedestrian count and congestion levels. We used the accuracy summary and recall summary noted the following:

- Decision Tree: Accuracy score - 32.5%
  Recall score - 32.6%

- Random forest : Accuracy score - 34%
  Recall score - 33.5%

- XGBoost : Accuracy score - 32.85%
  Recall score - 32.9%

FEATURE IMPORTANCE

Random forest

For Random forest model temperature, location and pedestrian count are the features that heavily affected congestion levels while weather condition as a feature came last

2) XGBoost

For XGBoost model pedestrian count, temperature and location in that order are the features that heavily affected congestion levels while weather condition as a feature came last

Conclusions

- The project yielded the desired results and the objectives were meant
- .
- The best models for both pedestrian and vehicle count was ARIMA and Random forest for the congestion levels.
- The features that heavily affected congestion levels were temperature, location and pedestrian count.

Recommendations

- Encourage businesses and schools to adopt staggered start times
- During identified peak pedestrian congestion times,increase and encourage frequent use of trains.
- Encourage walking and cycling during these times by improving infrastracture.
- Run public awareness campaigns and educate on best times to travel
- Consider implementing congestion charges during peak hours.

Next steps

Analyze how extreme weather impacts traffic patterns.

Continue to collect and analyze traffic data and try refining our models.