**📊 Project Title: Sales Prediction Using Python**

**🔍 Objective:**

To build a machine learning model that predicts sales revenue based on advertising budgets across different media channels (TV, Radio, Newspaper).

**📁 Dataset:**

* Dataset contains data for:
  + **TV Advertising** (USD spent)
  + **Radio Advertising**
  + **Newspaper Advertising**
  + **Sales** (units sold)
* Format: .csv
* Shape: 200 rows × 4 columns

**🧹 Data Preprocessing:**

* Checked for missing/null values
* All data was clean; no imputation was necessary

**📊 Exploratory Data Analysis (EDA):**

* **Correlation Heatmap** showed strong correlation between:
  + **TV** & **Sales** (highest)
  + **Radio** & **Sales**
* **Newspaper** showed a weaker correlation
* Pair plots and distribution plots were used for deeper understanding

**📈 Visualizations:**

* **Scatter plots** between Sales and features
* **Heatmap** of feature correlation
* **Regression line** plots for each feature vs. Sales
* **Actual vs. Predicted** sales line chart

**🤖 Model Building:**

* Used **Linear Regression** from sklearn
* Trained the model on 80% of the data
* Tested on 20%

**✅ Model Evaluation:**

| **Metric** | **Value** |
| --- | --- |
| R² Score | 0.90+ |
| MSE (Error) | Low (good) |
| Accuracy | ~90% |

**📌 Insights:**

* **TV budget** contributes the most to increased sales
* **Radio** has moderate influence
* **Newspaper** has minimal effect on sales

**🎯 Conclusion:**

Linear regression performs well in predicting sales. Businesses should prioritize spending on TV and Radio for higher sales ROI.

**🚀 Future Scope:**

* Apply **advanced regression models** (Random Forest, XGBoost)
* Create a **Streamlit dashboard**
* Include more features like:
  + Location
  + Product Category
  + Seasonal Trends
* Hyperparameter tuning for better accuracy
* Forecasting over time (Time Series models)