Black Friday Sales Data Analysis

BLACK FRIDAY SALES DATA Cleaning raw data and fetching insights

Table of Contents

1. Introduction
   1. Overview of Black Friday
   2. Purpose of Data Set
   3. Required Key findings
2. Prerequisites to the project
3. Workflow
4. KPIs
5. Insights and Other Key Findings
6. Conclusion
7. Introduction
   1. Overview of Black Friday:

* Black Friday is an annual shopping event occurring the day after Thanksgiving, known for its significant discounts and promotions.
* There increased sales on Black Friday.
* It marks the beginning of the holiday shopping season.
* Increased sales are driven by
  + heavy discounts
  + consumer anticipation
  + limited-time offers
  + and the availability of a wide range of products.
* Black Friday is known for its large crowds, long lines, and frenzied shopping atmosphere both in physical stores and online.
  1. Purpose of the Dataset:
* The dataset provides insights into consumer behaviour and purchasing patterns during Black Friday sales.
* It aims to analyse demographic trends, product preferences, and spending habits of customers participating in Black Friday shopping.
* The dataset that we are having contains 537578 rows and 12 columns.

Our main goal will be to do various kind of analysis and get inferences which will be able to help to provide valuable insights on the market and hence will be beneficial for the company.

* 1. Required Key findings (preliminary)
* How many customers participated in Black Friday sales
* What is the gender distribution among Black Friday shoppers?
* Which age groups are most active during Black Friday sales?
* What are the popular product categories during Black Friday sales?
* What is the average amount spent by customers per transaction?
* How are customers distributed across different city categories?
* What is the marital status distribution among Black Friday shoppers?

1. Prerequisites to the project
   1. Pandas
   2. Numpy
   3. Matplotlib
   4. Seaborn
2. Workflow:

To perform EDA on dataset following steps were taken at different stages -

 **Dataset Walkthrough:**

* Check basic dataset information like column names, data types, and presence of null values.
* Identify null values in columns like Product\_Category\_2 and Product\_Category\_3.
* Remove columns with many null values to avoid data loss.

 **Analyzing Columns:**

* Use unique() and nunique() functions to find unique values in each column.
* Utilize nunique() to count total customers and products.
* Use unique() to explore unique values in categorical columns like Gender and Age.

 **Analyzing Gender:**

* Analyze Gender column to observe the distribution of Male and Female values.
* Utilize groupby() function to compare purchasing behavior between genders.
* Visualize findings using pie charts and bar plots for better understanding.

 **Analysing Age & Marital Status:**

* Explore Age and Marital Status columns using groupby() to identify purchasing trends.
* Visualize insights using pie charts and bar plots.
* Analyze marital status distribution and its impact on purchasing behavior.

 **Multi Column Analysis:**

* Perform analysis using multiple columns like Age and Gender.
* Utilize Seaborn library for visualization and add legends for better interpretation.
* Explore various combinations of columns for deeper insights.

 **Occupation and Products Analysis:**

* Analyze Occupation, Product\_ID, and Product\_Category\_1 columns.
* Use countplot and groupby function to understand data distribution.
* Plot bar plots to identify the most purchased products and occupations.

 **Combining Gender & Marital Status:**

* Combine Gender and Marital Status columns for comprehensive analysis.
* Use Seaborn library, particularly countplot, to visualize combined data.
* Gain meaningful insights by analyzing the intersection of Gender and Marital Status.

1. KPI’s:

|  |  |
| --- | --- |
| Column | Unique Values |
| User\_ID | 5891 |
| Product\_ID | 3623 |
| Gender | 'F' (Female), 'M' (Male) |
| Age | '0-17', '18-25', '26-35', '36-45', '46-50', '51-55', '55+' |
| Occupation | 21 |
| City\_Category | 'A', 'B', 'C' |
| Stay\_In\_Current\_City\_Years | '0', '1', '2', '3', '4+' |
| Marital\_Status | 0, 1 |
| Product\_Category\_1 | 18 |
| Average Purchase Amount | $9333.86 |

1. Insights and Other Key Findings

* Gender Distribution: Male customers constitute around 75% of the dataset, while female customers make up approximately 25%, indicating a higher representation of males in Black Friday sales.
* Age Distribution: Majority of customers fall within the age groups of 26-35 and 36-45, followed by 18-25, with older and younger age groups being less represented.
* Occupation Distribution: The dataset includes customers from various occupations, suggesting a diverse customer base participating in Black Friday sales.
* City Category Distribution: Customers are spread across urban, semi-urban, and rural areas (City Categories A, B, and C), with urban areas (Category A) having the highest representation.
* Marital Status Distribution: Both married and unmarried customers are present in the dataset, with a slightly higher proportion of unmarried customers, indicating a mix of marital statuses among Black Friday shoppers.
* Product Category Analysis: Purchases are made across various product categories (Product\_Category\_1), highlighting diversity in product preferences among customers.
* Purchase Analysis: The average amount spent per transaction by customers on Black Friday is approximately $9333.86, with variations observed across different demographic groups.
* Purchase Distribution by Age: Customers aged 26-35 make the highest number of purchases, followed by the 36-45 age group, while younger and older age groups make fewer purchases.
* Age and Gender Distribution: Both genders are well-represented across different age groups, with the 26-35 age group having the highest representation for both males and females.
* Gender and Marital Status Distribution: The dataset contains a mix of gender and marital status combinations, offering insights into how these factors are distributed within the dataset.
* City Category Distribution: Most customers are from City Category B, followed by Category C and then Category A, revealing the distribution across urban, semi-urban, and rural areas.
* City Category and Purchase Analysis: Purchase patterns vary across different city categories, as illustrated by pie charts showing total and average purchases made in each category.

Other Important Findings:

|  |  |
| --- | --- |
| Finding | Values |
| Total Customer Participated | 537,577 |
| Gender distribution | Male = 75% (405,380) of the total |
| Female =25% (132,197) |
| Most active age group | 26-35 and 36-45 |
| Popular product categories | Product Category 1 has 18 unique values |
| Average Amount spend by Customer | Approximately $9333.86 |
| Customer Distribution | Customers are distributed across three city categories: A, B, and C. Most customers are from City Category B, followed by Category C and then Category A. |

1. Conclusion

* Customer Participation: The dataset contains information about customer demographics, preferences, and spending habits during Black Friday sales.
* Gender Distribution: Male customers are more represented compared to females, suggesting a gender imbalance in Black Friday sales.
* Age Distribution: Majority of customers belong to specific age groups, indicating the primary participants in Black Friday shopping.
* Product Preferences: Analysis of product categories reveals diverse preferences among customers, highlighting popular items during Black Friday sales.
* Purchase Analysis: The average amount spent per transaction varies across different demographic groups, such as age, gender, and marital status.
* Regional Differences: Customers are distributed across urban, semi-urban, and rural areas, with potential differences in spending behavior based on geographic location.
* Marital Status: Understanding the demographics of marital status can help tailor marketing strategies and promotions during Black Friday sales.

Overall, analyzing the dataset provides valuable insights into consumer behavior and preferences during Black Friday sales, aiding retailers in optimizing their marketing strategies and offerings for future sales events.