

Research on the Influencing Factors of Consumer Buying Behavior

Jiaqi Cai *

Department of Mathematics and Computer Science, New York University, New York, USA

* Corresponding author: jc10236@nyu.edu

Abstract. Today's businesses depend greatly on the buying habits of its customers. Understanding consumer behavior is essential for businesses to optimize their marketing strategies and boost customer satisfaction in today's fiercely competitive industry. This study examines customer personalities and their influence on purchasing behavior by using the purchase histories and demographic information of 2,240 clients. 10 research sources from Google Scholar are used to analyze the key factors that affect consumer behavior and the effectiveness of different clustering algorithms in identifying distinct customer groups in order to improve the accuracy of the study. The study emphasizes how important lifestyle, psychographic, and demographic characteristics are for affecting customer behavior. Businesses may better understand their clients and create marketing strategies that are tailored to their requirements by utilizing these characteristics. The work also raises the possibility of using clustering techniques to improve marketing efforts and increase client loyalty. Businesses may more effectively target their marketing efforts and boost customer satisfaction by categorizing customers into several categories based on their preferences and behavior patterns. As a result, this research offers insightful information about customer purchase habits. Businesses may better understand their clients and target their marketing initiatives to fit their demands by employing demographic and behavioral data. In doing so, companies may boost client pleasure and loyalty, eventually leading to long-term success.

Keywords: Consumer behavior, demographic data, purchase behavior analysis, linear regression.

1. Introduction

Understanding client purchasing behavior has emerged as a crucial component of effective business operations in today's market, which is changing quickly. The many activities, ideas, and feelings that buyers experience when choosing a product are referred to as customer buying behavior. A range of internal and external elements, including as one's own beliefs, societal standards, marketing messages, and product characteristics, all have an impact on this intricate and dynamic process. Businesses may create efficient marketing strategies that satisfy customer requirements and expectations and enhance sales and profitability by understanding the elements that affect consumer purchasing behavior.

The paper by Voramontri and Klieb presents findings from their empirical study on the impact of social media on decision-making process of consumers for intricate purchases, and the study reveals that social media usage affects consumer satisfaction in the information search and alternative evaluation stages for internet-savvy consumers in South-East Asia; the satisfaction level increases as consumers move towards the final purchase decision and post-purchase evaluation [1]. Palalic, Ramadani, Gilani, Gërguri-Rashiti, and Dana conducted a study to investigate the effect of social media on the purchasing behavior of Pakistani consumers. They collected data from 396 respondents through an online survey and found that among consumers under the age of 40, word-of-mouth and content credibility are the most significant factors influencing their decision-making process. These findings highlight the importance for business owners to consider these factors while devising their marketing strategies [2]. With the results showing that "personality" had a significant impact on consumer behavior while the economic situation had a negative relationship, Muniady, Al- Mamun, Permarupan, and Zainol's study sought to investigate the impact of personal factors on the consumer behavior of Malaysian university students [3]. Guo's study uses a questionnaire survey and AHP to analyze the factors influencing consumers' online purchasing decisions and finds that security, prices,

and commercial credit are the most important factors, while gender, educational attainment, and store design come in second. Online retailers are given advice based on these findings [4]. Qazzafi's qualitative study seeks to identify and comprehend the factors influencing consumer purchasing behavior toward goods and services, which include personal, psychological, social, and economic factors, and the study emphasizes that different ages, life cycle stages, motivation, and perceptions of consumers, as well as their surroundings, can influence their purchasing decision [5].

Lawan and Zanna's study investigate cultural, economic, and personal factors that affect clothing purchasing decisions in Borno state, Nigeria; it finds that cultural factors, particularly societal norms, have the greatest influence, followed by economic factors like income, with age being the only personal factor variable influencing purchasing decisions. As a result, marketing managers are advised to take socio-cultural factors into account when designing clothing for their markets [6]. According to the Rehman, Yusoff, Zabri, and Ismail's report, Pakistani consumers consider themselves trendy, value personal values and traits over sales promotions when purchasing fashion products, and like to shop in attractive and spacious malls; they also prefer recommendations from friends and family [7]. Orji, Sabo, Abubakar, and Usman's research revealed that social character, compliance, aggressiveness, and ethnocentrism have a substantial impact on the purchasing behavior of textile materials in South Eastern Nigeria, and they suggest that manufacturers and marketers should tailor their products to these personality factors, target social influencers in specific age groups, use celebrities for promoting products during festivals, and prioritize locally made textiles to promote economic growth [8]. Onu, Emmanuel, and Garvey's study concluded that social character, complaint, aggression, and ethnocentrism significantly positively influence purchasing decisions in South Eastern Nigeria and advises marketers to target social influencers and concentrate on advertising messages that acknowledge social status [9]. Al-Azzam's study examines consumer decision-making and how it affects the buying of home furnishings in Jordan, and it discovers a strong relationship between the choice to buy and the reference group, family, price, quality, and color [10].

In-depth analysis of customer purchasing behavior is provided in this research study, along with an examination of the numerous elements that affect it and the consequences for businesses attempting to create successful marketing plans. The Consumer Personality Analysis dataset, which is accessible on Kaggle, will be examined in particular, and a complete examination of the correlation between purchasing behavior and the people attributes contained in the dataset will be performed.

2. Method

2.1. Data Source and Basic Data Processing

The Kaggle dataset Customer Personality Analysis consists of 2239 rows and 28 columns, where each row represents a unique customer, and each column represents a unique customer attribute. First checking to see if there is any null values or empty cells in the dataset and find that 'Income' has 24 null values. Later, Basic Descriptive Statistics will be calculated, and median of 'Income' will be imputed in those 24 empty cells, since the median is a measure of central tendency that is less susceptible to outliers than the mean. In other words, high values in the dataset can significantly distort the mean value, but the median is less susceptible to these effects. Using the median to fill in missing values can assist preserve the general distribution of the data and limit the influence of outliers on the research. Moreover, using the median to impute missing values can be a straightforward and rapid method that needs minimum data processing. Z_CostContact and Z_Revenue columns will be deleted since they consist of same values for each row and do not contribute to the data set modeling. Some other columns will also be deleted since they will not be used in the data analysis later. Table 1 shows all the variables that will be used and their corresponding references.

Table 1. Variable description

Variables	References
Year_Birth	Customer's birth year
Education	Customer's education level
Marital_Status	Customer's marital status
Income	Customer's yearly household income
Kidhome	Number of children in customer's household
Teenhome	Number of teenagers in customer's household
MntWines	Amount spent on wine in last 2 years
MntFruits	Amount spent on fruits in last 2 years
MntMeatProducts	Amount spent on meat in last 2 years
MntFishProducts	Amount spent on fish in last 2 years
MntSweetProducts	Amount spent on sweets in last 2 years
MntGoldProds	Amount spent on gold in last 2 year
TotalKidAndTeen	Total number of kids in customer's household
TotalAmountSpent	Total amount spent on all categories of items

In Table 1, there are two new variables ‘TotalKidAndTeen’ and ‘TotalAmountSpent.’ Since both ‘Kidhome’ and ‘Teenhome’ calculate number of kids in the household, the data could be combined into one and call it ‘TotalKidAndTeen.’ Similarly, customer’s expenditure on specific category of items is not needed, so all spendings could be combined into one and call it ‘TotalAmountSpent.’ Furthermore, since both ‘Education’ and ‘Marital_Status’ contain texts which cannot be represented as numbers in graph or table, the input in those two variables need to be transformed into numbers. For ‘Education,’ the inputs are split into 2 categories: Use “1” to represent higher education level, which are Graduation, PhD, Master Education, 2n Cycle; Use “2” to represent lower education level, which is Basic. For ‘Marital_Status’, the input could be split into 2 categories based on their status: Use “1” to represent people that are with someone, such as Married and Together; Use “2” to represent people that are only by themselves, such as Single, Divorced, Widow, Alone, YOLO, and Absurd.

2.2. Research Method

The purpose of this study was to look at the connection between a person's characteristics and their overall product spending. The study employed numerous statistical methods to investigate the correlations between the variables. The correlations between the various variables were first shown using a correlation heatmap. The heatmap made it possible to assess the direction and intensity of the relationships between the characteristics of individuals and those of products.

After the correlation analysis, the research calculated out the descriptive statistics for the dataset. Those include mean, standard deviation, minimum, maximum, and so on, which were presented as part of this research to help understand how the data were distributed. Afterwards, a univariate analysis on both the demographic data and the overall cost of the product was performed.

The study then conducted a multivariate linear regression analysis to look at the connection between people's characteristics and overall product expenditures. People attribute factors served as the independent variables in the regression model, while the total costs served as the dependent variable.

3. Results and Discussion

3.1. Correlation Analysis

In order to study the relationship between all the variables, a Correlation Heat Map should be generated first to examine the positive and negative relationships.

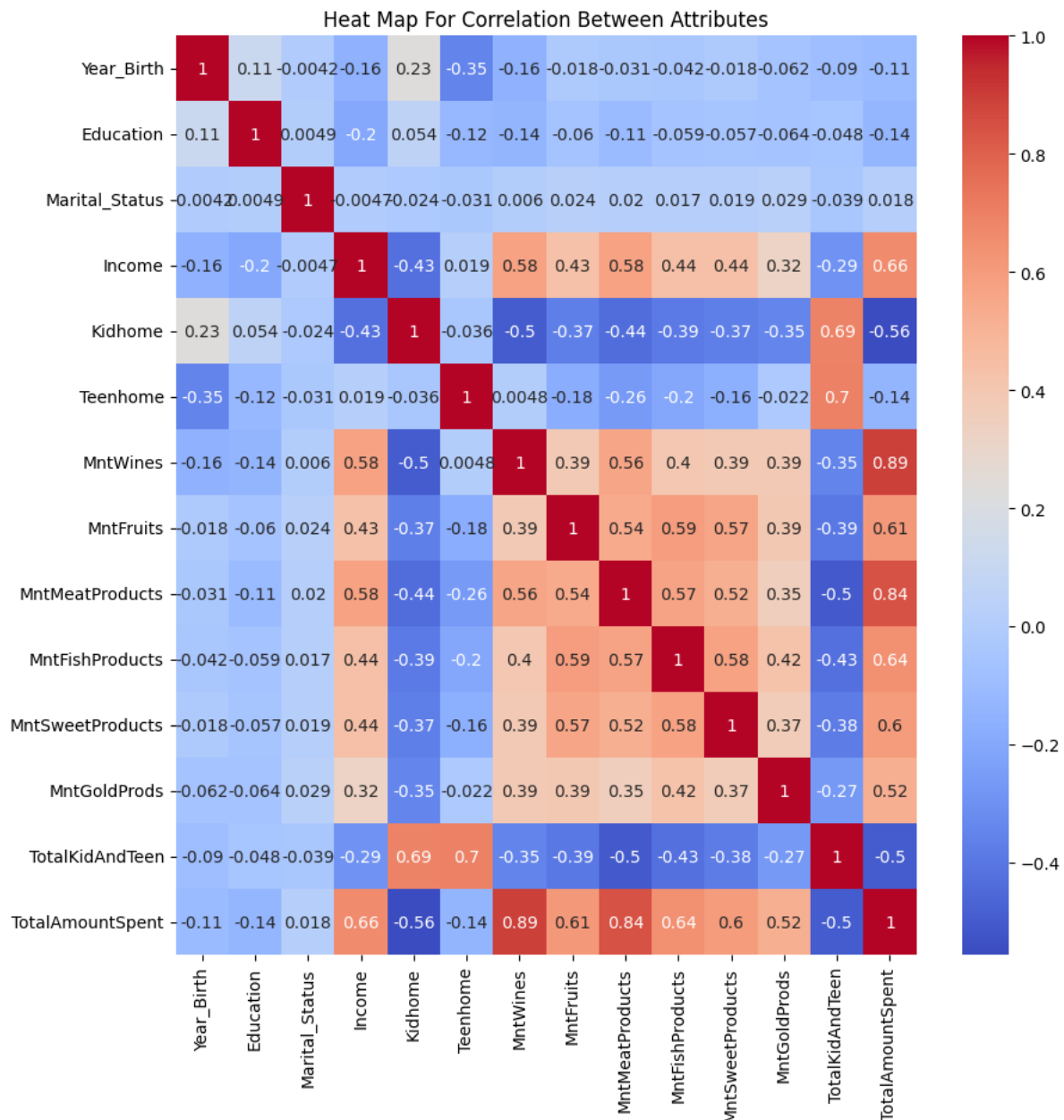


Figure 1. Heat Map for Correlation Between Variables

As shown in Figure 1, there is a strong positive correlation between ‘MntWines’ and ‘TotalAmountSpent’, ‘MntMeatProducts’ and ‘TotalAmountSpent’. However, both ‘MntWines’ and ‘MntMeatProducts’ are part of ‘TotalAmountSpent’, the strong positive correlation relationship does not contribute anything significant. Besides those two sets of strong correlations, there are no other two variables displaying strong correlation, so none of the columns in the dataset should be dropped.

3.2. Basic Descriptive Statistics of Dataset

Calculating basic descriptive statistics of the dataset is necessary to obtain a better understanding of the variables. From Table 2, several observations could be made. Most customers were born in between 1959 to 1977. Majority of customers belong to category 1 of education level with a high education standard. Many of them also belong to category 1 of marital status which means many of them are with someone. After filling in the empty cells in ‘Income’ with the median value, the average income becomes \$52237.975 and 75% of them fall below \$68289.75. Compared to the maximum value of ‘Income’ of \$666666, it is about 10 times greater than the average and 75th percentile value, so the maximum value for income is an outlier. Most of the customers also have 1 kid or teenager at

home. At last, for the total amount spent on products, the average amount is around \$605.798. Compared to median of \$396, the distribution of 'TotalAmountSpent' is not normally distributed but skewed.

Table 2. Basic descriptive statistics

	Mean	Std	Min	25%	50%	75%	Max
Year_Birth	1968.806	11.984	1893	1959	1970	1977	1996
Education	1.024	0.153	1	1	1	1	2
Marital_Status	1.355	0.479	1	1	1	2	2
Income	52237.975	25037.956	1730	35538.75	51381.5	68289.75	666666
Kidhome	0.444	0.538	0	0	0	1	2
Teenhome	0.506	0.545	0	0	0	1	2
TotalKidAndTeen	0.95	0.752	0	0	1	1	3
TotalAmountSpent	605.798	602.249	5	68.75	396	1045.5	2525

3.3. Univariate Analysis

In order to have a deeper understanding of the people attributes, the graph for each variable will be displayed. A histogram of the variable "Year Birth" is shown in Figure 2. The graph is seen to have outliers for the years 1893, 1899, and 1900. According to calculations, the median value of around 1970 and the mean value of roughly 1968 are very similar. Note, while the highest value of 1996 is closer to the center, the minimum value of 1893 deviates greatly from both the mean and median.

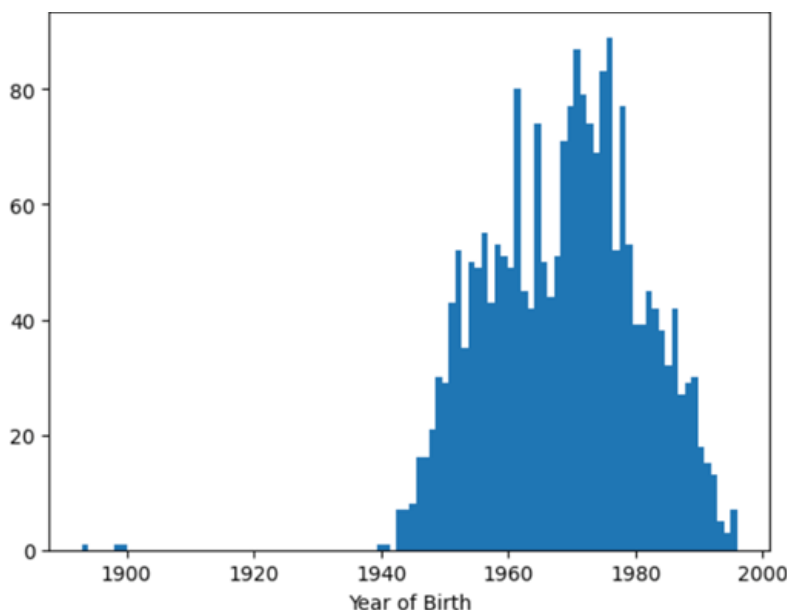


Figure 2. Histogram for Year of Birth

A histogram of the variable "Education" is shown in Figure 3. According to the graph, the majority of people in the dataset have a "Graduation" level of education, while a "Basic" level of education is held by a smaller percentage of people. It is worth noting that the dataset has been separated into two categories depending on educational attainment. The studies show that 97% of those in the sample had higher education levels, indicating that there is a proportionately greater population of people with higher education levels.

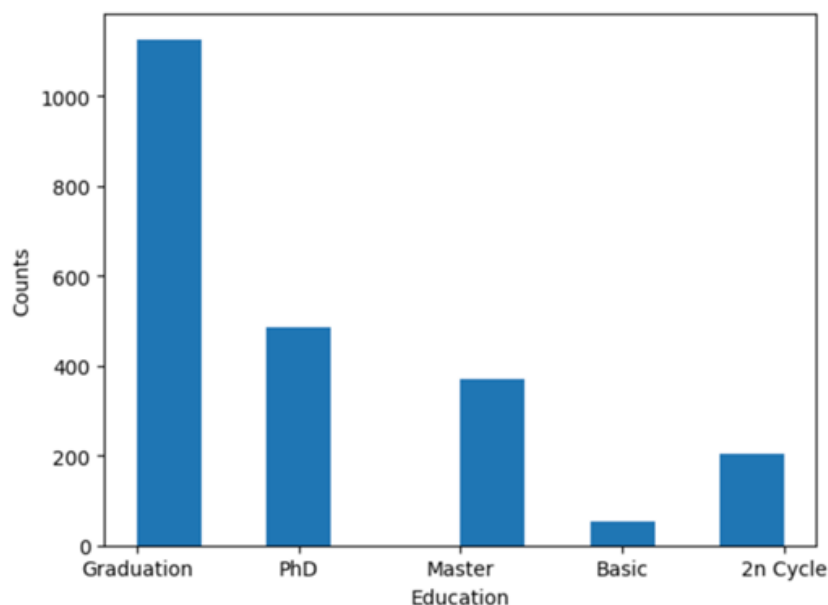


Figure 3. Histogram for Education

A histogram of the variable "Marital Status" is shown in Figure 4. The research shows that "Married" people make up the bulk of the dataset's participants, whilst "Absurd" and "YOLO" people are underrepresented. The dataset has been divided into two groups: one contains single people, while the other represents people who are in a relationship. From the statistics, it can be inferred that 36% of people are single and about 64% of people are in relationships.

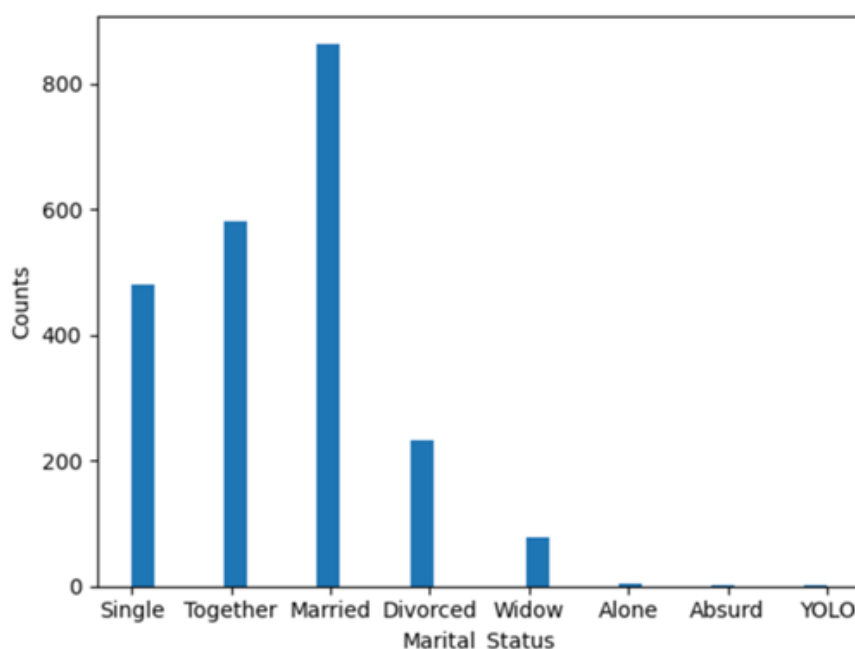


Figure 4. Histogram for Marital Status

Figure 5. depicts a histogram of the variable 'Income.' As previously discussed in the Basic Descriptive Statistics of the Dataset section, the maximum value is considerably distant from the other values. Additionally, the graph indicates several outliers, including the maximum value. Despite the presence of outliers, the remaining values exhibit a central tendency.

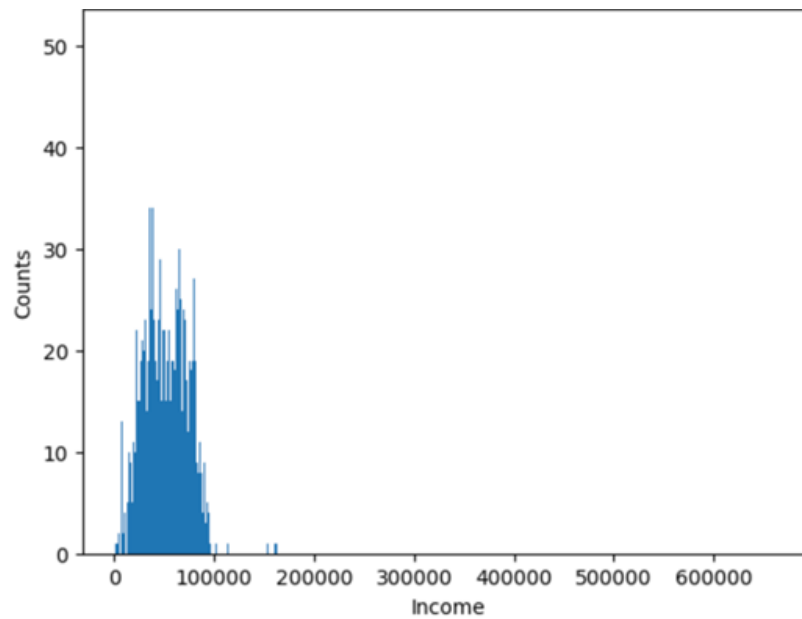


Figure 5. Histogram for Income

Figure 6. displays a histogram of the variable 'TotalKidAndTeen.' The analysis indicates that approximately 28.48% of individuals have no children or teenagers, while around 50.35% of individuals have one child or teenager. Furthermore, around 18.79% of individuals have two children or teenagers, while approximately 2.36% of individuals have three children or teenagers.

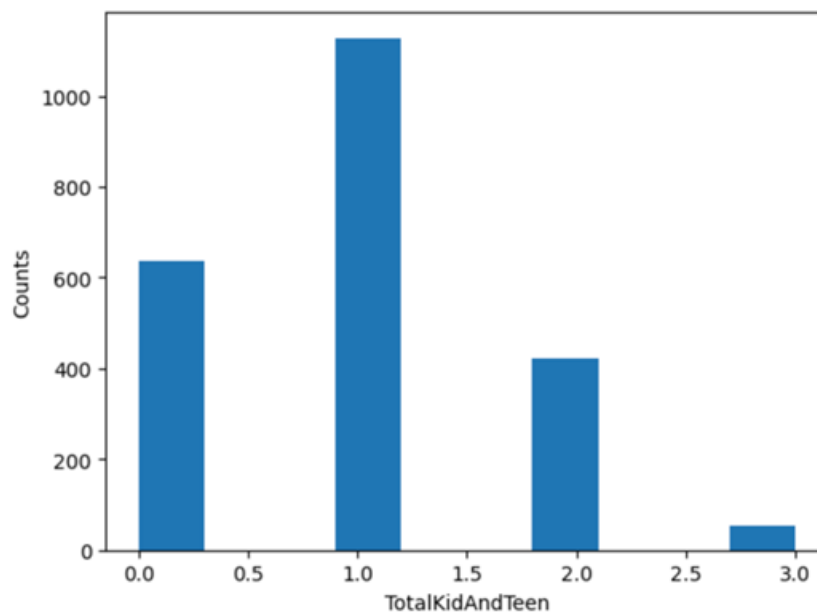


Figure 6. Histogram for Total Numbers of Kids and Teenagers

Figure 7 illustrates a histogram of the variable 'TotalAmountSpent.' As discussed earlier in the Basic Descriptive Statistics of the Dataset section, the distribution of the variable is positively skewed. The histogram indicates that the majority of data points fall within the range of \$0 and \$250, which deviates significantly from the mean and median values. The analysis suggests that the distribution is heavily skewed towards the right, implying that most individuals tend to spend relatively little on goods.

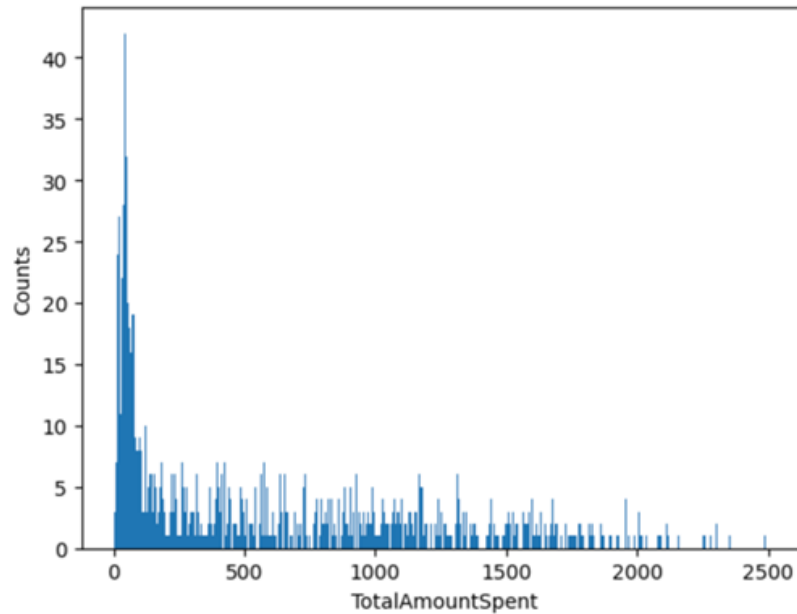


Figure 7. Histogram for Total Amount Spent on All Categories of Goods

3.4. Multivariate Linear Regression

A statistical technique known as Multivariate Linear Regression measures the strength of the linear relationship between several independent and dependent variables. The intercorrelation between the variables determines if the connection is linear. Multivariate Linear Regression is used to anticipate the behavior of the response variable based on the related predictor variables after being applied to a dataset.

Table 3. OLS Regression Result

Variables	Coef.	Std. err	t	P > t	[0.025	0.975]
const	5250.681	1447.979	3.626	0	2411.155	8090.207
Year_Birth	-2.505	0.734	-3.414	0.001	-3.944	-1.066
Education	-147.597	57.506	-2.567	0.01	-260.368	-34.827
Marital_Status	8.515	17.913	0.475	0.635	-26.614	43.644
Income	0.013	0	35.395	0	0.012	0.014
TotalKidAndTeen	-276.599	12.113	-22.835	0	-300.353	-252.845

From the results in Table 3, several conclusions could be made. The general equation for Multivariate Linear Regression is:

$$\begin{aligned} \widehat{TotalSpent} = & 5250.681 - 2.505 (Year_Birth) - 147.597(Education) + \\ & 8.515(Marital_Status) + 0.013(Income) - 276.599(TotalKidAndTeen) \end{aligned} \quad (1)$$

When holding constant all other predictor variables, the coefficient for each individual represents the average increase in the response variable for every unit increase in that specific predictor variable.

R-squared is a statistical metric that shows how much of the variance in the response variable can be assigned to the predictor variable. R-squared equals to 0.548, which indicates that 54.8% of the variance in TotalAmountSpent can be explained by the Year_Birth, Education, Marital_Status, Income, and TotalKidAndTeen.

If the significance level is set to be 0.05 and compare the p-values with it, 4 statistically significant variables could be found. Year_Birth, Education, Income, and TotalKidAndTeen all have p-values less than 0.05, so they would make the regression model fits better.

4. Conclusion

This paper studies the relationship between some consumer attributes and total spending on different products. A dataset with 2239 distinct clients and 28 characteristics was utilized in the study. The methods for data processing and cleaning included deleting unnecessary columns and imputing missing values using the median. Correlation heat map, basic descriptive statistics, univariate analysis, and multivariate linear regression analysis were then carried out as part of the study. The findings indicate that Year Birth, Education, Income, and TotalKidsAndTeen are statistically significant predictors of TotalAmountSpent, however Marital Status does not seem to be a relevant predictor. These results can be used to guide marketing plans aimed at certain demographic groups, such as younger or older consumers, customers with or without children, customers with higher or lower incomes or educational levels.

References

- [1] Voramontri D, Klieb L. Impact of social media on consumer behaviour. *Information and Decision Sciences*, 2019, 11 (3).
- [2] Palalic R, et al. social media and consumer buying behavior decision: what entrepreneurs should know. *Management Decision*, 2021, 59 (6): 1249 - 1270.
- [3] Muniady R, et al. Factors Influencing Consumer Behavior: A Study among University Students in Malaysia. *Asian Social Science*, 2014, 10 (9).
- [4] Guo Li. A Research on Influencing Factors of Consumer Purchasing Behaviors in Cyberspace. *International Journal of Marketing Studies*, 2011, 3 (3).
- [5] Qazzafi Sheikh. Factor Affecting Consumer Buying Behavior: A Conceptual Study. *International Journal for Scientific Research & Development*, 2020, 8 (2).
- [6] Lawan L, Zanna R. Evaluation of Socio-Cultural Factors Influencing Consumer Buying Behaviour of Clothes in Borno State, Nigeria. *International Journal of Basic and Applied Science*, 2013, 1 (3): 519 - 529.
- [7] Rehman F, et al. Determinants of personal factors in influencing the buying behavior of consumers in sales promotion: a case of fashion industry. *Young Consumers*, 2017, 18 (4): 408 - 424.
- [8] Orji M, et al. Impact of Personality Factors on Consumer Buying Behaviour Towards Textile Materials in South Eastern Nigeria. *International Journal of Business and Economics Research*, 2017, 6 (1): 7 - 18.
- [9] Onu A, et al. Assessing the Relationship between Personality Factors and Consumer Buying Behavior in South Eastern Nigeria. *International Journal of Business and Social Science*, 2014, 5 (11).
- [10] Al-Azzam Abdel. Evaluating Effect of Social Factors Affecting Consumer Behavior in Purchasing Home Furnishing Products in Jordan. *British Journal of Marketing Studies*, 2014, 2 (7): 80 - 94.