GROUP 15

ML MODEL TO PREDICT TRAFFIC FLOW

INTRODUCTION

Machine Learning Goal :

To build a regression model that predicts traffic volume at a specific junction (Cross_6) as our Target Variable using data from surrounding intersections (Cross_1 to Cross_5) and additional features such as weekday and time.

The model will help traffic authorities or smart city planners anticipate congestion and easily manage road usage.

Dataset: We used a Dataset from https://zenodo.org/records/3653880 and manipulated it to add some missing values.

Github
https://github.com/Oweci/GROUP-10-REGRESSION-ML-MODEL/edit/main/README.md

Our data has 11,519 rows and 8 columns

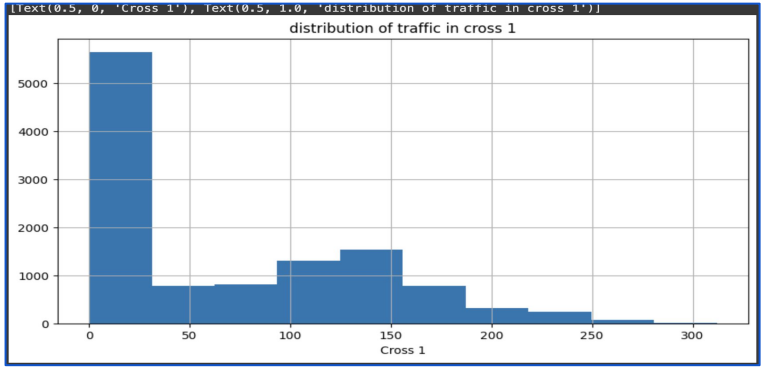
DATASET DESCRIPTION

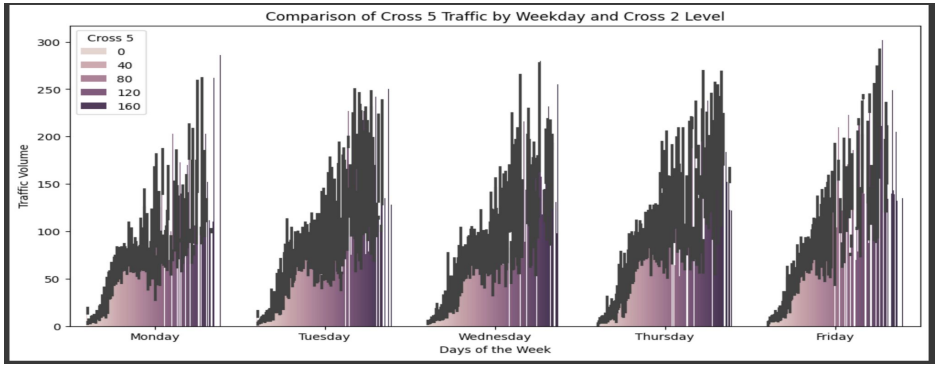
- Data represents traffic patterns across different junctions, time and days.
- Each row corresponds to traffic readings collected at one time interval.
- Traffic patterns vary depending on day of week and time.
- Tabular database (Numerical and Categorical)
- **Task :** Regression
- Our data has 11,519 rows and 8 columns

Columns	**Description**
Timestamp	Time durations
Weekday	Days if the week
Cross 1 to Cross 5	Cross roads (junctions)
Cross 6	Target Variable (Trying to predict the volume of cars)

EDA and VISUALIZATION (Salha & Sisco)

1. **Univariate Visualization :** Histogram
2. **Multivariate Visualization :** Grouped Bar chart





The histogram of Cross 1 shows that the majority of traffic counts are higher on the end,The road at Cross 1 is mostly busy. High traffic volume is the normal condition, and low traffic volume is the rare condition.

The multivariate bar chart shows how traffic at Cross 2 varies across the days of the week, while also comparing it with cross 5. We then use the colors inside each bar to show how the traffic volume at Cross 2 is affected by the traffic levels at a different intersection, Cross 5.

PREPROCESSING THE DATA (SISCO)

Used pipeline

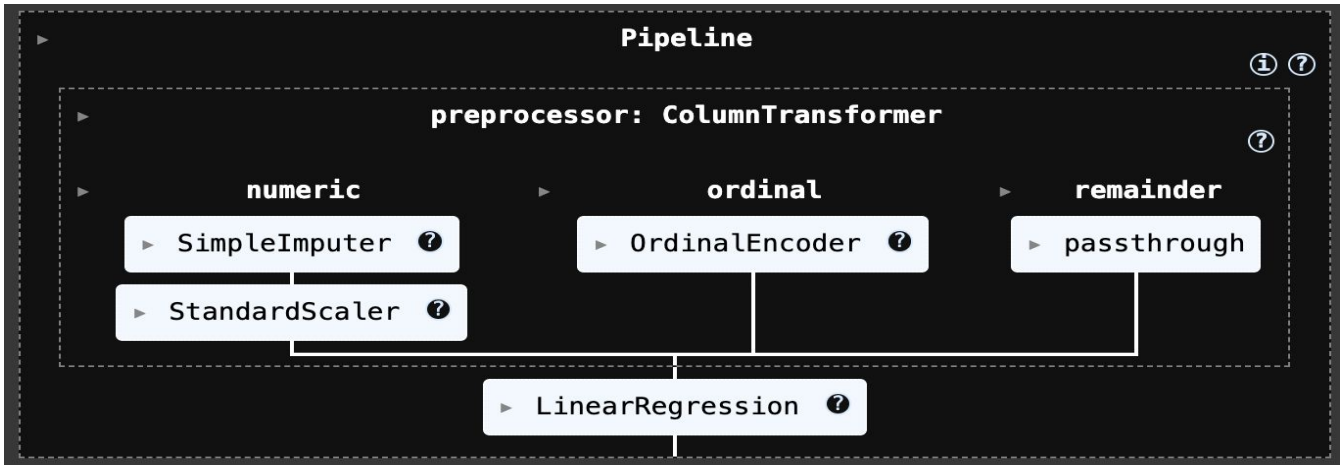
```
#creating transformers
num_transformer = Pipeline(steps = [
    ('imputer', numerical_meadian),
    ('scaler', scaler)
])

ordinal_transformer=Pipeline(
    steps=[('ord_enc',ordinal_encoder)]
    )
```

```
#combining transformers
df_traffic_preprocessor = ColumnTransformer(
    transformers =[
        ('numeric', num_transformer, num_cols),
        ('ordinal', ordinal_transformer,cat_cols)
    ],
    remainder='passthrough'
)
```

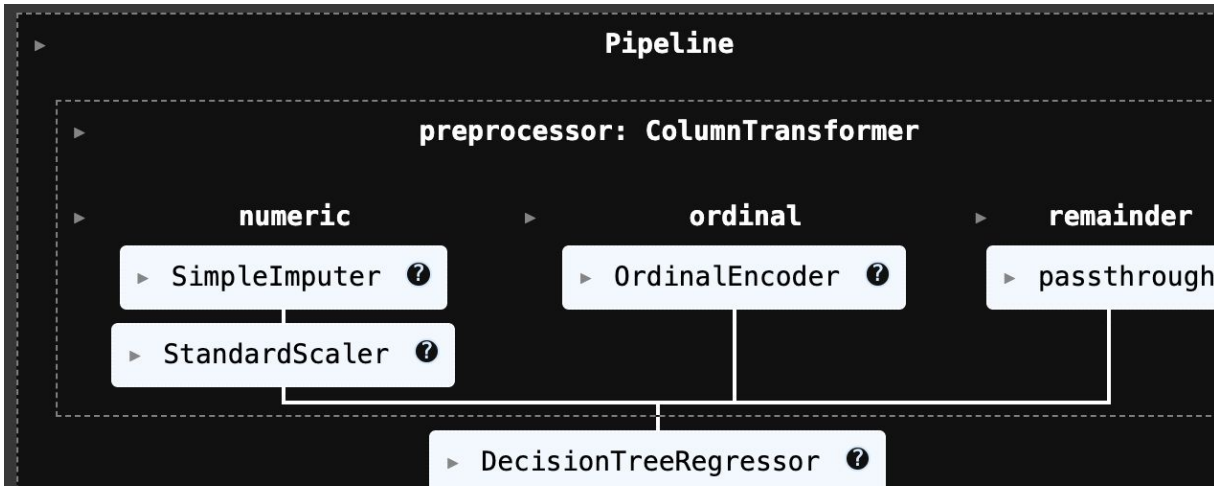
Model Development - (SISCO)

First Algorithm (Linear Regression)



Model Development cont.. (FLORENCE)

Decision Tree Regressor



Evaluation metrics (FLORENCE)

R squared score (Linear regression)

```
# Evaluate model performance on test data using R² score (coefficient of determination)
r2 = r2_score(y_test, y_predict)
print("R²:", r2)
```

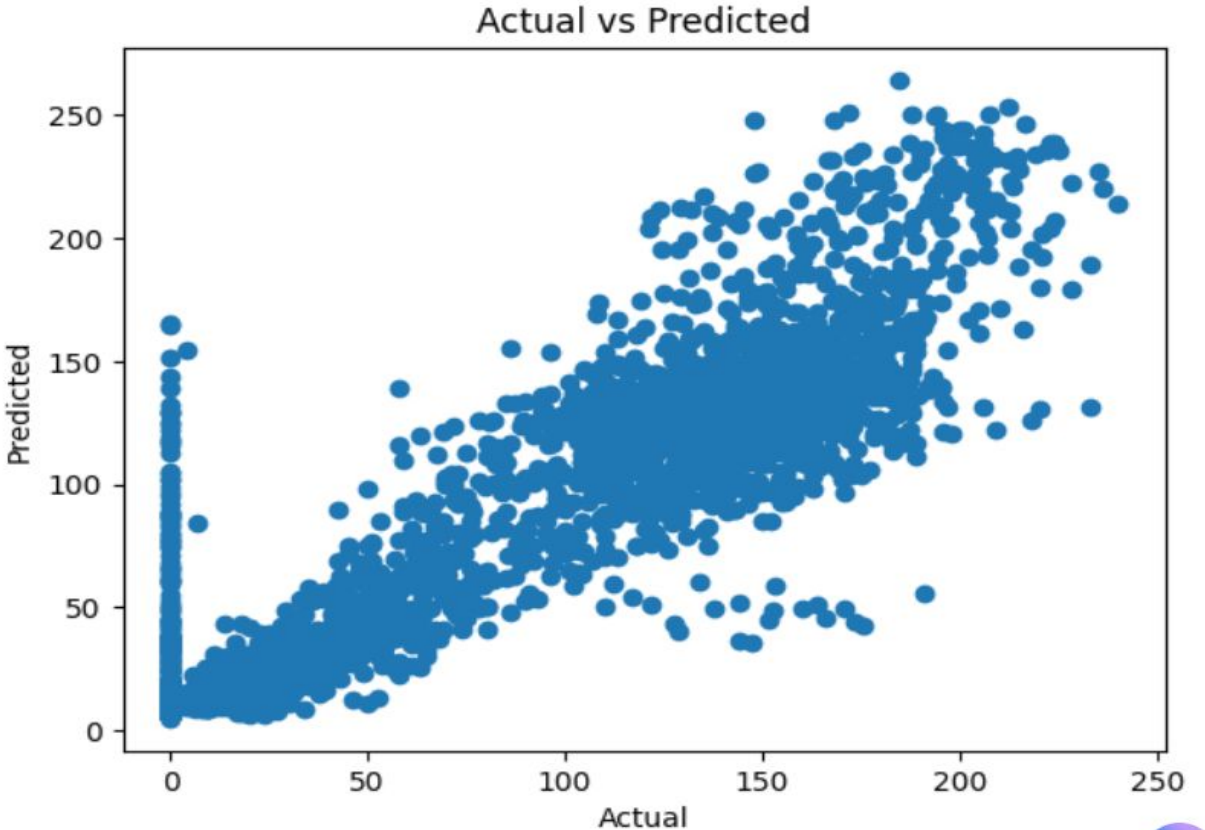
```
R²: 0.8540013437773113
```

R squared score (Decision tree regressor)

```
r2 = r2_score(y_test, y_predict_tree)
print("R²:", r2)
```

```
R²: 0.8476499006480733
```

Scatter plot for actual and predicted values. (FLORENCE)



THE END