**Walmart Sales Forecasting Project**

**Introduction**

This report provides a comprehensive analysis of the Walmart Sales Forecasting project. The primary objective of this project is to predict the sales for Walmart stores using historical data and various machine learning techniques.

**Libraries and Tools**

The following libraries are utilized throughout the project:

• **Data Manipulation and Analysis**: numpy, pandas

• **Visualization**: matplotlib, seaborn, plotly

• **Machine Learning**: sklearn

These libraries are essential for data preprocessing, visualization, and the implementation of machine learning models.

**Data Loading**

The project involves loading four CSV files:

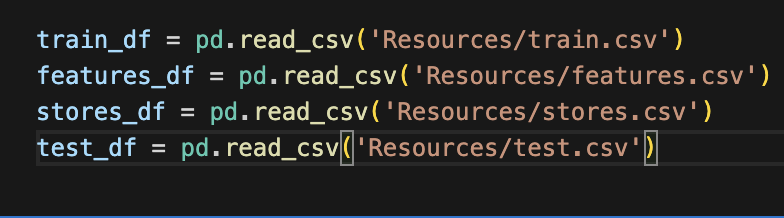
• train.csv: Contains the historical sales data used for training the models.

• features.csv: Includes additional features that might affect sales such as promotions and holidays.

• stores.csv: Provides information about the stores.

• test.csv: The dataset on which predictions are to be made.

**Code Snippet for Data Loading:**



**Data Exploration**

Initial exploration of the datasets involves the following steps:

**Displaying the First Few Rows**:

• Using the head() method to get a glimpse of the data structure and contents.

**Checking Columns and Shape**:

• Identifying the columns present in each dataset.

• Checking the dimensions of the datasets using the shape attribute.

**Basic Statistics and Information**:

• Using info() and describe() methods to get basic statistics and data types of each column.

**Handling Missing Values**:

• Identifying missing values using isnull() and isnull().sum() to ensure data quality and completeness.

**Code Snippet for Data Exploration:**

# Display the first few rows of train dataset

train\_df.head()

# Check columns and shape

train\_df.columns

train\_df.shape

# Basic statistics

train\_df.info()

train\_df.describe()

# Missing values

train\_df.isnull().sum()

**Feature Engineering and Selection**

Feature engineering is critical to improve model performance. Key steps include:

• **Transformations**: Applying necessary transformations to the data, such as scaling and normalization.

• **Handling Categorical Variables**: Converting categorical variables into numerical format using techniques like one-hot encoding.

• **Feature Creation**: Creating new features from existing ones that might help the model better understand the underlying patterns.

**Model Training and Evaluation**

Several machine learning models are trained and evaluated, including:

• **Ridge Regression**: Used to handle multicollinearity in the data.

• **Lasso Regression**: Helps in feature selection by shrinking less important feature coefficients to zero.

Models are evaluated using metrics such as:

• **Root Mean Squared Error (RMSE)**

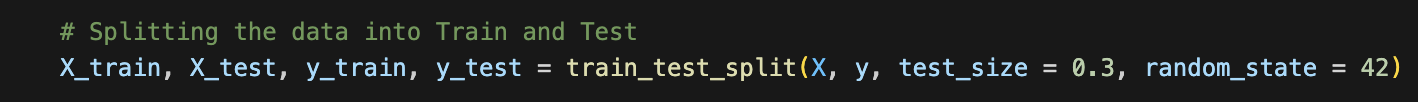
• **R² Score**

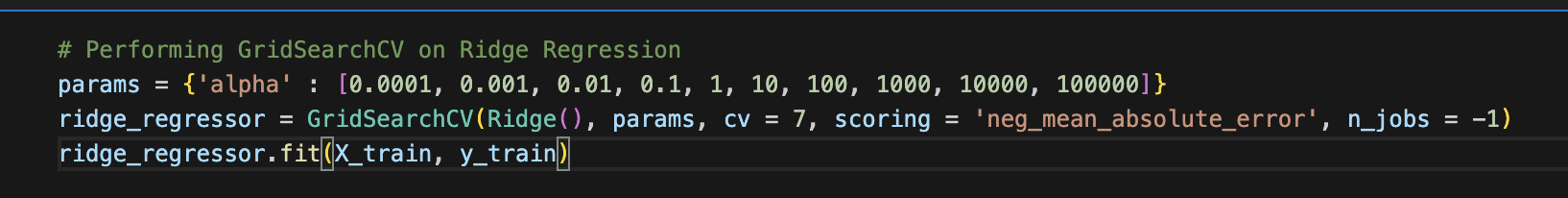
Hyperparameter tuning is performed using:

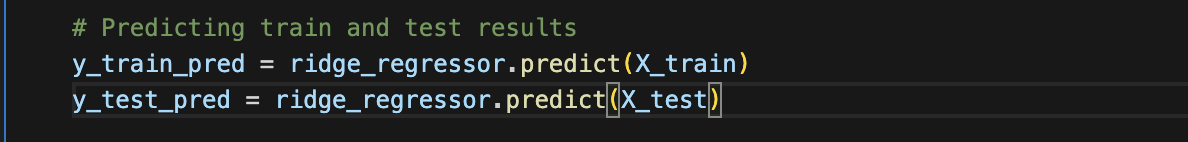
• **GridSearchCV**: Exhaustive search over specified parameter values.

• **RandomizedSearchCV**: Random search over parameters, which can be more efficient.

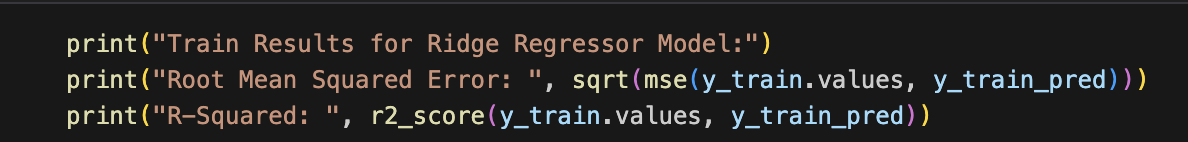
**Code Snippet for Model Training:**





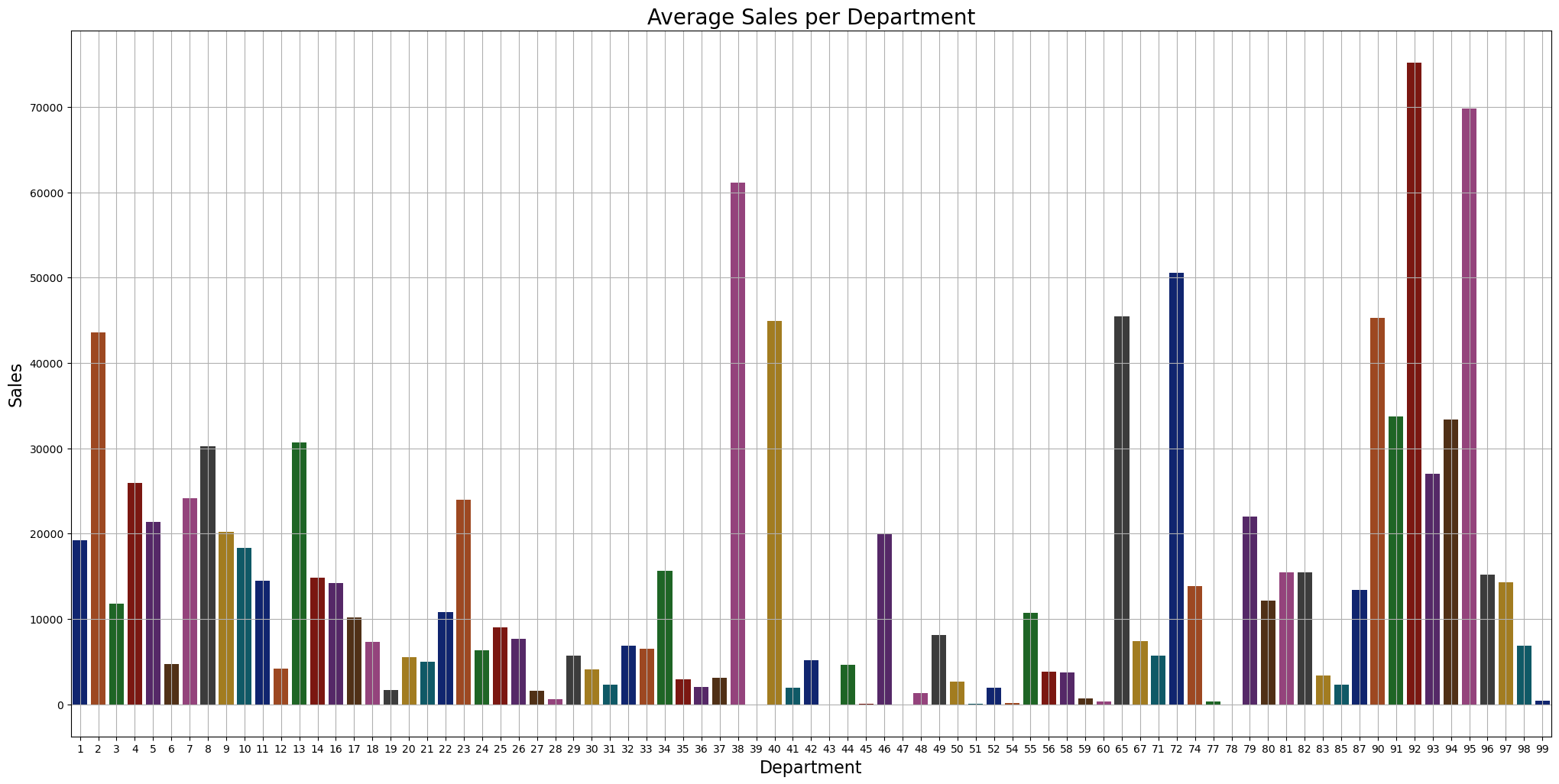


# Evaluating the model

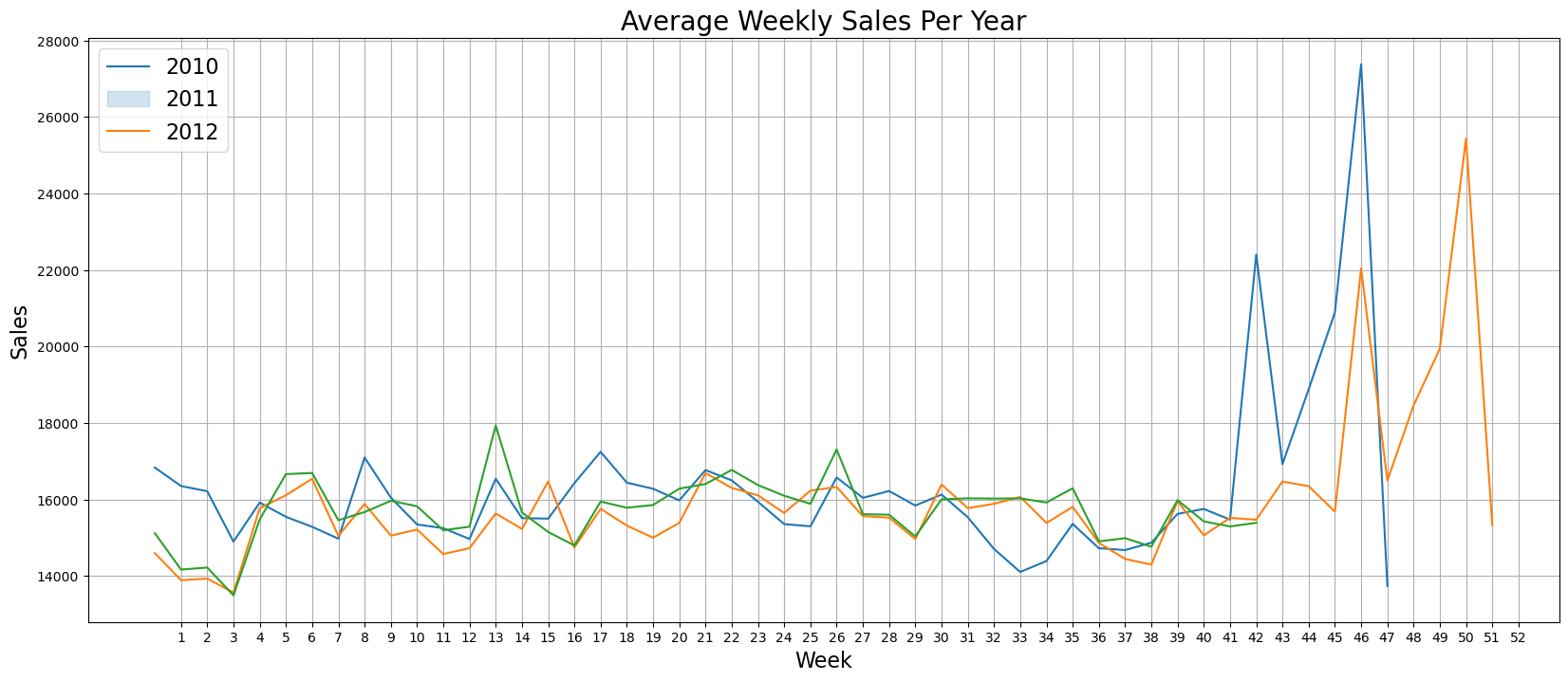


**Visualization**

Various plots are generated to understand data distribution, trends, and model performance:

• **Histograms**: To visualize the distribution of numerical features such as sales

• **Department Pie Charts**: To identify relationships between departments.

• **Time Series Plots**: To analyze sales trends over time.

Visualization helps in gaining insights and making informed decisions during feature engineering and model selection.

**Conclusion**

The Walmart Sales Forecasting project demonstrates a structured approach to predicting sales, starting from data loading and exploration to model training and evaluation. The project provides a detailed workflow that can be adapted and extended for similar forecasting tasks in different domains. The use of advanced machine learning techniques and thorough data analysis ensures robust and accurate predictions.