Capstone Project Submission

Team Member's Name, Email and Contribution:

Name: Anas Malik
Email: anasmalik081@gmail.com
Project Contribution:
    ● Contributed in Notebook, connect drive to notebook.
    ● Contributed to presentation.
    ● Data Inspection
    ● Data Wrangling
    ● Exploratory Data Analysis
    ● Model Implementation
    ● Prepared Presentation
    ● Prepared Technical Documentation

Please paste the GitHub Repo link.
https://github.com/anasmalik081/Bike-Sharing-Demand-Prediction-Capstone-Project.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)
Seoul is the capital of South Korea. I have a one year data of renting bikes in Seoul. In urban cities, rental bikes have been implemented and the use of rental bikes is increasing in our day to day life. The most challenging part for any rental company is to make rental bikes available to the public at the right time and right place. In this project our main aim is to solve this problem and predict the number of bikes demands across different hours and places. The dataset contains 8760 rows and 14 columns. The columns are Date, Rented Bike Count, Hour, Temperature, Humidity, Wind Speed, Visibility,

Dew Point Temperature, Solar Radiation, Rainfall, Snowfall, Seasons, Holiday, Functioning Day.

My approaches are as follows:
1. Understand the problem statement and dataset.
2. Data Inspection
3. Data Wrangling
4. Exploratory Data Analysis
5. Data Pre-processing
6. Model Implementation

Conclusion:
- EDA Conclusion:
    1. More number of bikes are rented in summer season.
    2. Chances of renting a bike increases when the temperature is between 15 to 30 degree celsius.
    3. When humidity in the air is between 30-70 percent, renting on bikes increases.
    4. People prefer to rent a bike on a working day rather than on holiday.
    5. More bikes are rented during daytime around 7-8AM in morning and 6-7PM in evening.

- ML Model Conclusion:
    1. Random Forest Algorithm gave best performance with an accuracy score of 98 % on training dataset and 91 % on test dataset.
    2. Linear and Ridge regression shows 77 % accuracy on training and 78 % accuracy on testing.
    3. The Lasso and ElasticNet Regression shows 76 % accuracy on training and 78 % accuracy on testing.