**Black Friday Sales EDA**

**Final Report – Group 2**

Mihir Patel, Srivamsi Thokkudubiyyapu, Vindhya Veeramachaneni, Sairam Akkineni, and Syam Sundar Akina (Everyone contributed equally)

Department of Information Systems Technologies and Business Analytics, University of North Texas

BCIS 5110: Programming for Business Analytics

Dr. Melody White

Assignment Due Date: Nov 29, 2023, 23:59

**Black Friday Sales EDA**

**Final Report – Group 2**

**Executive Summary**

The Black Friday sales event stands as a pivotal moment in the retail calendar, attracting a massive influx of shoppers eager to capitalize on lucrative deals. This report delves into the rich dataset sourced from Kaggle, offering a comprehensive analysis of Black Friday sales. The primary objectives are to unravel consumer trends, dissect purchasing behavior, and evaluate retailer strategies during this annual shopping extravaganza. To analyze and display the data, the project makes use of the Python programming language and several libraries for data analysis and visualization, including pandas, NumPy, matplotlib, and seaborn. To better understand the patterns and trends of customer behavior on Black Friday Sales, the project contains a variety of data visualizations, including bar charts, pie charts, and line graphs.

**Project Motivation/ Background**

Understanding consumer behavior during Black Friday is crucial for retailers aiming to tailor their strategies to meet the evolving demands of a dynamic market. This project seeks to unravel the patterns of product preferences, purchase timing, and the influence of demographic factors on consumer spending. By doing so, it aspires to provide retailers with actionable insights that can inform targeted marketing campaigns and product offerings. Beyond consumer behavior, the project delves into the strategies employed by retailers during Black Friday. From pricing dynamics and product categorization to the impact of promotions and advertising, the analysis aims to highlight successful tactics and areas for improvement. The goal is to equip retailers with the knowledge needed to refine their approach, enhance customer satisfaction, and maximize the potential of this high-stakes shopping event.

**Data Description**

The dataset has the following features:

1. It has 550068 rows.

2. It has 12 columns.

A screenshot of a computer

Description automatically generated

3. It contains integer, float, and object datatypes.

A screenshot of a computer code

Description automatically generated

Integer: “User\_ID”, “Occupation”, “Marital\_Status”, “Product\_Category\_1”, “Purchase”

Float: “Product\_Category\_2”, “Product\_Category\_3”

Object: “Product\_ID”, “Gender”, “Age”, “City\_Category”, “Stay\_In\_Current\_City\_Years”

A screenshot of a computer

Description automatically generated

4. “Product\_Category\_2” and “Product\_Category\_3” are the only two columns that have null values in them.

A screenshot of a computer

Description automatically generated

5. There are around 70% missing values in “Product\_Category\_3” and around 32% missing values in “Product\_Category\_2”

A screenshot of a computer program

Description automatically generated

* There are 8.44% of missing values in the Dataset.

7. There are no duplicate values in the Dataset.

A screen shot of a computer

Description automatically generated

**Data transformation/Exploratory data analysis**

Since the data was in good format. We didn’t do any transformation.

* With the describe function on "Object" data type, we get the following interesting insights:
  + There are 3631 Products in this Dataset.
  + There are 7 unique Age categories.
  + There are 3 different City categories, and the B category is where most customers are from.
  + The majority of buyers are M (Male) Gender.
  + Stay\_In\_Current\_City\_Years has 5 unique values with 1 year being the most common which means most buyers have been in that city for about a year and are new at the city.
  + The top product sold is P00265242 with 1880 sales.

A screenshot of a computer

Description automatically generated

* Product\_ID P0099942 is the most sold product by City category.

A screenshot of a computer

Description automatically generated

* In this dataset, we have 5891 unique users.
* There are 21 different occupations.

A screenshot of a computer

Description automatically generated

**Analysis**

* The top buyer by age is from 26-35 years accounting for 40% while the bottom buyers are from age 0-17 which is only 3%.
* 78% of the purchases come from people whose ages are between 18 and 45.

A pie chart with numbers and a number

Description automatically generated

A screenshot of a computer program

Description automatically generated

* Buyer that is Male in gender is 414259 (about 75%) while the Female count of buyers is 135809 (25 %).

A screen shot of a computer code

Description automatically generated

A screen shot of a graph

Description automatically generated

* There are 231173 buyers from the "B" category city which is about 42%; while the "A" category city accounts for 147720 buyers which is only 27%.

A computer code on a black background

Description automatically generated

A pie chart with numbers and a few percentages

Description automatically generated

* Marital\_Status 0 = 324731 || Marital\_Status 1 = 225337
* The buyer's Marital Status 0 is more when compared to status 1.

A graph with a bar

Description automatically generated

**A screen shot of a computer

Description automatically generated**

* Occupation 8 customers are lowest and occupation 4 are highest.

A graph with orange bars

Description automatically generated

A screen shot of a computer screen

Description automatically generated

* Most customers have been in the city for about a year.

A screen shot of a computer

Description automatically generated

A graph with green bars

Description automatically generated

**Findings and managerial implications**

***Demographic Insights:***

Through a meticulous examination of user demographics, we identified distinct patterns in participation across gender, age groups, occupations, and city categories. These insights provide a nuanced understanding of the diverse consumer base engaged in Black Friday sales.

***Product Preferences and Categories:***

Analysis of product details revealed specific categories that dominated Black Friday sales. By understanding these preferences, retailers can strategically curate their product offerings to align with consumer demand.

***Purchase Dynamics:***

The study of purchase patterns, including timing and frequency, uncovered trends that shed light on the peak periods of consumer activity. This knowledge is crucial for retailers aiming to optimize their sales strategies.

***Consumer Spending Behavior:***

A deep dive into consumer spending patterns illuminated variations across different segments. Understanding the factors influencing spending can empower retailers to tailor promotions and enhance the overall shopping experience.

***Impact on Retailers:***

An assessment of retailer performance during Black Friday highlighted successful strategies and areas for improvement. The findings offer actionable insights for retailers aiming to refine their approach and maximize the benefits of this high-stakes sales event.

**Conclusions**

By unraveling the dynamics of Black Friday sales, this project seeks to contribute valuable insights to retailers, marketers, and industry stakeholders. In a rapidly evolving retail landscape, understanding the nuances of consumer behavior and optimizing retail strategies are critical for success. Through this analysis, we aim to shed light on the factors that drive Black Friday's success and provide practical recommendations for stakeholders navigating the challenges and opportunities of this extraordinary sales event.

This analysis not only enriches our understanding of Black Friday sales but also provides practical recommendations for both retailers and consumers. By leveraging the revealed insights, stakeholders can make informed decisions, refine marketing strategies, and ultimately enhance the Black Friday experience for all participants.

**Appendix: python codes with proper documentation**

All Python code snippets including the outputs are submitted in another submission (Jupyter NoteBook) in “.ipynb” format.

**References**

1. <https://www.kaggle.com/datasets/pranavuikey/black-friday-sales-eda>