Sales Prediction Using Python

Project Overview

This project uses linear regression to predict future sales based on advertising spending across three platforms: TV, Radio, and Newspaper. The goal is to help businesses understand the impact of each advertising channel and optimize their marketing strategies for better ROI.

Objectives

- Predict Sales using advertising spend data
- Clean and preprocess data
- Visualize relationships using charts
- Train a regression model (Linear Regression)
- Evaluate the model performance
- Extract business insights from feature impact

Dataset Description

The dataset consists of the following columns:
- TV: TV ad spend (in thousands of USD)
- Radio: Radio ad spend (in thousands of USD)
- Newspaper: Newspaper ad spend (in thousands USD)
- Sales: Units sold (in thousands)

Tools & Libraries Used

- Python 3.x
- Pandas
- Seaborn & Matplotlib
- Scikit-learn

Steps

1. Data Cleaning & Preparation
2. Exploratory Data Analysis (EDA)
3. Feature & Target Selection
4. Train-Test Split
5. Model Training
6. Prediction & Evaluation
7. Feature Impact Analysis
8. Visualization of Predictions

Model Evaluation Results

- R^2 Score: 0.899
- Mean Squared Error: 3.17

Feature Coefficients:

Sales Prediction Using Python

TV: 0.0447
Radio: 0.1892
Newspaper: 0.0027

Insights

- Radio has the strongest influence on sales
- TV has moderate impact
- Newspaper has negligible effect

Recommendations:
- Focus more on Radio
- Use TV strategically
- Reduce Newspaper ad spend

Future Improvements

- Try advanced models like Ridge or Random Forest
- Use cross-validation
- Deploy as a web app using Streamlit or Flask

Author

Muhammad Yasir
BSCS Student
Sales Prediction Internship Task