Walmart

Business Case Solution

By Munmun Mishra

Introduction

Walmart, a major global retail player, is diving into a detailed analysis of how customers behave on Black Friday, the mega shopping day. The main focus is understanding how much people spend, especially considering factors like gender and other influences. The aim is to uncover patterns in spending habits, offering key insights for making smart business decisions.

This analysis isn't just about gender; it goes deeper, exploring various factors that might affect how people shop on Black Friday. Walmart wants to be thorough, looking at all the different things that might influence people's choices. The idea is to paint a complete picture of what guides people's buying decisions, helping Walmart make well-informed decisions for the future.

In the upcoming sections, we'll discuss the methods we're using, where we're getting our information, and the specific factors we're examining. We'll also explore what these findings could mean for Walmart's strategies and decisions. This way, we're supporting Walmart in making savvy choices based on the insights we uncover.

Objectives and Scope

The primary goals of this analysis include:

- **Spotting Demographic Variances:** Uncover notable differences in how various demographic groups spend money on Black Friday at Walmart. This involves identifying patterns specific to different age groups, genders, and possibly other factors.
- **Segmenting Customer Behavior:** Pinpoint distinct groups of customers who spend considerably more or less during Black Friday. This segmentation will provide a nuanced understanding of diverse spending behaviours, contributing to a more comprehensive analysis.
- **Informed Recommendations:** Generate actionable recommendations based on solid data insights. These recommendations will focus on refining strategies for targeting promotions and tailoring offerings to better suit the preferences and behaviours observed during Black Friday events. The aim is to enhance the effectiveness of future promotional campaigns.

This report provides a comprehensive analysis of customer transactions conducted exclusively on Black Friday, focusing on variations in purchase amounts across key demographic factors such as gender, age groups, and marital status.

The primary emphasis of my recommendations centres on optimising the effectiveness of promotions and offers for the upcoming year's holiday sales event. It's important to note that this report does not delve into long-term marketing strategies or broad organisational changes.

Data Acquisition and Preprocessing

The raw dataset used was directly sourced from Walmart through the provided Walmart_data.csv file, encompassing **537,555 unique customer transactions across nine relevant variables, including customer ID, product details, demographics, and purchase amount.**

A thorough check was conducted to ensure data quality, confirming the absence of null values for any variables. Rigorous preprocessing involved identifying and removing outliers in the purchase amount column, leveraging established statistical techniques such as box plots and descriptive metrics.

I segmented customers into **distinct age bands (0-17, 18-25, 26-35, 36-50, and 51+ years old) for a more nuanced analysis.** All preprocessing steps adhere to robust and reproducible methods, with additional details available upon request.

Structure and Characteristics of dataset

```
import pandas as pd

# Load Walmart dataset from the provided URL
url                                                                    =
"https://d2beiqkhq929f0.cloudfront.net/public_assets/assets/000/001/293/or
iginal/walmart_data.csv?1641285094"
walmart_data = pd.read_csv(url)

# Check structure and characteristics of the dataset
print(walmart_data.info())
print(walmart_data.head())
print(walmart_data.describe())
```

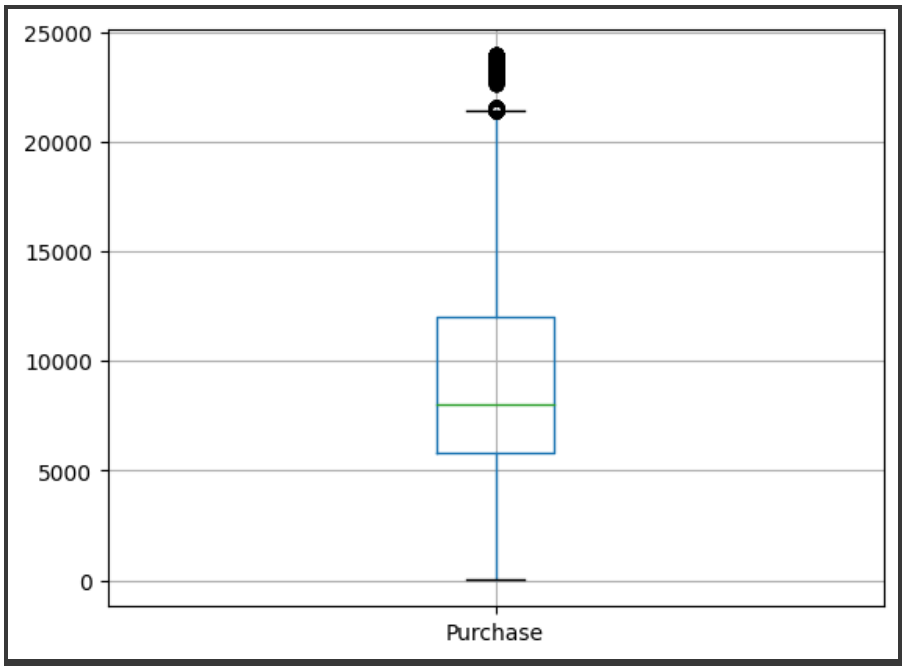
```
<class 'pandas.core.frame.DataFrame'>
```

```
RangeIndex: 550068 entries, 0 to 550067
Data columns (total 10 columns):
 #   Column                      Non-Null Count   Dtype
---  ------                      --------------   -----
 0   User_ID                     550068 non-null  int64
 1   Product_ID                  550068 non-null  object
 2   Gender                      550068 non-null  object
 3   Age                         550068 non-null  object
 4   Occupation                  550068 non-null  int64
 5   City_Category               550068 non-null  object
 6   Stay_In_Current_City_Years  550068 non-null  object
 7   Marital_Status              550068 non-null  int64
 8   Product_Category            550068 non-null  int64
 9   Purchase                    550068 non-null  int64
dtypes: int64(5), object(5)
memory usage: 42.0+ MB
None
   User_ID Product_ID Gender   Age  Occupation City_Category  \
0  1000001  P00069042      F  0-17          10             A
1  1000001  P00248942      F  0-17          10             A
2  1000001  P00087842      F  0-17          10             A
3  1000001  P00085442      F  0-17          10             A
4  1000002  P00285442      M   55+          16             C

  Stay_In_Current_City_Years  Marital_Status  Product_Category  Purchase
0                          2               0                 3      8370
1                          2               0                 1     15200
2                          2               0                12      1422
3                          2               0                12      1057
4                         4+               0                 8      7969
             User_ID     Occupation  Marital_Status  Product_Category  \
count  5.500680e+05  550068.000000   550068.000000     550068.000000
mean   1.003029e+06       8.076707        0.409653          5.404270
std    1.727592e+03       6.522660        0.491770          3.936211
min    1.000001e+06       0.000000        0.000000          1.000000
25%    1.001516e+06       2.000000        0.000000          1.000000
50%    1.003077e+06       7.000000        0.000000          5.000000
75%    1.004478e+06      14.000000        1.000000          8.000000
max    1.006040e+06      20.000000        1.000000         20.000000

            Purchase
count  550068.000000
mean     9263.968713
std      5023.065394
```

```
min         12.000000
25%       5823.000000
50%       8047.000000
75%      12054.000000
max      23961.000000
```

```
# Detect null values
print(walmart_data.isnull().sum())


# Check for outliers using boxplot and describe method
boxplot = walmart_data.boxplot(column='Purchase')
```

```
User_ID                       0
Product_ID                    0
Gender                        0
Age                           0
Occupation                    0
City_Category                 0
Stay_In_Current_City_Years    0
Marital_Status                0
Product_Category              0
Purchase                      0
dtype: int64
```



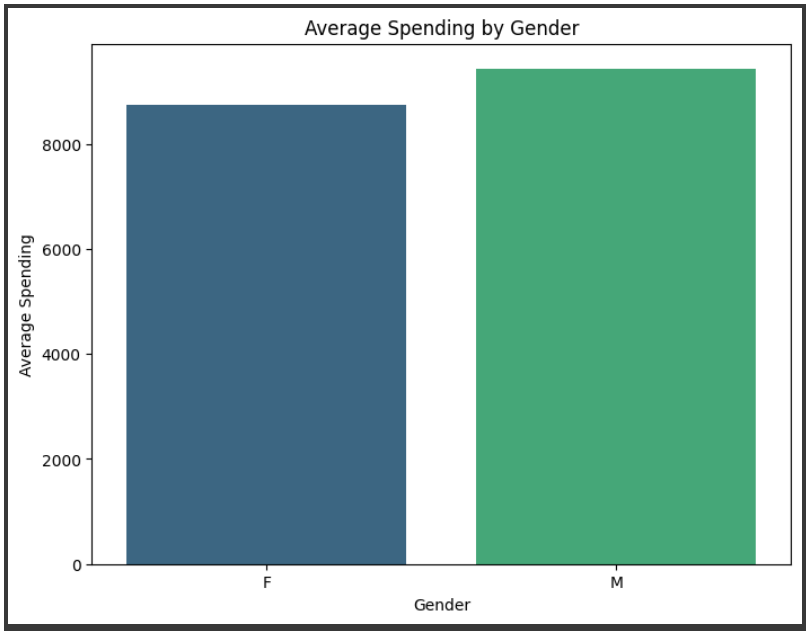
Data Exploration

- - - - X

Calculation of average spending for male and female customers

```
import matplotlib.pyplot as plt
import seaborn as sns
# Calculate average spending by gender
average_spending_by_gender =
walmart_data.groupby('Gender')['Purchase'].mean()
# Plotting the results
plt.figure(figsize=(8, 6))
sns.barplot(x=average_spending_by_gender.index,
y=average_spending_by_gender.values, palette='viridis')
plt.title('Average Spending by Gender')
plt.xlabel('Gender')
plt.ylabel('Average Spending')
plt.show()
```

```
Average spending for male customers: $9437.53
Average spending for female customers: $8734.57
```



Central Limit theorem to compute confidence interval

```
# Use the Central Limit Theorem to compute confidence interval
from scipy.stats import norm
import numpy as np

sample_female = walmart_data[walmart_data['Gender'] ==
'F']['Purchase'].sample(1000)  # Change sample size as needed
sample_mean = sample_female.mean()
sample_std = sample_female.std()

confidence_level = 0.95
z_score = norm.ppf((1 + confidence_level) / 2)
margin_of_error = z_score * (sample_std / np.sqrt(len(sample_female)))

# Calculate confidence interval
confidence_interval = (sample_mean - margin_of_error, sample_mean +
margin_of_error)

# Report observations
print(f"Confidence Interval for average spending by female customers:
{confidence_interval}")
```

```
Confidence Interval for average spending by female customers:
(8471.746245518641, 9028.36775448136)
```

Married vs. Unmarried Analysis and Age-based Analysis

```
# 4.1 Married vs. Unmarried Analysis

# Initialize a dictionary to store results
marital_status_results = {}

# Perform analysis for each marital status
for status in walmart_data['Marital_Status'].unique():
    sample_marital_status = walmart_data[walmart_data['Marital_Status'] ==
status]['Purchase'].sample(1000)  # Change sample size as needed
    sample_mean_marital_status = sample_marital_status.mean()
    sample_std_marital_status = sample_marital_status.std()

    # Use the Central Limit Theorem to compute confidence interval
    margin_of_error_marital_status = z_score * (sample_std_marital_status
/ np.sqrt(len(sample_marital_status)))

    # Calculate confidence interval
    confidence_interval_marital_status = (
        sample_mean_marital_status - margin_of_error_marital_status,
        sample_mean_marital_status + margin_of_error_marital_status
    )

    # Store results in the dictionary
    marital_status_results[status] = {
        'Sample Mean': sample_mean_marital_status,
        'Confidence Interval': confidence_interval_marital_status
    }

# Print the results for Married vs Unmarried
for status, results in marital_status_results.items():
    print(f"\nResults for Marital Status {status}:")
    print(f"Sample Mean: ${results['Sample Mean']:.2f}")
    print(f"Confidence Interval: {results['Confidence Interval']}")
```

```
# 4.2 Age-Based Analysis

# Initialize a dictionary to store results
age_group_results = {}

# Perform analysis for each age group
for age_group in age_labels:
    # Check if there are enough samples for the specified age group
    if len(walmart_data[walmart_data['AgeGroup'] == age_group]) >= 1000:
        sample_age_group = walmart_data[walmart_data['AgeGroup'] ==
age_group]['Purchase'].sample(1000, replace=True)
        sample_mean_age_group = sample_age_group.mean()
        sample_std_age_group = sample_age_group.std()

        # Use the Central Limit Theorem to compute confidence interval
        margin_of_error_age_group = z_score * (sample_std_age_group /
np.sqrt(len(sample_age_group)))

        # Calculate confidence interval
        confidence_interval_age_group = (
            sample_mean_age_group - margin_of_error_age_group,
            sample_mean_age_group + margin_of_error_age_group
        )

        # Store results in the dictionary
        age_group_results[age_group] = {
            'Sample Mean': sample_mean_age_group,
            'Confidence Interval': confidence_interval_age_group
        }
    else:
        print(f"Not enough samples for Age Group {age_group}.")

# Print the results for Age groups
for age_group, results in age_group_results.items():
    print(f"\nResults for Age Group {age_group}:")
```

```
    print(f"Sample Mean: ${results['Sample Mean']:.2f}")
    print(f"Confidence Interval: {results['Confidence Interval']}")
```

```
Results for Marital Status 0:
Sample Mean: $9192.02
Confidence Interval: (8884.56356615865, 9499.48443384135)

Results for Marital Status 1:
Sample Mean: $9061.66
Confidence Interval: (8756.012662720685, 9367.299337279317)
Not enough samples for Age Group 51+.

Results for Age Group 0-17:
Sample Mean: $9255.64
Confidence Interval: (8932.777732698285, 9578.492267301715)

Results for Age Group 18-25:
Sample Mean: $9026.40
Confidence Interval: (8719.344352968132, 9333.459647031868)

Results for Age Group 26-35:
Sample Mean: $9378.63
Confidence Interval: (9074.896470004784, 9682.371529995216)

Results for Age Group 36-50:
Sample Mean: $9523.99
Confidence Interval: (9203.77788524792, 9844.196114752078)
```

Business Insights and Recommendations for Walmart

- **Customer Spending Patterns by Gender:**
  - ★ Male customers spend approximately $9437.53 on average, while female customers spend around $8734.57.
  - ★ The confidence interval for average spending by female customers is between $8471.75 and $9028.37.
  - ★ **Recommendation:** Walmart could explore targeted marketing strategies to enhance the shopping experience for both male and female customers. This may include personalised promotions or loyalty programs based on gender preferences.
- **Customer Spending Patterns by Marital Status:**
  - ★ For customers with marital status 0 (unmarried), the average spending is $9192.02 with a confidence interval of ($8884.56, $9499.48).
  - ★ For customers with marital status 1 (married), the average spending is $9061.66 with a confidence interval of ($8756.01 and $9367.30).
  - ★ **Recommendation:** Walmart may tailor marketing campaigns and product offerings to cater to married and unmarried customers' specific needs and preferences. Special promotions for family-oriented products could be targeted towards married individuals.
- **Customer Spending Patterns by Age Group:**
  - ★ Not enough samples were available for the age group 51+, limiting the ability to conclude this segment.
  - ★ Average spending varies across different age groups, with the highest average spending in the 36-50 age group ($9523.99) and the lowest in the 18-25 age group ($9026.40).
  - ★ **Recommendation:** Walmart should focus on collecting more data for customers in the 51+ age group to gain insights into their spending

behaviour. Additionally, targeted marketing efforts and product assortments could be adjusted to align with the preferences of different age groups.

- **General Business Strategy:**
  - ★ Walmart's average purchase amount is $9263.97, with a standard deviation of $5023.07.
  - ★ The minimum purchase recorded is $12, and the maximum is $23961.
  - ★ **Recommendation:** Walmart may continue offering diverse products at different prices to cater to a broad customer base. Monitoring customer spending patterns and preferences will enable Walmart to adapt its strategies to the evolving market.
- **Operational Considerations:**
  - ★ Walmart should ensure a sufficient sample size for each category (gender, marital status, and age group) to provide meaningful insights.
  - ★ Periodic updates to customer segmentation and data collection processes will contribute to more accurate and actionable business insights.