

# PROJECT-5

## **Project Title: Predicting Profit Based on Startup Data**

**Project Description:** This project aims to analyze and predict the profitability of startups based on their spending patterns in different areas such as R&D, Administration, and Marketing. The dataset used contains information about 50 startups including their expenditures and location (State). The main goal is to build a predictive model using linear regression to understand which factors most significantly influence a startup's profit.

**Summary:** The project starts with data exploration and preprocessing using Python libraries like Pandas, Seaborn, and Matplotlib. The dataset is loaded and explored to understand its structure, distributions, and correlations among variables. Categorical data (State) is encoded numerically for modeling purposes.

Visualizations such as heatmaps and pair plots are used to identify correlations between variables. Key observations reveal that R&D Spend and Marketing Spend show strong positive correlations with Profit, while Administration shows minimal correlation.

Following feature selection based on correlation analysis, a linear regression model is chosen and trained using a subset of features (R&D Spend and Marketing Spend). The model is evaluated using metrics like Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared.

Lastly, the model is used to make predictions on both test data and a randomly selected sample from the dataset. The results indicate that the linear regression model performs well in predicting profits based on spending patterns.

**In conclusion**, this project demonstrates how data analysis and machine learning techniques can be applied to understand and predict the financial success of startups based on their operational expenditures.