Data Science 3

Task 1:

For the visualization of the data, I have created the following figures:

Figure [1]: Box plots for each numerical column to identify outliers

As can be seen in the figure it seems there are no outliers in the data and the distribution of data is quite uniform.

Figure [2]: Frequency Distribution of Gender

As can be seen in the figure the distribution of data is quite uniform with slightly more males.

Figure [3]: Frequency Distribution of Education Level

As can be seen in the figure the distribution of data is quite uniform with slightly more graduates.

Figure [4]: Frequency Distribution of Marital Status

As can be seen in the figure the distribution of data is quite uniform with slightly more widowed.

Figure [5]: Frequency Distribution of Number of Children

As can be seen in the figure the distribution of data is quite uniform with slightly more people with two children.

Figure [6]: Frequency Distribution of Shopping Persona

As can be seen in the figure the distribution of data is quite uniform with slightly more minimalists.

Figure [7]: Distribution of Age

As can be seen in the figure the distribution of data is quite uniform based on the KDE, with more people in the ages of 18, 28, 38, 49, 59 and 69.

Figure [8]: Annual Income Distribution by Gender

As can be seen in the figure the distribution of data is quite uniform with annual income ranging from $100K - $200K having an average per group of $150K.

Figure [9]: Relationship between Gender and Customer Lifetime Value

As can be seen in the figure the distribution of data is quite uniform with customer lifetime values ranging from about 27 to 77 having an average per group of 50.

Figure [10]: Relationship between Age and Customer Lifetime Value

As can be seen in the figure the distribution of data is quite uniform with customer lifetime values ranging from about 23 to 77 having a total average of about 50.

Figure [11]: Relationship between Online Shopping Score and Purchase Diversity

As can be seen in the figure the distribution of data is quite uniform with different purchase diversity characteristic distributing evenly between the full shopping score range.

Task 2:

Using the elbow technique in figure [12], I have chosen to use 14 clusters numbered from 0-13. The cluster's mean numerical columns can be seen in figure [13]. Below I will list the main insights on each cluster and how I think its best to target each cluster:

Cluster 0: This cluster has relatively **low online shopping scores** and **purchase diversity**, but **higher social media influence**. To target this cluster, **focus on social media** marketing campaigns and **offer personalized discounts** to increase online engagement.

Cluster 1: This cluster has **higher** values **for purchase diversity** and customer **lifetime value**, indicating a diverse and valuable customer base. Target them with **personalized offers** and incentives to maintain their **loyalty**.

Cluster 2: Customers in this cluster have **higher online shopping scores** and **discount sensitivity**. Target them with **attractive discounts** and **promotions** to encourage frequent purchases.

Cluster 3: This cluster has **high** values for **customer lifetime** value and **online engagement**. Focus on building strong **customer relationships** through **personalized experiences** and **rewards programs** to maximize their lifetime value.

Cluster 4: Customers in this cluster have **lower** values for **online shopping scores** and **social media influence**, but **higher discount sensitivity**. Target them with **targeted advertising** campaigns and **exclusive discounts** to drive their engagement.

Cluster 5: This cluster shows **lower online engagement** and **social media influence** but **higher purchase diversity**. Target them by improving their **online shopping experience** and providing a **wide range of product options**.

Cluster 6: Customers in this cluster have **high online engagement** and relatively **low purchase diversity**. Target them by introducing **new product** offerings and **personalized recommendations** to enhance their shopping experience.

Cluster 7: This cluster exhibits **high** values for **online shopping scores** and **social media influence**. Target them through **influencer collaborations** and **social media marketing** campaigns to capitalize on their engagement.

Cluster 8: This cluster represents customers with **high** values for **online shopping scores** and **customer lifetime value**. Target them by providing a **seamless online shopping** experience and **personalized recommendations** to increase their loyalty.

Cluster 9: Customers in this cluster have **lower** values for **online shopping scores** and **purchase diversity**. Target them with **targeted advertising** campaigns and **incentives** to encourage their engagement.

Cluster 10: This cluster shows **higher** values for **purchase diversity** and **discount sensitivity**. Target them with **personalized offers** and **discounts based on their preferences** to maximize their satisfaction.

Cluster 11: This cluster exhibits **higher online engagement** and **social media influence**. Target them through **social media marketing** campaigns and **interactive content** to enhance their engagement.

Cluster 12: Customers in this cluster have **higher** values for **online shopping scores** and **customer lifetime value**. Target them with **loyalty programs** and **personalized experiences** to maintain their loyalty.

Cluster 13: This cluster represents customers with relatively **higher** values for **online engagement** and **social media influence**. Target them through **social media advertising** and **influencer partnerships** to leverage their influence.

Figures:

Figure [1]: Box plots for each numerical column to identify outliers

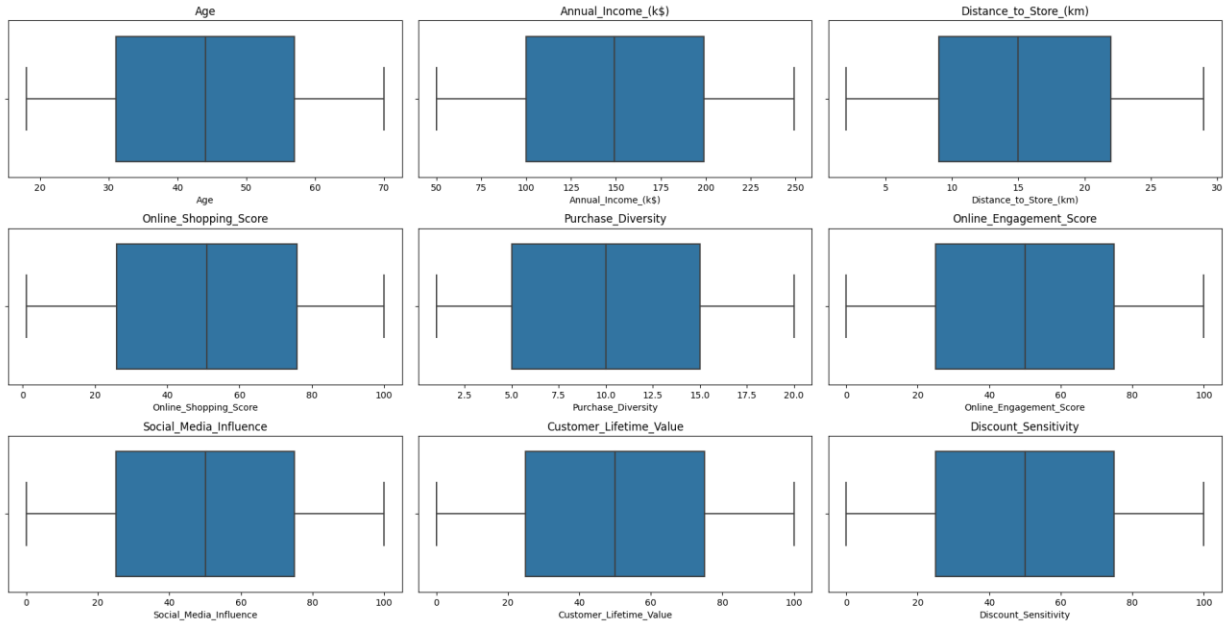

Figure [2]: Frequency Distribution of Gender

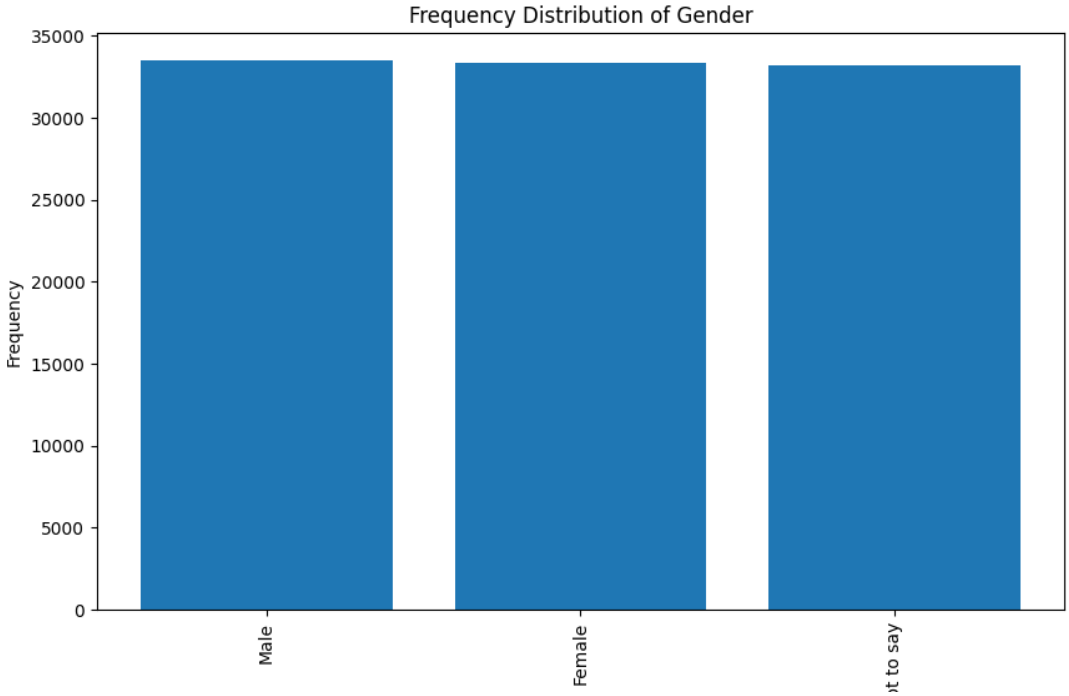

Figure [3]: Frequency Distribution of Education Level

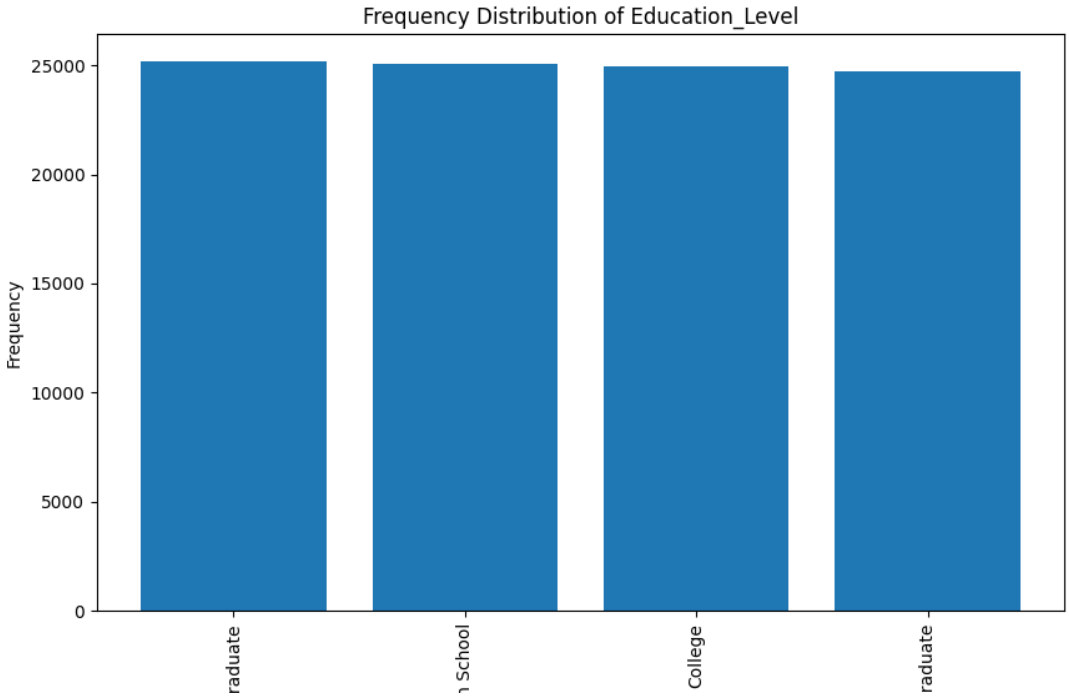

Figure [4]: Frequency Distribution of Marital Status

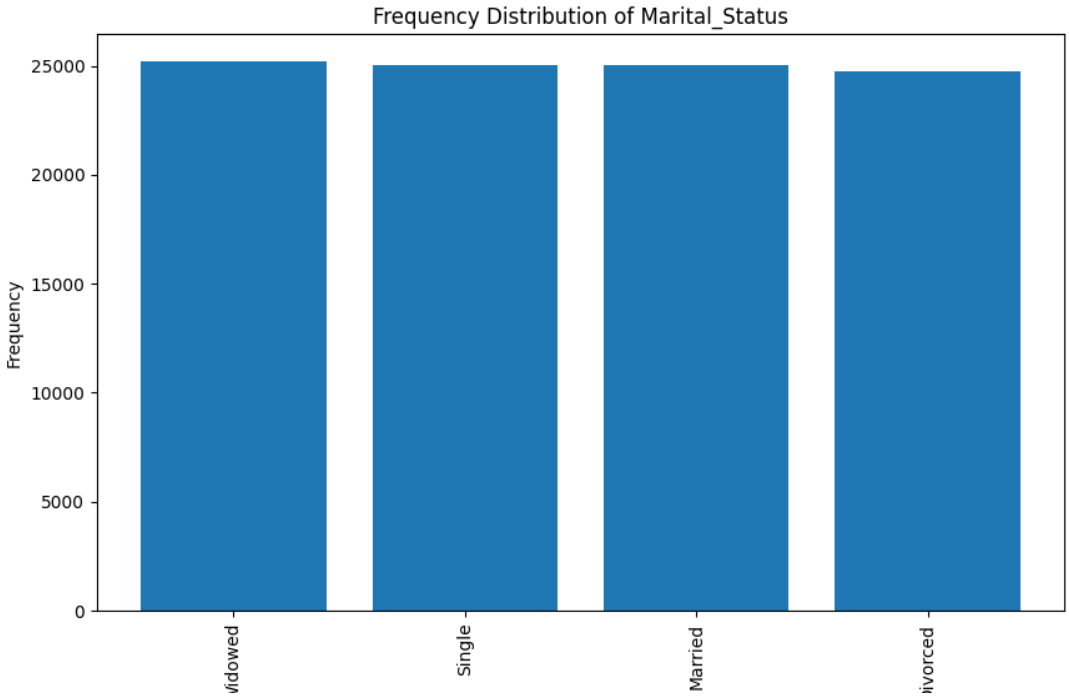

Figure [5]: Frequency Distribution of Number of Children

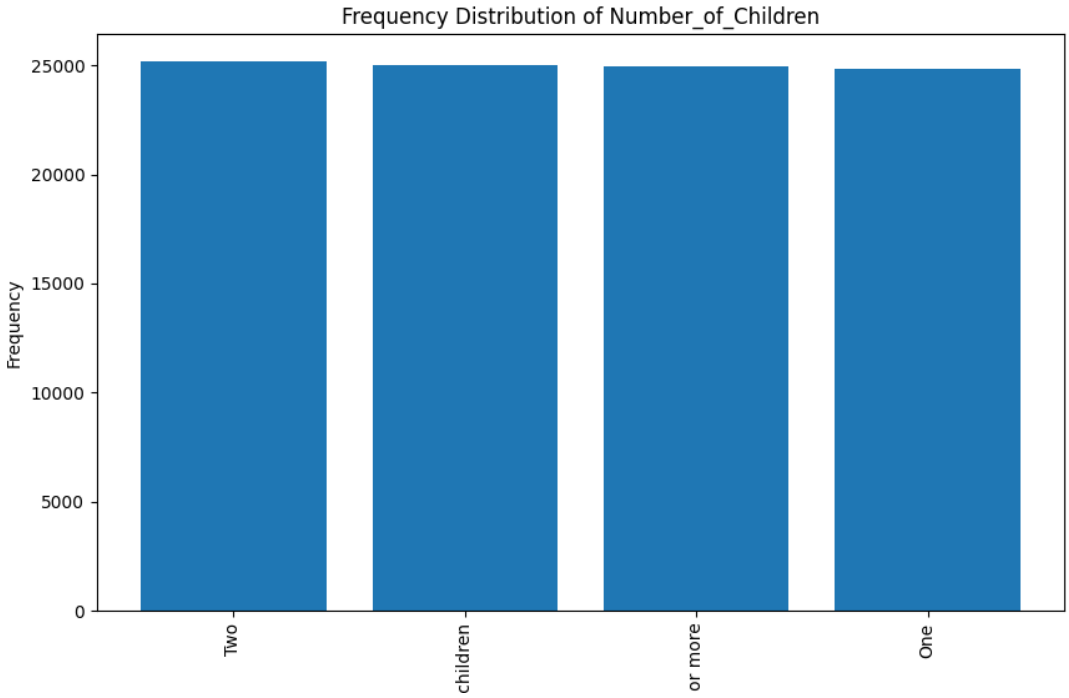

Figure [6]: Frequency Distribution of Shopping Persona

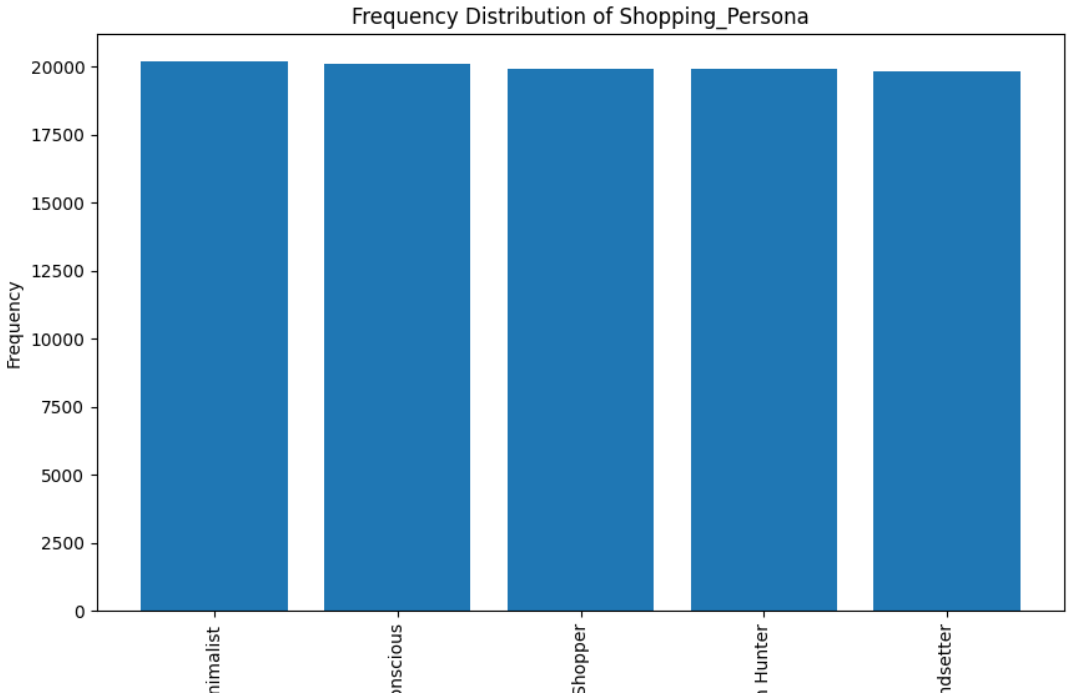

Figure [7]: Distribution of Age

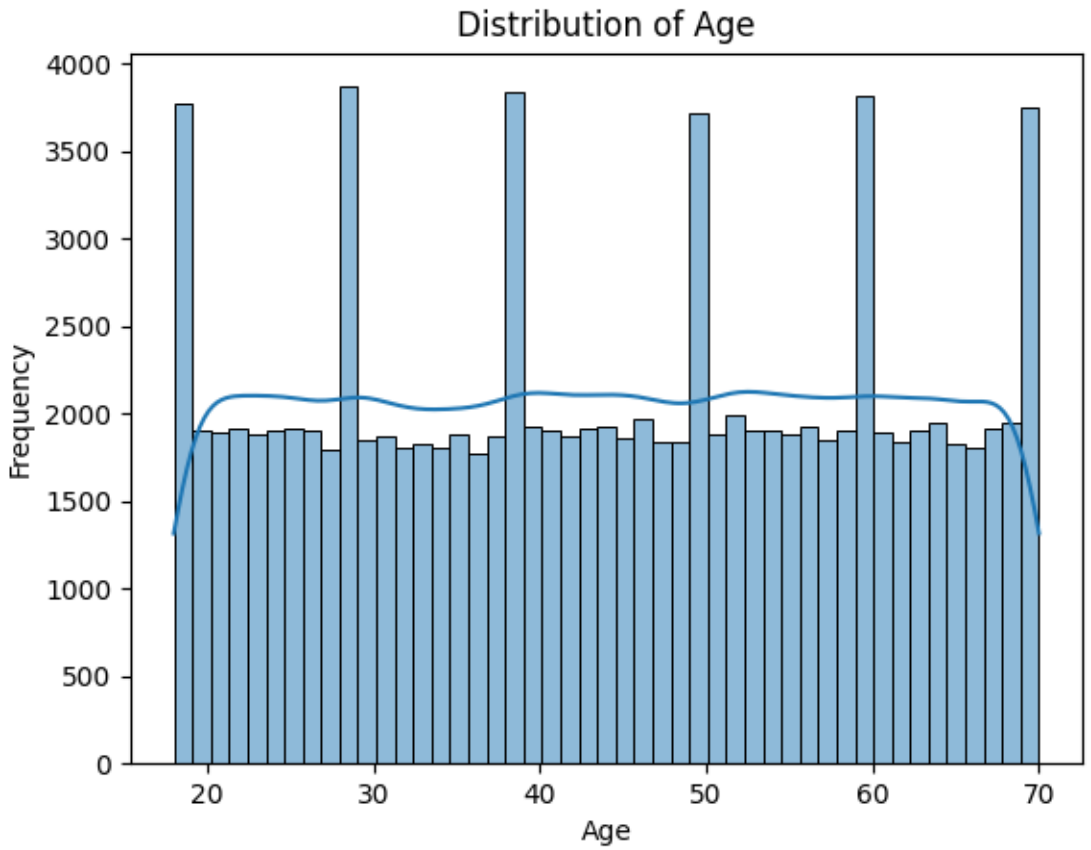

Figure [8]: Annual Income Distribution by Gender

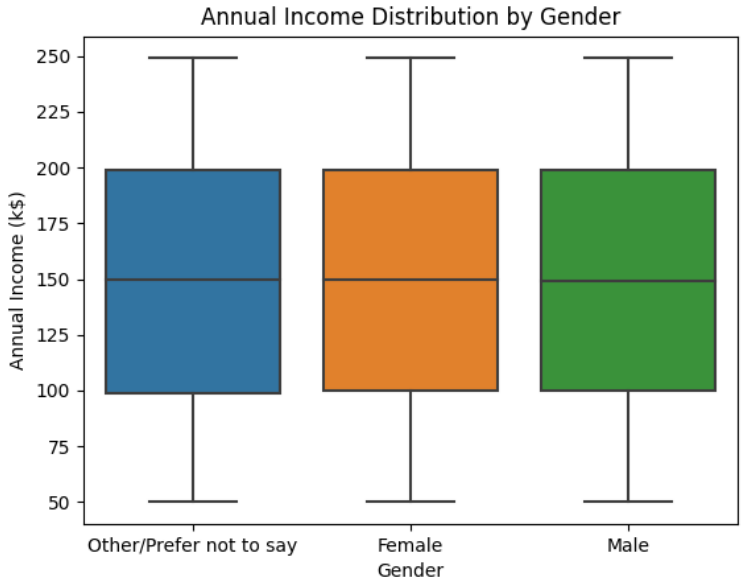

Figure [9]: Relationship between Gender and Customer Lifetime Value

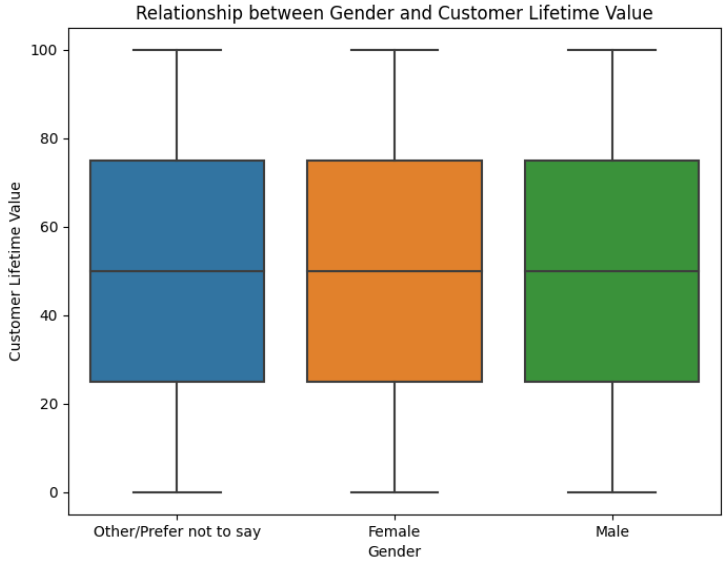

Figure [10]: Relationship between Age and Customer Lifetime Value

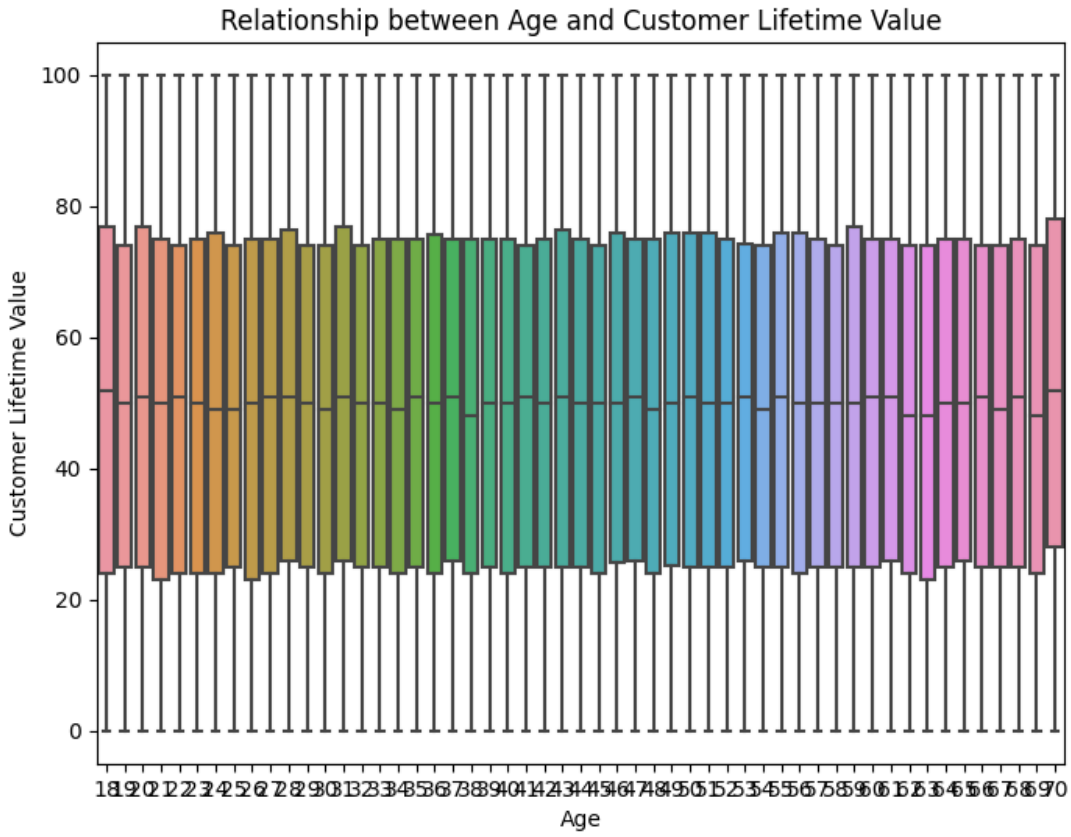

Figure [11]: Relationship between Online Shopping Score and Purchase Diversity

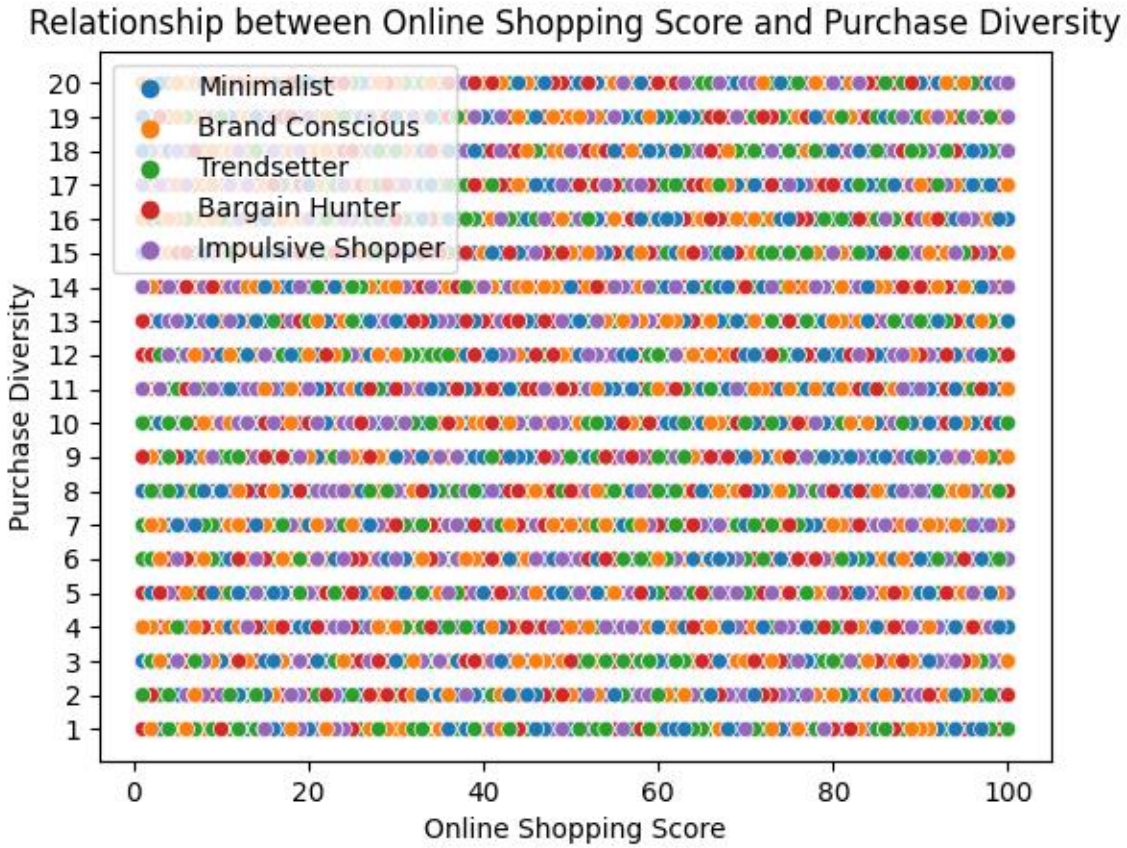

Figure [12]: Elbow Curve to Determine Optimal Number of Clusters

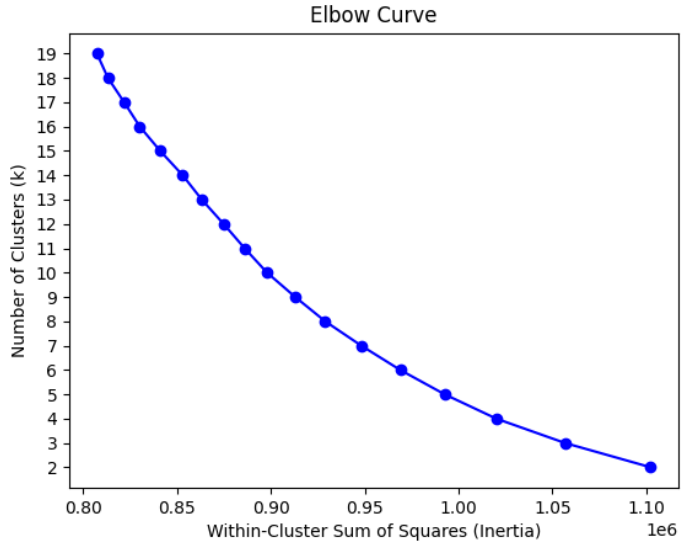

Figure [13]: Cluster Analysis of numerical columns based on mean