
SEOUL BIKE DEMAND
PREDICTION

PROBLEM STATEMENT

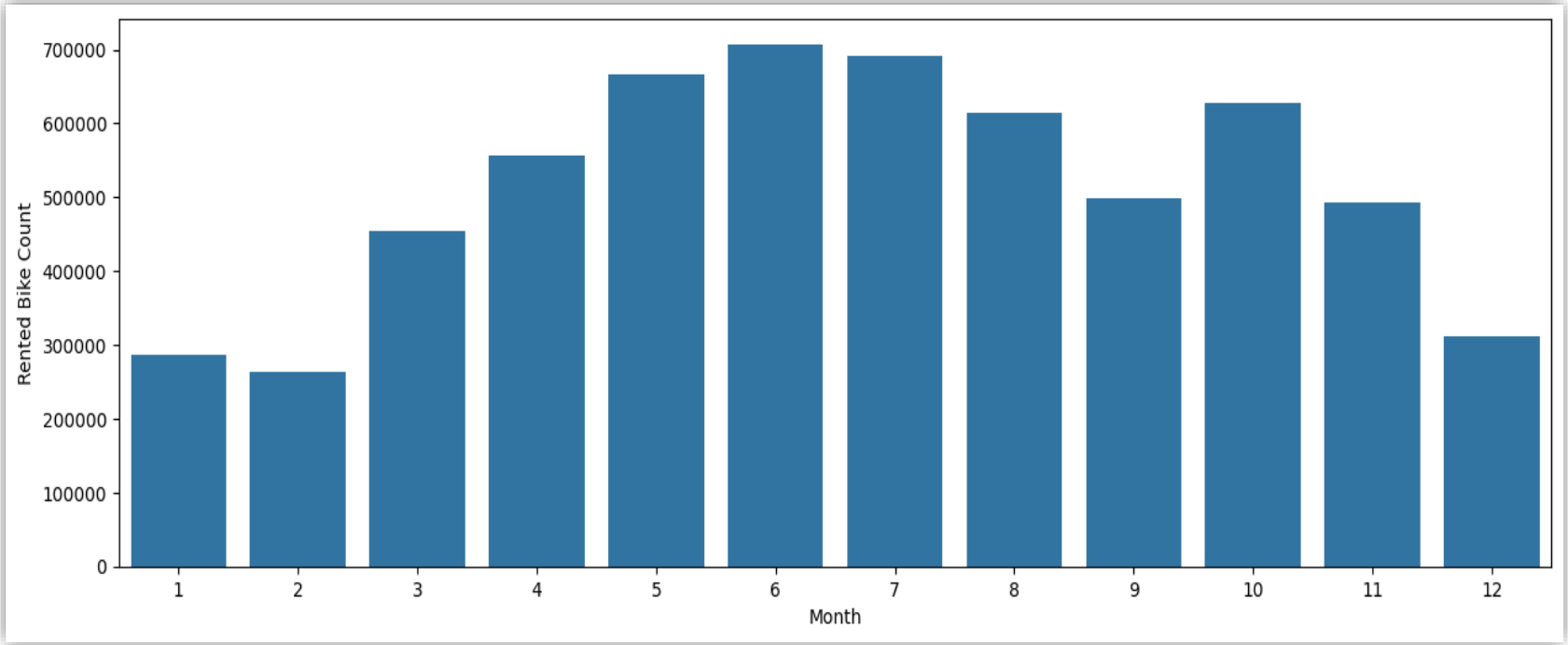
- The objective of this project is to accurately predict bike rental demand in Seoul, South Korea, using machine learning techniques. By analyzing factors such as weather conditions, time of day, and other relevant variables, we aim to develop robust predictive models.
- Exploratory data analysis (EDA) will provide insights into the dataset, and model tuning will optimize performance. Deep learning methods may be explored for complex pattern recognition, ensuring reliable predictions.

- Date
- Hour
- Temperature(°C)
- Humidity (%)
- Wind speed
- Dew point temperature
- Visibility
- temperature(°C)
- Solar Radiation
- Rainfall(mm)
- Snowfall (cm)
- Seasons Holiday
- Functioning Day
- Rented Bike Count

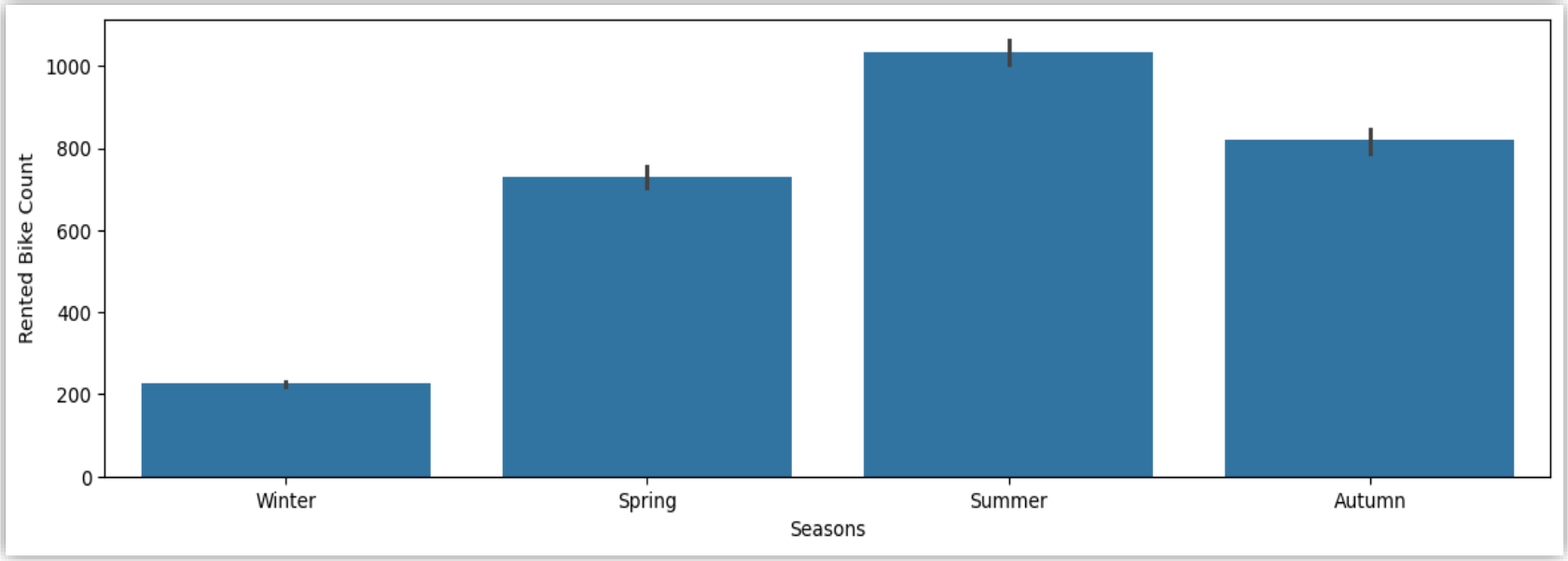
Data Preprocessing

1. The overall dataset does not contain any missing values.
2. Relevant features were created using the date column.
3. Applied Standardization technique to scale the features, ensuring equal contribution of all variables to the model.
4. Categorical variables were encoded:
   - Manually encoded: Holiday, Functioning Day
   - Dummy encoding: Seasons, Week Day.

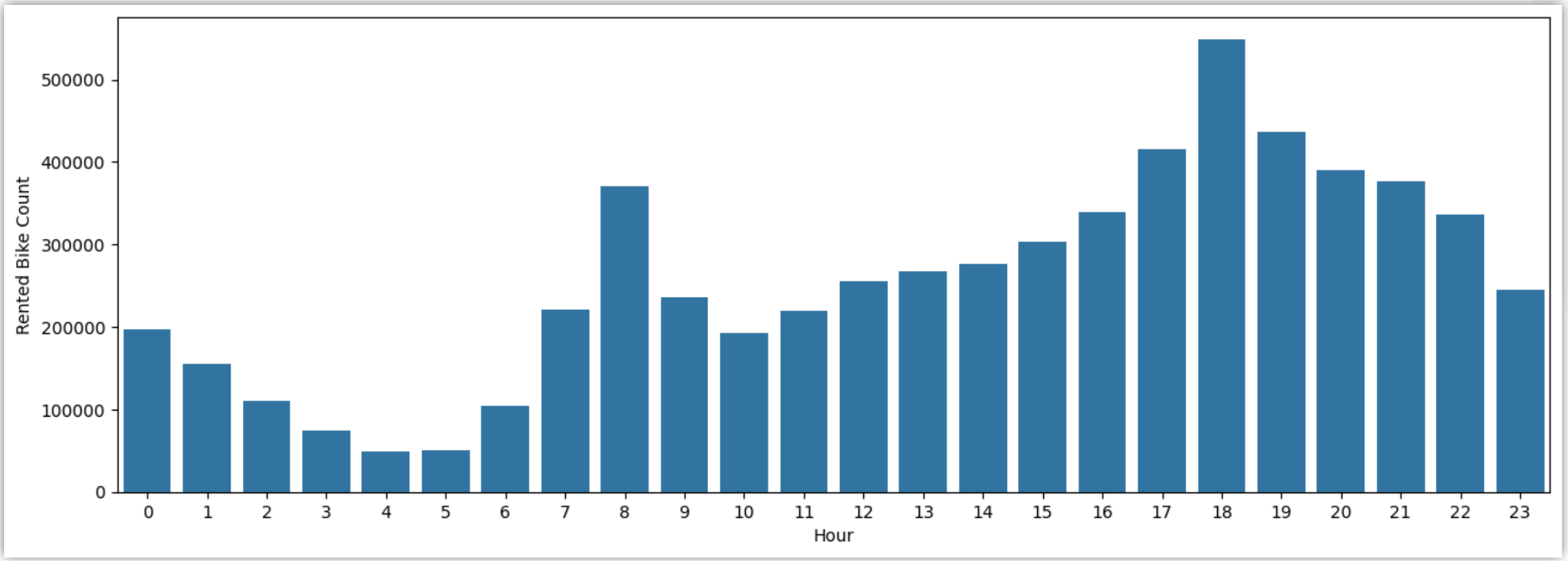
Data Analysis



During the summer months of June and July, bike demand reaches its peak.

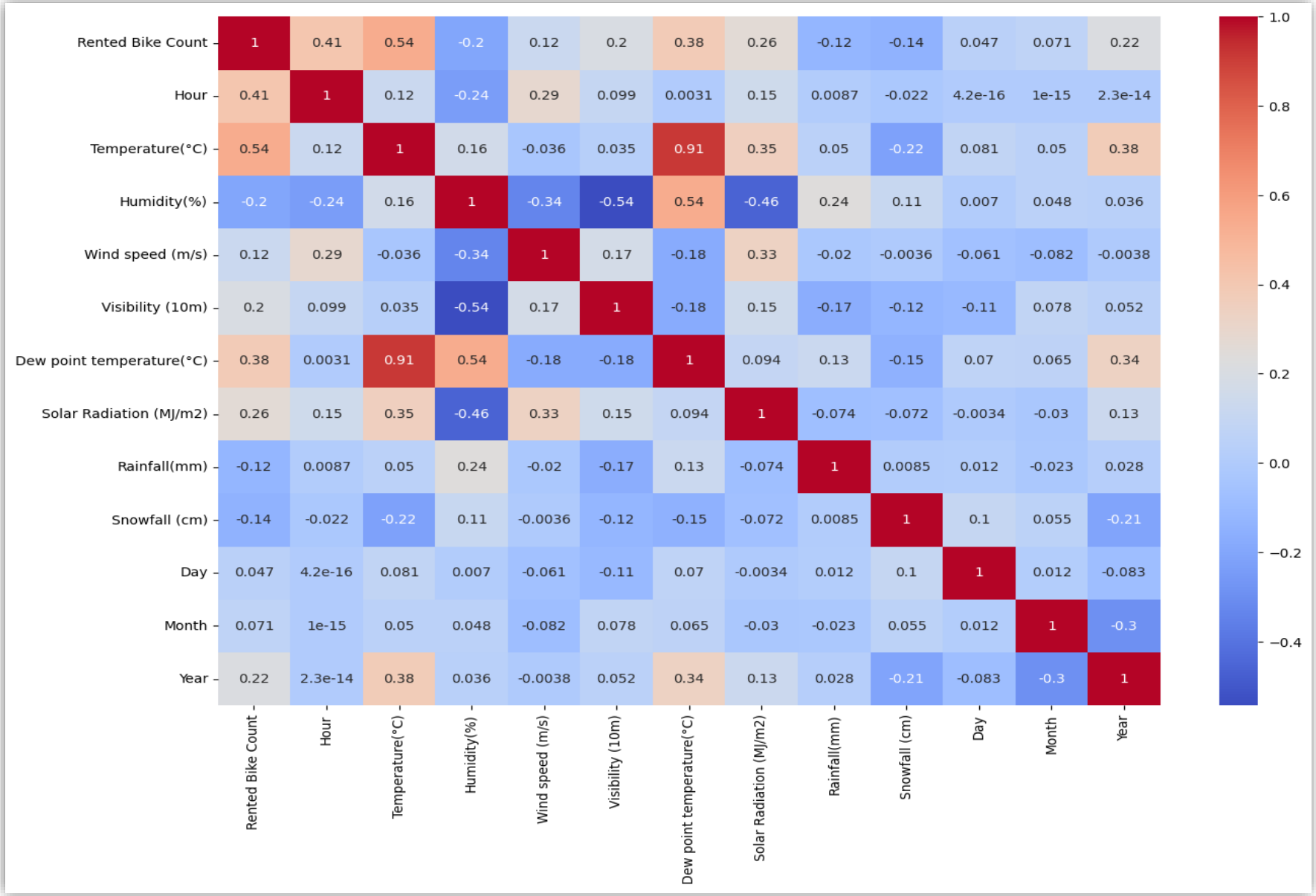


The highest bike demand occurs during summer, but demand significantly drops in winter.



- Bike demand is minimum at 4 am during the night.
- Demand increase in the early morning at 8 am as many people go for work.
- The highest demand for bikes happens at 6 pm when lots of people are traveling.

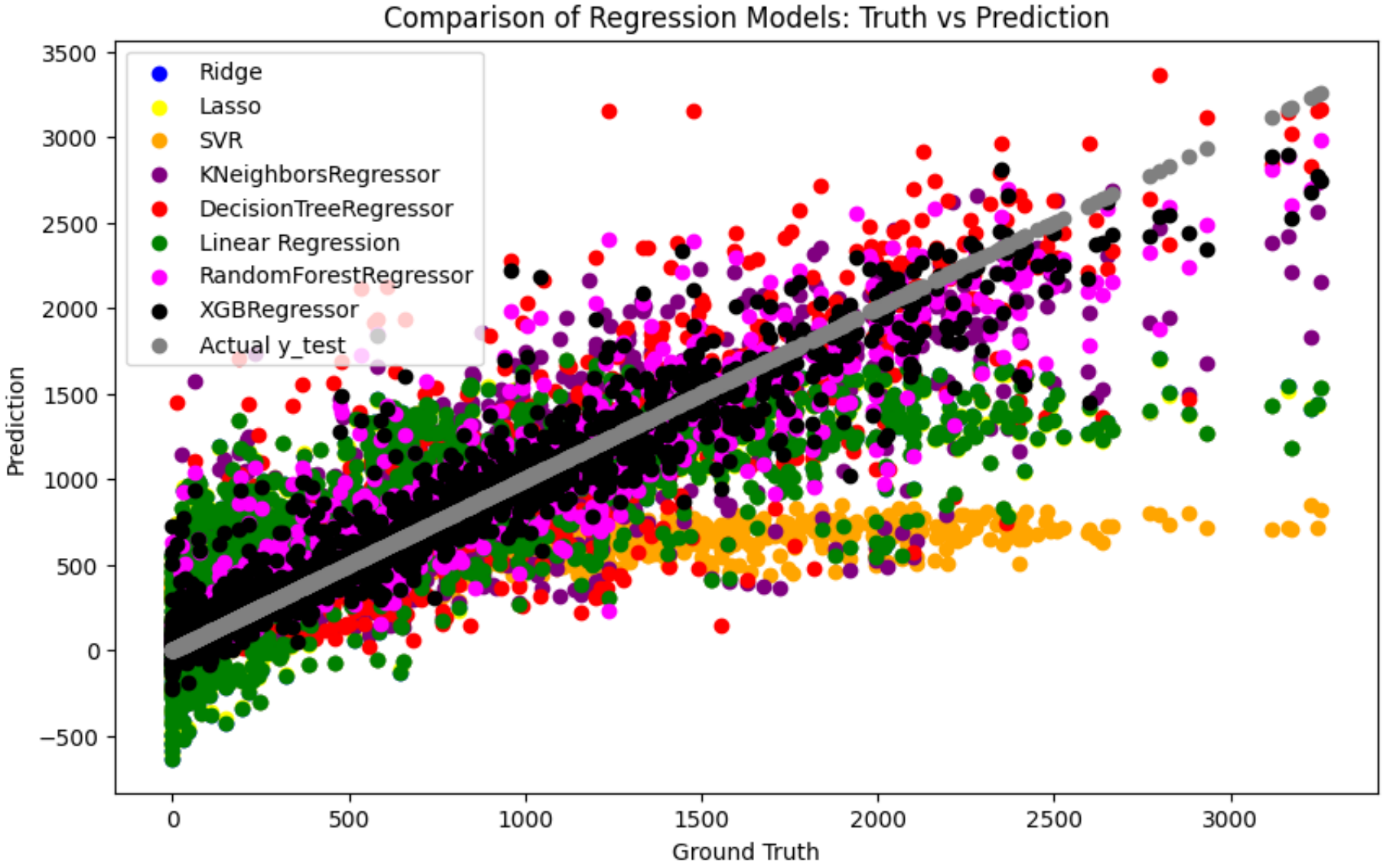
Heatmap :


Rented Bike Count 1 0.41 0.54 -0.2 0.12 0.2 0.38 0.26 -0.12 -0.14 0.047 0.071 0.22
Hour 0.41 1 0.12 -0.24 0.29 0.099 0.0031 0.15 0.0087 -0.022 4.2e-16 1e-15 2.3e-14
Temperature(°C) 0.54 0.12 1 0.16 -0.036 0.035 0.91 0.35 0.05 -0.22 0.081 0.05 0.38
Humidity(%) -0.2 -0.24 0.16 1 -0.34 -0.54 0.54 -0.46 0.24 0.11 0.007 0.048 0.036
Wind speed (m/s) 0.12 0.29 -0.036 -0.34 1 0.17 -0.18 0.33 -0.02 -0.0036 -0.061 -0.082 -0.0038
Visibility (10m) 0.2 0.099 0.035 -0.54 0.17 1 -0.18 0.15 -0.17 -0.12 -0.11 0.078 0.052
Dew point temperature(°C) 0.38 0.0031 0.91 0.54 -0.18 -0.18 1 0.094 0.13 -0.15 0.07 0.065 0.34
Solar Radiation (MJ/m2) 0.26 0.15 0.35 -0.46 0.33 0.15 0.094 1 -0.074 -0.072 -0.0034 -0.03 0.13
Rainfall(mm) -0.12 0.0087 0.05 0.24 -0.02 -0.17 0.13 -0.074 1 0.0085 0.012 -0.023 0.028
Snowfall (cm) -0.14 -0.022 -0.22 0.11 -0.0036 -0.12 -0.15 -0.072 0.0085 1 0.1 0.055 -0.21
Day 0.047 4.2e-16 0.081 0.007 -0.061 -0.11 0.07 -0.0034 0.012 0.1 1 0.012 -0.083
Month 0.071 1e-15 0.05 0.048 -0.082 0.078 0.065 -0.03 -0.023 0.055 0.012 1 -0.3
Year 0.22 2.3e-14 0.38 0.036 -0.0038 0.052 0.34 0.13 0.028 -0.21 -0.083 -0.3 1
Rented Bike Count
Hour
Temperature(°C)
Humidity(%)
Wind speed (m/s)
Visibility (10m)
Dew point temperature(°C)
Solar Radiation (MJ/m2)
Rainfall(mm)
Snowfall (cm)
Day
Month
Year
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4

Model Building

Accuracy on Validation Set :

Model	Training	Validation Set	RMSE
Ridge	0.556	0.537	434.67
Lasso	0.556	0.537	434.57
Linear Regression	0.556	0.537	434.67
SVR	0.208	0.213	566.72
K-Neighbors	0.840	0.748	320.52
Decision Tree	1.0	0.748	297.04
Random Forest	0.984	0.891	210.59
XGB Regressor	0.989	0.914	186.89



- XGBRegressor outperforms all other models, making it the top choice.

Hyperparmeter Tunning :

Top Model	Without tune	Fining Tunning
Random Forest	0.891	0.901
XGBRegression	0.914	0.923

- So I confirmed that the XGBRegressor model is the best option.
- I then used the Joblib function to save both the trained model and the feature scaling process for future inference tasks.

Thank You