Apple Sales Prediction

By
Yamijala Praneetha

Brief

This project aims to develop a predictive model for estimating apple sales based on weather conditions and the week number. By leveraging historical sales data, the model can provide insights into future sales trends, helping businesses make informed decisions regarding inventory management, marketing strategies, and resource allocation.

Objectives

- To analyze the relationship between weather conditions and apple sales.
- To create a regression model that accurately predicts apple sales based on weather and week number.
- To evaluate the performance of the model using various statistical metrics.

Steps Involved

- Data Collection: Gather historical sales data, including information on weather conditions and week numbers.
- Data Preprocessing: Clean the data by handling missing values, encoding categorical variables, and scaling numerical features.
- Feature Selection: Identify and select relevant features that influence apple sales.

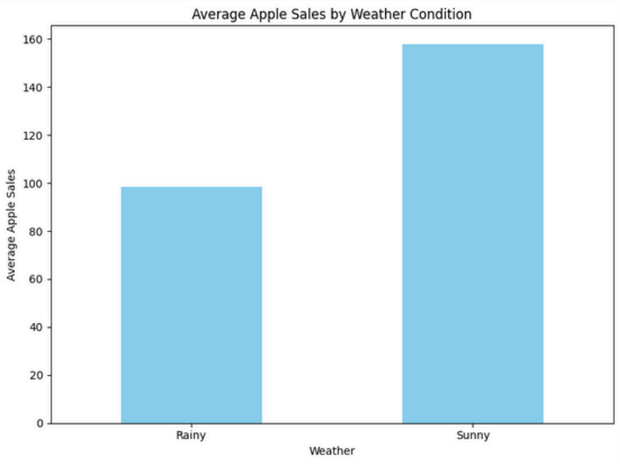
Steps Involved Continues...

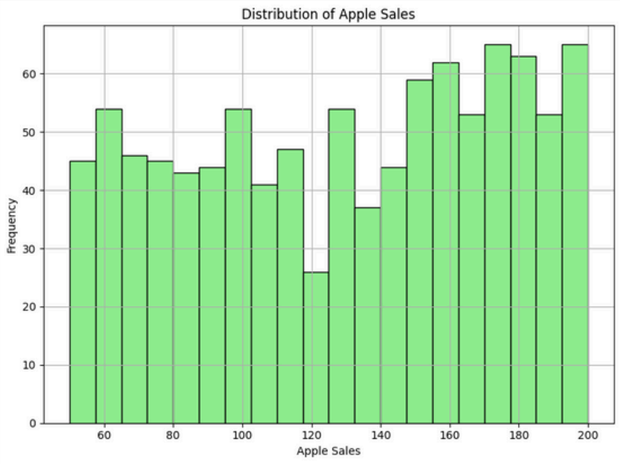
- Model Training: Train a linear regression model using the processed data.
- Model Evaluation: Assess the model's performance using metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R^2 Score.
- Prediction: Use the trained model to predict apple sales and evaluate its accuracy.

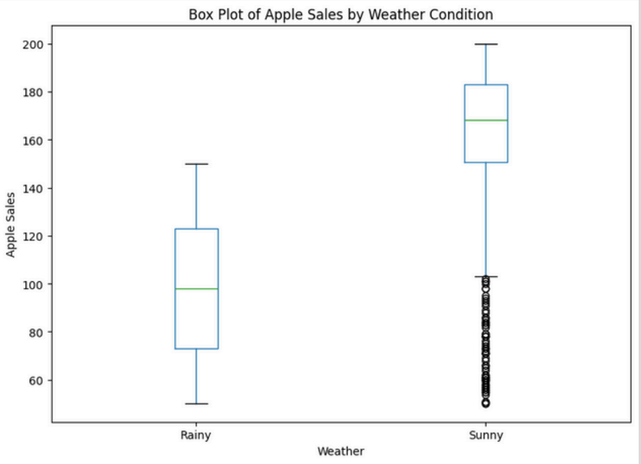
Data Preprocessing

- Handling Missing Values:
  Forward fill method to handle any missing values
- Encoding Categorical Data:
  Used LabelEncoder to convert 'Weather' feature into numerical labels
- Feature Engineering:
  Extracted Week_Number from the 'Week' feature using regular expressions
- Feature Scaling:
  Applied StandardScaler to normalize the 'Apple Sales' feature

Exploratory Data Analysis







Feature Selection

- Selected Features:
  - Weather
  - Week_Number
- Target Variable:
  - Apple Sales

Splitting the Data

- Train-Test Split:
  - Ratio: 80% training, 20% testing
  - Mention the use of train_test_split from sklearn

Model Training

- Experimented Models:
  - Linear Regression
  - Random Forest Regressor
  - Gradient Boosting Regressor
  - XGBoost Regressor
  - LightGBM Regressor
  - CatBoost Regressor

Final Model Chosen

- Chosen due to its simplicity and ability to provide a clear, interpretable formula for predicting apple sales based on weather and week number.

Prediction Formula

- Sales = -0.6124 + (1.2906 * Weather) + (-0.0001 * Week_Number)

Model Evaluation

- Evaluation Metrics:
  - Mean Squared Error (MSE)
  - Root Mean Squared Error (RMSE)
  - Mean Absolute Error (MAE)
  - R^2 Score

Results

	Train	Test
MSE	0.5827	0.4717
RMSE	0.7634	0.6868
MAE	0.5955	0.5609
R^2	0.4177	0.5264

Future Work

- Potential Improvements:
  - Exploring more features
  - Using more advanced models

Conclusion

- Objective:
  - Predict apple sales based on weather and week number
- Process
  - .Preprocessed data (handled missing values, encoded 'Weather', extracted Week_Number, scaled 'Apple Sales').
  - Selected Weather and Week_Number as features.
  - Chose Linear Regression for its simplicity and clear prediction formula.
- Findings
  - Linear Regression provided an interpretable prediction formula.
  - Achieved good performance metrics.
- Implication
  - Assists in forecasting apple sales, aiding inventory and strategic planning.

Thank You!