Black Friday Deals Predictions

(By Anoop Garige Marissa Hamilton Gifty Suneeth Vinay Kastur)

# Introduction

Black Friday is an annual shopping event that takes place on the Friday following Thanksgiving. This shopping phenomenon has become deeply established in American culture which markets the beginning of the holiday shopping season. Retailers all over plan for this one day all year. Which sales will bring in the most consumers, or what are going to be the popular items for the year?

Prior to the past couple of years Black Friday was known for opening on Thanksgiving night and everyone waiting at the doors for the best deals. The most popular item for years was a new flatscreen tv at the lowest price you would see all year. But those Black Friday years have changed for the better, with stores staying closed on Thanksgiving so people will spend more time with their families, and in recent years moving to Cyber Monday or Cyber week to promote more online shopping. Electronics, especially televisions, gaming consoles, and smartphones, have traditionally been some of the most popular items on Black Friday. Additionally, clothing, home appliances, and toys are also in high demand.

9 billion dollars in most recent years is spent on Black Friday and the week after, which makes a huge economic impact on those retailers that participate and ultimately leads to larger profits. For most retailers, this one week of the year is monumental for their annual sales and provides a huge opportunity to increase revenue, clear out old inventory, and attract some new customers. This economic boost also ripples through various sectors, including logistics, marketing, and employment, contributing to the overall vitality of the economy during the holiday season.

In recent years, there has been a shift toward online shopping due to the convenience and safety concerns associated with in-person shopping during the holiday season, especially during the COVID-19 pandemic. The biggest questions these retailers have are what old inventory can we put at a discounted rate to get rid of it? And, with recent inflation within the past year will there be a decline in Black Friday transactions?

# /

**Problem Statement 1:** Some problems the retailers might face are excessive markdowns and overstocking of certain products during Black Friday due to inaccurate sales forecasts. This leads to reduced profitability and increases operational challenges. When retailers inaccurately predict the demand for certain products, they may find themselves with excess inventory that needs to be cleared quickly. This often results in excessive markdowns, as retailers’ resort to price reductions to entice customers to purchase these surplus items.Markdowns can significantly impact profit margins of retailers. Selling products at lower prices than originally planned not only decreases potential profits but also diminishes the value of the item in the eyes of the customer.

**Problem Statement 2:** Another problem faced by retailers are difficulties in effectively allocating marketing budgets across various channels during the Black Friday sales period, resulting in sub optimal promotional strategies. This is a multichannel complexity that retailers often utilize multiple marketing channels such as online advertising, social media, and traditional advertising to reach their target audience. The makes more a complex allocation as one must have a good understanding of the customer behavior and preferences of platform.

To make a complex situation even more so the Black Friday period is highly competitive, with numerous retailers trying to obtain the attention of customers. The effectiveness not only relies of the channels of marketing but also how well its stands against the competition.

# Data Characteristics:

We will be taking a dataset of sales transactions captured at a retail store during Black Friday and comparing different features, seeing what trends we're able to analyze with the data provided, and aiming to extract valuable insights that can inform future marketing strategies and inventory management decisions for both retailers and consumers.

First, let's delve into the characteristics of our data sets, which will serve as the foundation for our analysis. The features of our data set are diverse and encompass a wide range of variables that allow us to gain valuable insights into customer behavior during Black Friday.

The dataset comprises 550,069 rows and 12 columns. In the data overview, it is specified that the dataset consists of 537,577 rows (transactions) and 12 columns (features), detailed below: User\_ID, which represents the unique identifier of the user, has a total of 5891 unique users in the dataset. Product\_ID, serving as the unique identifier for the product, is also included. These features include User ID, which acts as a unique identifier for individual customers, enabling us to track and analyze the purchasing patterns of each customer. Product ID is another key feature, providing a specific reference to the items that were purchased, allowing us to discern which products are more popular during this shopping event. Gender, age, marital status, and occupation are customer-related features, and their inclusion in the data set allows us to explore how these demographic factors influence the purchasing decisions of our clientele.

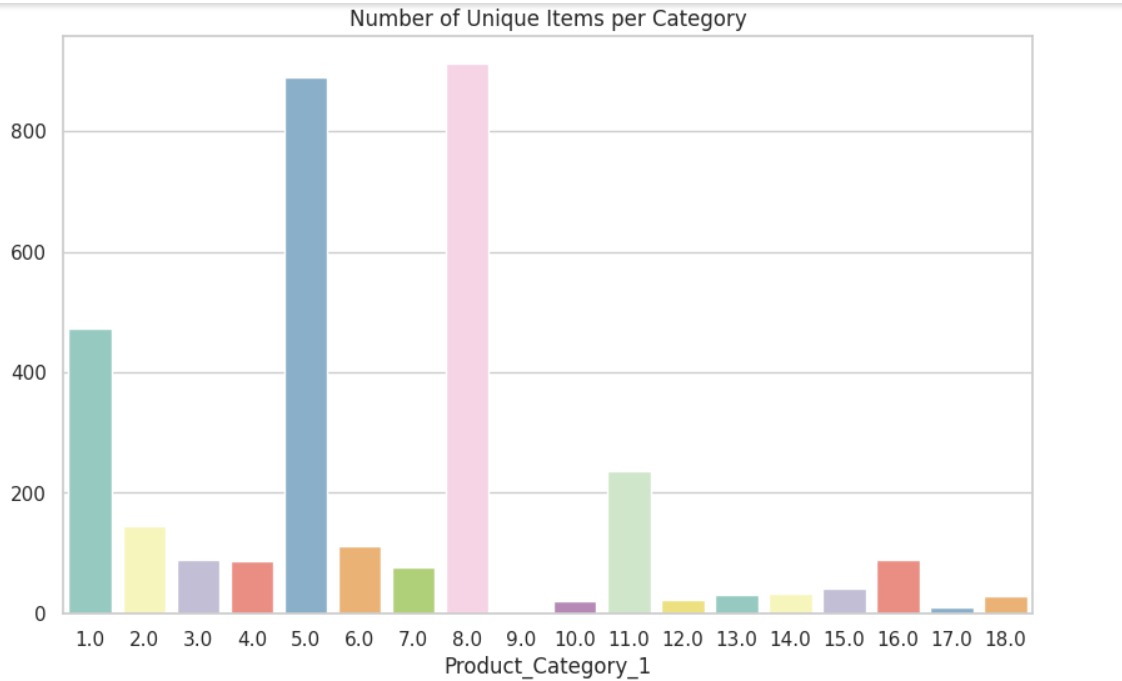
In addition, the data set includes information on cities, which are categorized into three distinct groups ‘A’ which is Tampa,’ ‘B,’ represents St. Louis, and ‘C’ is New York City, and the Stay In current city feature, which informs us about the duration of time each customer has resided in their current city. These geographic aspects offer insights into how location and the length of residence might impact purchasing behavior. Lastly, the two features, product category, and purchase amount, are central to our analysis, as they provide a direct link to the items being purchased and the monetary value of those transactions. Together, these features will empower us to unravel the intricate dynamics of Black Friday shopping trends.

# DATA VISUALIZATION:

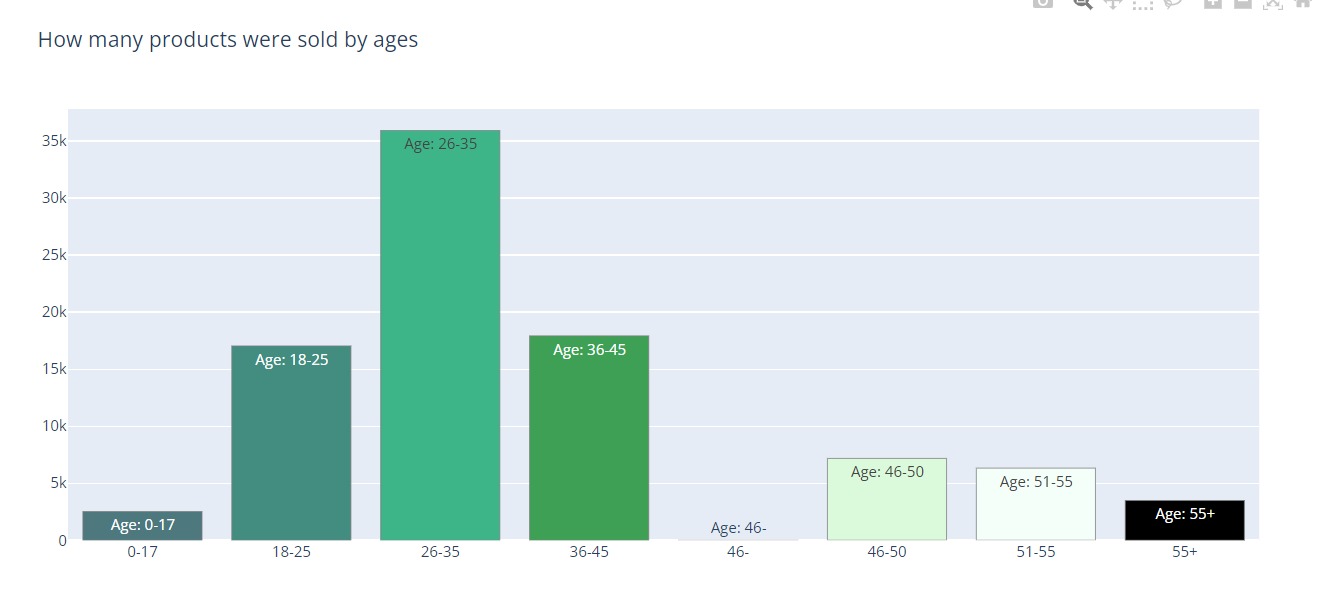
As we progress in our analysis, the rich tapestry of features in our data set will enable us to uncover patterns, correlations, and meaningful associations that can guide retailers and marketers in making informed decisions, optimizing their product offerings, and enhancing the Black Friday shopping experience for customers. By mining this data, we aim to gain a deeper understanding of the factors that drive consumer behavior during this critical shopping season, ultimately contributing to the success of both retailers and shoppers alike.

With these features, we're hoping to make Black Friday purchase predictions based on the customer specifications and where the customer is located, thereby enhancing the shopping experience by tailoring recommendations and deals to individual preferences and geographic factors. We set parameters for various customer-related features within our dataset. One straightforward parameter involved split up purchases by gender, revealing a surprising find. It was determined that, on Black Friday, the number of transactions performed by males was almost three times that of females. Throughout the year, marketing predominantly targets women and their spending habits while shopping. This insight could be highly valuable for retailers as they plan the items to feature in their Black Friday ads going forward. It may lead them to emphasize products such as grills, electronics, and games.

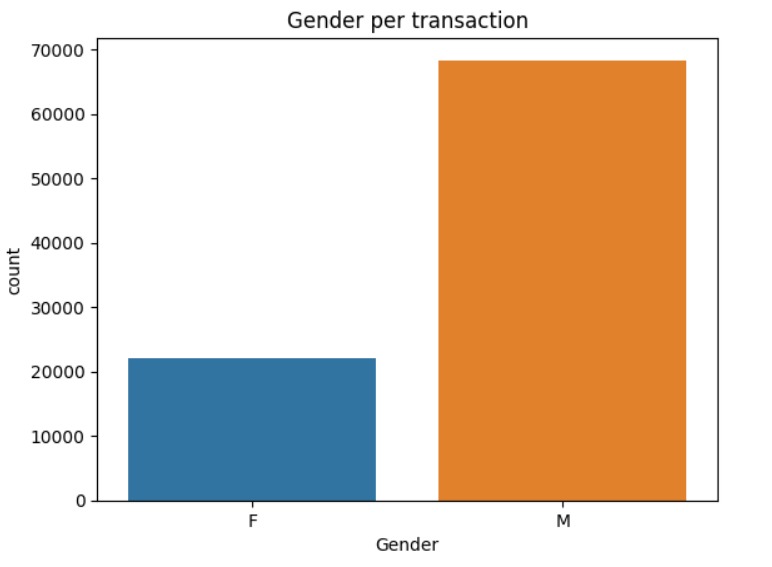
# ANALYSING OF THE DATA:



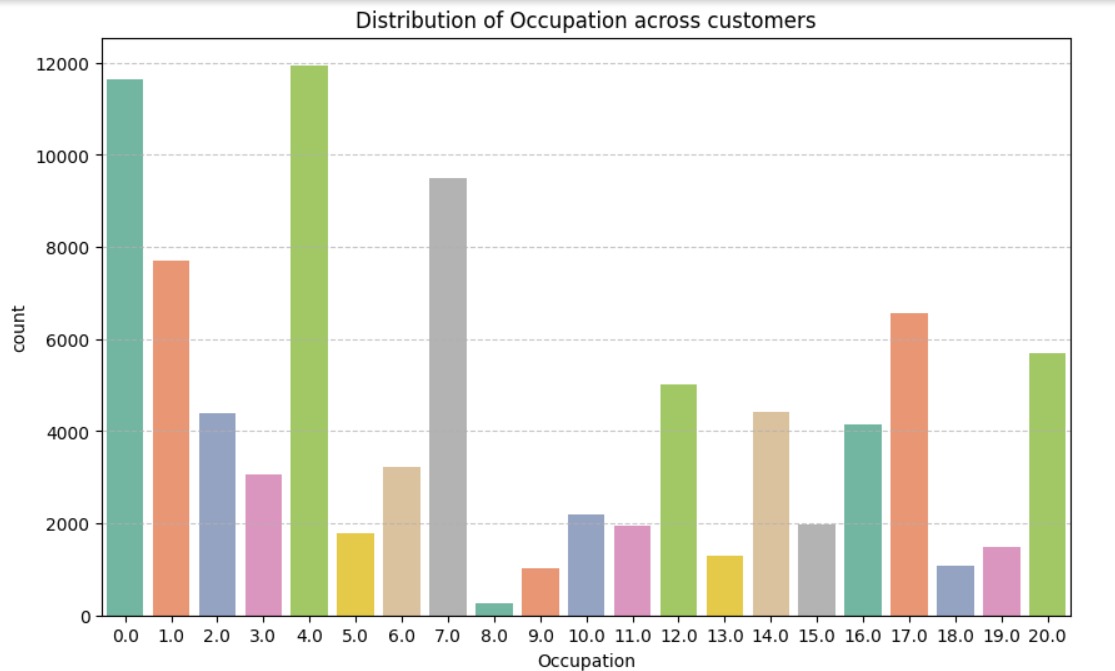
Another feature we set parameters on was age range. We used the following parameters: Ages 0-17, 18-25, 26-35, 36-45, 46-50, 51-55, and 55+. This analysis showed the most products sold to ages between 26-35. This aligns with the common perception that a significant portion of spending occurs in one’s late 20s and early 30s.

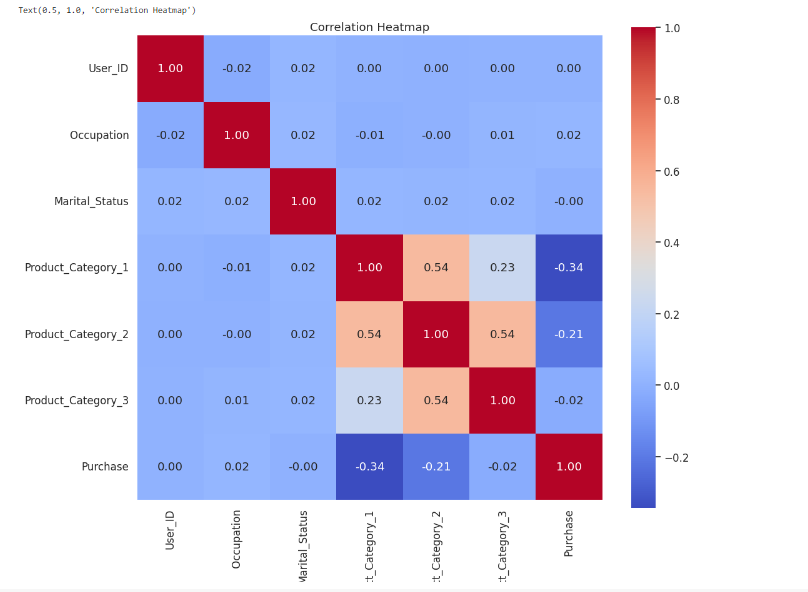


Some of the data we set parameters for were customer related features. A simple parameter of splitting up the purchases by gender. A little surprising to find that gender per transaction was almost triple Male to female in the number of transactions performed on Black Friday.



Another feature we set parameters on is distribution of occupation across customers. In short some of the occupations were Nurse, Sales Representative, Police Officer, Restaurant Server, and Bus Driver. This data was analyzed, and the most transactions performed were done so by individuals that worked as Nurses. After doing these basic analyses to identify simple trends, such as which gender, age, and occupation performed the most transactions on Black Friday, we were able to delve into more complex models to gain deeper insights and refine our understanding of consumer behavior during this retail phenomenon.





# Regression Models

One of the models we used to perform our analysis was a linear regression model. In this analysis, we utilized a Linear Regression model to predict Black Friday sales based on the features in our dataset. Here are the key takeaways:

LINEAR REGRESSION:

The linear regression model allowed us to establish a relationship between the various features in our dataset and the total Black Friday sales. By diving into the factors such as User ID, Product ID, Gender, Age, Marital Status, Occupation, City category, stay in current city, product category, and purchase amount, we were able to identify which variables had a significant impact on sales figures.

**Model Training:** We trained a Linear Regression model using the provided dataset. The model aimed to establish a linear relationship between the input features and the purchase amount, making it a straightforward choice for initial analysis, which allowed us to assess the impact of these key factors on Black Friday sales with a clear and interpret-able approach.

**Mean Absolute Error (MAE):** The model's MAE, a measure of the average absolute prediction error, was found to be 3575.52. This means that, on average, the model's predictions were off by approximately $3575.52 concerning the actual purchase amounts.

**Mean Squared Error (MSE):** The model's MSE, which quantifies the magnitude of prediction errors, was calculated to be 21,849,890.00. This metric provides an understanding of the squared differences between predictions and actual values.

***Summary***

In summary, the Linear Regression model provided a baseline for Black Friday sales prediction. While it showed potential for prediction, it is essential to note that more complex models and feature engineering may be required to further improve accuracy. The identified MAE and MSE values serve as reference points for evaluating the performance of more advanced models in our analysis. The other model we used to perform our analysis was a decision tree regression model. The Decision Tree Regression model was employed to predict Black Friday sales using the provided dataset. Here are the key points regarding the results:

DECISION TREE MODEL:

**Model Training:** We created and trained a Decision Tree Regression model to capture complex, non-linear relationships within the dataset. Decision trees can partition the feature space into segments and assigning values to those segments.

**Model Score:** The Decision Tree model achieved a score of 0.5647 when evaluated on the test data. This score, also known as the coefficient of determination (R-squared), indicates that the model explains approximately 56.47% of the variance in the Black Friday sales, which is a reasonably good fit.

**Mean Absolute Error (MAE):** The MAE for the Decision Tree model was found to be 2344.73. This metric measures the average absolute prediction error, indicating that, on average, the model's predictions deviated by approximately $2344.73 from the actual purchase amounts.

**Mean Squared Error (MSE):** The Decision Tree model's MSE was calculated to be 10,938,299.59. MSE quantifies the squared differences between predictions and actual values, providing insights into the magnitude of prediction errors. Lastly, we utilized a MLP regressor which is a neural network model with potential for complex patterns. However, it shows convergence issues.

MLP REGRESSOR:

The MLPRegressor model has been trained and evaluated for Black Friday sales prediction. The warning suggests that the model may not have fully converged. The model's performance is measured in terms of R-squared, MAE, and MSE. The MLP Regressor had a MAE of 2998.32 and a MSE of 16839780.71.

By using this format, you can present the key information related to your MLPRegressor model concisely and clearly to your audience.

# Comparsion:

Now that we've done more complex models with our Black Friday sales dataset, we can put those results together to compare.

Decision Tree model with a test data R-squared score of 0.5647, Mean Absolute Error (MAE) of 2344.73 and a Mean Squared Error (MSE) of 10,93. The Linear regressions model MAE was 3575.52 and the MSE was 21,849,890.00. Lastly, the MLP Regressor had a MAE of 2998.32 and a MSE of 16839780.71.

# EVALUATION:

Our evaluation was that the decision tree regression stood out as a strong Black Friday sales prediction model. The decision tree regression model achieved the lowest MAE and MSE, suggesting it had the best predictive accuracy amount the models for this given task. The MLP Regressor, while having a lower R-squared score, may require further optimization due to the convergence warning.

# Solutions:

Some of the solutions our group came up for our problems were as follows:

In the effort to enhance operational efficiency and maximize the revenue during Black Friday shopping event, our goal is to construct an experienced regression model. This model will not only leverage an analysis of historical sales data, customers preferences, and external factors of the past but also capture the nature of customer behavior and the impact of external variables. Some of these external variables might include economic indicators, seasonal trends, and marketing initiatives which will enrich the models’ predicting capabilities.

The envisioned regression model serves as a valuable tool for retailers to anticipate which products are likely to experience larger demand during Black Friday. A retailer that holds these insights will be able to adjust pricing strategies and maintain the correct amount of inventory.

Additionally, the insights derived from the model will allow the retailers to make informed decisions on allocation of resources, providing them with the resources to streamline the logistics. By aligning the stock levels with demand there would be a perfect arrangement of supply and demand in their inventory.

Implementing a regression model to analyze the impact of different marketing channels on Black Friday sales, by considering factors such as online advertising, social media promotions, and in-store campaigns, the model can provide insights into the most effective channels. Retailers can use this information to allocate marketing budgets strategically, ensuring maximum reach and impact. By relying on data and statistical analysis, retailers can mitigate the risk of making decisions based on opinion and assumptions.

# Conclusion:

In conclusion, having the models like decision tree, multi-layer perception regressor, and linear regressor only promotes a more logical solution for problems that retailers face when participating in Black Friday week. In all it would allow them to allocate the budget efficiency and enhance their decision-making processes. The synergy of advanced models empowers retailers to make intelligent decisions across inventory management, advertising, and customer interactions.