**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.** |
| **GitHub Link:-**  https://github.com/Nitesh7179/Capstone-Project--Retail-Sales-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)** |
| **Retail sales prediction is done by 3 group members M Sameer Ahamed, Ayush Goyal, and Nitesh Bhowmick. In this project, we got Retail sales prediction as a CSV file.**  **As we downloaded the data as a CSV file from the alma better capstone project dashboard we encoded the file in the colab notebook by mounting the drive & all the members of the group participated throughout the project with great efforts.**  **The data was huge and some of the columns were not needed so we drop them from the data.**  **The cleaning data was done and created the new cleaned data frame consisting of the column were compared to gain the knowledge for the prediction. Worked individually gaining the same insights by doing some EDA.**   * **In the Rossmann sales prediction project there is a dataset which contains sales information** * **The sales column contain172817rows with 0 sales. So we created a new data frame in which we removed 0 sales rows and tried to train our model we used various algorithms and got an accuracy score of around 74%** * **The total dataset sale =0 rows. So we trained another model using various algorithms with accuracy near about 92%which is far better than the previous model.** * **The removing sales =0 rows actually removes a lot of information from the dataset as it has 172817rows which is quite large** * **\*\*Rossmann Sales Dataset\*\* - This dataset is a live dataset of Rossmann Stores. In analysing this problem we observe that the Rossmann problem is a regression problem and our primary goal is to predict the sales figures of the Rossmann problem. In this Notebook, we work on the following topics Analysing the dataset by using Exploratory Data Analysis using exponential moving averages analysing trends and seasonality in the Rossmann dataset Analyse Regression using the following prediction analysis. A) Linear Regression Analysis B)Elastic Regression (Lasso & Ridge Regression). C) Decision tree**   **Contributors Roles:**   1. **Ayush Goyal:**   1. Data Wrangling:  1. work on data handling  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 holidays and school holidays  3. Visualizing based on the day of the week and open promo |