**DIWALI SALES DATA ANALYSIS**

**Business Problem**

*A company has shared their Diwali sales data, seeking an exploratory data analysis to unveil valuable insights. With the ultimate objective of enhancing customer experience and boosting revenue, the analysis is aimed to deliver a summary of their sales data. By delving into their sales performance, identify trends, customer preferences, and potential opportunities for optimization.*

*Our exploration will encompass a multifaceted examination of sales by region, product categories, and customer behaviours. We will pinpoint top-performing states and categories. The ultimate goal is to provide actionable recommendations that enable the company to refine strategies, enhance customer engagement, and create tailored initiatives that resonate with their audience. By leveraging the insights derived from the exploratory data analysis, the company can not only elevate customer satisfaction but also strategically increase their revenue, thereby charting a course toward sustainable growth.*

**Pre – Gathering and Analysis**

1. **Imported Libraries:** Libraries such as numpy, which facilitated intricate mathematical operations critical to the examination. Additionally, the imported pandas, enabling seamless data manipulation, comprehensive cleansing, and thorough exploration. Matplotlib and seaborn are also imported to further enriched the analysis, empowering the creation of compelling visual representations that will illuminate key insights within the dataset, thereby fostering a deeper understanding of the Diwali sales trends and dynamics.
2. **Imported DataFrame:** The second step included the importing of the dataframe utilizing the pd.read\_csv() function. To eliminate any potential encoding errors, the 'unicode\_escape' parameter was incorporated, ensuring the seamless importation of the dataset.
3. **Dataframe Check:** To check whether the proper file has been imported with the exact data several codes, such as df.shape (returns the nrow, and ncolumns), and df.head() (Gives the overview of the first five rows), were passed.

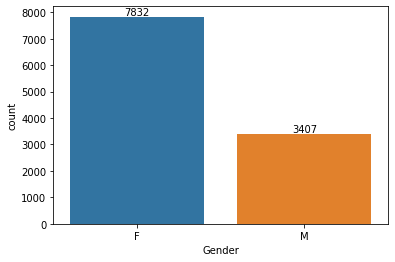
**Data Cleaning**

1. **Passed df.info():** By using df.info(), an assessment of the dataset was done, revealing key information on column data types, counts, and the existence of null values. Upon examination it was found that some values of 'orders' column are missing, and all the values of 'status' and 'unnamed1' column are missing.
2. **Dropping Columns:** Dropped ‘Status’ and ‘Unnamed1’ column, as they do not contain any value using the df.drop() function.
3. **Dropping Null Values:** The 12 rows in the 'orders' column were removed which had null values. The choice of dropping those rows were done because the number of rows were insignificant compared to the total length of rows.
4. **Changing DType:** The 'Amount' column's datatype was modified from float64 to int32, as decimal precision was unnecessary for the dataset's needs.
5. **df.describe():** This function facilitated a deeper dive into the mathematical parameters of the data, enabling a more informed and insightful exploratory data analysis phase.

**Exploratory Data Analysis**

* **Gender:**
  + For the purpose of dissecting the customer demography within the sales data, a gender-based order count was done. This examination revealed a predominant proportion of orders attributed to 'Female' customers. A visual representation of this variation is illustrated in **Figure 1.1**.

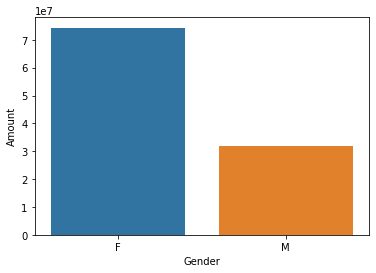
**Orders Created by Genders**

* + ****

**Figure 1.1**

* + After the initial analysis, I delved deeper into the data to understand the revenue contribution from each gender. The results revealed that 'Female' customers have a notably higher contribution, almost double that of 'Male' customers (Check **Figure 1.2**), emphasising their significant role in driving the store's revenue.

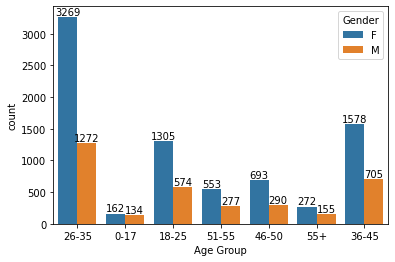
**Revenue by Gender**

* + ****

**Figure 1.2**

* + The above graphs reveals that most of the buyers are females and even the purchasing power of females are higher than men.
* **Age:**
  + Age group plays an important part in data analysis as it is crucial and helps to identify target demographics, tailor marketing strategies, and understand consumer behaviour based on different life stages.
  + Analysing orders generated by different age groups provides insights into purchasing patterns. This information can guide marketing efforts, and customer engagement strategies tailored to each age segment's needs and behaviours. The **Figure 2.1** shows the number of orders different age groups and genders generated.

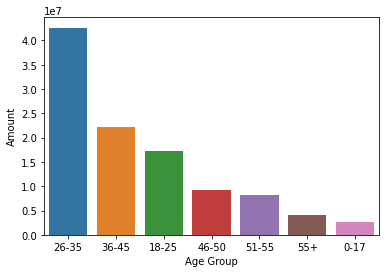
**Order Count by Age Group**



**Figure 2.1**

* + As the number of orders does not truly represent the revenue generation, revenue contribution was also taken into account and analysed (**Figure2.2**).

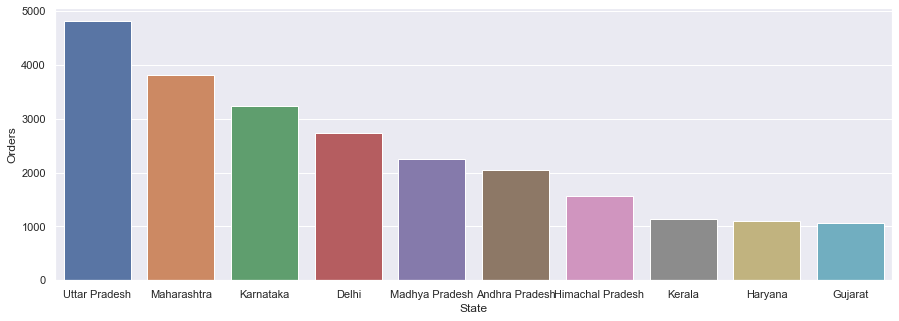
**Revenue by Age Group**

* + 

**Figure 2.2**

* + The graphs clearly show that the majority of buyers fall within the 26-35 age range, with a significant presence of female customers.
* **Top States:**
  + Identifying top-performing areas in terms of order generation helps allocate resources for a stronger supply chain, and focus on areas with high customer demand for optimized business outcomes. Taking this into consideration, and to analyse order geography, a bar chart representing the order count of 10 top performing states is plotted (**Figure 3.1**).

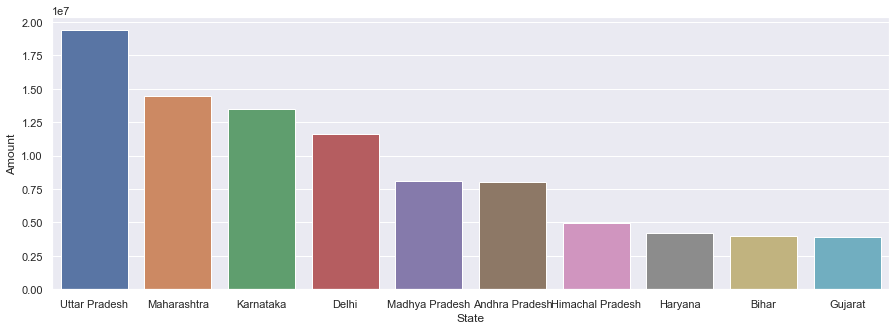
**Orders Created by States**

* + 

**Figure 3.1**

* + The best performers are Uttar Pradesh, followed by Maharashtra, Karnataka, Delhi, Madhya Pradesh and so on.
  + As previously highlighted, the count of orders alone might not always provide a comprehensive insight. Recognizing the significance of revenue as a key performance metric for a store, the analysis was extended to correlate the data with revenue contributions. This association is visually presented in **Figure 3.2.**

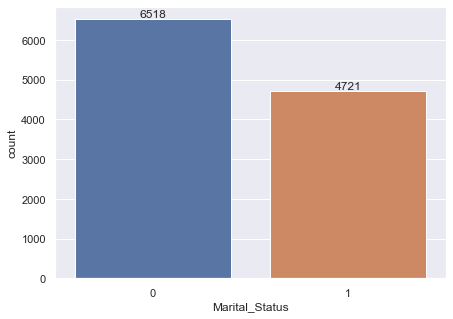
**Amount Spend by States**

* + 

**Figure 3.2**

* + There is not much of a difference between the **Figure 3.1** and **3.2** graphs, expect the **8th** and **9th** place.
* **Marital Status:**
  + Does marital status effect the purchase behaviour of the customers? And if yes, then how does it affect the behaviour of the customers?

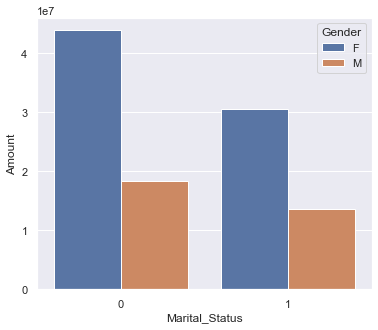
**Orders Created by Marital Status**

* + 

**Figure 4.1**

* + The observation reveals that individuals categorized as ‘Married’, ‘0' exhibit a greater count of orders, totalling 6518, in comparison to 'Single' customers labelled as '1', who recorded a lower order count of 4721. This discrepancy might arise from the augmented combined income of a family unit, which inherently translates into higher purchasing capability. Alternatively, this trend could stem from an increased demand for various products to meet the diverse needs of multiple family members.
  + Does the married person generate more revenue for the store also?

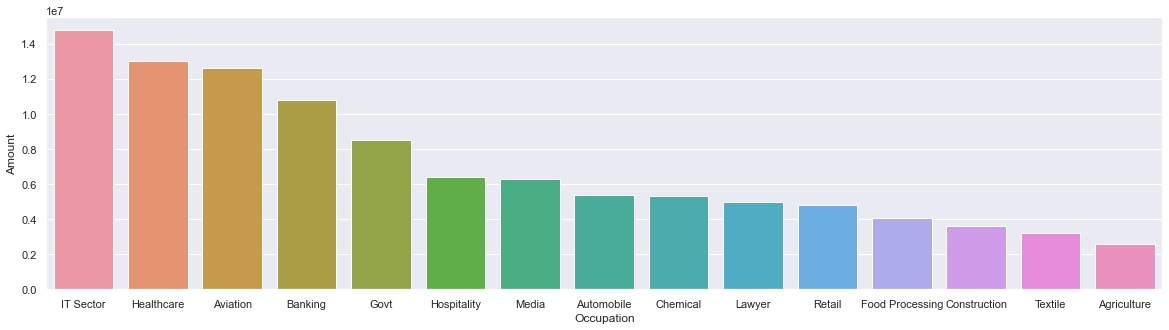
**Amount Spend by Marital Status**

* + 

**Figure 4.2**

* + This pattern is also evident in the value of goods they purchase (**Figure 4.2**). Most of the buyers are married (Female).
* **Customer Occupation:**
  + A customer's occupation can significantly impact their purchasing power, as it directly correlates with the amount they spend. Customers with more lucrative occupational backgrounds tend to have a greater expenditure capacity compared to those in less lucrative fields. This trend is effectively depicted in the graph provided below (**Figure 5.1**).

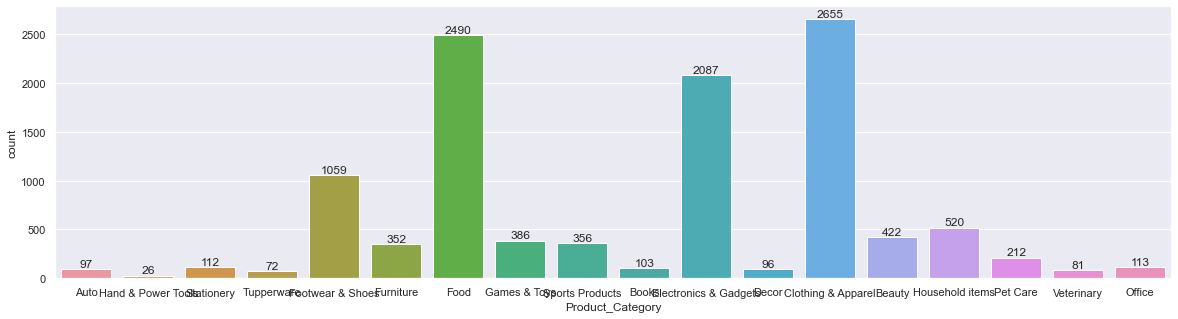
**Amount Spend by Occupation**

* + ****

**Figure 5.1**

* + From above graph we can see that most of the spenders are working in IT, Healthcare and Aviation sector.
* **Product Category:**
  + It's important to ensure a consistent stock availability for product categories that exhibit rapid sales turnover, even if these categories might not be the highest revenue generators. The graph depicted in **Figure 6.1** illustrates the sales performance across various product categories, highlighting those that experience swift turnover.

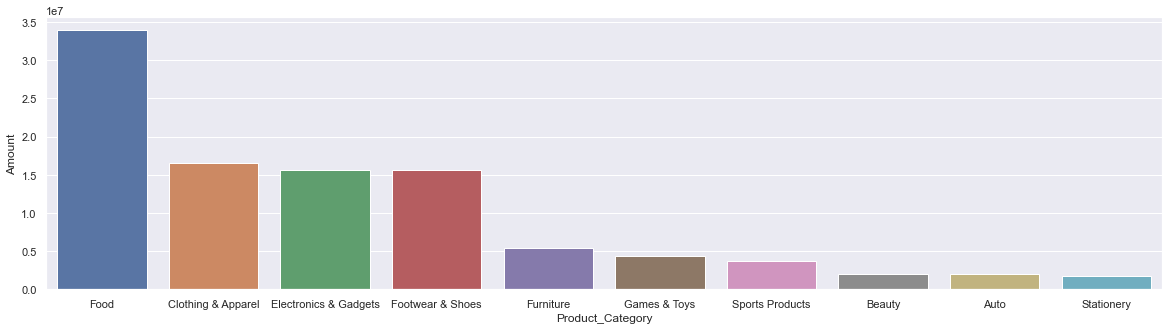
**Orders By Different Categories**



**Figure 6.1**

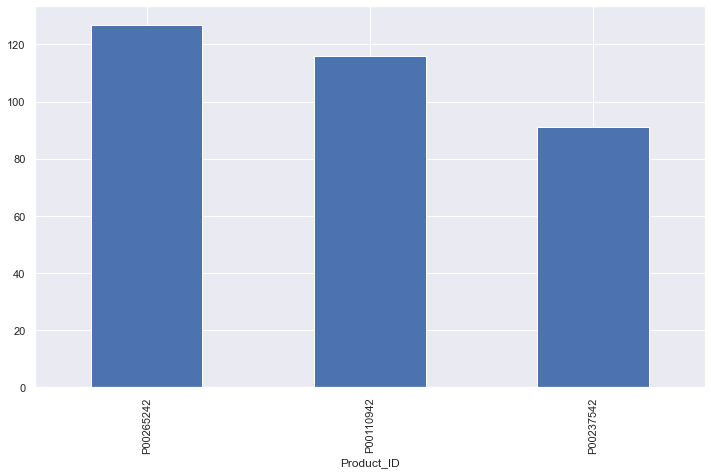
* + Now, the categories with the highest sales are (Figure 6.2):

**Revenue By Product Categories**

****

**Figure 6.2**

**Products With the Highest Sales**

****

**Figure 6.3**

* + Food, clothing and electronics are the top 3 highest revenue generating categories, followed by footwear. While the 3 best selling products are Product ID: P00265242, P00110942, P00237542.

**Suggestions**

In the pursuit of enhancing customer experience and driving revenue growth, an extensive analysis of the Diwali sales data has yielded valuable insights.

1. **Gender Dynamics and Sales:** The analysis highlights a notable gender-based trend, where 'Female' customers emerge as the dominant contributors to both order count and sales volume. This indicates an opportunity to tailor marketing strategies and product offerings to cater to the preferences and needs of female customers.
2. **Age Segmentation:** The age group between 26 and 35 stands out as a pivotal demographic, boasting the highest order frequency and revenue generation. Conversely, as age increases, revenue generation tends to decrease. To maximize profitability, consider targeting marketing efforts and promotions towards the lucrative 26-35 age bracket.
3. **Regional Insights:** Regions like Uttar Pradesh, Maharashtra, and Karnataka exhibit strong order contributions and revenue generation. Conversely, Gujarat and Haryana lag behind. Leveraging these regional insights can aid in refining supply chain and distribution strategies to cater to high-performing areas and tapping into untapped markets.
4. **Marital Status and Purchasing Power:** Married female customers are revealed as possessing the highest purchasing power. This group could potentially be targeted with premium products and personalized offers, thereby maximizing revenue potential.
5. **Occupational Affiliations:** Customers employed in IT, Healthcare, and Aviation sectors demonstrate a robust propensity for higher sales value. Tailored offerings and engagement strategies for these segments could yield significant revenue gains.
6. **Product Category Dynamics:** Identifying the best-performing categories such as Food, Clothing, and Electronics can inform inventory management and stocking decisions. On the contrary, categories like Beauty, Auto, and Stationary should be scrutinized for improvement opportunities.
7. **High-Performing Products:** A few products, with IDs P00265242, P00110942, and P00237542, stand out as the best-selling. Ensuring the availability and promotion of these products can further boost overall sales and revenue.

**End of report**