IBM NAAN MUDHALVAN

**DATA ANALYTICS WITH COGNOS**

**Public Transportation Analysis**

|  |  |
| --- | --- |
| **DOMAIN :** | **Data Analytics With Cognos** |
| **PROJECT TITLE :** | **Public Transportation Analysis** |
| **TEAM MEMBERS AND REGISTER NUMBER** | **1.Bathri Narayanan S - 420421104011**  **2.Dinesh G - 420421104016**  **3.Dhilip Kumar M - 420421104015**  **4.VinothKumar S - 420421104303** |

**Dataset Link:**

**https://www.kaggle.com/datasets/rednivrug/unisys?select=20140711.CSV**

**1. Data Collection:**

Gather relevant data sources: This could include data on bus routes, stops, schedules, passenger counts, weather data, traffic conditions, and more.

Data preprocessing: Clean and prepare the data, handle missing values, and ensure data consistency.

2. Data Analysis:

Exploratory Data Analysis (EDA): Understand the characteristics of the data through statistical summaries, data visualization, and correlation analysis.

Identify key performance indicators (KPIs): Define the metrics that will be used to assess the efficiency of the public transportation system. KPIs may include on-time performance, passenger load, wait times, etc.

Analyze historical trends: Look for patterns and trends in the data that might indicate areas of improvement.

3. Model Building:

Build a predictive model: Create a machine learning model to predict key metrics like on-time performance, passenger demand, or travel times.

Consider feature engineering: Create new features that might improve the predictive power of your model, such as time of day, weather conditions, special events, etc.

Implement optimization models: Use optimization techniques to improve route planning and scheduling.

4. Model Evaluation:

Split the data into training and testing sets.

Evaluate the model's performance using appropriate metrics, such as accuracy, F1-score, or mean absolute error, depending on the specific task.

Fine-tune the model: Adjust hyperparameters, feature selection, and data preprocessing to improve model performance.

Cross-validation: Perform k-fold cross-validation to ensure the model's generalizability.

5. Visualization and Reporting:

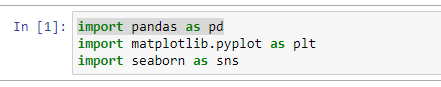
Create visualizations: Use tools like Matplotlib, Seaborn, or Plotly to visualize the results of your analysis and model predictions. Visualization can include geographical plots of routes, line charts showing performance trends, and heatmaps of passenger demand.

Generate reports: Summarize your findings, including insights from data analysis, model performance, and recommendations for improving public transportation efficiency.

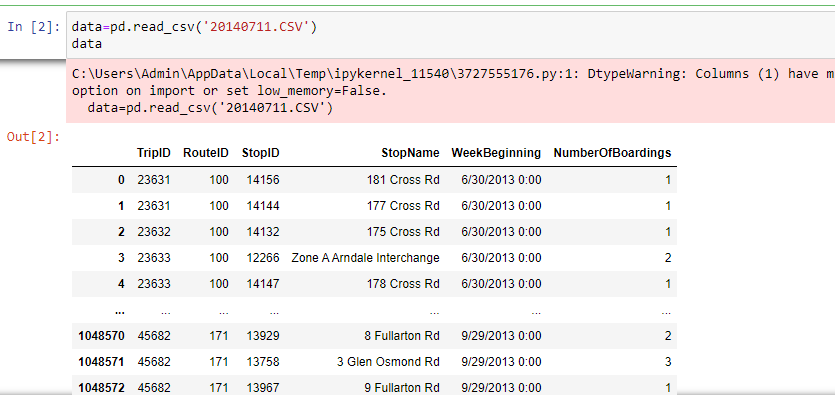
Interactive dashboards: Create web-based dashboards using tools like Tableau, Power BI, or Dash to allow stakeholders to interact with the data and model outputs

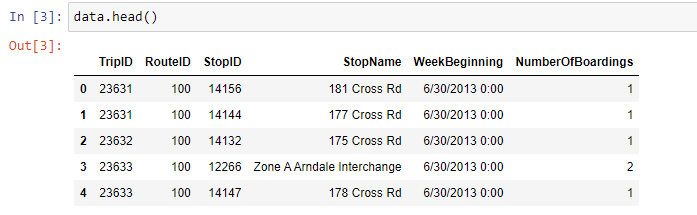
**Program:**

**1.Import the Libraries:**

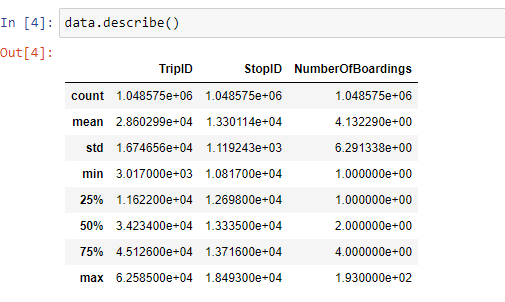
****

**2. Import the Dataset:**

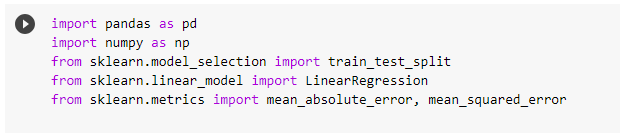
****

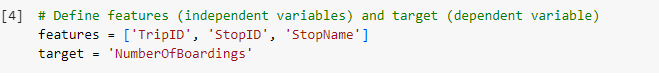
****

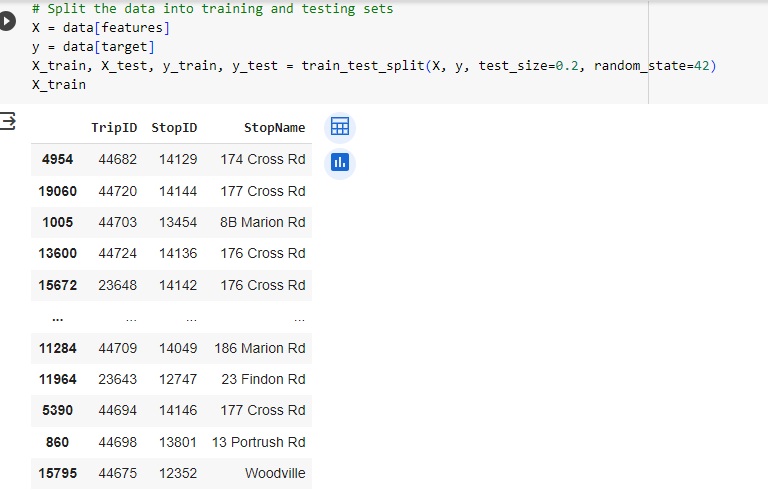
**3.Describe the Dataset:**

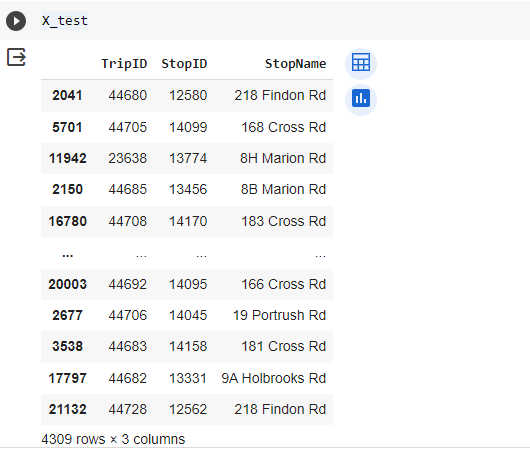
****

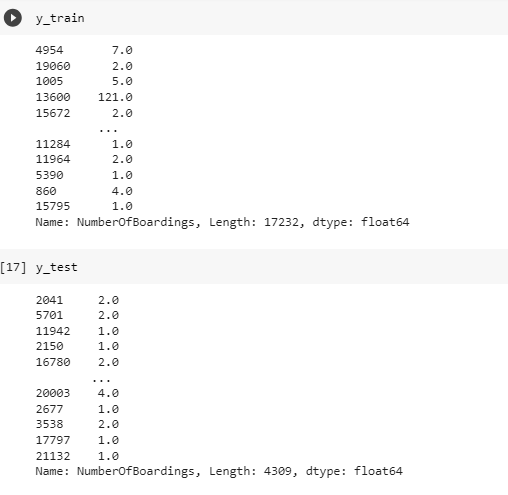
**4. Public Transport Analysis:**

****

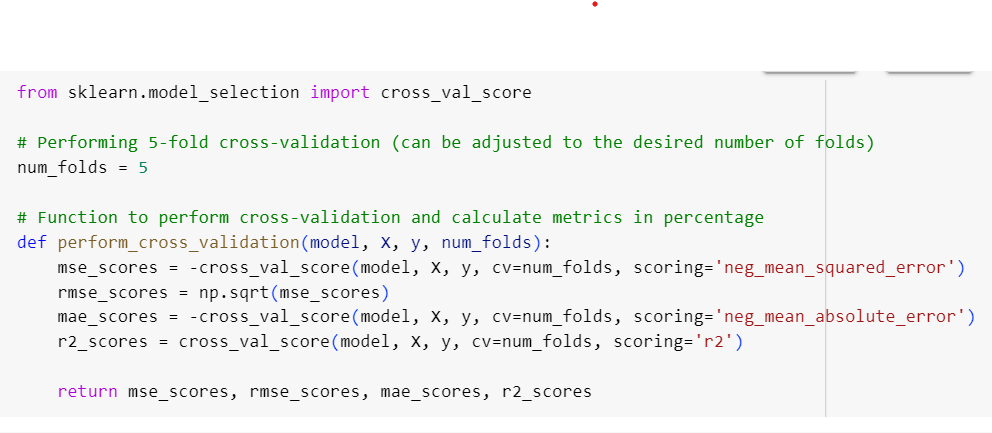
****

****

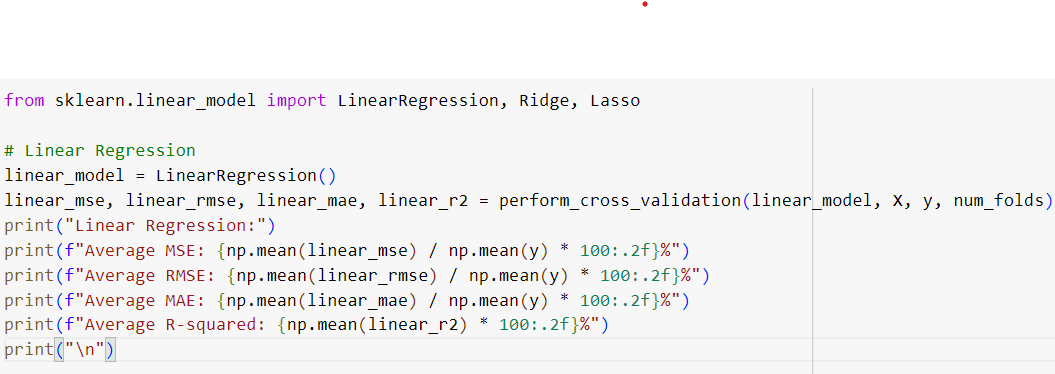
****

****

MODEL EVALUATION:



LINEAR REGRESSION:



OUTPUT:

Linear Regression:

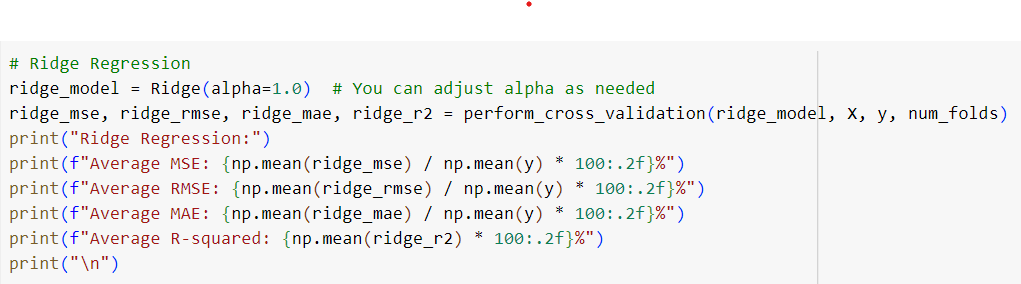
Average MSE: 21.90%

Average RMSE: 13.01%

Average MAE: 10.38%

Average R-squared: 87.53%

RIDGE REGRESSION:



OUTPUT:

Ridge Regression:

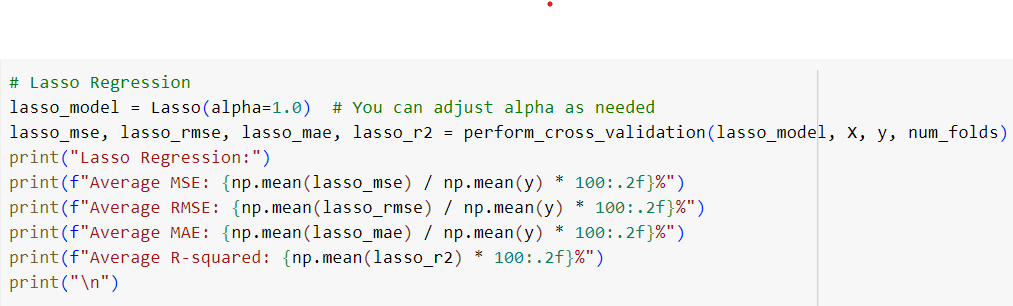
Average MSE: 19.67%

Average RMSE: 11.20%

Average MAE: 8.54%

Average R-squared: 89.19%

LASSO REGRESSION:



OUTPUT:

Lasso Regression:

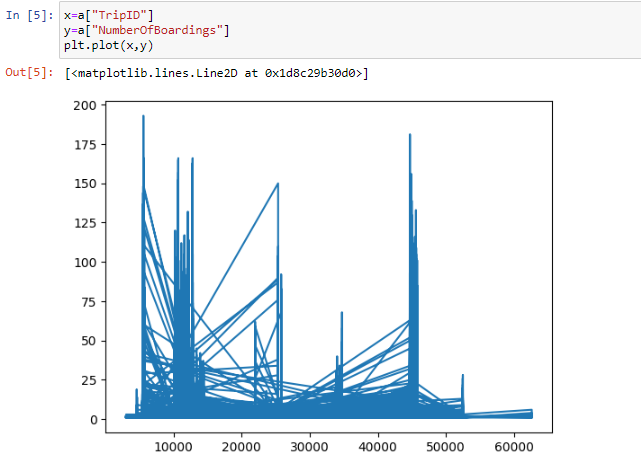
Average MSE: 115.55%

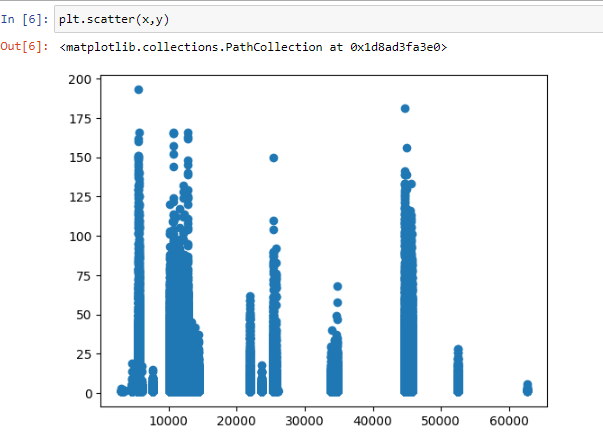
Average RMSE: 27.51%

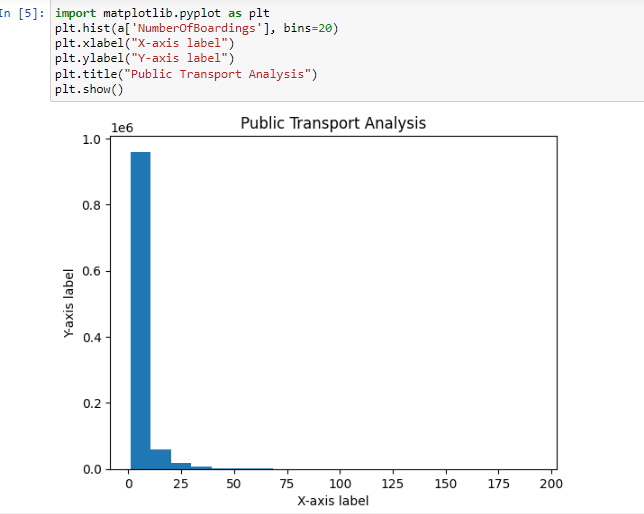
Average MAE: 22.39%

Average R-squared: 35.98%

VISUALIZATION:

****

****

****