**Capstone Project Submission**

**Instructions:**

i) Please fill in all the required information.

ii) Avoid grammatical errors.

|  |
| --- |
| **Name, Email, and Contribution:** |
| Name: Mayank Ghai  Email: [mayankghai1195@gmail.com](mailto:mayankghai1195@gmail.com)  Contribution:   1. Business Problem 2. Data Collection and Pre-processing 3. Data Manipulation 4. Exploratory Data Analysis 5. EDA Conclusion and Validating Hypotheses 6. Modelling 7. Model Performance and Evaluation 8. Store-wise Sales Predictions 9. Conclusion and Recommendations |
| **Please paste the GitHub Repo link.** |
| Github Link:- https://github.com/Mayank0195/Capstone\_Project\_ML\_Regression\_Retail\_Sales\_Prediction  Drive link:- https://drive.google.com/drive/folders/1SY\_9lGtXdt2ZeP75Cgn3FwvigW\_c0cuR?usp=sharing |
| **Please write a summary of your Capstone project and its components. Describe the problem statement, your approaches, and your conclusions. (200-400 words)** |
| **Problem Statement:** Rossmann operates over 3,000 drug stores in 7 European countries. Rossmann store managers are currently tasked with predicting their daily sales up to six weeks in advance. Store sales are influenced by many factors, including promotions, competition, school and state holidays, seasonality, and locality. With thousands of individual managers predicting sales based on their unique circumstances, the accuracy of results can be quite varied.  You are provided with historical sales data for 1,115 Rossmann stores. The task is to forecast the "Sales" column for the test set. Note that some stores in the dataset were temporarily closed for refurbishment. **Approach:** The approach followed here is to first check the sanctity of the data and then understand the features involved. The events followed were in our approach:  Understanding the business problem and the datasets   * **Data cleaning and preprocessing-** Finding null values and imputing them with appropriate values.   + Converting categorical values into appropriate data types and merging the datasets provided to get a final dataset to work upon. * **Exploratory data analysis-** of categorical and continuous variables against our target variable. * **Data manipulation-** Feature selection and engineering, feature scaling, outlier detection and treatment, and encoding of categorical features. * **Modeling**- To know which model will be best for the prediction we have split the data into Test-Train data and with its help we checked the parameter that is used to choose the best model among the given predictive models. * **Model Performance and Evaluation** * **Store-wise Sales Predictions** * **Conclusion and Recommendations**   **Conclusion:**  Businesses use sales forecasts to determine what revenue they will be generating in a particular timespan to empower themselves with powerful and strategic business plans. Important decisions such as budgets, hiring, incentives, goals, acquisitions, and various other growth plans are affected by the revenue the company is going to make in the coming months, and for these plans to be as effective as they are planned to be it is important for these forecasts to also be as good.  Some important conclusions drawn from the analysis are as follows:   * The positive effect of promotion on Customers and Sales. * Most stores have competition distance within the range of 0 to 10 km and had more sales than stores far away probably indicating competition in busy locations vs remote locations. * Store type B though being few in number had the highest sales average. The reasons include all three kinds of assortments especially assortment level b which is only available at type b stores and is open on Sundays as well. * The outliers in the dataset showed justifiable behaviour. The outliers were either of store type b or had a promotion going on that increased sales.   **Recommendations:**   * More stores should be encouraged for promotion. * Store type B should be increased in number. * There's a seasonality involved, hence the stores should be encouraged to promote and take advantage of the holidays. |