Title: Profit Prediction Using Multiple Regression Models

# **Subtitle:** Project by RANJITH AMBEERA

## **Date**: OCT-04-2024

Executive Summary:

**Objective**: Predict company profit based on key features like R&D Spend, Administration Cost, and Marketing Spend using various machine learning models.

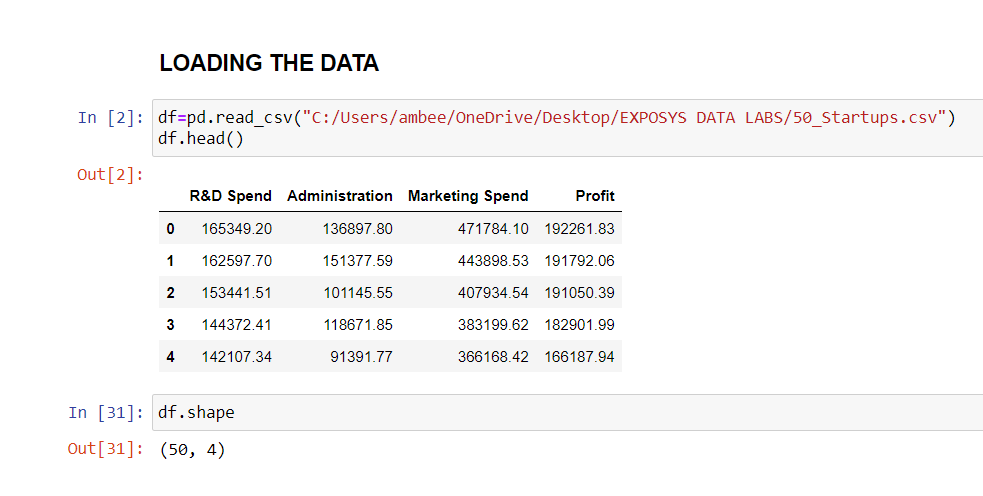
**Key Findings**: Linear Regression provided the best predictions with an R² score of 0.9001.

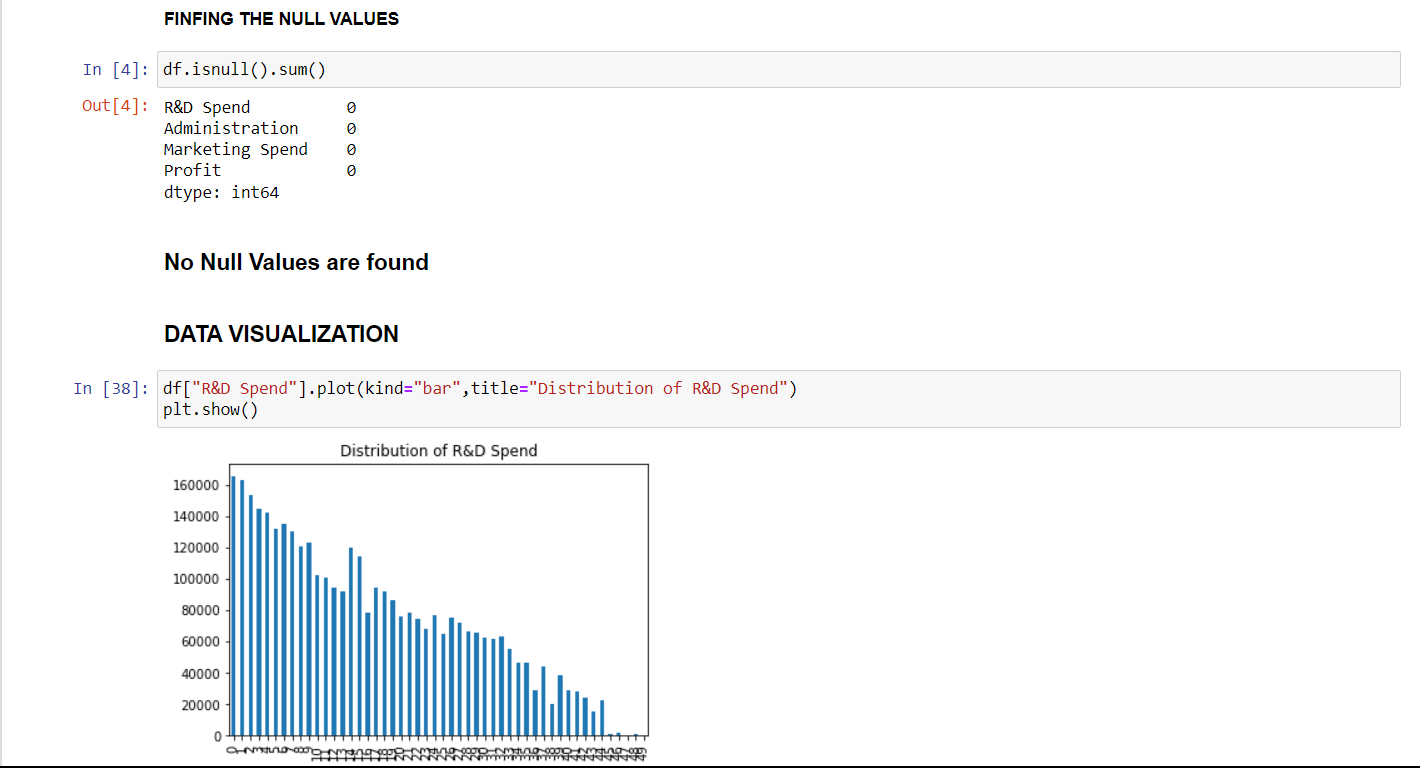
**Introduction:**

**Problem Statement:** Businesses need accurate profit predictions to make informed financial decisions. This project uses multiple regression models to predict company profit.

**Objective:** Evaluate different regression models to find the most accurate one for predicting company profit.

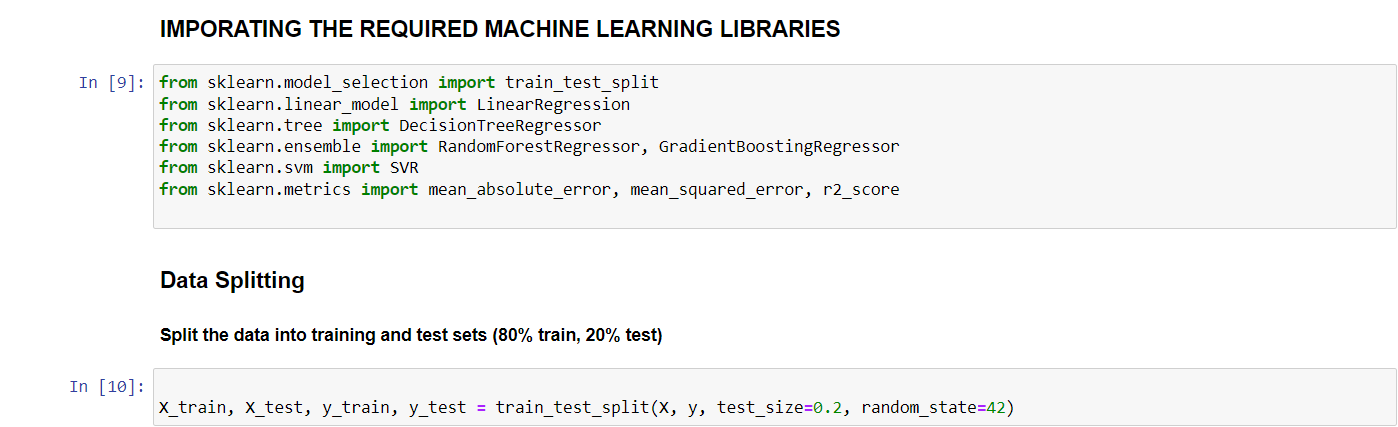
**Data Description**

* **Dataset**: A hypothetical dataset with features: 
  + **R&D Spend**
  + **Administration Cost**
  + **Marketing Spend**
  + **Target Variable**: Profit
* **Data Preprocessing**:



* + Checked for missing values.
  + Split into training and test sets.

**Modeling**

****

* **Models Implemented**:
  + Linear Regression
  + Decision Tree
  + Random Forest
  + Support Vector Regression (SVR)
  + Gradient Boosting
* **Training and Evaluation**:
  + Models were evaluated based on MAE, MSE, RMSE, and R² scores using the test set.

### **Model Comparison**

Compare the performance of each model.

| **Model** | **MAE** | **MSE** | **RMSE** | **R² Score** |
| --- | --- | --- | --- | --- |
| **Linear Regression** | 6979.15 | 80,926,321 | 8995.91 | 0.9001 |
| Decision Tree | 10,926.72 | 319,876,763 | 17,885.10 | 0.6050 |
| Random Forest | 6510.51 | 93,275,923 | 9657.95 | 0.8848 |
| SVR | 22,844.11 | 955,479,565 | 30,910.83 | -0.1799 |
| Gradient Boosting | 8600.02 | 85,846,621 | 9265.35 | 0.8940 |

**Insights and Observations**

* **Linear Regression**: Best overall performance due to strong correlation between features and target variable.
* **Random Forest**: Performed well but had slightly higher RMSE.
* **SVR**: Poor performance, likely due to the complexity of the dataset.

**Profit Prediction Function**

* Implemented a predict\_profit function that handles various input scenarios.
* **Examples**:
  + All values provided: Predicted Profit = $183,218.23
  + Only Marketing Spend provided: Predicted Profit = $117,659.08
  + No values provided (using default/mean): Predicted Profit = $115,651.72

**Conclusion**

* **Best Model**: Linear Regression
* **Conclusion**: Linear Regression’s simplicity and effectiveness made it the best fit for the dataset.