S&P 500 Prediction using Multiple Linear Regression

## 📝 Problem Statement

This project aims to predict the \*\*S&P 500 stock index\*\* based on two key macroeconomic indicators:

- \*\*Interest Rate\*\* (%)

- \*\*Employment Numbers\*\* (in thousands/millions)

The dataset is used to build a \*\*Multiple Linear Regression (MLR)\*\* model to understand the relationships between these economic factors and the S&P 500 index price.

---

## 📂 Table of Contents

1. [Introduction](#introduction)

2. [Dataset Description](#dataset-description)

3. [Multiple Linear Regression Overview](#multiple-linear-regression-overview)

4. [Model Development](#model-development)

5. [Results & Interpretation](#results--interpretation)

6. [Conclusion](#conclusion)

7. [How to Run](#how-to-run)

8. [Dependencies](#dependencies)

---

## 📖 Introduction

Financial markets are influenced by multiple macroeconomic factors, with \*\*interest rates\*\* and \*\*employment\*\* being two of the most impactful.

The \*\*S&P 500 index\*\* is widely regarded as a benchmark for U.S. stock market performance.

By modeling these relationships, we aim to provide insights that could support investment strategy and market analysis.

---

## 📊 Dataset Description

The dataset contains:

- \*\*Interest Rates\*\*: Cost of borrowing money, expressed as a percentage.

- \*\*Employment\*\*: Labor market strength (number of employed people).

- \*\*S&P 500 Price\*\*: The target variable (closing price/index value).

---

## 📐 Multiple Linear Regression Overview

Multiple Linear Regression models the relationship between one dependent variable and two or more independent variables.

The mathematical form is:

\[

\hat{y} = \beta\_0 + \beta\_1(\text{Interest Rate}) + \beta\_2(\text{Employment}) + \epsilon

\]

Where:

- \(\beta\_0\) = Intercept

- \(\beta\_1, \beta\_2\) = Coefficients

- \(\epsilon\) = Error term

---

## 🛠 Model Development

1. \*\*Data Preprocessing\*\*:

- Checked for missing values and outliers

- Normalized variables if necessary

2. \*\*Exploratory Data Analysis (EDA)\*\*:

- Scatter plots and correlation matrix

3. \*\*Model Training\*\*:

- Fitted a multiple linear regression model

4. \*\*Model Evaluation\*\*:

- Used R², Adjusted R², RMSE to measure performance

- Checked p-values for statistical significance

---

## 📈 Results & Interpretation

- \*\*Interest Rate\*\*: Negative correlation with S&P 500 — higher interest rates may lower stock prices.

- \*\*Employment\*\*: Positive correlation with S&P 500 — strong labor markets often align with higher index levels.

- Model performance was evaluated using statistical metrics and found to capture a significant portion of variance.

---

## ✅ Conclusion

The MLR model demonstrates:

- A \*\*negative impact\*\* of interest rates on the S&P 500 index.

- A \*\*positive impact\*\* of employment on the index.

While useful as a baseline predictor, the S&P 500 is influenced by other variables (inflation, global events, earnings reports), so more advanced models or additional features could improve accuracy.