**COFFEE SALES**

**Project Objective:**

To analyze the coffee sales data, identify sales trends & purchasing patterns of customer and to build a machine learning model by using ml algorithm.

**Tools used:**

1. Python- Spyder-5.3.1
2. Machine Learning
3. Excel-2019

**Steps involved:**

1. Problem Statement
2. Data Collection
3. Data Cleaning
4. Exploratory Descriptive Analysis (EDA)
5. Machine Learning Modeling
6. Model Evaluation & Conclusion

**Problem Statement:**

To analyze the coffee sales data, find out the monthly sales, day sales, weekly sales & purchasing patterns of customer and also do the time series EDA.

**Data Collection:**

The coffee sales dataset is collected from the data source.

The coffee sales dataset consists of 6 columns. Columns are date, datetime, cash\_type, card, money, and coffee\_name.

Date – Date of the purchased coffee product.

Datetime- The datetime provides the information of date along with time of purchased coffee product.

Cash\_Type- The cash\_type column has two unique values- cash, card. That means purchased coffee product is either in cash mode or card mode.

Card – The card column gives information about if we purchased coffee product in card mode it will read the card number.

Money - The money column gives information about cost of purchased coffee products.

Coffee\_Name- The name of the coffee product. 8 different types of coffee products are present.

**Data Cleaning:**

After Loading the coffee sales dataset, we must clean the dataset. We found zero duplicates in the dataset and 89 missing values in the column ‘card’. All 89 missing values in the column ‘card’ are from cash users only. After that we followed different data issues like feature engineering, etc. We cleaned the dataset.

**Exploratory Data Analysis:**

Conducted the EDA to visualize the sales trends.

We plotted bar plots and line charts to analyze the time series dataset.

We used matplotlib and seaborn libraries to visualize the data.

**Machine Learning Modeling:**

After EDA we must select the ml algorithm and build a machine learning model.

We choose **the Multiple Linear Regression Model**. Because Linear Regression is a statistical method used to model the relationship between a dependent variable and one or more independent variables. It makes predictions for continuous or numeric variables such as sales, salary, stock price, money, etc. Multiple Linear Regression is an extension of linear regression.

And other supervised learning algorithms like SVM (Support Vector Machine), K-Nearest Neighbour, also used for predictive modeling, but linear regression model is simple and easy way to do prediction of dependent variable.

After Choosing Machine Learning Algorithm, we must split the data into train data and test data. Then define the features and target variable.

After that train the model and make the predictions.

**Model Evaluation & Conclusion:**

Evaluating a Multiple linear regression model is crucial to understand how the model fits the data and how accurate its predictions are. Here are some common evaluation metrics and techniques for linear regression.

1. Mean Square Error

MSE measures the average squared difference between actual and predicted values. Low MSE indicates a better fit.

1. R-Squared

R-Squared indicates the proportion of the variance in the dependent variable that is predictable from the independent variable.

**Conclusion:**

After the time series EDA and builded a machine learning model, we concluded and gained some insights.

* **The most customers are preferred for the** **Americano with Milk and Latte coffee products.**
* **Americano with Milk, Latte and Cappuccino are top selling coffee products.**
* **Latte coffee product contributed high revenue and Espresso coffee product generated low revenue.**
* **Tuesday and Wednesday has the highest sales of the week, and on other days sales are similar.**
* **In hourly sales analysis, we observed three peak hours within each day. The peak hours are 10:00am, 11:00 am and 7:00pm.**