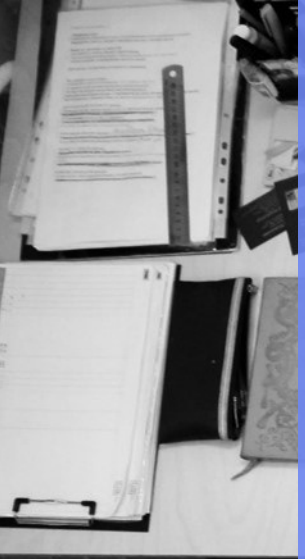
Amazon Sales Data Analysis Project



Objective

- Development of a predictive model for predicting sales.
- Perform ETL (Extract-Transform-Load) on dataset.
- Develop dashboard by using tableau.

Benefits

- Better understand and optimise revenue generation in future
- Maximize forecasting accuracy
- Make current sales experience our top priority

Architechture

Project is done with 4 steps

Data Collection

The Dataset provided by ineuron. Collected from their website in csv fromat.

01

Data Analysis

This steps is a combination of Exploratory data analysis, Statics and Label Encoding

02

Business Analysis

This Step made with Machine Learning algorithms. Used Pycaret library and Data visualization.

03

Business Intelligence

Used Tableu for creating Dashboard and combined Data Analytics with Business KPI's.

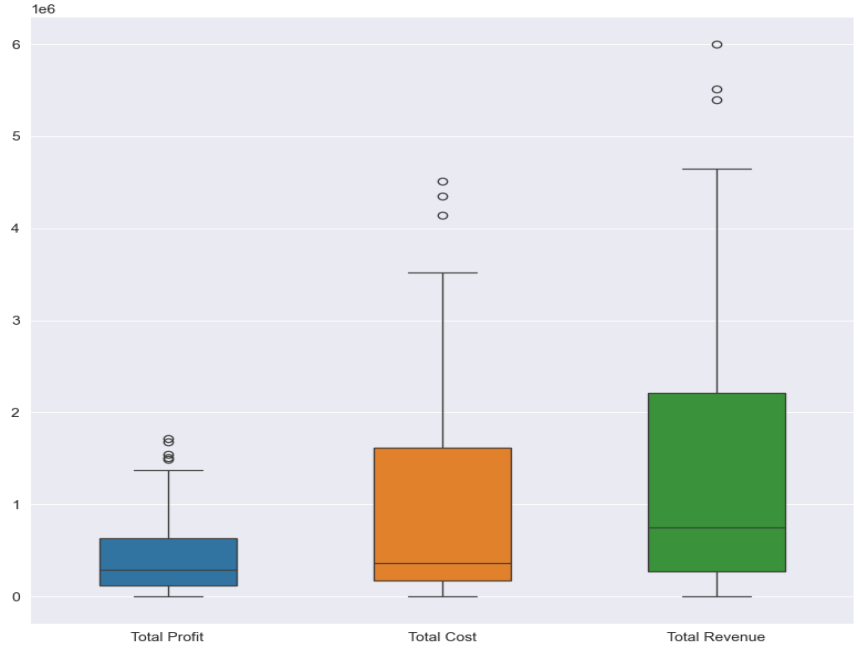
04

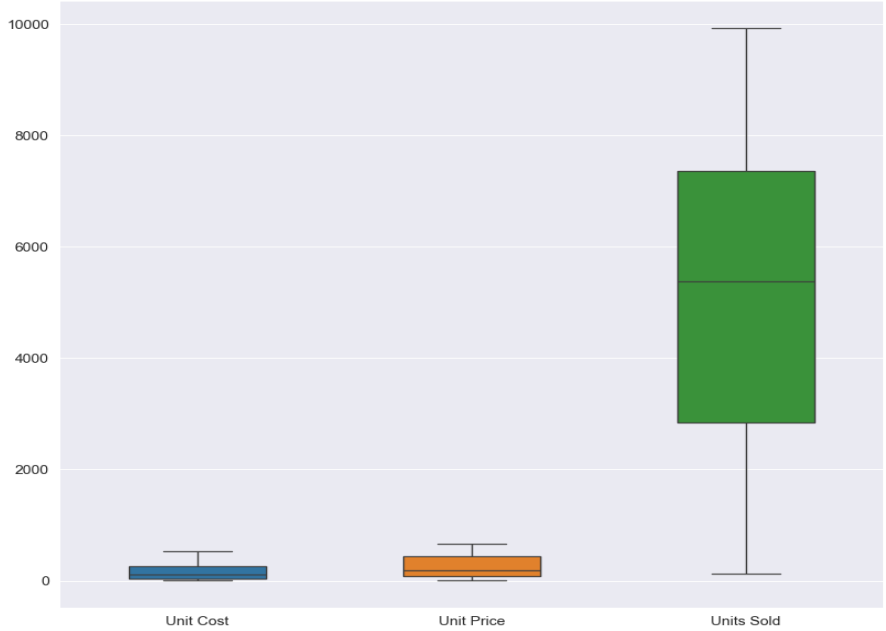
Data Preprocessing

- Importing necessary libraries for data analysis such as : Pandas, Numpy, Matplotlib & Seaborn etc.
- Using pd.read_csv() function stores the data in pandas dataframe named data.
- Using data.column showing columns present in dataframe.
- info() function show basic information of dataframe like null value count of each column and their data type
- Changing the data type of different column for model training and analysis
- Using describe function on dataframe for getting basic stats of numerical dataset
- Adding extra column to dataframe which contain only month, year and month with year
- Using isnull().sum() checking out total null value in all the column of dataframe

Exploratory Data Analysis

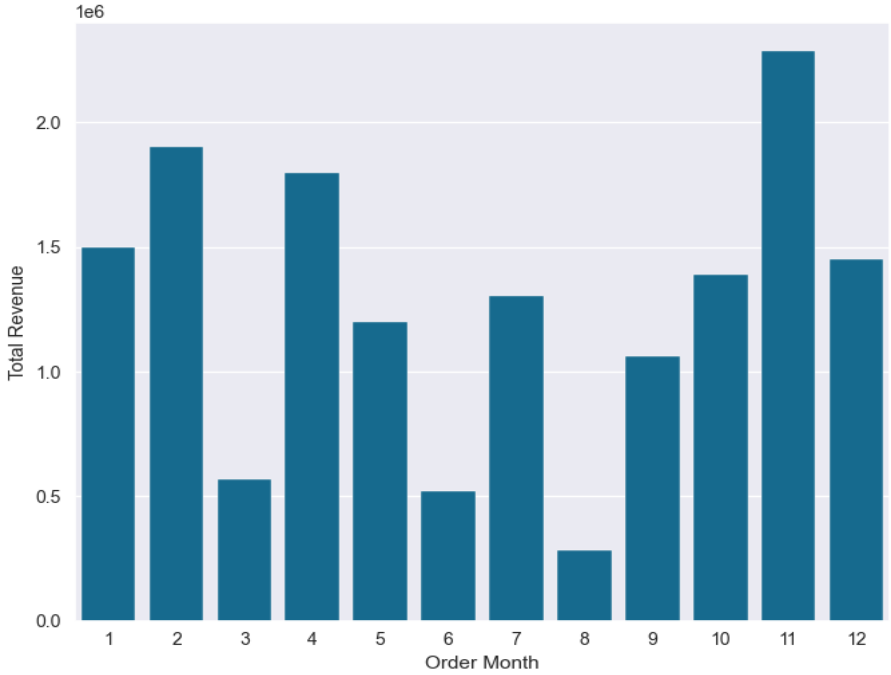
Checking Outliers in the dataframe by using Box Plot

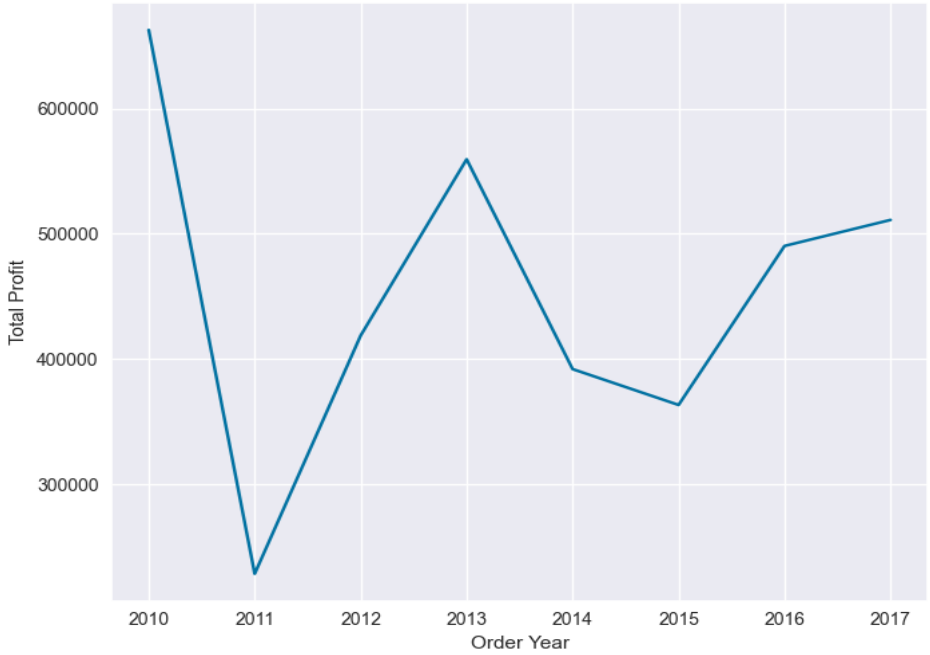




There are some outliers in Total Profit, Total Cost and Total Revenue colums. No outliers in Unit Cost, Unit Price and Unit Sold Columns.

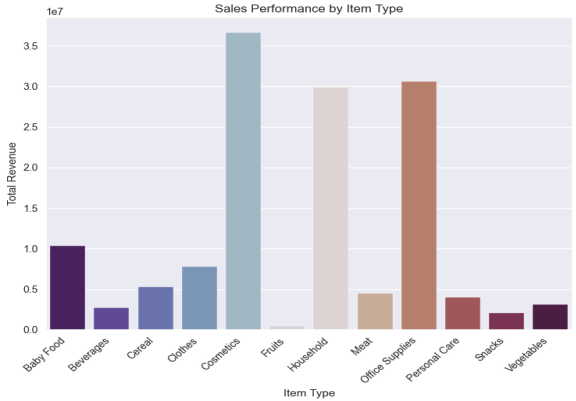
Understanding the Total Revenue per Month with bar chart and a line chart for trend

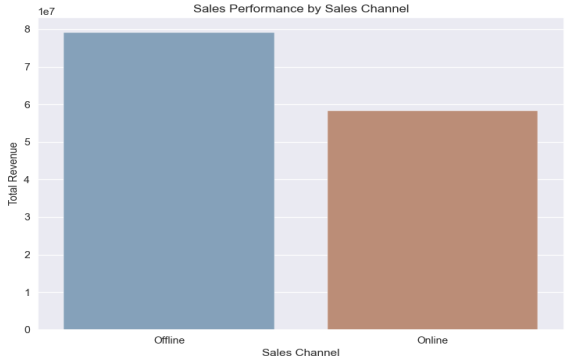


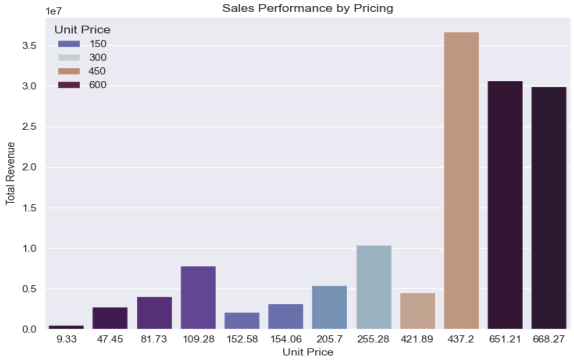


The total revenue shows variations from year to year. For instance, there was a significant increase in revenue from 2011 to 2013, followed by a decline in 2015. Each month's revenue varies across different years, indicating potential seasonal trends or fluctuations in consumer behavior. August has lowest rrevenue and November has highest revenue.

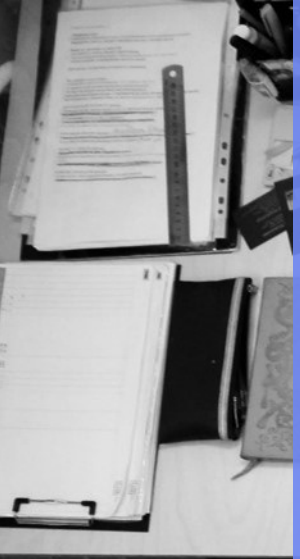
Factor Analysis







- The total revenue and units sold vary significantly across different regions and countries.
- The item types 'Cosmetics', 'Household', and 'Office Supplies' contribute significantly to total revenue.
- Offline sales generate higher total revenue and units sold compared to online sales.
- Orders with order priority 'H' (high) contribute the most to total revenue and units sold.
- There is a variation in total revenue and units sold based on the unit cost of products.



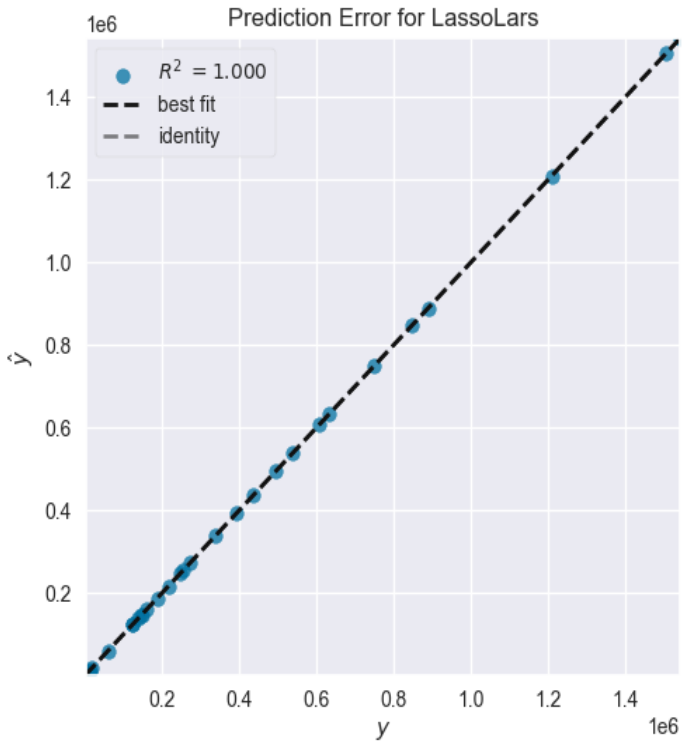
Predictive Analysis

- ILabel Encoding of Item Type, Sales Channel and Order Priority for modeltraining.
- Dropping columns Region, Country, Order Date MonthYear, Order ID and ShipDate.

Pycaret Library

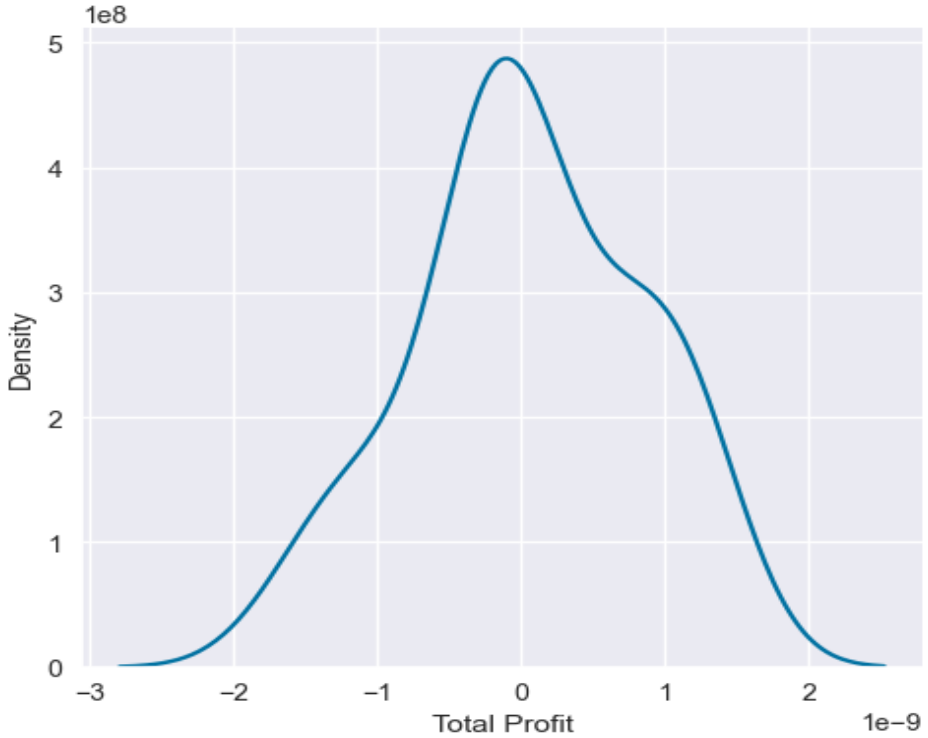
- PyCaret is an open-source, low-code machine learning library in Python.
- Allows users to quickly and easily build, compare, and deploy machine
- learning models on structured and tabular data.
- Reduce the amount of code needed to build a model.
- It provides preprocessing and feature engineering functions.
- Automatic model selection and hyperparameter tuning.
- Support for a wide range of machine learning algorithms

Plotting residuals and prediction error for Lasso Least Angle Regression based trained model

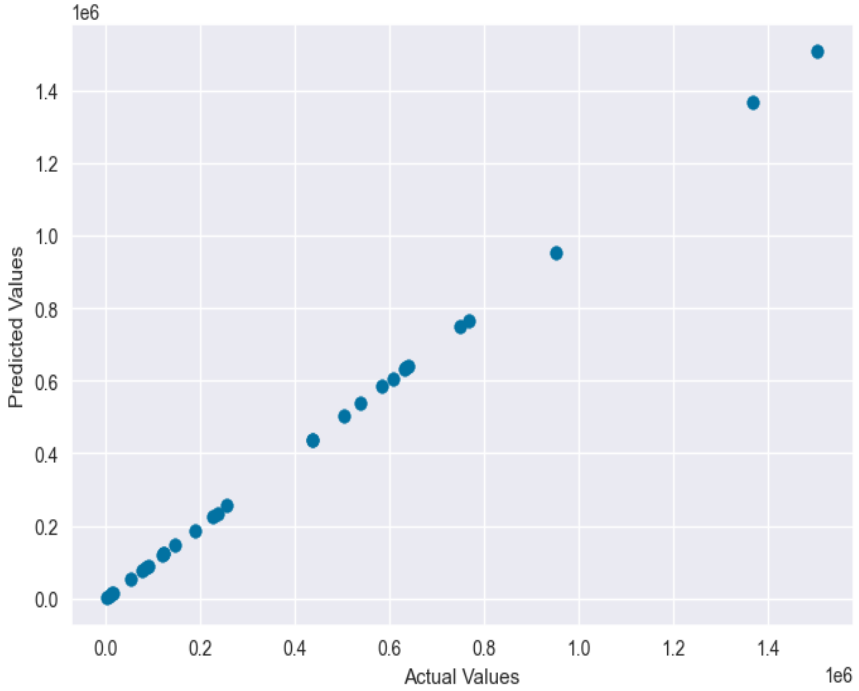


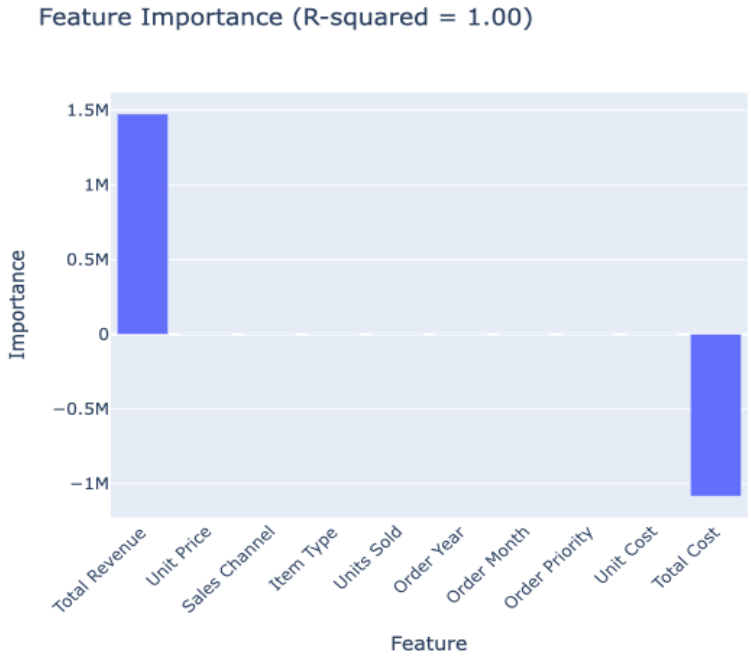
Implementation of linear regression

- Selecting the independent variables and target variable.
- Splitting the data into training and testing datasets.
- Standardizing the dataset.
- Performing fit transform on X_train dataframe.
- Performing fit transform on X_test dataframe.
- Applying Linear Regression on X_train and y_train.
- Calculating mean squared error.
- Creating kernel density estimate plot



Plotting the predicted values against the actual values to visualize how well the model is fitting the data.





Key Findings and Recommendation

Sales Performance

Certain product categories, such as cosmetics and office supplies, generated higher total revenue and units sold compared to others.

Recommendation:

Invest in popular product categories to capitalize on existing demand. Consider expanding product offerings within high-performing categories to further boost sales.

01

Time Trends:

Monthly and yearly sales trends indicated seasonal fluctuations and long-term growth patterns.

Recommendation:

Anticipate seasonal demand variations and adjust inventory levels and marketing efforts accordingly. Implement strategies to sustain long-term growth and capitalize on emerging market trends.

02

Relationship Analysis:

Correlation analysis identified strong positive correlations between variables such as total revenue, units sold, and total profit.

Recommendation:

Continuously monitor and analyze key metrics to identify potential areas for improvement. Implement data-driven strategies to enhance sales performance and profitability.

03