**Capstone Project Submission**

**Instructions:**

i) Please fill in all the required information.

ii) Avoid grammatical errors.

|  |
| --- |
| **Team Member’s Name, Email and Contribution:** |
| **Team member name:**Chetan Patil  **Email:** chetanpatil4160@gmail.com  **Team contributions**   1. Find the question 2. How many bikes rented by season wise? 3. Which season has maximum number of bikes rented? 4. How many holidays in dataset? 5. Did some work on data visualization 6. Plot Regression charts 7. Did some work on data cleaningand changed the column names? 8. Add some new columns (weekdays and weekends) 9. Break the date column into (day month and year) 10. Find solutions for questions 11. Worked on conclusion 12. Linear Regression ML model 13. Lasso Regression ML model 14. Plot Distribution of data   **Team member name:** Rajesh Patil  **Email:**rajesh.patil775607@gmail.com  **Team contribution:**   1. Question forming 2. How long customer would like to rent customers? 3. which month has maximum and minimum number ofBike rented? 4. Are people use rented bikes on Functioning Day? 5. data cleaning 6. data visualization by using line chart 7. Rented bike vs integer features 8. technical document preparation 9. Outlier detection and removal 10. Helped to make Presentation ppt 11. Decision Tree 12. Random Forest 13. conclusion   **Team member name:**Sachin Chaudhari  **Email:sachinchaudhari4141@gmail.com**  **Team contributions**   1. Correlation between data 2. With the group help prepare presentation 3. Worked on Presentation preparation and its interpretation. 4. Normalization of data 5. Worked on OLS model 6. One Hot Coding 7. Ridge Regression ML model 8. Elastic Net Regression ML model   **Team Member’s Name: -**Mrunal Badgujar  **Email: -mrunalbadgujar18@gmail.com**  **Contribution: -**   1. Imported all the libraries for data exploration, Sorting, Cleaning and Visualization. 2. Imported and mounted data set require for analysis from google drive to google co lab. 3. Exploring data set like number of columns and row with heading and what is data type of each value using python libraries methods like Info (), shape, describe (), Head (), Tail (). 4. Checked Null Value and enriched/dropped/replace with appropriate value using basic business knowledge and requirement with help of mathematics rules. 5. With all the group member brainstorming finalize some observation and their visualization with interpretation. 6. With the group help prepare presentation. 7. Gradient Boosting ML algorithm 8. Conclusion |
| **Please paste the GitHub Repo link.** |
| GitHubLink: - https://github.com/chetanpatil4160/Bike-Shearing |
| **Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)** |
| The Bike Sharing System is an emerging mode of transport in the world and most of the developing countries are on the path of following the western model of Bike Sharing Systems. In India, some entrepreneurs have tried to setup a bike share system and have failed in the past as they have failed to use data analytics properly. There is a possibility that bike stations can be full or empty when a traveller comes to the station. Thus, predicting the use of such predictions can be helpful for users to plan their travels and also for entrepreneurs to set up the system properly. This paper presents different ways to predict the number of bikes that can be rented in such a system. For case study purposes, we have used a public data set. The predictions are made for every hour of the day.    As the first step, performing data wrangling over the raw data. Furthermore, we have analysed data by EDA and took the help of data visualization to analyse data. The main objective of this project is to find them. Currently, rental bikes are introduced in many urban cities for the enhancement of mobility and to be comfortable. It is important to make the rental bike available and accessible to the public at the right time as it lessens the waiting time. Eventually, providing the city with a stable supply of rental bikes has become a major concern. The crucial part is the prediction of the bike count required at each hour for the stable supply of rental bikes.  While completing this project, we mainly focused on the number of rented bikes and its dependent features, which gives the highest number of bookings, such as holiday and temperature, season etc    In the season rent analysis, we came to know many customers prefer renting bikes in summer. Many visitors prefer renting a bike on non-holidays rather than on a holiday. Furthermore, we analysed data Month wise Bike booking, so June is the month that has a large number of bookings, followed by July and May. As we came to know from one analysis, many customers would like to rent bikes for 18 hrs. followed by 17hrs and 19hrs. One of our analyses shows people like to use bikes on a functioning day. As we correlated all the features, we came to know the dew point temperature had the same values as the temperature due to the multicollinearity we had dropped.  After we checked, our data was not distributed normally, so we found some outliers in our data and, after that, we removed detected outliers and plotted the data with updated values. After training the data and testing the data, we applied some ML Algorithms like Linear, Lasso, Ridge, Random Forest etc. and after applying this model, we came to the best fit model is Random Forest Regressor and Gradient Boosting gridsearchcv gives the highest R2 score of 99% and 95% receptively for Train Set and 92% for Test set. We can deploy this model. |