Capstone Project Submission

Team Member's Name, Email and Contribution:
1. Ayush Goyal   erayushgoyal96@gamil.com   2. M Sameer Ahamed   sameerm8095@gmail.com   3. Nitesh bhowmick   nitesh.gnit@gmail.com
Please paste the GitHub Repo link.
https://github.com/Ayugoyal/Retail-sales-prediction**-**
Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)
Retail sales prediction is done by 3 group members M Sameer Ahamed, Ayush Goyal, Nitesh bhowmick . In this project we got   Retail sales prediction as csv file.   **As we downloaded the data as csv file from alma better capstone project dashboard we encoded the file in colab notebook through mounting the drive we have no idea about the project because this was our first project .all the member from the group participated throughout the project with great efforts.**   **The data was huge and some of the column were not needed so we drop it from the data.**   **The cleaning data was done and created the new cleaned data frame consists of the column were compared to gain the knowledge for the prediction. Worked individually gaining same insights doing some EDA .**   • **In the Rossmann sales prediction project there is a dataset which contains sales information.**   • **The shape of Rossmann dataset = 1017209, 8 and Store dataset = 1115, 10.**   • **In both dataset 'Store' column is common. So we do inner join on the basis of column 'Store'.**   • **This dataset is a live dataset of Rossmann Stores. On analyzing this problem we observe that Rossmann problem is a regression problem and our primarily goal is to predict the sales figures of Rossmann problem. In this Notebook we work on following topics Analyzing the dataset by using Exploratory Data Analysis using exponential moving averages analysis trends and seasonality in Rossmann dataset Analyze Regression using following prediction analysis. A) Linear Regression Analysis B)Elastic Regression (Lasso & Ridge Regression). C) Decision tree and D) Random Forest Regression**

- **We trained a model using various algorithms accuracy near about 92%.**
- **The removing sales =0 rows actually removes lot of information from dataset as it has 172817rows which is quite large and therefore we didinot remove those values.**

From all the above experiment we can conclude that Random Forest Regression got the best Result.

Accuracy Score is 92%

Contributors Roles:
1. **Ayush Goyal:**
    1. Data Wrangling:
        1. work on data handing
    2. Visualizing based on sales
3. Visualizing based on sales on dependent variable
    4. Deploy & Run Linear Regression Model
    5. Run & Deploy Lasso Regression
2. **M Sameer Ahamed:**
1. Data Wrangling:
    1. work on merging the DataFrame

2. Visualizing based on store types
3. based on sales between assortment and store types
    4. Scaling the data, Train & Test the data
    5. Run & Deploy Decision Tree Regression model

3. **Nitesh Bhowmick:**
1. Data Wrangling:
    1. work on changing different dtypes into int types

2. Visualizing based on state holiday and school holiday
3. Visualizing based on day of week and open promo
    4. Run & Deploy Random Forest Regression Model
    5. . Run & Deploy Ridge Regression