

# OLA BIKE RIDE REQUEST DEMAND FORCAST PROJECT REPORT

***Submitted in partial fulfilment of the requirements for the award of the degree of***

# BACHELOR OF TECHNOLOGY (INT-254)

**IN**

# COMPUTER SCIENCE AND ENGINEERING

**Under the Guidance of:**

# Dr. Pawan Kumar (Professor)

**Submitted By**

**Hrishikesh Bharti (12014036)**

**Kanishk Rao (12015255)**

**CANDIDATE’S DECLARATION AND CERTIFICATE**

We hereby certify that the work, which is being presented in this report entitled, **OLA BIKE RIDE REQUEST DEMAND FORCAST PROJECT REPORT**, in partial fulfilment of the

requirements for the degree of **Bachelor of Technology, Course Code INT-254** submitted in the **Computer Science and Engineering , Lovely Professional University, Punjab**; by **MD Hrishikesh Bharti (12014036), Kanishk Rao (12015255)** is the authentic record of our own work carried out under the supervision of **Dr. Pawan Kumar Mall, Professor, Computer Science and Engineering, Lovely Professional University, Punjab.**

We further declare that the matter embodied in this report has not been submitted by us for the award of any other degree.

Candidate(s) Signature

This is to certify that the above statement made by the candidate is correct to the best of my knowledge and belief.

Signature of Supervisor **Dr. Pawan Kumar** Date:

# ACKNOWLEDGMENT

It is our pleasure to acknowledge the contributions of all who have helped us and supported us during this Project report.

First, we thank God for helping us in one way or another and providing strength and endurance to us. We wish to express my sincere gratitude and indebtedness to our supervisor Mr. Pawan Kumar Mall, Computer Science and Engineering, Lovely Professional University, Punjab; for his intuitive and meticulous guidance and perpetual inspiration in completion of this report. In spite of his busy schedule, he rendered help whenever needed, giving useful suggestions and holding informal discussions. His invaluable guidance and support throughout this work cannot be written down in few words. We also thank him for providing facilities for my work in the department.

We are also humbly obliged by the support of our group members and friends for their love and caring attitude. The sentimental support they rendered to us is invaluable and everlasting. They have helped us through thick and thin and enabled us to complete the work with joy and vigour. We thank the group members for entrusting in each other and following directions, without them this report would never have been possible.

We are also thankful to our parents, elders and all family members for their blessing, motivation and inspiration throughout our work and bearing with us even during stress and bad temper. They have always provided us with high moral support and contributed in all possible ways in completion of this report.

# TABLE OF CONTENTS

* **CANDIDATE’S DECLARATION AND CERTIFICATE**
* **ACKNOWLEDGMENT**

# INTRODUCTION

# OBJECTIVE

# MACHINE LEARNING

* **OLA BIKE REQUEST DEMAND**
* **PARAMETERS**

# ML ALGORITHMS

# OBSERVATIONS

# CONCLUSION

# REFERENCES

# INTRODUCTION

Machine learning is almost everywhere nowadays. It’s become more and more necessary day by day. From the recommendation of what to buy to recognizing a person, robotics, everywhere is machine learning. So in this project, we’ll

create the “Hello World” of machine learning which means Iris flower classification.

From telling rickshaw-wala where to go, to tell him where to come we have grown up. Yes, we are talking about online cab and bike facility providers like OLA and Uber. If you had used this app some times then you must have paid some day less and someday more for the same journey. But have you ever thought what is the reason behind it? It is because of the high demand at some hours. this is not the only factor but this is one of them.

In this project, we will try to predict ride-request for a particular hour using machine learning. One can refer to the below explanation for the column names in the dataset and their values as well.

What is machine learning?

Machine learning is about learning to predict something or extracting knowledge from data. ML is a part of artificial intelligence. ML algorithms build a model based on sample data or known as training data and based upon the training data the algorithm can predict something on new data.

Machine learning is about learning to predict something or extracting knowledge from data. ML is a part of artificial intelligence. ML algorithms build a model based on sample data or known as training data and based upon the training data the algorithm can predict something on new data.

Categories of Machine Learning :

Supervised machine learning: Supervised machine learning are types of machine learning that are trained on well-labeled training data. Labeled data means the training data is already tagged with the correct output.

Unsupervised machine learning: Unlike supervised learning, unsupervised learning doesn’t have any tagged data. It learned patterns from untagged data. Basically, it creates a group of objects based on the input data/features.

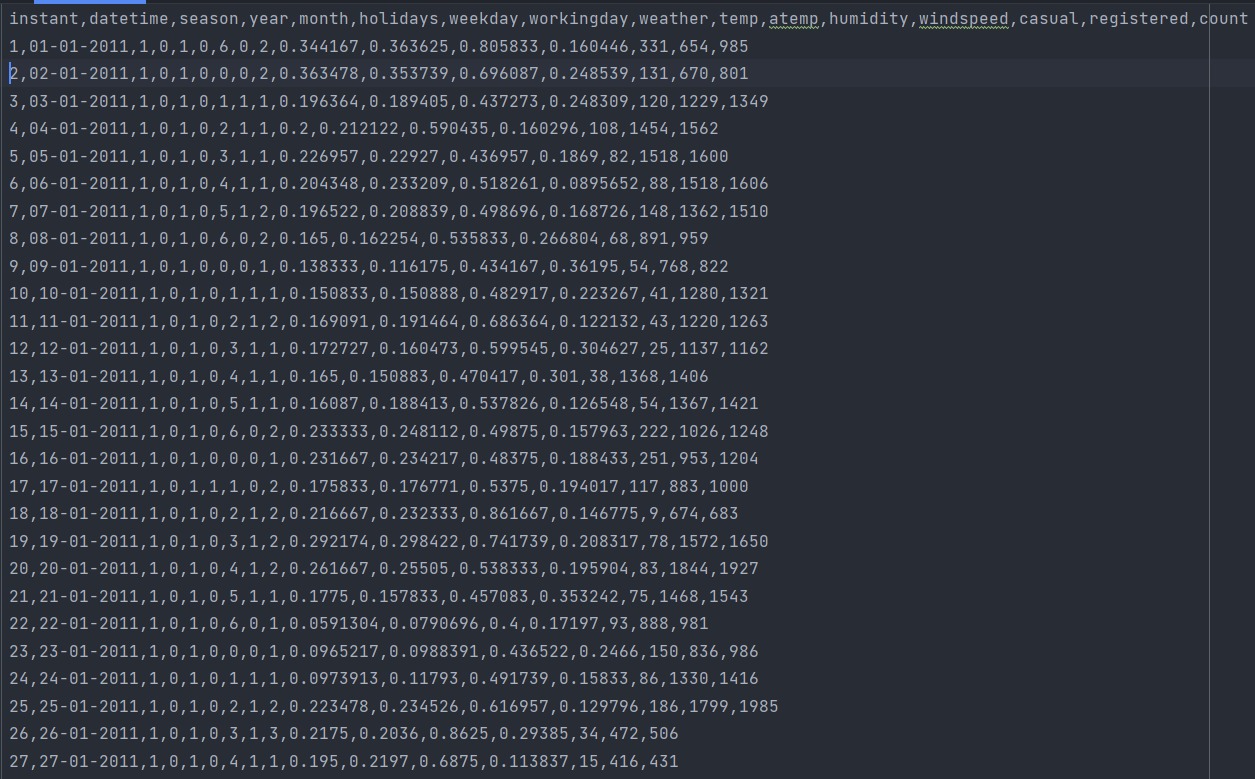
Semi-supervised machine learning:

Semi-supervised learning falls between supervised and unsupervised learning. It has a small amount of tagged data and a large amount of untagged data.

# OBJECTIVE

1. To develop an algorithm to predict ride-request demand.
2. Apply different algorithms to develop an ideal model.
3. we will try to predict ride-request for a particular hour using machine learning. One can refer to the below explanation for the column names in the dataset and their values as well.

**Dataset used for deriving the modules: -**

****

**Importing the Modules: -**

Python libraries make it easy for us to handle the data and perform typical and complex tasks with a single line of code.

Pandas – This library helps to load the data frame in a 2D array format and has multiple functions to perform analysis tasks in one go.

Numpy – Numpy arrays are very fast and can perform large computations in a very short time.

Matplotlib/Seaborn – This library is used to draw visualizations.

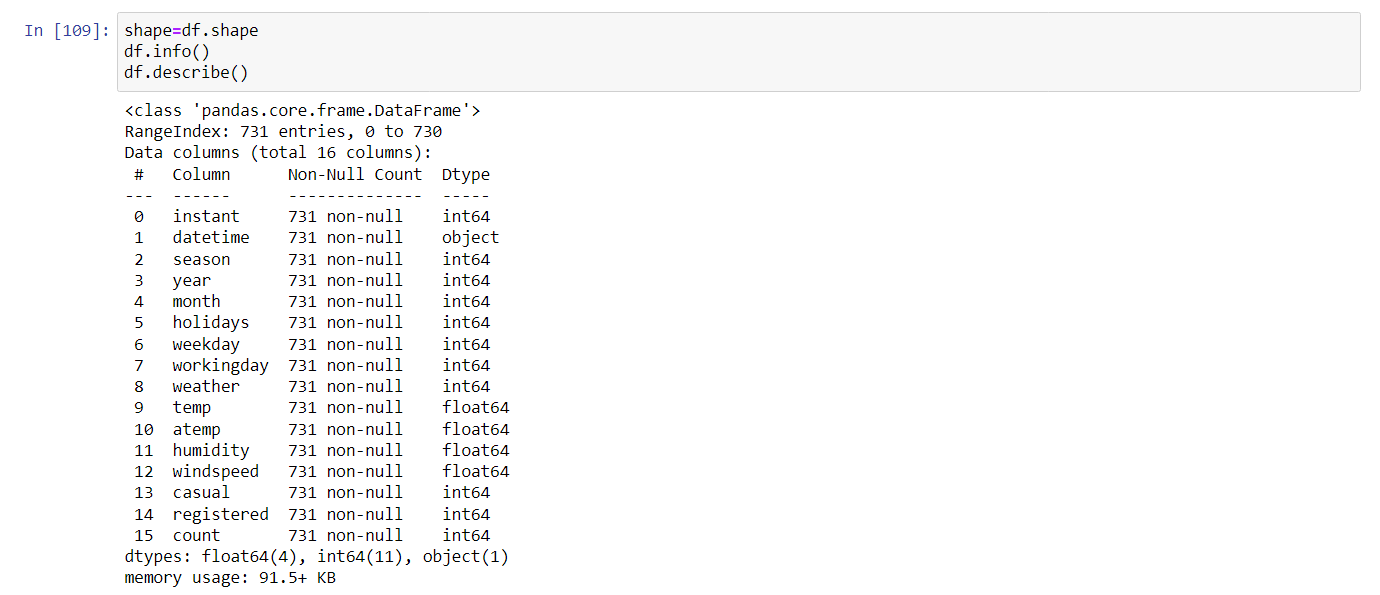
Sklearn – This module contains multiple libraries are having pre-implemented functions to perform tasks from data preprocessing to model development and evaluation.

XGBoost – This contains the eXtreme Gradient Boosting machine learning algorithm which is one of the algorithms which helps us to achieve high accuracy on predictions.

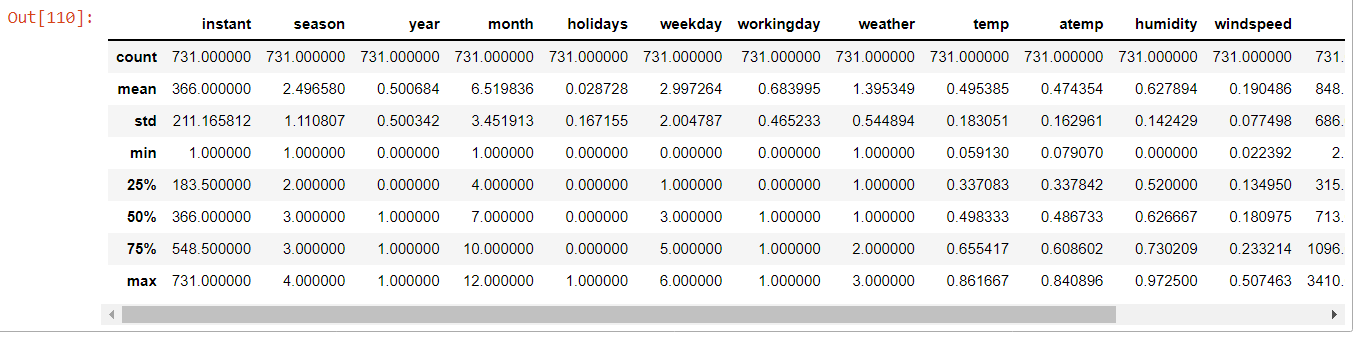
****

**Determining the shape and size of dataset: -**

Let’s check which column of the dataset contains which type of data

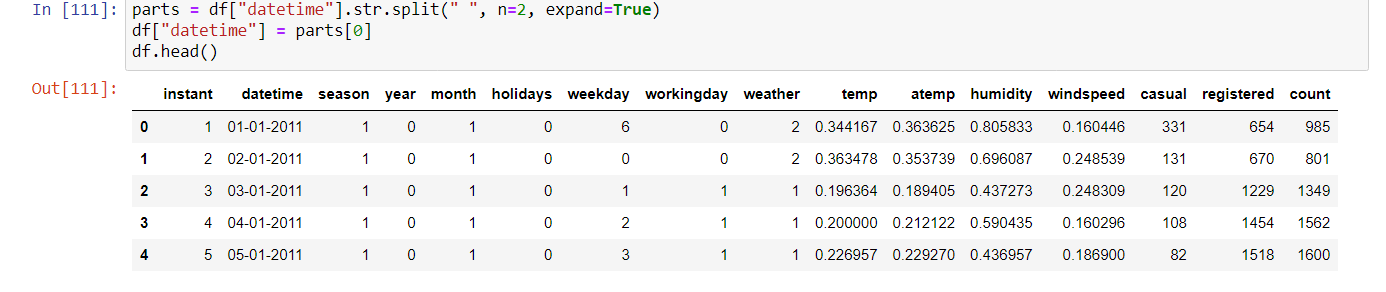
****

As per the above information regarding the data in each column we can observe that there are no null values.

****

## **Feature Engineering**

There are times when multiple features are provided in the same feature, or we have to derive some features from the existing ones. We will also try to include some extra features in our dataset so, that we can derive some interesting insights from the data we have. Also if the features derived are meaningful then they become a deciding factor in increasing the model’s accuracy significantly.



In the above step, we have separated the date and time. Now let’s extract the day, month, and year from the date column A picture containing table

Description automatically generated

Whether it is a weekend, or a weekday must have some effect on the ride request count.

A picture containing table

Description automatically generated

It would be nice to have a column which can indicate whether there was any holiday on a particular day or not.

A picture containing text

Description automatically generated

Now let’s remove the columns which are not useful for us

Graphical user interface, text, application

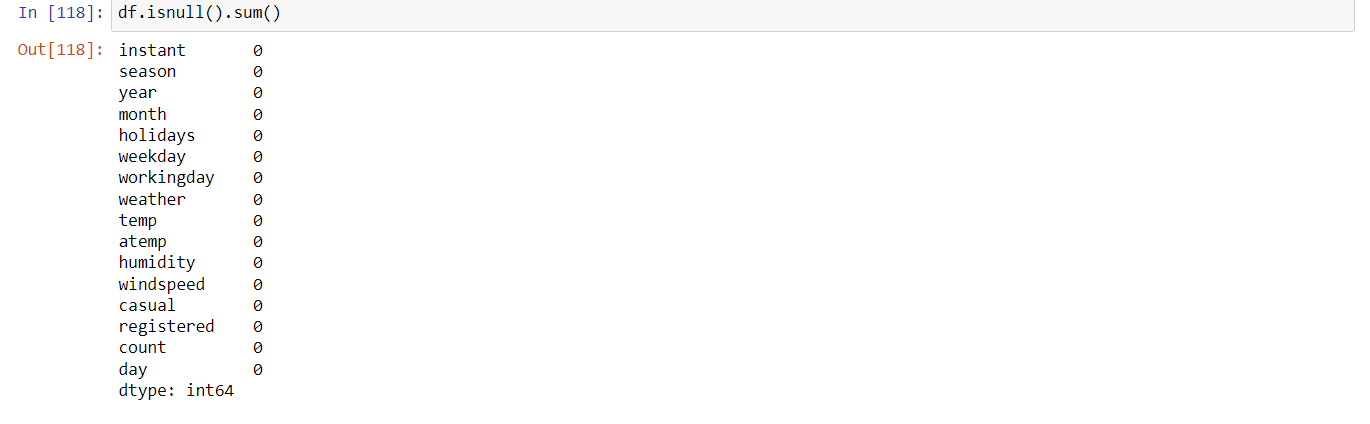
Description automatically generated

There may be some other relevant features as well which can be added to this dataset but let’s try to build a build with these ones and try to extract some insights as well.

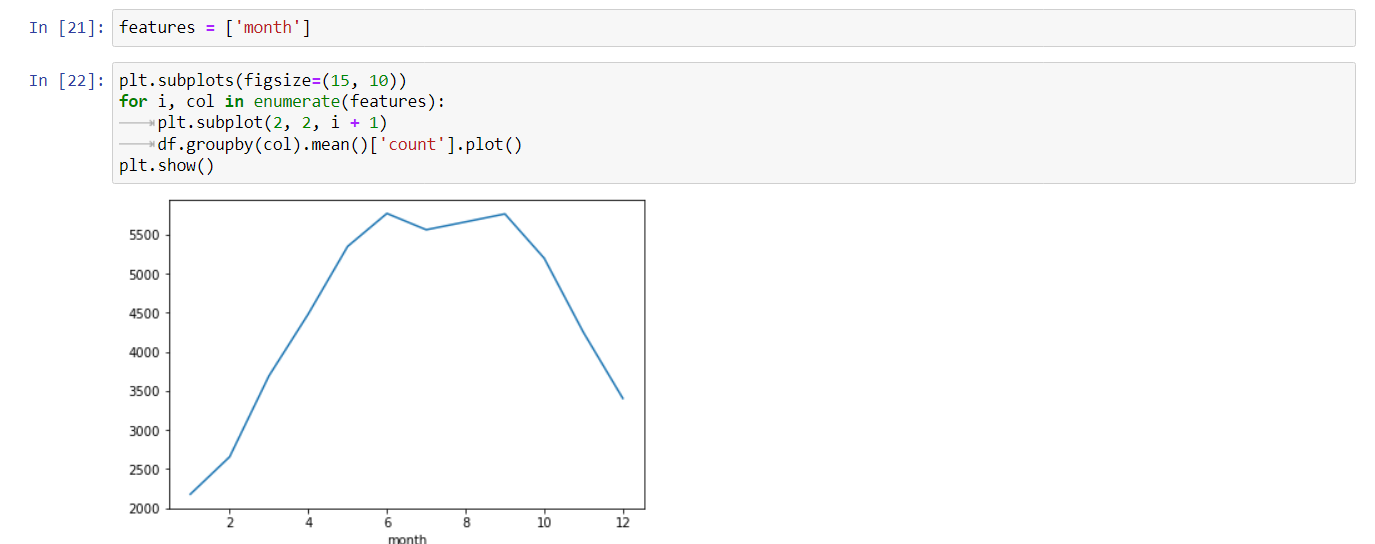
**Exploratory Data Analysis**

EDA is an approach to analyzing the data using visual techniques. It is used to discover trends, and patterns, or to check assumptions with the help of statistical summaries and graphical representations.

We have added some features to our dataset using some assumptions. Now let’s check what are the relations between different features with the target feature.



Now, we will check for any relation between the ride request count with respect to the month.



From the above [line plots](https://www.geeksforgeeks.org/line-chart-in-matplotlib-python/) we can confirm some real-life observations:

* The average ride request count has dropped in the month of festivals that is after the 7th month that is July that is due to more holidays in these months.

|  |
| --- |
| From the above bar plots, we can confirm some real-life observations:   * Ride request demand is high in the summer as well as season. * The third category was extreme weather conditions due to this people avoid taking bike rides and like to stay safe at home. * On holidays no college or offices are open due to this ride request demand is low. * More ride requests during working hours as compared to non-working hours. * Bike ride requests have increased significantly from the year 2011 to the year 2012.     Temperature values are normally distributed but due to the high number of 0 entries in the windspeed column, the data distribution shows some irregularities. |
| Ah! [outliers](https://www.geeksforgeeks.org/machine-learning-outlier/)let’s check how much data we will lose if we remove outliers.    We can remove this many rows because we have around 0 rows of data so, this much data loss won’t affect the learning for our model.      Now let’s check whether there are any highly correlated features in our dataset or not.    Here the registered feature is highly correlated with our target variable which is count. This will lead to a situation of data leakage if we do not handle this situation. So, let’s remove this ‘registered’ column from our feature set.  Now, we have to remove the outliers we found in the above two observations that are for the humidity and wind speed.   **Model Training** Now we will separate the features and target variables and split them into training and the testing data by using which we will select the model which is performing best on the [validation data](https://www.geeksforgeeks.org/training-vs-testing-vs-validation-sets/).    [Normalizing](https://www.geeksforgeeks.org/normalization-vs-standardization/)the data before feeding it into machine learning models helps us to achieve stable and fast training.  Now we have to , We have split our data into training and validation data also the normalization of the data has been done. Now let’s train some state-of-the-art machine learning models and select the best out of them using the validation dataset.      Now we predict the values of test set.      Here we check that how accurate our data is:    The predictions made by the XGBRegressor are really amazing compared to the other model. In the case of XGBRegressor, there is a little bit of overfitting but we can manage it by hyperparameter tuning. |

# REFERENCES

# GOOGLE.COM

* **TUTORIALSPOINT.COM**

# GEEKFORGEEKS.COM

# WIKIPEDIA

# YOUTUBE