Capstone Project Submission

Team Member's Name, Email and Contribution:

1. **Mrityunjay Singh Chandel(mrityunjaychandel98@gmail.com)**
    - Reading the Dataset
        - Comprehending each column in the dataset
        - Raising the bar of questions from the dataset
        - Gaining worthful inference from the dataset for analytical conclusions.
    - Exploratory Data Analysis
        - Statistical Based Analysis
        - Using the missing no to identify the null value via matrix visualization.
        - Checking Missing Values.
        - Calculating VIF.
    - Data Visualization
        - Plotting the insights in the graph using: Distribution Plot, Histogram, Regression Plot, Bar Plot, Strip Plot, Heatmaps, etc.
    - Training the Model and Calculating the Score
        - Training the model to create the list of matrix.
        - Calculating MSE, RMSE, R2_Score, Adjusted_r2
    - Train-Test Split Using Different Models
        - Linear Regression
        - Lasso Regression
        - Ridge Regression
        - Elastic Net Regression
        - Decision Tree Regression
        - Gradient Boosting
        - Extreme Gradient Boosting
    - Conclusion

2. **Nimisha Jadhav (jnimisha21@gmail.com)**
    - Reading the Dataset
        - Comprehending each column in the dataset
        - Raising the bar of questions from the dataset
        - Gaining worthful inference from the dataset for analytical conclusions.
    - Exploratory Data Analysis
        - Statistical Based Analysis
        - Using the missing no to identify the null value via matrix visualization.
        - Checking Missing Values.
        - Calculating VIF.
    - Data Visualization
        - Plotting the insights in the graph using:

Distribution Plot, Histogram, Regression Plot, Bar Plot, Strip Plot, Heatmaps, etc.

- Training the Model and Calculating the Score
  - Training the model to create the list of matrix.
  - Calculating MSE, RMSE, R2_Score, Adjusted_r2
- Train-Test Split Using Different Models
  - Linear Regression
  - Lasso Regression
  - Ridge Regression
  - Elastic Net Regression
  - Decision Tree Regression
  - Gradient Boosting
  - Extreme Gradient Boosting
- Conclusion

3. **Vishal Chakrabarty (vishalchakrabarty20@gmail.com)**
   - Reading the Dataset
     - Comprehending each column in the dataset
     - Raising the bar of questions from the dataset
     - Gaining worthful inference from the dataset for analytical conclusions.
   - Exploratory Data Analysis
     - Statistical Based Analysis
     - Using the missing no to identify the null value via matrix visualization.
     - Checking Missing Values.
     - Calculating VIF.
   - Data Visualization
     - Plotting the insights in the graph using: Distribution Plot, Histogram, Regression Plot, Bar Plot, Strip Plot, Heatmaps, etc.
   - Training the Model and Calculating the Score
     - Training the model to create the list of matrix.
     - Calculating MSE, RMSE, R2_Score, Adjusted_r2
   - Train-Test Split Using Different Models
     - Linear Regression
     - Lasso Regression
     - Ridge Regression
     - Elastic Net Regression
     - Decision Tree Regression
     - Gradient Boosting
     - Extreme Gradient Boosting
   - Conclusion

4. **Sagar Tikmani (sagartikmani900@gmail.com)**

   - Reading the Dataset
     - Comprehending each column in the dataset
     - Raising the bar of questions from the dataset
     - Gaining worthful inference from the dataset for analytical conclusions.

- Exploratory Data Analysis
  - Statistical Based Analysis
  - Using the missing no to identify the null value via matrix visualization.
  - Checking Missing Values.
  - Calculating VIF.
- Data Visualization
  - Plotting the insights in the graph using: Distribution Plot, Histogram, Regression Plot, Bar Plot, Strip Plot, Heatmaps, etc.
- Training the Model and Calculating the Score
  - Training the model to create the list of matrix.
  - Calculating MSE, RMSE, R2_Score, Adjusted_r2
- Train-Test Split Using Different Models
  - Linear Regression
  - Lasso Regression
  - Ridge Regression
  - Elastic Net Regression
  - Decision Tree Regression
  - Gradient Boosting
  - Extreme Gradient Boosting
- Conclusion

5. **Divyesh Dhanani (divyeshdhanani143@gmail.com)**
   - Reading the Dataset
     - Comprehending each column in the dataset
     - Raising the bar of questions from the dataset
     - Gaining worthful inference from the dataset for analytical conclusions.
   - Exploratory Data Analysis
     - Statistical Based Analysis
     - Using the missing no to identify the null value via matrix visualization.
     - Checking Missing Values.
     - Calculating VIF.
   - Data Visualization
     - Plotting the insights in the graph using: Distribution Plot, Histogram, Regression Plot, Bar Plot, Strip Plot, Heatmaps, etc.
   - Training the Model and Calculating the Score
     - Training the model to create the list of matrix.
     - Calculating MSE, RMSE, R2_Score, Adjusted_r2
   - Train-Test Split Using Different Models
     - Linear Regression
     - Lasso Regression
     - Ridge Regression

- Elastic Net Regression
- Decision Tree Regression
- Gradient Boosting
- Extreme Gradient Boosting

- Conclusion

Please paste the GitHub Repo link.

GithubLink:-https://github.com/Link/to/Repo

Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)

Research says!

Several bike/scooter rides sharing facilities have started up lately especially in metropolitan cities like San Francisco, New York, Chicago and Los Angeles, and one of the most important problem from a business point of view is to predict the bike demand on any particular day. While having excess bikes results in wastage of resource (both with respect to bike maintenance and the land/bike stand required for parking and security), having fewer bikes leads to revenue loss (ranging from a short-term loss due to missing out on immediate customers to potential longer-term loss due to loss in future customer base), Thus, having an estimate on the demands would enable efficient functioning of these companies.

Bike sharing systems are a means of renting bicycles where the process of obtaining membership, rental, and bike return is automated via a network of kiosk locations throughout a city. Using these Bike Sharing systems, people rent a bike from one location and return it to a different or same place on need basis. People can rent a bike through membership (mostly regular users) or on demand basis (mostly casual users). This process is controlled by a network of automated kiosk across the city.

Here, we forecasted bike rental demand of Bike sharing program in Seoul based on historical usage patterns using supervised ML Regression.

We thoroughly understood the relationship between the variables leading us to focus on generating the hypothesis followed by Data Exploration part.

This dataset contains approximately 8760 observations with 14 columns and a combination of category and numerical variables.

Performing Exploratory Data Analysis based on Statistical, Market Based, Univariate, Bivariate and Categorical analysis.

As said a data can be only converted into an information when molded from a clustered form to a stack structured form

The Procedure counts in Steps of Importing the dataset and cleaning the data viewing through the lens of statistical analysis and revising the dataset for further analysis and gaining the inference to move further on Dataset Modelling.

Splitting the test | train data Dataset splits into these training, validation and test dataset, we'll take the test data for analyzing
the performance of training data set and apply modelling to it.

We created machine learning models to predict for casual and registered users separately and then combine them to generate the overall prediction for the counts.

Hyperparameter optimization or tuning is the problem of choosing a set of optimal hyperparameters for a learning algorithm. By contrast, the values of other parameters (typically node weights) are learned. To test different hyperparameter configurations we used GridSearchCV for this configuration test.

The Models we tested are listed as below:

1. Linear Regression
2. Lasso Regression
3. Ridge Regression
4. Elastic Net Regression
5. Random Forest
6. Gradient Boosting
7. Extreme Gradient Boosting

Concluding with shortlisting 3 models which worked precisely in this scenario were: Random Forest, Gradient Boosting and Extreme Gradient Boosting with 92%, 93% and 95% accuracy rate and deployable.