**Capstone Project 2 : Retail Sales Prediction**

**Contributor Role**

**Suraj Kad**

**(**[**suraj.kad.90@gmail.com**](mailto:suraj.kad.90@gmail.com)**)**

**Shubham Sawant (**[**shubhamsawant248@gmail.com**](mailto:shubhamsawant248@gmail.com)**)**

**Tanvir Patel**

**(**[**tanviripatel1998@gmail.com**](mailto:tanviripatel1998@gmail.com)

**Import Packages**

* **Data Preparation**
* **Load Dataset**
* **Dealing With Missing Values**
  1. **Count missing values in each dataset**
  2. **Remove features with high percentages of missing values**
  3. **Replace missing values in features with low percentages of missing values**
* **Date Extraction**
* **Joining Tables**
* **Drop Subsets Of Data Where Might Cause Bias**
* **Feature Engineering**
* **Create new variable "AvgSales"**
* **Create new variable "AvgCustomer"**
* **Transform Variable "StateHoliday"**
* **Exploratory Data Analysis**
* **Correlation Heatmap**
* **Sales Distribution**
* **Customer Distribution**
* **Sales Over Days Of A Month**
* **Sales Over Weeks**
* **Sales By Store Type**
* **Sales By Assortment**
* **Sales vs. Number Of Customers**
* **Sales vs. Competition Distance**
* **Sales By Promotion**
* **Pair Plot**
* **Store Sales Prediction (Regression Models)**
* **Linear Regression (OLS)**
* **LARS Lasso Regression**
* **Decision Tree Regression**
* **Random Forest Regression**
* **Model Selection**

**GitHub Link:**

**https://github.com/TanvirPatel/Retail-Sales-Prediction.git**

**Drive : https://drive.google.com/drive/folders/1ZZ3kB7N4c4ckLt7F3S8OJcJK\_MzXlBQr?usp=share\_link**