**Diwali Sales Analysis using Python**

# OBJECTIVES OF THE PROJECT :

* To analyse Diwali sales on the basis of different products, gender, state, age, occupation and zone area of the customers.
* Improving customer experience by analysing sales data.
* Increase revenue.

# Overview:

**Exploratory Data Analysis** (EDA) is primarily used to see what data can reveal beyond the formal modelling or hypothesis testing task and provides a better understanding of data set variables and the relationships between them. It can also help determine if the statistical techniques you are considering for data analysis are appropriate. Originally developed by American mathematician John Tukey in the 1970s, EDA techniques continue to be a widely used method in the data discovery process today.  
In this project, we'll explore the Diwali Sales dataset. The dataset contains various customers who have purchased different products, their age, gender, marital status, and occupation. Through this analysis, we aim to uncover hidden trends, outliers, correlations, and other important aspects of the data.

1. The project is structured as follows:
2. Exploratory\_Data\_Analysis-Diwali\_Sales.ipynb - Here you will find Jupyter notebooks used for data cleaning, visualization, and analysis.
3. Output - You can place any images generated during your analysis in this folder.
4. Diwali Sales Data.csv - This directory contains the dataset files.

# Steps performed to execute the project:

Performed the below steps to analyse Diwali sales data to improve customer experience and sales.

1.Project Learning and understanding the data.

2. Performed data cleaning and manipulation

3. Performed exploratory data analysis (EDA) using pandas, matplotlib and seaborn libraries

4. Improved customer experience by identifying potential customers across different states, occupation, gender and age groups

5. Improved sales by identifying most selling product categories and products, which can help to plan inventory and hence meet the demand

# [Requirements:](https://github.com/DivyanshSuryawanshi02/Diwali-Sales-Analysis-EDA/blob/main/README.md#requirements)

To run the notebooks and analysis in this project, you'll need the following:

Python

Jupyter Notebook or Jupyter Lab

Required Python libraries: Pandas, NumPy, Matplotlib, Seaborn

# Loading .csv file using pandas

load the csv file containing all the raw data using pandas function pd.read\_csc()

# Data Cleaning

To check the datatype of each column we are going to use df.info() [we have load our csv file in df] df.info() show the information such as column name, datatype, non-null count of rows, etc To delete any useless columns we run command df.drop([‘C1’,’C2'], axis=1, inplace=True)……….axis=1 is used for total row and inplace=True used for saving the changes.

After removal of useless columns, we have the raw data file but it is uncleaned and contains null values. To check that null values we have to run code using pandas as pd.isnull(df).sum().

When any row has null values then to eliminate such null values we call command of pd.dropna(inplace=True)

If we want to find min, max, avg, total, standard deviation, 25%, 50%, 75% etc we use df.describe()

Now after cleaning data its time to analyze data.

# Exploratory Data Analysis (EDA)

# Gender

Now by using seaborn we will analyse data and plot graph of [count vs gender] We get to know that female count is more than that of male.

Then plot graph of [amount vs gender] we get to know that females spent more amount than male.

# Age

Plotted graph of [count vs age group] and applied legend of gender to separate bars on the basis of gender we analyse that females of age group 26–35 did most buying.

Plotted graph of [amount vs age group] we analyse that most of the buyers are of age group 26–35 years female.

# State

Now we are going to anlyse [orders vs state] using seaborn From graph we can see that most of the orders are from Uttar Pradesh, Maharashtra and Karnataka respectively.

We plotted graph of [amount vs state] From the graph we can see that most of the total amount are from Uttar Pradesh, Maharashtra and Karnataka respectively.

# Marital Status

Firstly we analyse the number of married and unmarried customers for that, need to plot graph of [count vs marital status], after plotting graph we can see that, most of the buyers are married.

Then we plotted graph of [amount vs marital status] and applied legend of gender to separate bars on the basis of gender. We can conclude that most of the buyers are married women’s.

# Occupation

We have to anlyse occupation wise count and amount of customers. We have to plot graph of [count vs occupation] and also plot graph of [amount vs occupation]. From the graph we can see that most of the buyers are working in IT, Healthcare and Aviation sector.

# Product Category

Need to plot graph of [count vs product category] and [amount vs product category], from the graph we can see that most of the sold products are from Food, Clothing and Electronics category.

# CONCLUSION OF THE PROJECT

Married women of age group 26–35 years from Uttar Pradesh, Maharastra and Karnataka working in IT, Healthcare and Aviation are more likely to buy products from Food, Clothing and Electronics category. Also, the product with product id P00265242 is sold the most.