

# **Bike Rental Prediction using Python and Data Science**

**A case study on predicting bike rentals  
with historical data**

**Name: Sarah Moin**

**Date: 19/09/2024**



# Objective & Dataset Overview

► **Objective:** Predict future bike rentals using historical data from a bike-sharing system.

► **Dataset Description:**

► **hour.csv** – 17,379 hourly records

► **day.csv** – 731 daily records

► **Features Overview:**

► **Features include:** *season, yr, mnth, hr, weekday, weathersit, temp, atemp, hum, windspeed, and bike counts (casual, registered, cnt).*

```
day_df.head(10)
```

	instant	dteday	season	yr	mnth	holiday	weekday	workingday	weathersit	temp	atemp	hum	windspeed	casual	registered	cnt
0	1	01/01/2011	1	0	1	0	6	0	2	0.344167	0.363625	0.805833	0.160446	331	654	985
1	2	02/01/2011	1	0	1	0	0	0	2	0.363478	0.353739	0.696087	0.248539	131	670	801
2	3	03/01/2011	1	0	1	0	1	1	1	0.196364	0.189405	0.437273	0.248309	120	1229	1349
3	4	04/01/2011	1	0	1	0	2	1	1	0.200000	0.212122	0.590435	0.160296	108	1454	1562
4	5	05/01/2011	1	0	1	0	3	1	1	0.226957	0.229270	0.436957	0.186900	82	1518	1600
5	6	06/01/2011	1	0	1	0	4	1	1	0.204348	0.233209	0.518261	0.089565	88	1518	1606
6	7	07/01/2011	1	0	1	0	5	1	2	0.196522	0.208839	0.498696	0.168726	148	1362	1510
7	8	08/01/2011	1	0	1	0	6	0	2	0.165000	0.162254	0.535833	0.266804	68	891	959
8	9	09/01/2011	1	0	1	0	0	0	1	0.138333	0.116175	0.434167	0.361950	54	768	822
9	10	10/01/2011	1	0	1	0	1	1	1	0.150833	0.150888	0.482917	0.223267	41	1280	1321

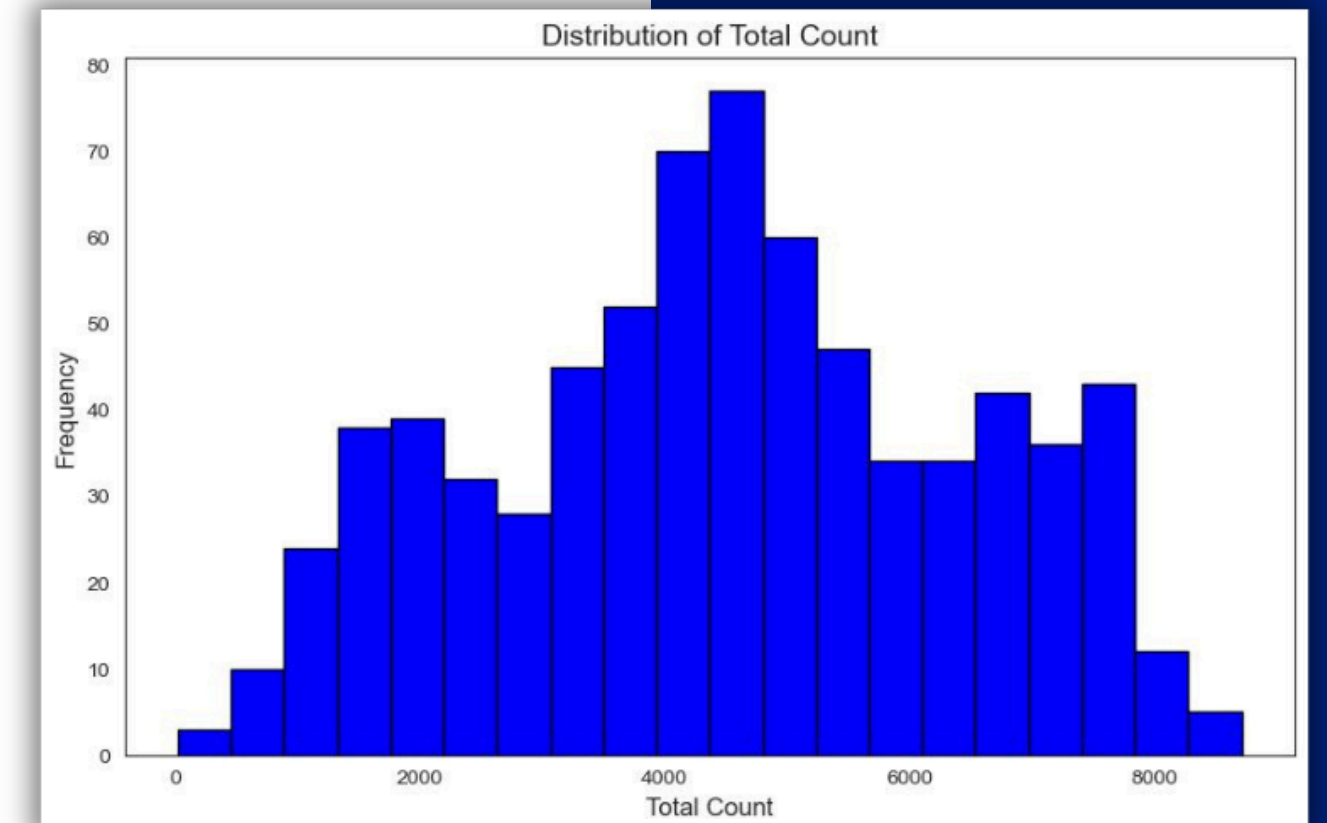
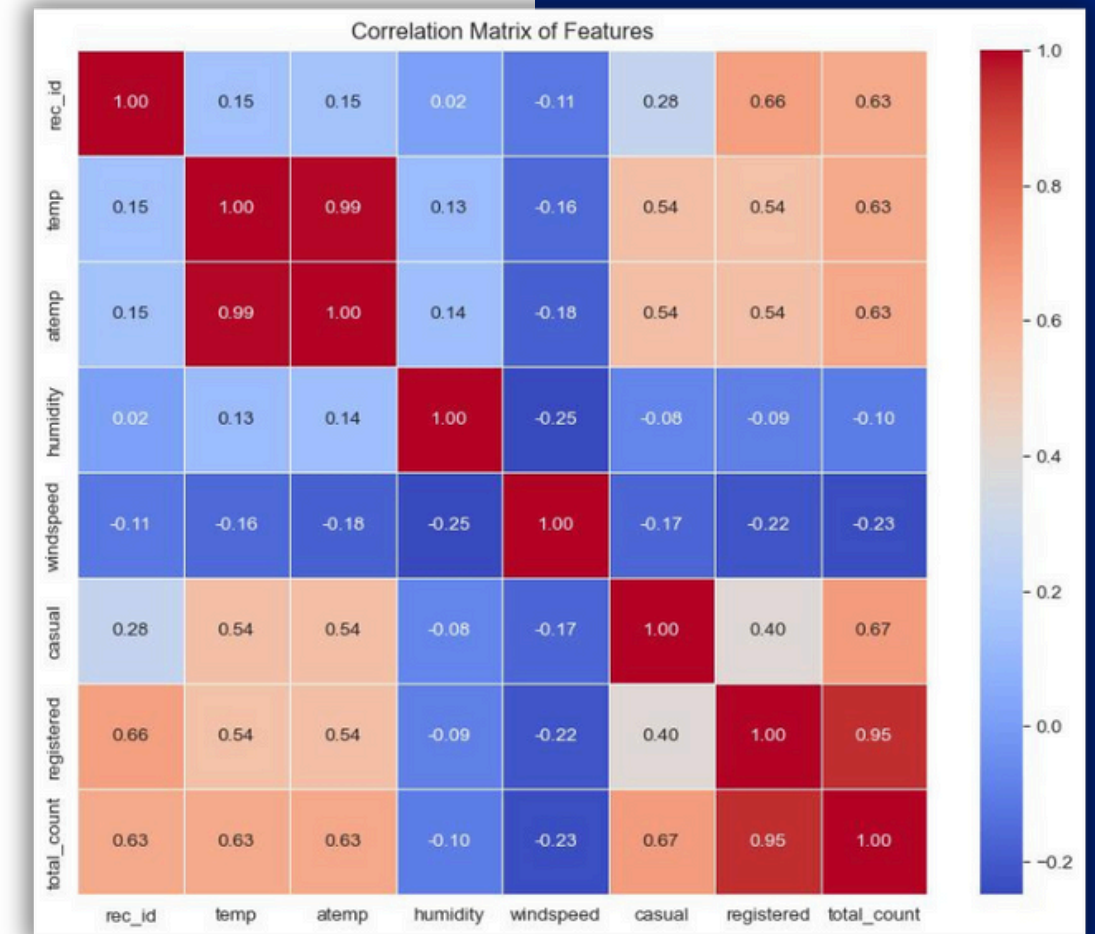
# Data Preprocessing & Exploration

## Cleaning

- Renamed columns to improve readability and make the dataset easier to work with.
- Checked for missing data across all columns.
- Removed any rows with null values to ensure data integrity.
- Identified and deleted duplicate rows to avoid redundant data that could skew the analysis.

## Feature Engineering

- Handled categorical features like season, year, month, holiday, weekday etc. using label encoding.
- Created lags in data in interval of 7 days .

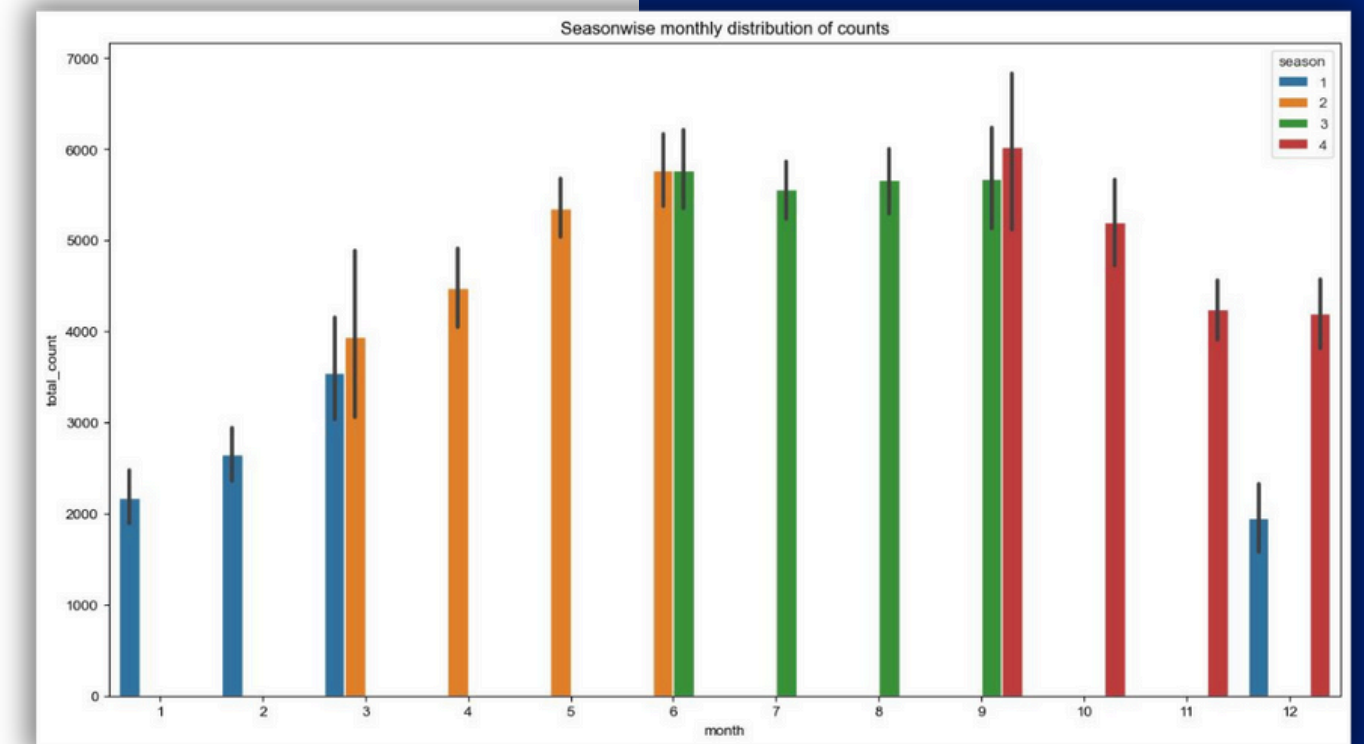
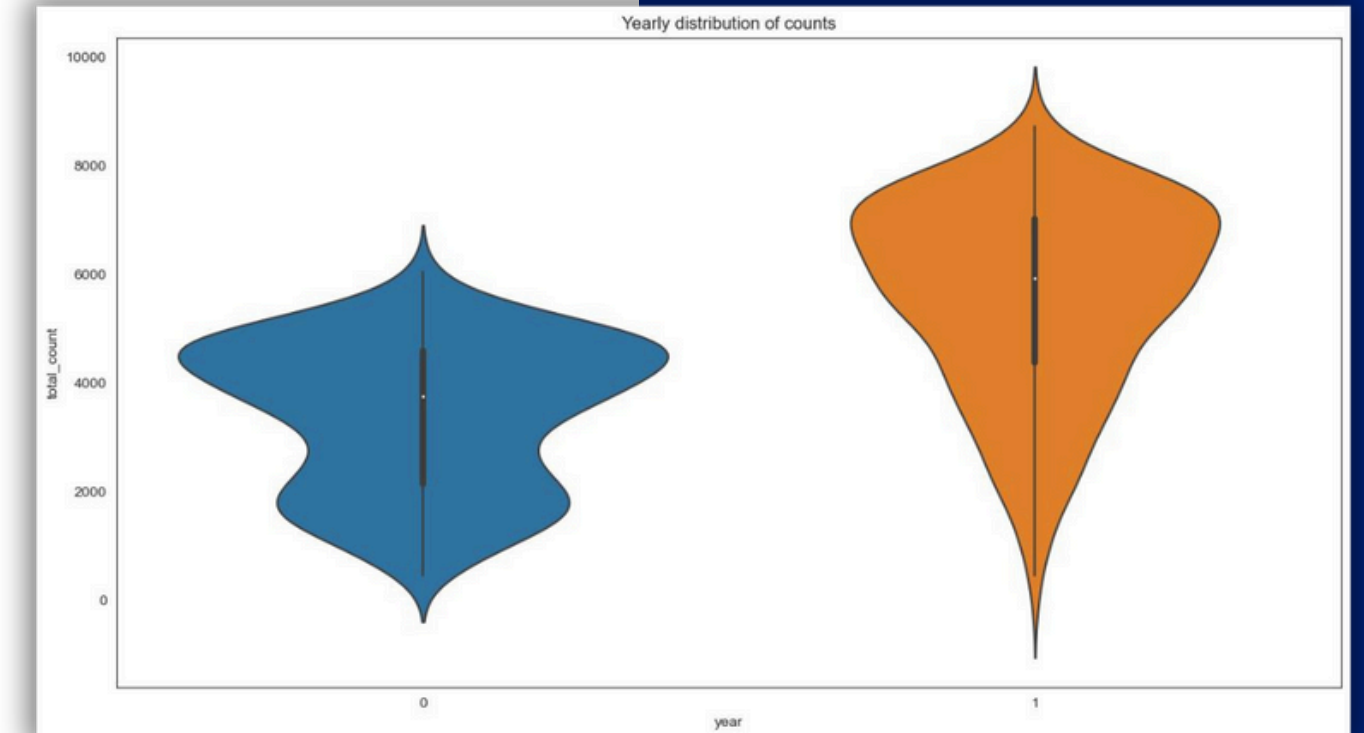
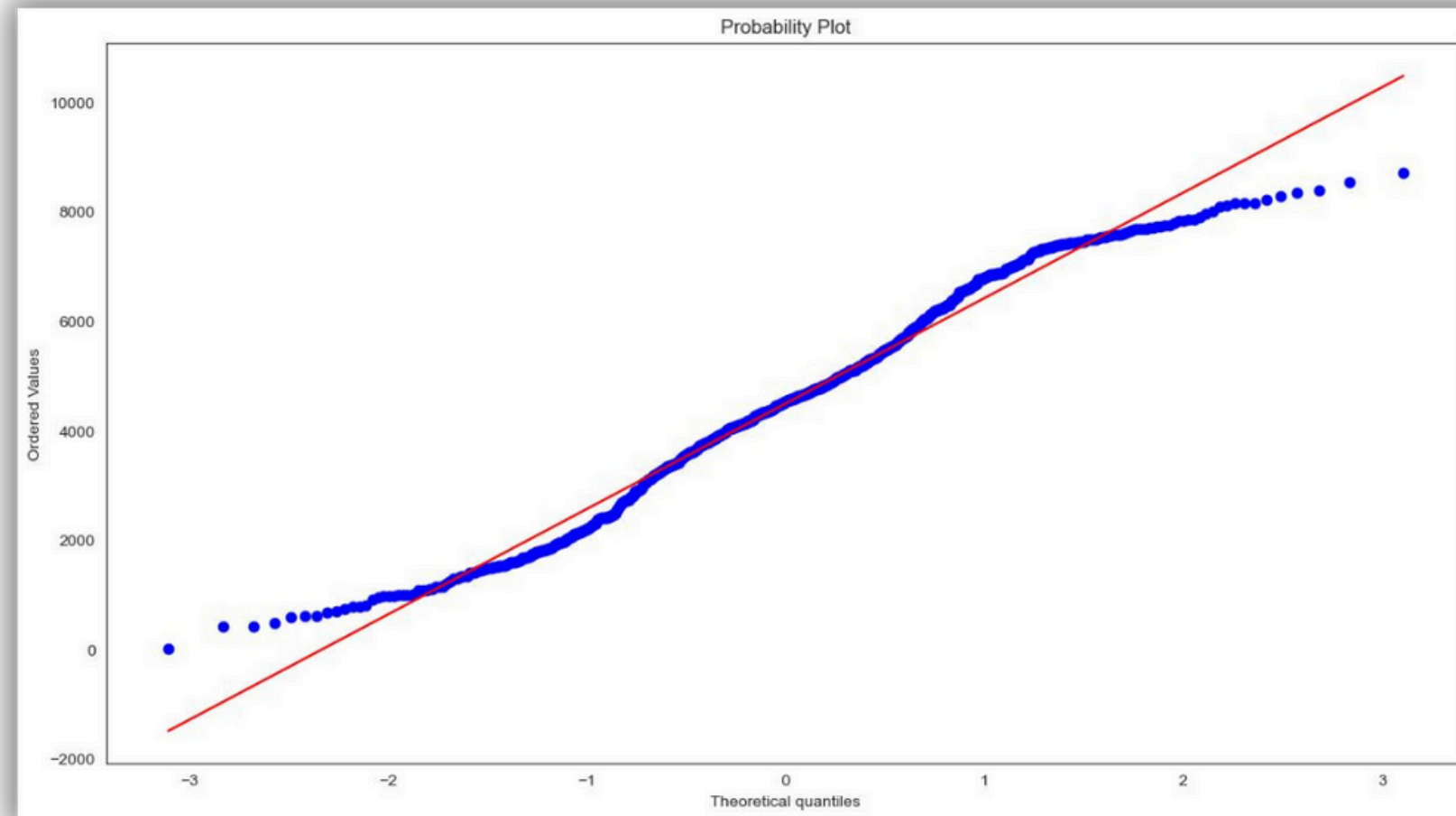




# Data Preprocessing & Exploration

## Exploratory Data Analysis (EDA)

- Explored dataset by plotting different graphs and tables.



# Modeling Approach

## Train-Test Split

Split the data 80-20 for training and testing.

## Algorithms

Regression Algorithms are used like

- Linear Regression
- Random Forest
- Gradient Boosting Regressor

## Feature Selection

Features: Removed humidity and windspeed as they were not making any impact which can be seen in correlation matrix.



# Model Evaluation & Results

## Evaluation Metrics

MSE, R2 Score

## Results

### Linear Regression:

Mean Squared Error (One-Hot):

8.263703816779052e-25

R<sup>2</sup> Score (One-Hot): 1.0

### Random Forest:

rf: Mean Squared Error:

111814.21645379311

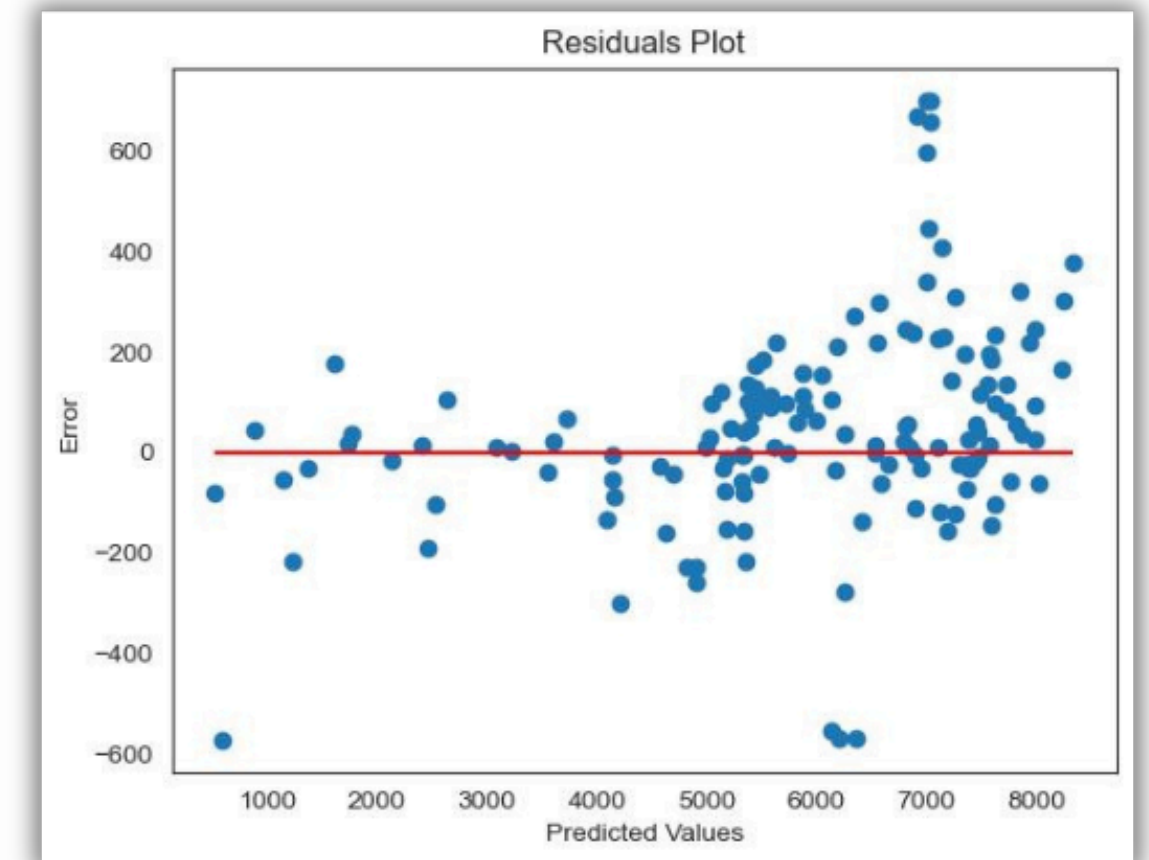
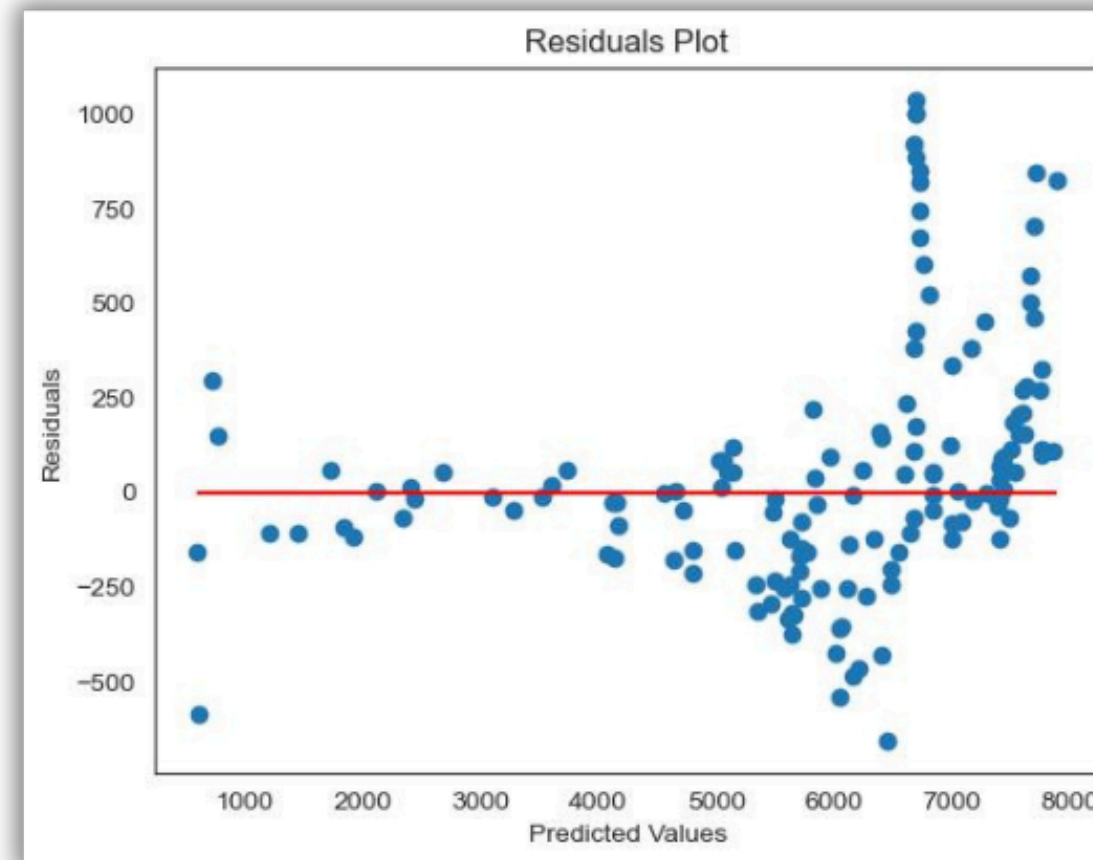
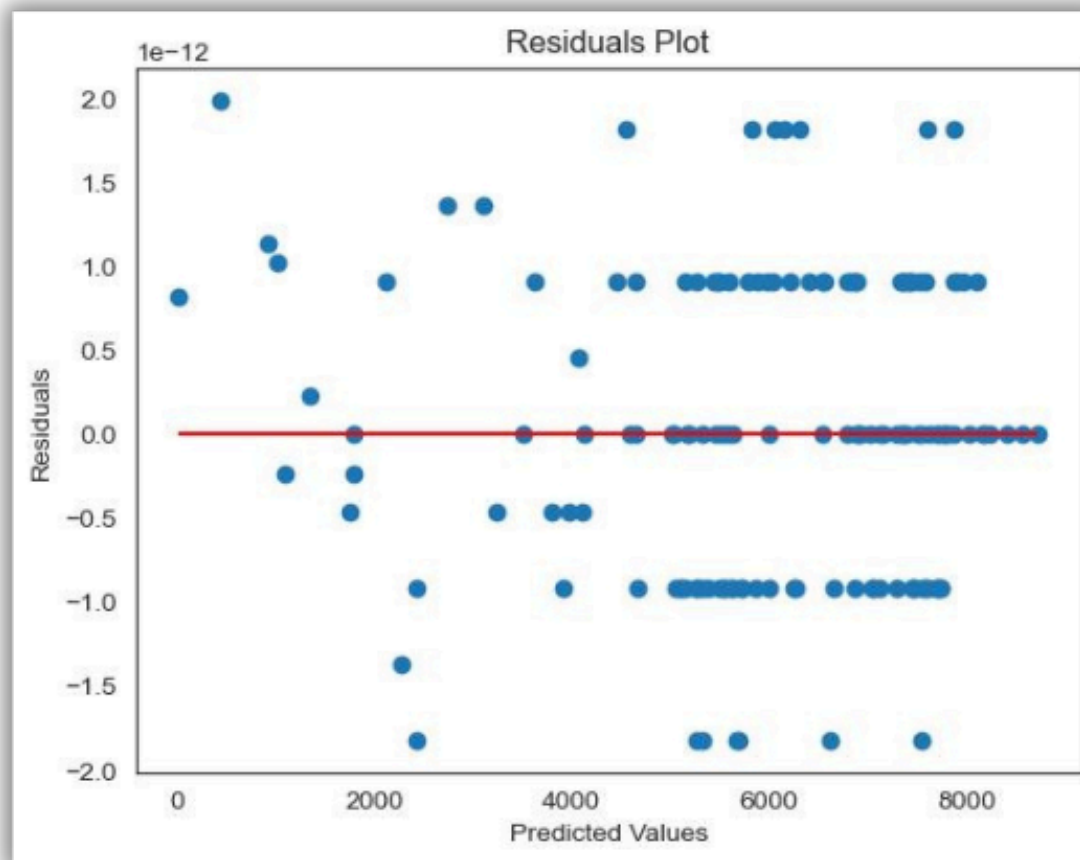
R-squared: 0.9683311835688072

### Gradient Forest:

gbr Mean Squared Error:

44104.11061910983

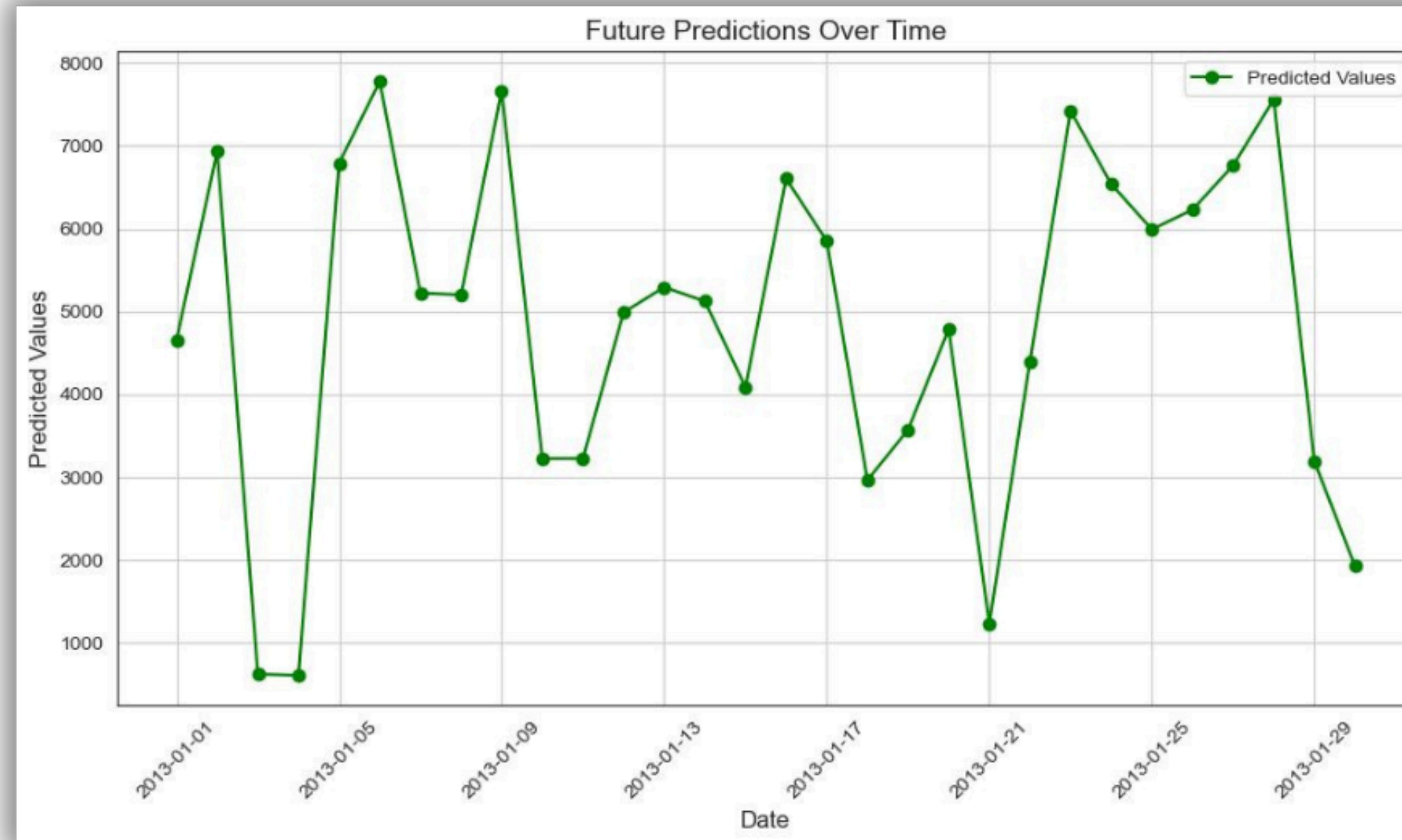
R-squared: 0.9875085205857093



# Model Evaluation & Results

## Results

Prediction fo next 30 days



## Best Model Selection

Linear Regression

# Conclusion & Next Steps

## Conclusion

By using historical data, we successfully predicted future bike rentals. Our best model achieved an RMSE and  $R^2$ .

## Next Steps

01

Further optimization of the model (e.g., hyperparameter tuning).

02

Deploying the model for real-time predictions.

03

Integrating external data like weather forecasts.



# THANK YOU!

