**Capstone Project Submission**

**Instructions:**

i) Please fill in all the required information.

ii) Avoid grammatical errors.

|  |
| --- |
| **Team Member’s Name, Email and Contribution:** |
| • **Abhash Jain (**abhashjain9@gmail.com)  **• Data Manipulation**   * + - * Handling null and duplicate values       * Checking for Correlation - With the help of heatmap.       * Dropping some columns - Due to high correlation       * Removing Outliers   **• EDA-Analysis of Categorical Variables**   * + - * Analysis of Rented Bike Count with respect to Month       * Analysis of Rented Bike Count with respect to Week       * Analysis of Rented Bike Count with respect to Hour       * Analysis of Rented Bike Count with respect to Functioning day       * Analysis of Rented Bike Count with respect to Holiday       * Analysis of Rented Bike Count with respect to Seasons   **• Analysis of Numerical Variables**   * + - * Density distribution among the columns like Rented Bike Count, Temperature, Humidity, Wind Speed, Visibility, Solar Radiation, rainfall and Snowfall.       * Count of bikes rented in different temperature.       * Count of bikes rented in different visibility ranges.   **• One Hot Encoding**   * + - * The main aim of One Hot encoding is to produce binary integers of 0 and 1 to encode our categorical features.   **• Regression Plot**   * + - * Regression analysis is a set of statistical processes for estimating the relationships between a dependent variable and one or more independent variable.   **• Machine Learning Algorithms(Regression)**   * + - * Linear Regression Model(L1 and L2 Regularisation)       * Random Forest Regressor with GridSearchCV       * Gradient Boosting Regressor with GridSearchCV   • Other than this I prepared summary, PPT report and Technical document for this study.   * **Mounika Dontula (**[**mounikadontula2795@gmail.com**](mailto:mounikadontula2795@gmail.com)**)**   **• Data Manipulation**   * + - * Handling null and duplicate values       * Checking for Correlation - With the help of heatmap.       * Dropping some columns - Due to high correlation       * Removing Outliers   **• EDA-Analysis of Categorical Variables**   * + - * Analysis of Rented Bike Count with respect to Month       * Analysis of Rented Bike Count with respect to Week       * Analysis of Rented Bike Count with respect to Hour       * Analysis of Rented Bike Count with respect to Functioning day       * Analysis of Rented Bike Count with respect to Holiday       * Analysis of Rented Bike Count with respect to Seasons   **• Analysis of Numerical Variables**   * + - * Density distribution among the columns like Rented Bike Count, Temperature, Humidity, Wind Speed, Visibility, Solar Radiation, rainfall and Snowfall.       * Count of bikes rented in different temperature.       * Count of bikes rented in different visibility ranges.   **• One Hot Encoding**   * + - * The main aim of One Hot encoding is to produce binary integers of 0 and 1 to encode our categorical features.   **• Regression Plot**   * + - * Regression analysis is a set of statistical processes for estimating the relationships between a dependent variable and one or more independent variable.   **• Machine Learning Algorithms(Regression)**   * + - * Linear Regression Model(L1 and L2 Regularisation)       * Linear regression with elastic net       * Decision tree with grid search cv       * Gradient Boosting Regressor with GridSearchCV   • Other than this I prepared PPT report and Technical document for this study. |
| **Please paste the GitHub Repo link.** |
| Github Link:-  <https://github.com/abhashjain9/Bike-Sharing-Demand-Prediction>  https://github.com/MOUNIKADONTULA/Bike-Sharing-Demand-Prediction-.git |
| **Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)** |
| We first decided to take up this project solely due to our mutual interest in Seoul Bike Sharing Demand Prediction. Then we decided first work individually gaining insights doing some eda etc. I started to form the questions and we discussed with team member. After doing some random EDAs. I gained some confidence.  **Short summary:**   * Hour of the day holds most importance among all the features for prediction of dataset. * It is observed that highest number bike rental counts in Autumn/fall Summer Seasons and the lowest in Spring season. * We observed that the highest number of bike rentals on a clear day and the lowest on a snowy or rainy day. * As we can see the top 5 important features of our dataset are: Season\_winter, Temperature, Hour, Season\_autumn and Humidity. * Peoples do not use rented bikes in no functioning day. * People tend to rent bikes when the temperature is between -5 to 25 degrees. * People tend to rent bikes when the visibility is between 300 to 1700. * Linear Regression, Lasso and Ridge are not at its best. * The above experiments we can conclude that gradient boosting and random forest regressor with using hyperparameters we got the best results |