# 

# 

# Customer-Segment-and-Shopping-Pattern

**Casey Patel**

1. **Introduction**

Customer segmentation is a crucial strategy for businesses to gain a deeper understanding of their customers and their needs. By dividing customers into distinct groups based on specific characteristics, businesses can develop targeted marketing strategies, improve product development, and enhance customer service.

Marketing segmentation enables businesses to craft more effective marketing campaigns by tailoring messages to specific customer segments. For instance, geographic segmentation allows businesses to group customers by location, creating region-specific offers that cater to local preferences. Behavioral segmentation, on the other hand, groups customers based on their interactions with a brand, such as product consumption or website usage. This approach enables businesses to create more relevant marketing messages and campaigns that resonate with their target audience.

By leveraging customer segmentation, businesses can unlock the full potential of their marketing efforts, driving more conversions, and fostering stronger customer relationships.

1. **Problem Statement**

Our goal is to distinguish distinct customer segments based on their age and purchase amount (in USD). By analyzing these two key variables, we aim to identify clusters of customers with similar characteristics, enabling us to develop targeted marketing strategies and improve customer relationships.

1. **Stakeholders:**

The client for a mental health business-related issue could be sales team, customer service team, product development team, data analyst and external agencies. These stakeholders would benefit from the insights gained from this analysis, and their input and feedback would be valuable in refining the analysis and implementing the results.

1. **Datasets**

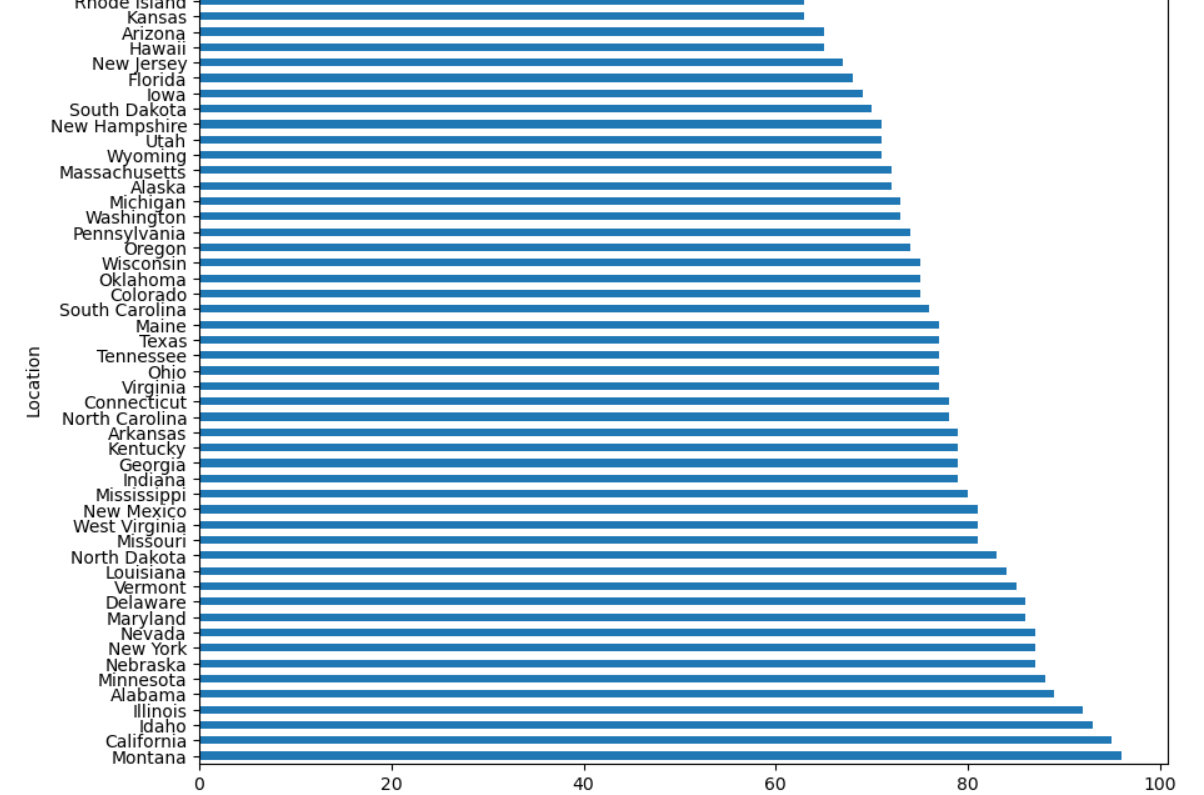
The dataset is obtained from Kaggle called [Consumer Behavior and Shopping Habits Dataset](https://www.kaggle.com/datasets/zeesolver/consumer-behavior-and-shopping-habits-dataset). The Consumer Behavior and Shopping Habits Dataset provides a detailed overview of consumer preferences and purchasing behaviors. It includes demographic information, purchase history, product preferences, and preferred shopping channels (online or offline). This dataset is essential for businesses aiming to tailor their strategies to meet customer needs and enhance their shopping experience, ultimately driving sales and loyalty.

1. **Data Wrangling**

Upon examining the imported CSV file, data consistencies were identified. There were no missing values in columns. Dataset is of high quality, with no missing values and no evidence of data corruption or inconsistencies.

**Distribution of Items Purchased by Location**

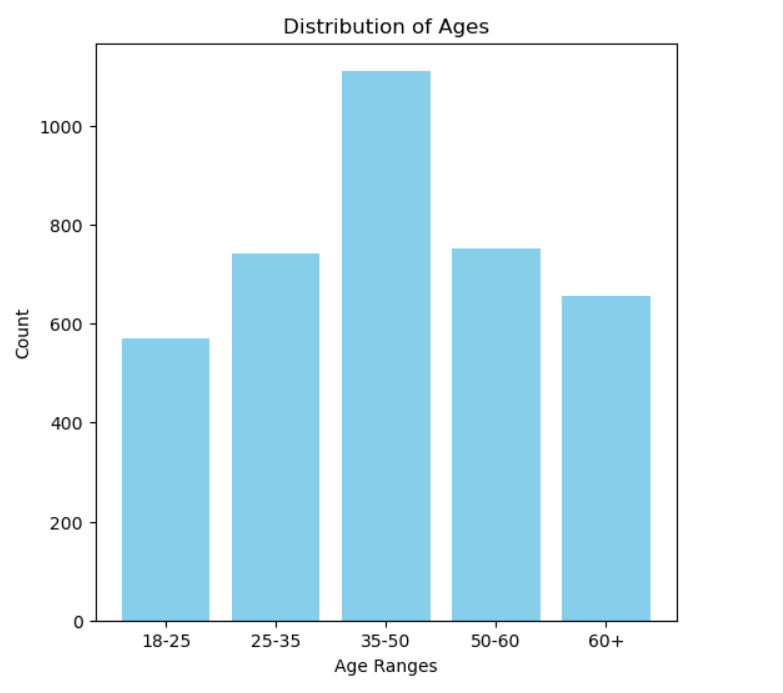
Next, I visualized Distribution Of item purchased By Location. The following figure presents the distribution of items purchased by location, providing insights into the most popular locations for shopping.



As shown in the figure, the top locations with the highest number of items purchased are Montana, California, Idaho and Illinois. This information can be used to inform marketing and sales strategies, such as targeting promotions and advertising to these locations.

**Distribution of Ages**

Next, analysis examines the distribution of ages among the shopping trends dataset. To facilitate this analysis, we defined five age ranges: 18-25, 25-35, 35-50, 50-60, and 60+. We then counted the number of individuals in each age range and visualized the results in the figure below.



As shown in the figure, the majority of individuals fall within the 35-50 age range, with more than 1000 individuals in this category. The 25-35 age range follows closely, with 750 individuals. The remaining age ranges have fewer individuals, with 600, 700, 650. These findings provide valuable insights into the age demographics of our shopping trends dataset, which can inform marketing and sales strategies.

**Average Number of Previous Purchases by Category**

Next, I analyzed the average number of previous purchases by category to gain insights into shopping trends. The results are presented in the table below.

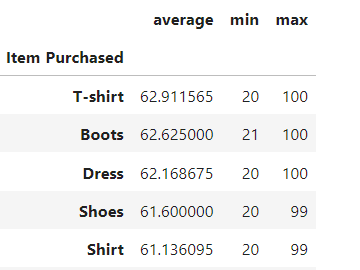


As shown in the table, the average number of previous purchases varies by category. On average, customers in the Accessories category have made the most previous purchases, with an average of 25.73 purchases. The Footwear and Clothing categories follow closely, with averages of 25.23 and 25.20 purchases, respectively. The Outerwear category has the lowest average number of previous purchases, with an average of 24.96 purchases.

These findings suggest that customers in certain categories, such as Accessories and Footwear, tend to make more repeat purchases than those in other categories.

**Average Purchase Amount by Item Purchased**

Then, I analyzed the average purchase amount by item purchased to gain insights into shopping trends. The results are presented in the table below, with the top 5 items shown.



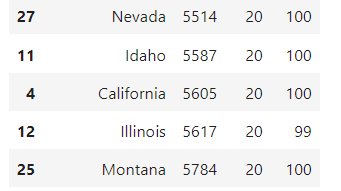
As shown in the table, the top 5 items with the highest average purchase amounts are T-shirts, Boots, Dresses, Shoes, and Shirts. On average, customers spend the most on T-shirts, with an average purchase amount of $62.91. Boots and Dresses follow closely, with average purchase amounts of $62.63 and $62.17, respectively.

These findings suggest that customers are willing to spend more on certain types of items, such as T-shirts and Boots, which may be considered essential or high-quality products. This information can be used to inform pricing and inventory strategies, as well as targeted marketing campaigns to promote these high-value items.

1. **Explanatory Data Analysis (EDA):**

**Regional Sales Performance**

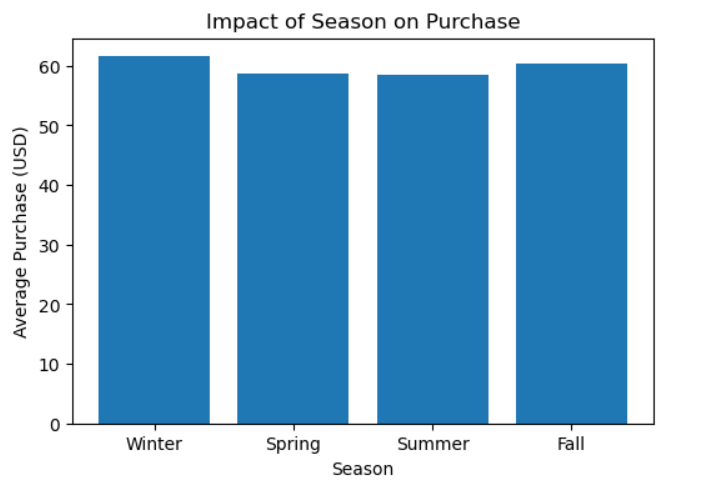
I examined the sales performance by region to gain insights into geographic trends. The results are presented in the table below, with the top 5 regions shown.



As shown in the table, the bottom 5 regions in terms of total sales are Nevada, Idaho, California, Illinois, and Montana. Nevada has the lowest total sales, with a total of $5,514, while Montana has the highest total sales among these 5 regions, with a total of $5,784.

**Impact of Season on Purchase**

Our analysis reveals that the season has a significant impact on the average purchase amount. As shown in the figure below, the average purchase amount varies across different seasons.

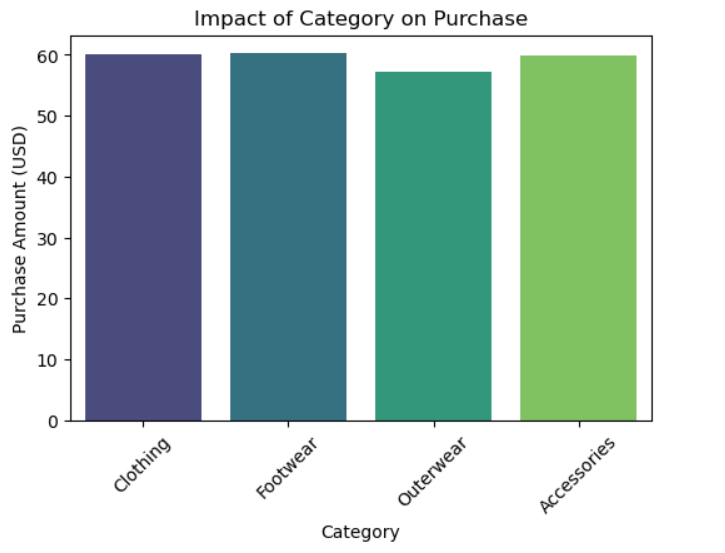


The results indicate that Winter has the highest average purchase amount, with an average of $60. On the other hand, Spring and Summer have the lowest average purchase amounts, both with an average of $55. Fall falls in between, with an average purchase amount of $57.

These findings suggest that businesses may want to consider seasonal fluctuations in purchase behavior when developing their marketing strategies.

**Impact of Category on Purchase**

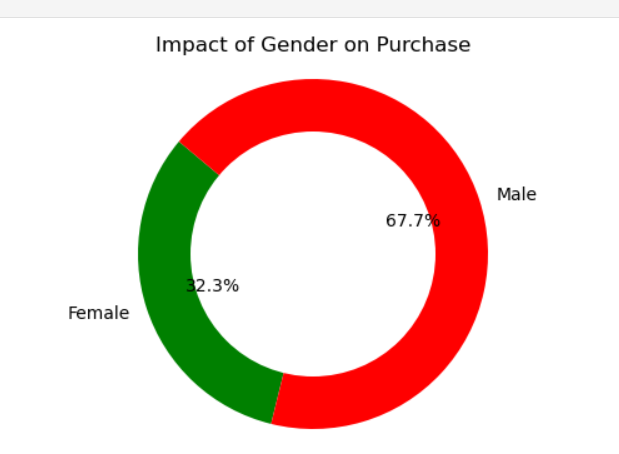
Our analysis reveals that the item category has a notable impact on the average purchase amount. As shown in the figure below, the average purchase amount varies across different categories.



The results indicate that Clothing, Footwear, and Accessories have the highest average purchase amounts, with an average of approximately $59. In contrast, Outwear has a relatively lower average purchase amount, with an average of around $55.

**Impact of Gender on Purchase**

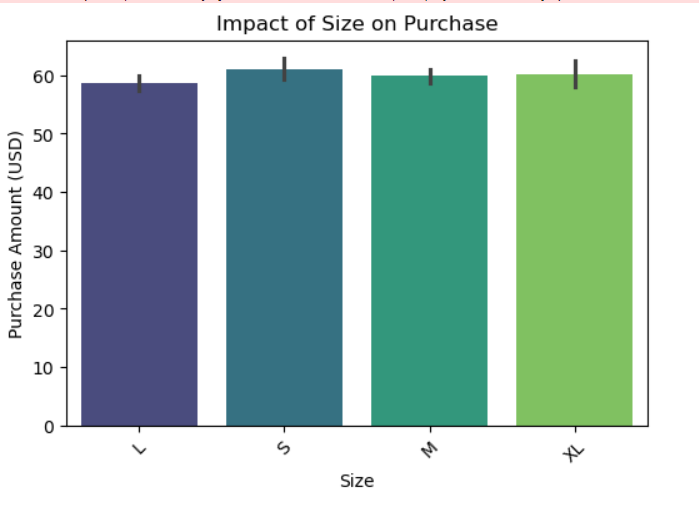
Our analysis reveals that gender has a significant impact on purchase behavior. As shown in the figure below, the proportion of total purchase amounts varies significantly between males and females.



The results indicate that Female customers account for approximately 32% of the total purchase amount, while Male customers account for a substantial 67%. This significant disparity suggests that businesses may want to consider tailoring their marketing strategies and product offerings to better cater to the male demographic, which appears to drive a larger share of sales.

**Impact of Size on Purchase**

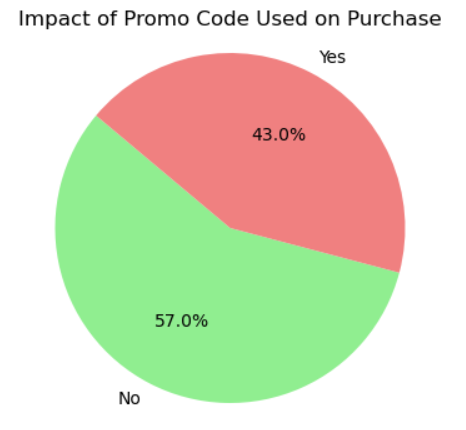
Our analysis reveals that item size has a notable impact on purchase behavior. As shown in the figure below, the average purchase amount varies across different sizes.



The results indicate that the L size has a significantly lower average purchase amount compared to the other sizes. In contrast, XL, Small, and Medium sizes have relatively higher average purchase amounts.

**Impact of Promo Code Used on Purchase**

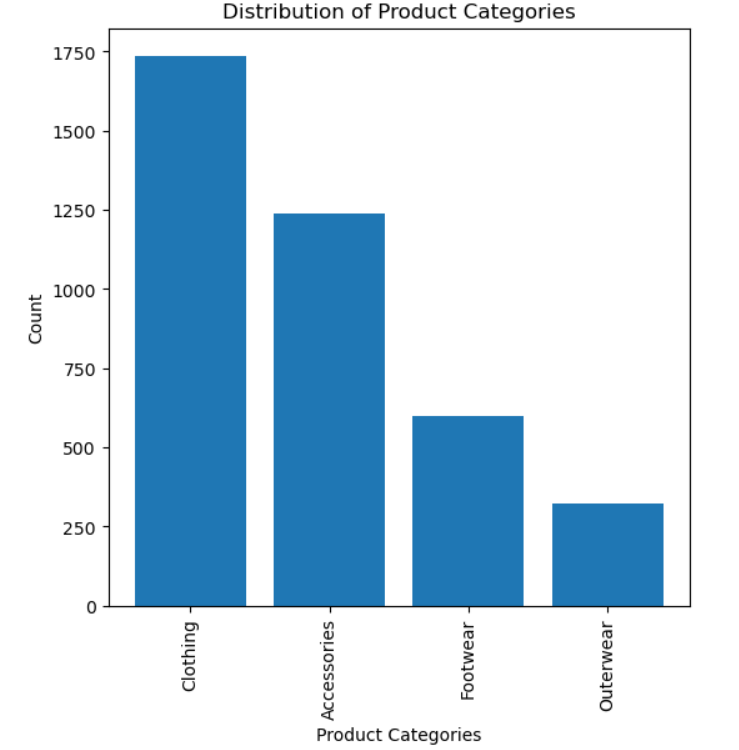
Then, I analyzed impact of promo code used on purchase. Our analysis reveals that the usage of promo codes has a relatively even split among customers. As shown in the figure below, the proportion of customers who used promo codes versus those who did not is nearly equal.



The results indicate that 43% of customers used promo codes, while 57% did not. Interestingly, this suggests that there may be no significant impact of promo code usage on purchase amounts.

**Distribution of Product Categories**

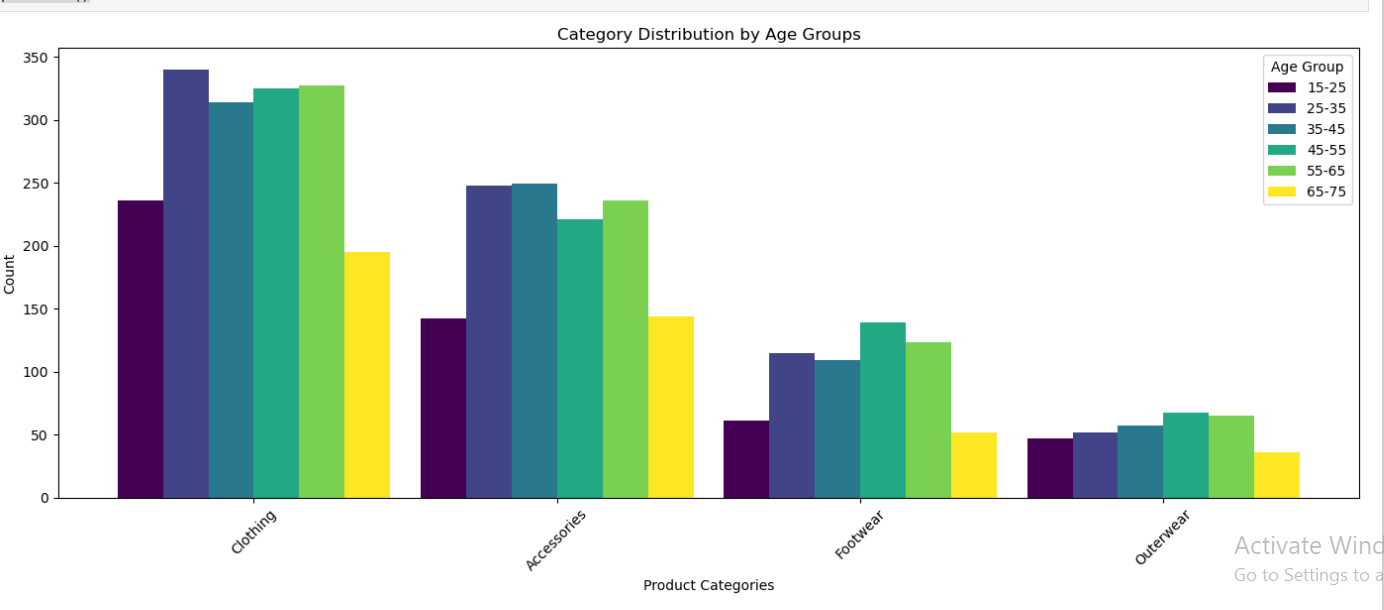
Next, I analyzed distributions of product categories. Our analysis reveals that product categories have varying levels of popularity among consumers. As shown in the figure below, the distribution of product categories is skewed, with some categories being significantly more popular than others.



The results indicate that Clothing is the most popular category, with a count of 1740. This suggests that clothing products are in high demand among consumers. On the other hand, Outwear is the least popular category, with a count of 270.

**Category Distribution by Age Groups**

Our analysis reveals that the distribution of product categories varies across different age groups. As shown in the figure below, the popularity of product categories changes as age increases.



The results indicate that Clothing is the most popular category across all age groups, with the highest count in every age group. This suggests that clothing products are consistently in high demand among consumers across different age groups.

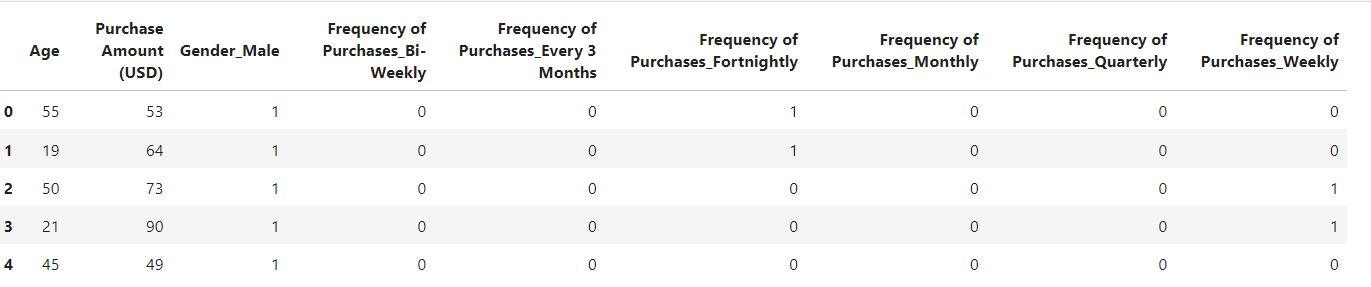
1. **Machine Learning Model:**

**Customer Segmentation Using K-Means Clustering**

I began by segmenting customers based on their demographic and behavioral characteristics to identify distinct groups with similar preferences and needs. To achieve this, we selected four relevant columns from the dataset: Age, Gender, Purchase Amount (USD), and Frequency of Purchases.

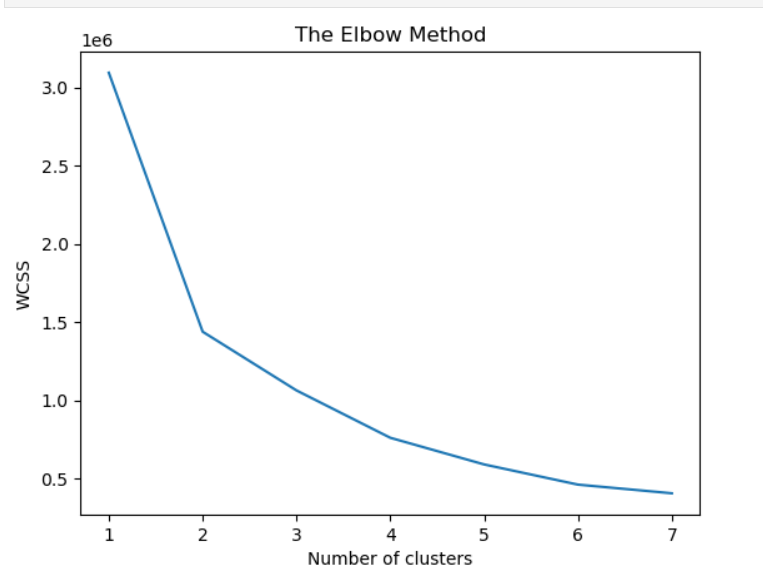
Then, I applied one-hot encoding to the categorical columns Gender and Frequency of Purchases to convert them into numerical variables. This step is essential for K-Means clustering, which requires numerical input features.

The resulting DataFrame, df\_segments, contains the transformed data with the following columns:



Then, I employed the Elbow Method to determine the optimal number of clusters for our K-Means clustering model. This method involves iterating over different numbers of clusters and calculating the Within-Cluster-Sum-of-Squares (WCSS) for each iteration.

Our results are visualized in the following plot:



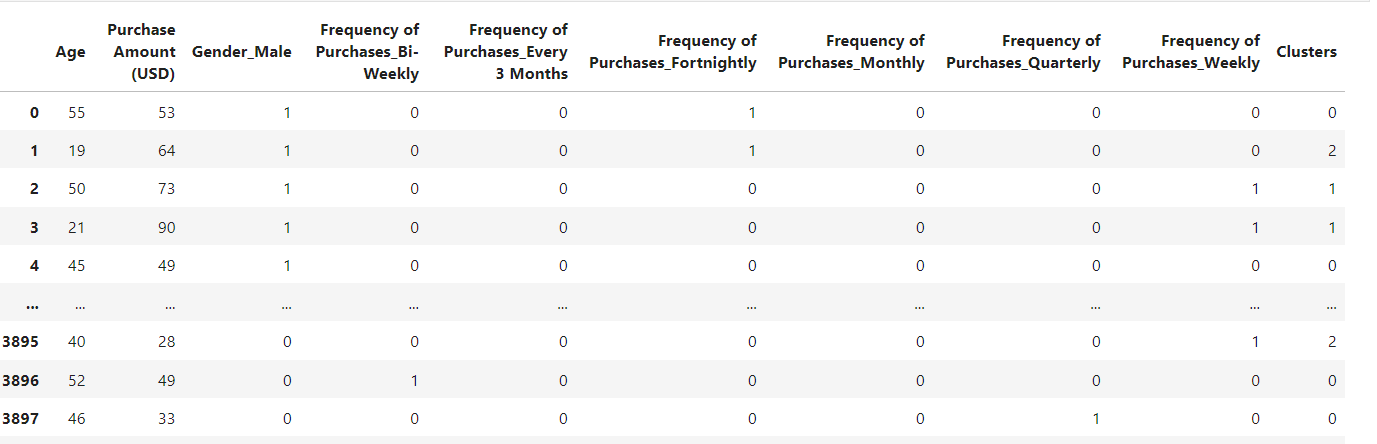
The plot shows the WCSS values for each number of clusters, ranging from 1 to 7. The Elbow Method suggests that the optimal number of clusters is the point at which the WCSS curve starts to flatten out, indicating that further increases in the number of clusters do not significantly improve the model's fit.

Based on our analysis, we conclude that 3 is a good number of clusters for our customer segmentation model. This suggests that our customer base can be divided into three distinct segments with similar characteristics and preferences.

Next, I created a KMeans instance with 3 clusters and fit the model to our preprocessed data, df\_segments.

This will enable us to analyze the characteristics of each cluster and identify patterns and trends in customer behavior.

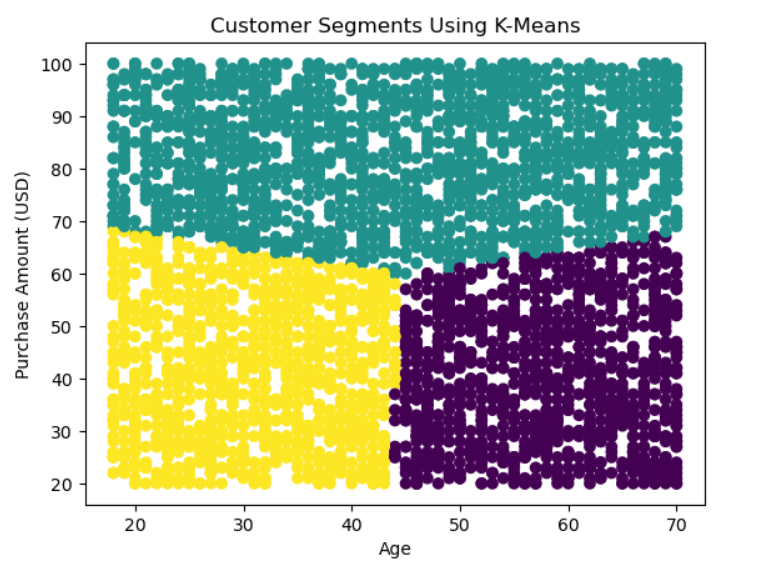
Here is the updated DataFrame:



To gain a deeper understanding of the customer segments, I visualized the clusters using a scatter plot.

The scatter plot displays the customers' Age on the X-axis and Purchase Amount (USD) on the Y-axis. Each point on the plot represents a customer, and the color of the point corresponds to the cluster label assigned by the K-Means model. We used the viridis color map to enhance the visualization.

The resulting plot provides a clear visualization of the three customer segments:



K-means clustering revealed three distinct customer segments, represented by different color codes.

**Cluster 0 (Yellow points):** This segment consists of younger customers (average age: 20-45) who tend to make lower-value purchases (average purchase amount: $50-$70).

**Cluster 1 (Purple points):** This segment comprises old-aged customers (average age: 45-70) who make lower-value purchases (average purchase amount: $50-$70).

**Cluster 2 (green points):** This segment is characterized by middle customers (average age: 30-60) who tend to make higher-value purchases (average purchase amount: $70-$100).

1. **Conclusion**

**Prioritize the 30-70 Age Group:** This demographic represents the largest and fastest-growing customer segment, making it a prime target for marketing efforts. Businesses should consider developing products and services specifically tailored to their needs.

**Nurture High-Spending Customers:** A significant portion of customers make substantial purchases, indicating an opportunity to increase revenue from existing clients. Businesses should develop strategies to target and retain these high-value customers.

**Target Women for Clothing and Accessories:** Businesses selling clothing and accessories to women can potentially boost sales by focusing marketing and sales strategies on this demographic.

**Engage with Male Customers:** Males tend to provide higher average review ratings than females, suggesting that targeting this demographic can enhance overall reputation.

**Focus on Pennsylvania, Alaska, and Arizona:** These states exhibit a clear preference for quality and style, indicating high-value customers with a strong willingness to pay. Businesses should prioritize campaigns focused on these regions.

**Tailor Marketing to Customer Segments:** K-means clustering revealed three distinct customer segments: young low-spenders, older lowaverage-purchasers, and higher-spending value-conscious customers. Businesses should tailor their marketing strategies to address the unique needs and preferences of each segment.