**Capstone Project Submission**

**Instructions:**

i) Please fill in all the required information.

ii) Avoid grammatical errors.

| **Team Member’s Name, Email and Contribution:** |
| --- |
| 1. Ayush Goyal [erayushgoyal96@gamil.com](mailto:erayushgoyal96@gamil.com)  2. M Sameer Ahamed [sameerm8095@gmail.com](mailto:sameerm8095@gmail.com)  3. Nitesh bhowmick [nitesh.gnit@gmail.com](mailto:nitesh.gnit@gmail.com) |
| **Please paste the GitHub Repo link.** |
| https://github.com/ [Ayugoyal](https://github.com/Ayugoyal)/[Retail-sales-prediction-](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 sales column contain172817rows with 0 sale. So we created a new data frame in which we removed 0 sales rows and tried train our model we used various algorithms and got accuracy score around 74%** * **The total dataset sale =0 rows . So we trained and another model using various algorithms accuracy near about 92%which is far better than previous model.** * **The removing sales =0 rows actually removes lot of information from dataset as it has 172817rows which is quite large** * **\*\*Rossmann Sales Dataset\*\* - This dataset is a live dataset of Rossmann Stores. On analysing 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 Analysing the dataset by using Exploratory Data Analysis using exponential moving averages analyse trends and seasonality in Rossmann dataset Analyse Regression using following prediction analysis. A) Linear Regression Analysis B)Elastic Regression (Lasso & Ridge Regression). C) Dession tree**   **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  **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  **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 |