**Project Report: Diwali Sales Data Analysis**

**1. Project Overview**

The Diwali Sales Data Analysis aims to understand the purchasing behaviors of customers during the Diwali season. This analysis looks at customer demographics and product category trends to extract insights on how different factors such as gender, age, marital status, state, occupation, and product preferences impact the sales during Diwali.

**Objective:**

* To analyze and visualize customer purchasing behavior during Diwali.
* To understand the demographic patterns influencing purchasing power.
* To categorize products and evaluate sales trends based on customer attributes.

**2. Data Collection and Preprocessing**

The dataset used for this project contains customer orders made during the Diwali season. The dataset contains various attributes including:

* **Gender**
* **Age Group**
* **State**
* **Marital Status**
* **Occupation**
* **Product Category**
* **Amount Spent**
* **Number of Orders**

The raw data is available in the file: [Diwali Sales Data](https://github.com/KoritalaBhargavi/DataAnalysis/blob/main/Diwali%20Sales%20Data.xls).

**Data Preprocessing Steps:**

1. **Loading Data**: The data was loaded using pandas.read\_excel() into a DataFrame.
2. **Handling Missing Values**: Any rows with missing data were removed using the dropna() method.
3. **Removing Irrelevant Columns**: Non-relevant columns such as Status and unnamed1 were dropped to simplify the dataset.
4. **Data Type Conversion**: The Amount column was converted to an integer type for consistency in calculations.
5. **Renaming Columns**: Some columns, such as Marital\_Status, were renamed to more meaningful names (e.g., Shaadi).

**3. Exploratory Data Analysis (EDA)**

**Gender Analysis:**

* **Visualization**: A countplot was used to visualize the distribution of genders in the dataset.
* **Insight**: Females account for the majority of buyers, and they have higher spending compared to males.

**Age Group Analysis:**

* **Visualization**: A countplot and barplot were used to analyze purchases by age group.
* **Insight**: The age group 26-35 years shows the highest purchasing activity, with females contributing significantly to total sales.

**State Analysis:**

* **Visualization**: Bar charts were created to display the total number of orders and total sales by state.
* **Insight**: Uttar Pradesh, Maharashtra, and Karnataka are the states with the highest number of orders and total sales.

**Marital Status:**

* **Visualization**: A countplot and grouped bar chart were used to analyze the spending behavior by marital status.
* **Insight**: Married women tend to have a higher purchasing power compared to other groups.

**Occupation Analysis:**

* **Visualization**: A countplot and barplot were used to visualize the sales distribution across different occupations.
* **Insight**: Customers working in the **IT**, **Healthcare**, and **Aviation** sectors tend to spend more.

**Product Category Analysis:**

* **Visualization**: Countplots and barplots were used to identify the most popular product categories by sales.
* **Insight**: **Food**, **Clothing**, and **Electronics** are the top-selling categories during the Diwali season.

**4. Results and Insights**

Based on the exploratory data analysis, the following key insights were observed:

1. **Gender**: Most of the buyers are **females**.
   * **Total Purchase Amount**: Female customers tend to spend more than male customers.
2. **Age Group**: The dominant buying age group is **26-35 years**.
   * **Female Buyers**: This age group shows higher spending, particularly among women.
3. **State**: **Uttar Pradesh**, **Maharashtra**, and **Karnataka** are the leading states in terms of orders and sales.
   * **Highest Orders**: Uttar Pradesh had the highest number of orders.
   * **Highest Sales**: Maharashtra and Karnataka had the highest total sales.
4. **Marital Status**: **Married women** have higher purchasing power.
   * **Insight**: Married women tend to make higher-value purchases compared to others.
5. **Occupation**: Customers working in **IT**, **Healthcare**, and **Aviation** sectors are more likely to make higher-value purchases.
   * **Spending Power**: These sectors show more consistent high purchasing activity.
6. **Product Category**: The most popular product categories are **Food**, **Clothing**, and **Electronics**.
   * **Sales Distribution**: These categories dominate in terms of total sales amount.

**5. Conclusion**

From the analysis, we can conclude that the most likely customers to purchase during the Diwali season are:

* **Married women**, aged between **26-35 years**, from **Uttar Pradesh**, **Maharashtra**, and **Karnataka**, and working in the **IT**, **Healthcare**, and **Aviation** sectors.
* These customers are primarily purchasing products in the **Food**, **Clothing**, and **Electronics** categories.

This analysis provides valuable insights for marketers and business owners to target the right demographic groups during festive sales.

**6. Future Work**

This analysis can be extended by:

* **Time Series Analysis**: Tracking the purchasing trends over time during the Diwali season to identify peak sales periods.
* **Customer Segmentation**: Using clustering techniques to create targeted marketing campaigns based on customer profiles.
* **Predictive Modeling**: Applying machine learning models to predict future sales and demand based on past trends.

**7. References**

* Dataset: [Diwali Sales Data](https://github.com/KoritalaBhargavi/DataAnalysis/blob/main/Diwali%20Sales%20Data.xls)
* Jupyter Notebook for Analysis: [Diwali Sales Analysis Notebook](https://github.com/KoritalaBhargavi/DataAnalysis/blob/main/Diwali_Sales_Analysis.ipynb)