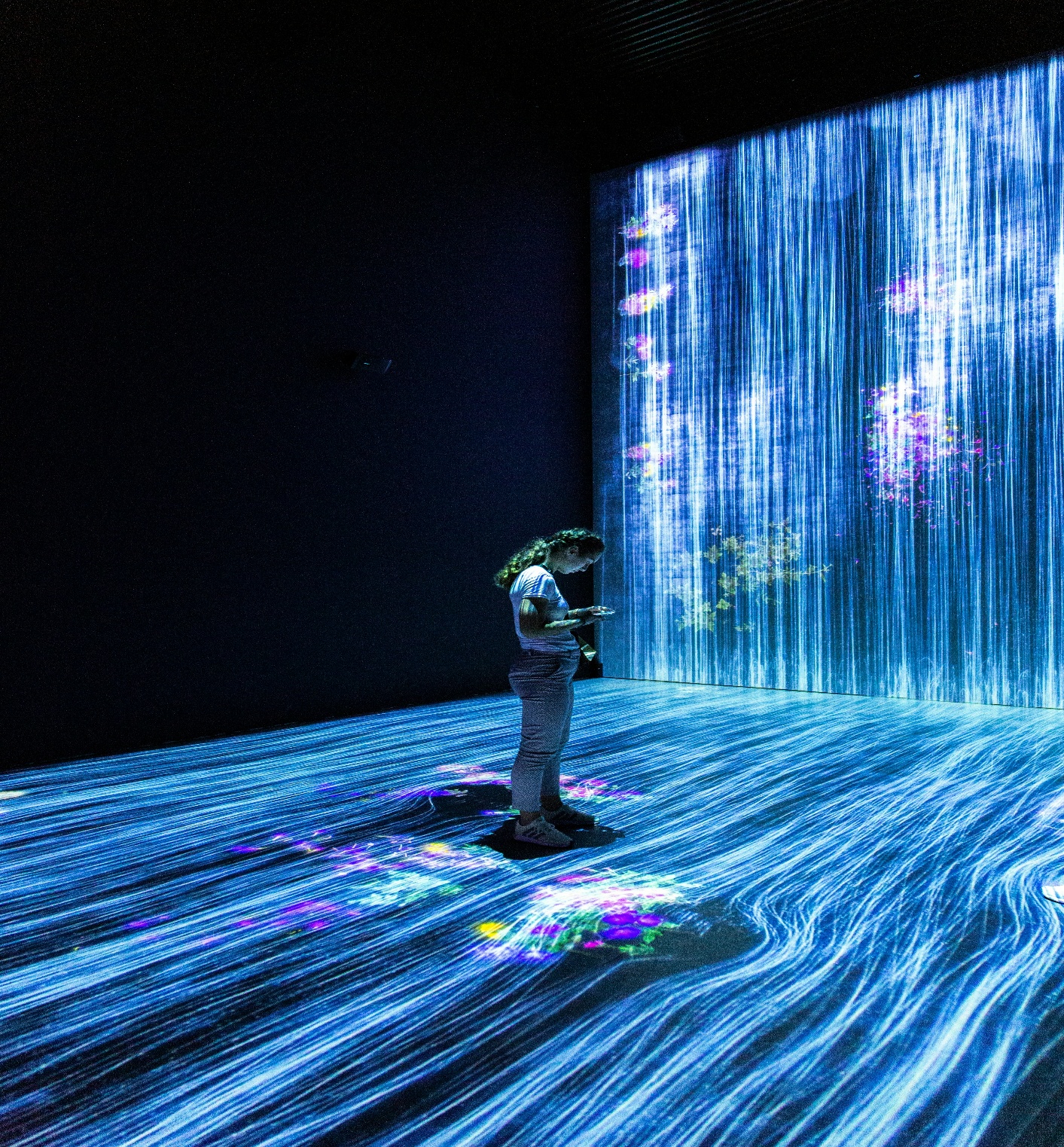
****

**Bike Sharing Demand Prediction Report**

**Prepared by:**  Zaryab Ahmad Khan

**Fellowship Program:**  ML/DL Fellowship

**Project Overview:** This is my monthly project

where I am learning Python basics, advanced

Python, data preprocessing, and model

development.

Contents

[Introduction 2](#_Toc171767753)

[Data Exploration 2](#_Toc171767754)

[ Shape of Dataset: 2](#_Toc171767755)

[ Variable Types: 2](#_Toc171767756)

[ Categorical Variable Analysis: 2](#_Toc171767757)

[Univariate and Bivariate Analysis 3](#_Toc171767758)

[Outlier and Missing Value Analysis 3](#_Toc171767759)

[Feature Engineering 3](#_Toc171767760)

[Model Development 3](#_Toc171767761)

[Hyperparameter Tuning 3](#_Toc171767762)

[Final Model 3](#_Toc171767763)

[Conclusion 3](#_Toc171767764)

# Introduction

The objective of this project is to predict bike-sharing demand using historical data. The dataset contains various features, including seasonal information, temperature, humidity, and wind speed. This analysis employs different machine learning models to determine the best predictors of bike rental counts.

# Data Exploration

The dataset, day.csv, was loaded for analysis, revealing its shape and structure. Key exploratory data analysis (EDA) steps included:

* Shape of Dataset: The dataset contains numerous records and features, giving insights into its complexity.
* Variable Types: Several variables were converted to appropriate data types (e.g., categorical and float) to facilitate analysis.
* Categorical Variable Analysis: Count distributions for categorical variables were visualized using bar plots, revealing patterns in bike usage across seasons and weekdays.

## Univariate and Bivariate Analysis

Descriptive statistics provided a summary of numerical features, while distribution plots visualized the data's spread. Notable findings included:

* Seasonal variations significantly affect bike rental counts.
* The relationship between temperature and demand was observed, suggesting temperature as a crucial feature.

## Outlier and Missing Value Analysis

Outlier detection was performed using box plots, identifying significant outliers in wind speed and humidity variables. These outliers were subsequently removed from the dataset. The analysis confirmed that there were no missing values present.

## Feature Engineering

New features were created through binning:

* Months were categorized into two groups, focusing on demand patterns.
* Weekdays were similarly binned to separate working days from weekends.

## Model Development

Multiple regression models were developed, including:

* Linear Regression
* K-Neighbors Regressor
* Support Vector Regressor
* Decision Tree Regressor
* Random Forest Regressor
* Gradient Boosting Regressor

Each model's performance was evaluated using cross-validation, with R² scores indicating predictive accuracy.

## Hyperparameter Tuning

The Random Forest model and Gradient Boosting model were fine-tuned to optimize their performance. The best parameters were identified through systematic experimentation.

## Final Model

The Gradient Boosting Regressor, utilizing the parameters learning\_rate=0.045, max\_depth=3, n\_estimators=300, and subsample=0.7, was identified as the most effective model.

## Conclusion

The project successfully demonstrated the steps involved in preprocessing data, feature engineering, and modeling using machine learning techniques. The results indicate that various factors influence bike-sharing demand, with the most notable being seasonal variations and temperature. This project has significantly enhanced my understanding of Python and machine learning methodologies.