**SUMMARY OF RETAIL SALES PREDICTION**

**LINEAR REGRESSION**

* **INTRODUCTION**

Rossman operates over 3,000 drug stores in 7 European countries. Currently, Rossman store managers are tasked with predicting their daily sales for 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 Rossman 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.

* **PROBLEM**

In this Retail Sales Prediction, machine learning models are created that predict sales of these 1115 drug stores across the European market and compare the results of these models. In addition to this, an effort has been made to analyse and find all the features that are contributing to higher sales and the features which are leading to lower sales, so that improvement plans can be worked upon.

* **TOOLS REQUIRED**

To do analysis on Ted Talk data we used language python. In python we have various libraries which we must import like NumPy, Pandas, Matplotlib and Seaborn**,** Scikit-Learn.

* **ACTIVITY**

1. First step involved is understanding the data and getting answers to some basic questions like; What is the data about? How many rows or observations are there in it? How many features are there in it? What are the data types? Are there any missing values? And anything that could be relevant and useful to our investigation.
2. Handling missing values is an important skill in the data analysis process.
3. Exploratory data analysis is a crucial part of data analysis. It involves exploring and analysing the dataset given to find out patterns, trends and conclusions to make better decisions related to the data, often using statistical graphics and other data visualization tools to summarize the results.
4. Data manipulation involves manipulating and changing our dataset before feeding it to various regression machine learning models. This involves keeping important features, outlier treatment, feature scaling and creating dummy variables if necessary.
5. In statistics, an outlier is a data point that differs significantly from other observations.
6. Feature Scaling is a technique to standardize the independent features present in the data in a fixed range. It is done to prevent biased nature of machine learning algorithms towards features with greater values and scale.

**Conclusion and Recommendations:**

**Conclusion:**

The main objective of sales forecasting is to paint an accurate picture of expected sales. Sales teams aim to either hit their expected target or exceed it.

When the sales forecast is accurate, operations go smoothly and future planning for the company's growth is done efficiently.

Upon having this analysis, it can be established that given the dataset, the model developed is able to explain 95.5878 % of the variations and is able to predict the sales values in a good range.

**Some important insights to draw from the analysis includes:**

* There were more sales on Monday. probably because shops generally remain closed on Sundays which had the lowest sales in a week. This validates the hypothesis about this feature.
* The positive effect of promotion on Customers and Sales is observable. Most stores have competition distance within the range of 0 to 10 kms 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 specially assortment level b which is only available at type b stores and being open on Sundays as well. The outliers in the dataset showed justifiable behaviour. The outliers were either of store type b or had promotion going on which 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.