

Black Friday Sales Insights: How Demographics Influence Purchasing Behavior

Dalvi Mayur, Dinesh Subashree, Nemade Suyog, Singh Aryaman

Abstract- This study explores the complex dynamics of consumer behavior around Black Friday deals, emphasizing the role that demographics play in shaping buying habits. To analyze and comprehend the patterns, we used statistical techniques including correlation analysis and hypothesis testing on the extensive black-Friday dataset, which contains a variety of client demographics and purchase information. A thorough investigation of the connections between age, gender, marital status, occupation, and purchasing behavior was made possible by the dataset's abundance of detailed information. Our results show notable differences in the product categories that different age groups and marital statuses choose to buy, as well as considerable variances in purchasing preferences among the different demographic groupings. For example, younger age groups exhibit clear preferences for product categories, and the volume and kind of purchases were influenced by marital status. These findings have important ramifications for retailers as they point to the possibility of more specialized marketing campaigns and inventory control based on client demographics. The study advances more than just our knowledge of customer behavior.

Table of contents

1. Introduction	2
2. Data	2
2.1. Table Schema	2
2.2. Source and Influence of Bias	2
3. Methods	3
3.1. Exploratory Data Analysis (EDA)	3
3.2. Methods Used	3
3.3. Hypothesis Test	4
4. Results	8
5. Conclusion	11
6. Reference	12

1. INTRODUCTION

The term "Black Friday," which has its roots in American culture, refers to the Friday after Thanksgiving and has traditionally signaled the start of the holiday shopping season. This is one of the biggest retail events of the year, recognized for its huge discounts and promotional deals. Due to its distinct characteristics, including a wide spectrum of customers and a high volume of transactions, Black Friday offers an unparalleled chance to thoroughly examine consumer behavior. This study explores how age, gender, and marital status affect customer purchase behavior during Black Friday discounts. It raises an important question: How do these demographic factors affect consumer spending behavior on Black Friday? The significance of this question extends beyond academic interest, as understanding consumer behavior is pivotal for retailers and marketers. This knowledge is especially vital on black Friday, a day characterized by intense competition among the retailers and heightened customers activity. While the previous research has provided valuable insights into consumer behavior in various retail contexts, there remains a gap in focused studies on black Friday, particularly regarding the impact of demographic factors on purchasing decisions.

2. DATA

The study's information was painstakingly gathered during a Black Friday sale, as reported by Chandel (2018). The dataset, which included a wide range of customer transactions unique to Black Friday, was created from the sales records of sizable retail establishments. The technique for gathering data, which mostly depends on the transactional systems of the shop, offers a comprehensive and deep record of customer purchases, but it has drawbacks and biases that should be considered.

2.1 TABLE SCHEMA

The dataset incorporates several critical features that are central to our analysis:

- a) Product ID: this identifier allows for the tracking and categorization of individual products, essential for analyzing purchasing trends.
- b) Gender: categorized as male or female, providing insights into gender-based purchasing preferences.
- c) Age: grouped into categories, this feature helps in understanding how age influence shopping behavior.
- d) Occupation: represented by 21 distinct categories, offering a glimpse into how occupation might relate to consumer choices.
- e) City category: labeled as 'A', 'B', or 'C', it reflects the type of areas the consumer resides in, which could influence shopping patterns.
- f) Marital status: indicated as single or married, this factor is analyzed to see if marital status affects purchasing decisions.
- g) Product category: encompasses various product types, crucial for identifying popular items or categories among different demographic groups.
- h) Purchase amount: the actual amount spent, providing direct insights into spending behavior.

2.2 SOURCE AND INFLUENCE OF BIAS

The dataset, while comprehensive, is subject to several potential biases:

- a) Regional bias: the dataset is indicative of the consumer demographic in the store's specific location. This could limit the applicability of the findings to other regions with different demographic compositions.
- b) Product range bias: the dataset's insights are influenced by the variety and types of products offered by the store. This could skew data towards certain categories, affecting the overall analysis of consumer preferences.

- c) Temporal bias: as the data represent sales from a specific year, it may not account for evolving consumer trends or economic factors influencing purchasing behavior in subsequent years.
- d) Gender bias: the dataset categorizes gender binary as male or female, which might overlook the purchasing behavior of non-binary or gender-nonconforming individuals. This binary classification could limit the understanding of a more nuanced gender influence on purchasing patterns.

3. METHODS

3.1 EXPLORATORY DATA ANALYSIS (EDA)

We carried out a thorough exploratory data analysis (EDA) before doing hypothesis testing. In this process, the distribution of the important variables was visualized, trends were found, and the dataset was examined for abnormalities or outliers. To better grasp the context of the next statistical tests, our EDA concentrated on describing the key data features, frequently utilizing graphical representations.

3.2 METHODS USED

In our analysis, we employed several statistical tests, each suitable for the type of data and the specific hypothesis being tested:

- a) T-Test: The t-test is a statistical method used to determine if there is a significant difference between the means of two groups.

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Where: \bar{X}_1 and \bar{X}_2 are the sample means of the two groups, s_p is the pooled standard deviation, n_1 and n_2 are the sample sizes of the two groups.

- b) Chi – Square Test: The Chi-Square Test is a statistical test used to determine if there is a significant association between categorical variables in a contingency table.

$$\chi^2 = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Where: O_{ij} is the observed frequency in cell (i,j), E_{ij} is expected frequency in a cell (i,j)

- c) ANOVA TEST: The Analysis of Variance (ANOVA) test is a statistical method used to assess whether there are any statistically significant differences between the means of three or more independent (unrelated) groups.

$$F = \frac{\text{Between-Group Variability}}{\text{Within-Group Variability}}$$

Where: Between-Group Variability is the variance among the means of different groups and Within-Group Variability is the average variance within each group.

3.3. HYPOTHESIS TESTS

3.3.1 EVALUATE SPENDING HABITS BETWEEN MALE AND FEMALE CUSTOMERS.

- a) Population: Black Friday Sales in the United States
- b) Sample: Purchases made by Male and Female customers
- c) Objective: Evaluate spending habits between male and female customers
- d) Null Hypothesis (H0): The average purchase amount for male and female customers is the same.

$$H_0: \mu_{male} = \mu_{female}$$

- e) Alternative Hypothesis (H1): There is a significant difference in the average purchase amount between male and female customers.

$$H_1: \mu_{male} \neq \mu_{female}$$

T-Test

Significance Value	T-statistic	P-value
0.05	46.3582	~ 0.00

Mean Purchase Amount by Gender

Gender	Mean Purchase Amount
Male	9437.5260
Female	8734.5657

Test Interpretation:

The computed t-statistic is 46.36, indicating a substantial difference between the average purchase amounts of male and female customers. The p-value is less than the chosen significance level ($\alpha = 0.05$). Therefore, we reject the null hypothesis. The direction of the difference in means is given by the sign of the t-statistic. In this case, the positive t-statistic (46.36) indicates that male customers, on average, have a higher purchase amount compared to female customers. Based on the t-test, we can conclude that, in the hypothetical data, there is a statistically significant difference in average purchase amounts between male and female customers, and male customers tend to have a higher average purchase amount

3.3.2 MARITAL STATUS AND CITY CATEGORY ARE INDEPENDENT

- a) Population: Here the population is the Black Friday Sales in United States
- b) Sample: The sample dataset is the purchases made by married and single customers in different city categories.
- c) Objective: We claim that Marital status and city category are independent.
- d) Null Hypothesis (H0): Marital status and city category are independent.
- e) Alternative Hypothesis (H1): Marital status and city category are not independent; there is an association between them.
- f) Test: We can use chi-square test as we need to find the extent of association between two categorical variables.

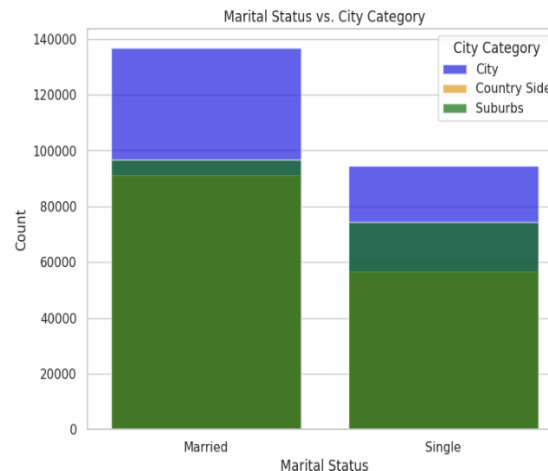
Chi-Square Test

Significance Value	Chi-Square Statistic	P-value
0.05	870.9131	7.64e-190

Test Interpretation:

The chi-square statistic is a measure of the extent of the association between the two categorical variables. In this case, the large chi-square value suggests a substantial association between Marital Status and City Category. The p-value is extremely small, significantly below the common significance level of 0.05. This indicates strong evidence against the null hypothesis. The rejection of the null hypothesis suggests that there is a significant association between Marital Status and City Category. In other words, these two variables are not independent of each other. While the test shows an association, it doesn't provide information on the strength or direction of the association. Additional analyses or visualization techniques will help with a more detailed understanding.

To Visualize this, we can create a stacked bar chart with the contingency table.



This visualization provides a clear picture of how the counts are distributed across marital status and city categories. It seems that the distribution of married individuals is more skewed towards the "City" category, while single individuals have a more balanced distribution across "City" categories.

3.3.3 IS THERE ANY SIGNIFICANT DIFFERENCE IN PURCHASE AMOUNTS BETWEEN DIFFERENT AGE GROUPS

- Population: Here the population is the Black Friday Sales in United States
- Sample: The sample dataset is the purchases made by people in different age groups.
- Objective: We claim that there is no significant difference in purchase amounts between different age groups
- Null Hypothesis (H0): There is no significant difference in purchase amounts across age groups.
$$H_0: \mu(0-17) = \mu(18-25) = \dots = \mu(55+)$$
- Alternative Hypothesis (H1): There is a significant difference in purchase amounts across at least one pair of age groups.
$$H_1: \text{At least one pair of means is different.}$$
- Test: We can use analysis of variance (ANOVA), which compares means across multiple groups.

ANOVA Test

Significance Value	F-Statistic	P-value
0.05	98.4737	~ 0.00

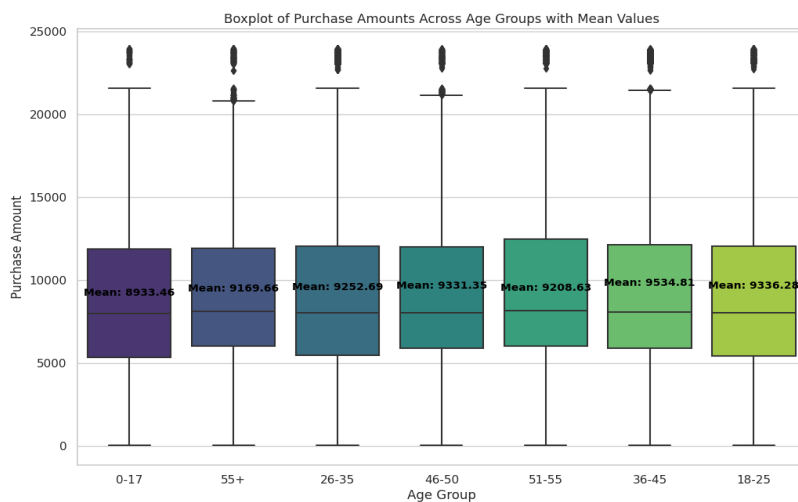
Mean Purchase Amount by Age Group

Age group	Mean purchase amount
0-17	8933.46
18-25	9336.28
26-35	9252.69
36-45	9534.81
46-50	9331.35
51-55	9208.63
55+	9169.66

Test Interpretation:

The result of the ANOVA test indicates a statistically significant difference in purchase amounts across different age groups. The F-statistic is a measure of the variability between groups relative to the variability within groups. In this case, the high F-statistic of 40.58 suggests that there is a significant difference in the average purchase amounts across different age groups. The p-value is extremely low, essentially zero. This indicates strong evidence against the null hypothesis. The rejection of the null hypothesis suggests that at least one pair of age groups has different average purchase amounts. In other words, the average purchase amounts are not the same across all age groups.

To visualize this, we can create a box plot.



The boxes are noticeably different in length and the medians are at different positions, it suggests variability in the central tendency of purchase amounts among age groups. The significant difference in means (as indicated by the hypothesis test results) is reflected in the visual differences in the boxplot. In summary, the boxplot with mean values serves as a complementary visualization to the hypothesis test, offering a more intuitive understanding of how purchase amounts vary across different age groups. The observed differences in central tendency and spread in the plot are consistent with the rejection of the null hypothesis, providing additional insight into the practical significance of the findings.

3.3.4 IS THERE ANY SIGNIFICANT DIFFERENCE IN PURCHASE AMOUNTS ACROSS DIFFERENT DURATIONS OF STAY IN THE CURRENT CITY

- a) Population: Here the population is the Black Friday Sales in United States
- b) Sample: The sample dataset is the purchases made by people with duration of stay in the current city.
- c) Objective: We claim that there is no significant difference in purchase amounts across different durations of stay in the current city.
- d) Null Hypothesis (H0): There is no significant difference in purchase amounts across different durations of stay in the current city.

$$H0: \mu(1 \text{ year}) = \mu(2 \text{ years}) = \dots = \mu(4+ \text{ years})$$
- e) Alternative Hypothesis (H1): There is a significant difference in purchase amounts across at least one pair of duration categories.

$$H1: \text{At least one pair of means is different.}$$
- f) Test: We can use analysis of variance (ANOVA), which compares means across multiple groups.

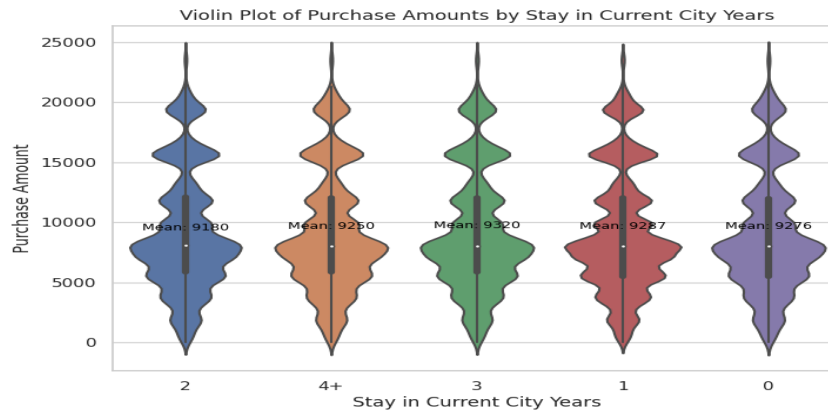
ANOVA Test

Significance Value	F-Statistic	P-value
0.05	40.5757	1.05e-49

Test Interpretation:

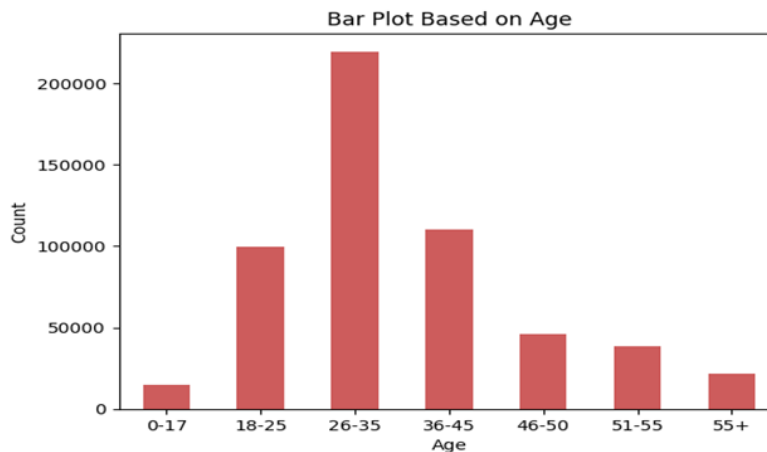
The result of the hypothesis test indicates a small difference in purchase amounts across different durations of stay in the current city. The F-statistic is a measure of the variance between groups relative to the variance within groups. In this case, it is 9.38. The p-value is below the significance level of 0.05. Since the p-value is below the significance level, we reject the null hypothesis. The small p-value suggests that the observed differences in purchase amounts among different duration categories are unlikely to have occurred by random chance alone. In summary, the statistical analysis provides strong evidence to conclude that there are significant differences in purchase amounts across different durations of stay in the current city.

This can be better visualized by using a violin plot.



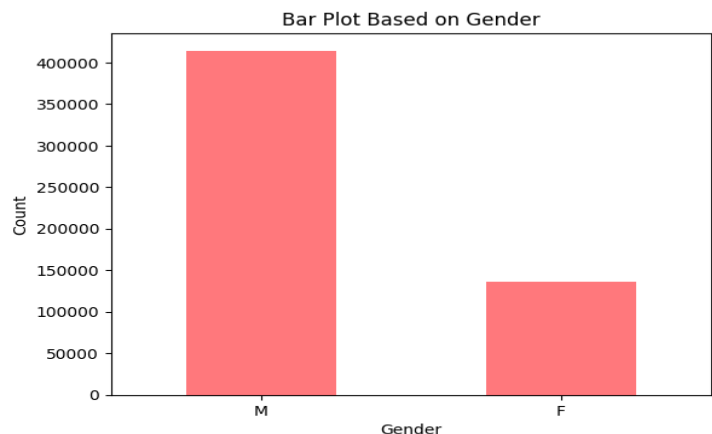
The F-statistic is calculated as part of the analysis of variance (ANOVA) test. It compares the variability between the group means to the variability within the groups. A large F-statistic suggests that the means of at least one pair of groups are significantly different. In our test the F-statistic is 9.382. In the context of the means in the plot above, if the F-statistic is large and the associated p-value is below the chosen significance level (commonly 0.05), it indicates that there are significant differences in purchase amounts across different durations of stay. Here, the difference with mean is not huge. While statistical significance is important, it's also crucial to consider practical significance. Even if the F-statistic is significant, we assess whether the observed differences are practically meaningful or substantial. In summary, the F-statistic helps assess whether there are significant differences in purchase amounts across different durations of stay in the current city. If the F-statistic is large and the associated p-value is below the significance level, it suggests that at least one pair of duration categories has different mean purchase amounts.

4. RESULTS



The image's bar plot displays the number of purchases made during the Black Friday sale by various age groups. The age group '26-35' has many more purchases than any other, suggesting that they are the most active in terms of participating in Black Friday discounts. Next are the age categories '36-45' and '18-25', with '36-45' having a little greater count than '18-25'. Although not nearly as prominent as the '26-35' age bracket, both make significant contributions to the sales. As we get older age groups, the number of purchases decreases dramatically. There is a modest difference in the count between the age categories '46-50' and '51-55', with the former having a lower count than the latter. Age groups '0-17' and '55+' have the fewest purchases overall; '55+' has more purchases than '0-17', but still has a significantly lower total than the other age groups. A merchant or marketer could deduce from this bar plot that younger adult—especially those between the ages of 26 and 35—

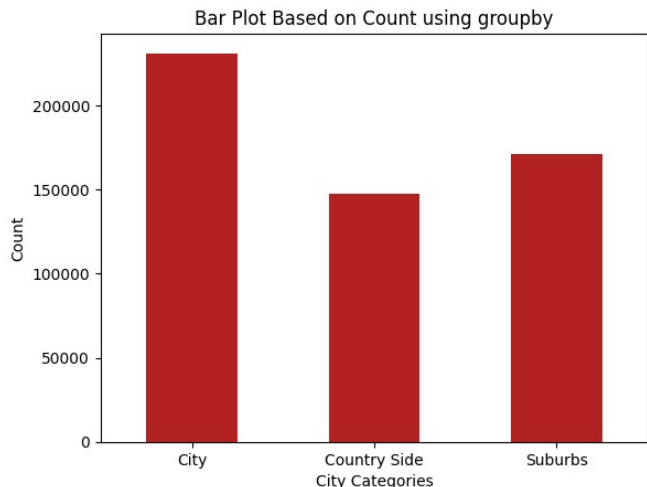
are the key purchasers during Black Friday sales and should be the focus of marketing initiatives. Furthermore, as they now make up a lower portion of sales, there is the opportunity to interact more with the older demographics ('46-55+' years) and the youngest age group ('0-17').



The bar plot depicts the count of purchases made by each gender during the black Friday sale. From this plot we can interpret that:

- a) Males (denoted by 'M') have a significant higher count of purchase compared to females (denoted by 'F').
- b) Females (denoted by 'F') have a significant lower count of purchase compared to males (denoted by 'M').
- c) The count for males is over twice as high as that for females, indicating that males were more active buyers during this black Friday event.

This observation implies that, given the increased engagement of male consumers during this sales time, marketing methods and product positioning may be maximized. Because men make up a bigger portion of purchases than women, retailers should think about adjusting their promotions and product offerings to suit their preferences. It also draws attention to a chance to investigate the reasons behind the lower involvement rate among women and create plans to increase interaction with female consumers.



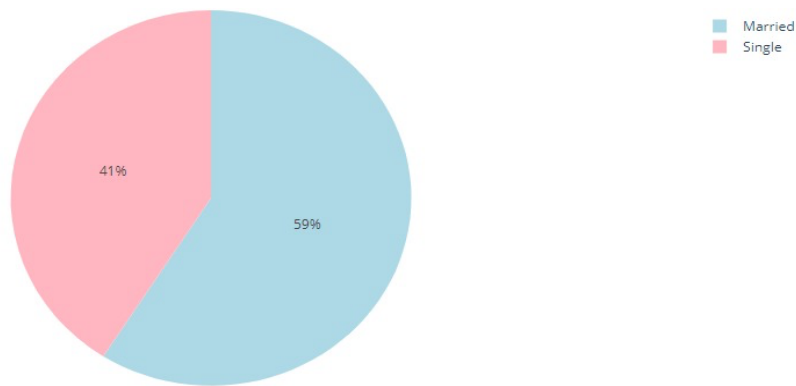
The bar plot shows the count of purchases grouped by city categories: city, countryside, and suburbs.

The insights from the graph:

- a) Consumers from the city category made the most purchases, with the count significantly higher than the other two categories.
- b) The countryside category has the lowest count of purchases, which indicates that the consumers from the county areas participated less in the black Friday sales event compared to those from cities and suburbs.
- c) The suburbs category has a count that is notably less than the city but still substantially higher than the countryside, suggesting a moderate level of participation in black Friday sales.

This data may indicate that a greater concentration of consumers in urban areas is taking advantage of Black Friday bargains, or that access to sales—whether through physical stores or online—may be more favorable in metropolitan areas. There could be several reasons for the lower participation rate from rural areas, including limited access to stores and a decreased focus on Black Friday sales. Retailers may use this information to help them customize their distribution networks and marketing plans to boost sales and reach various city categories.

Pie Chart Based Marital Status

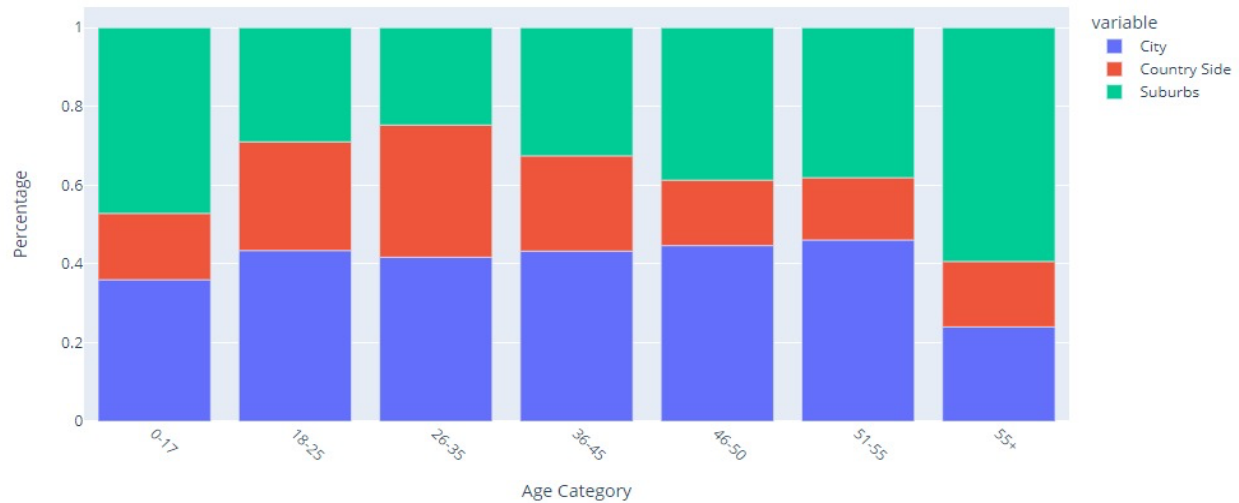


The pie chart represented the distribution of purchases made during the black Friday sales, categorized by marital status. The chart shows two segments:

- a) Individuals who are married represented 59% of the purchase, making them the majority in this dataset.
- b) Single individuals account for 41% of the purchases.

This insight suggests that married individuals may prioritize black Friday sales, potentially due to family-oriented purchases or combined spending power. This demographic could be a key target for business, particularly for products that cater to families or households. The data also indicates that while single individuals make up a smaller portion of the sales, they are a significant market segment. Marketing strategies that appeal to the unique need and preferences of single consumers could help business tap into this substantial consumer base. Tailoring promotions and products to meet the distinct needs of married and single shoppers could result in more effective sales strategies during major retail events like black Friday.

Normalized Stacked Bar Chart



The normalized stack bar chart displays the distribution of purchase made during black Friday across the various age categories, with the data further categorized by the shopper's city type: city, countryside, and suburbs.

From the graph, we can get insights the following:

- In every age category, the largest segment of purchase is consistently made by those from the city. This implies that city residents are the most active participants in black Friday sales regardless of the age.
- The suburbs contribute the second-largest segment of purchase across all age groups, suggesting that while suburban shoppers are less active than city shoppers, they still participate significantly in black Friday sales.
- The countryside has the smallest share of purchases across all age groups. However, this segment is not negligible, indicating that rural areas still contribute to black Friday sales, albeit to a lesser extent than urban and suburban areas.
- The proportion of purchases from the city, countryside, and suburbs appears relatively stable across different age groups, suggesting that the location influence on shopping behavior is consistent across different ages.

5. CONCLUSION

Our analysis of Black Friday sales data has given us a thorough understanding of consumer purchasing patterns and how they relate to various demographic variables. The study analyzed a dataset with many parameters, including age, gender, marital status, and city category, using sophisticated statistical techniques. Our analysis's findings showed several trends, including the following: city dwellers made the most purchases compared to those from the suburbs and countryside; single people made a higher percentage of purchases than married people; males outpaced females in terms of purchase volume; and younger adults, especially those between the ages of 26 and 35, were the most active shoppers.

The research topic, which aimed to comprehend how demographic characteristics influence purchase behavior during Black Friday discounts, is closely linked to these findings. According to the insights, age, gender, and marital status are important demographic factors that influence consumer behavior. For retailers, this means that focusing on demographic groups that are more likely to participate in Black Friday shopping could significantly optimize marketing strategies and inventory planning.

Future research could expand on this by utilizing a multi-year dataset to take trends over time into account, including more demographic variables like education level or income level, and contrasting the observed behaviors with other noteworthy sales times of the year. To further enhance the quantitative results of this study, qualitative research could be carried out to comprehend the reasons behind the identified purchasing trends. The intention would be to provide a more comprehensive picture of consumer behavior that could be used to improve the shopping experience before, during, and after Black Friday.

6. REFERENCES

- [1] Dataset: "Black Friday Sales - Train Dataset"
Source: Analytics Vidhya Black Friday Hackathon
URL: <https://datahack.analyticsvidhya.com/contest/black-friday/>
- [2] Goldman, Daniel. (2018). The Basics of Hypothesis Tests and Their Interpretations. 10.31219/osf.io/u2csn.
- [3] Emmert-Streib F, Dehmer M. Understanding Statistical Hypothesis Testing: The Logic of Statistical Inference. *Machine Learning and Knowledge Extraction*. 2019; 1(3):945-961.
<https://doi.org/10.3390/make1030054>
- [4] Kolawole, Ayotunde & Sekumade, Adelomo. (2017). Hypotheses and Hypothesis Testing. 10.13140/RG.2.2.28299.39202.
- [5] Anderson, D. R., Burnham, K. P., & Thompson, W. L. (2000). Null Hypothesis Testing: Problems, Prevalence, and an Alternative. *The Journal of Wildlife Management*, 64(4), 912–923.
<https://doi.org/10.2307/3803199>