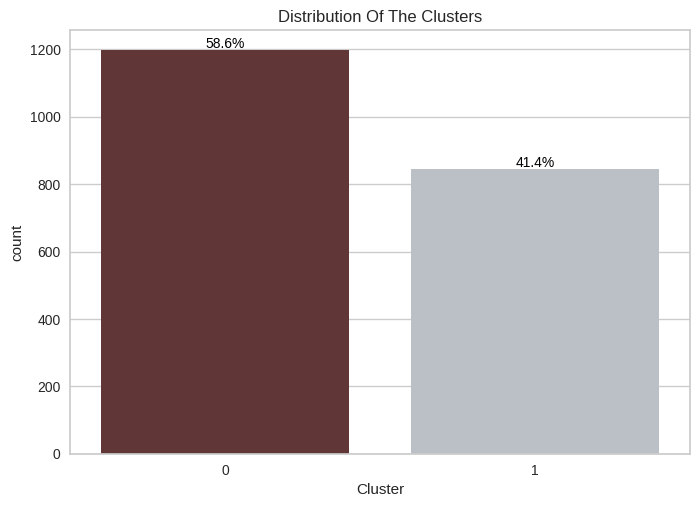
**Machine learning model report**

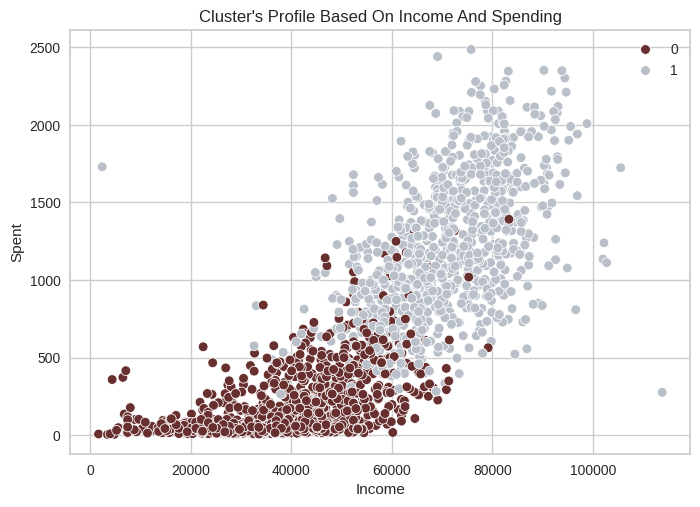
Number of clusters = 2



# Distribution of cluster 0 is larger than cluster 1

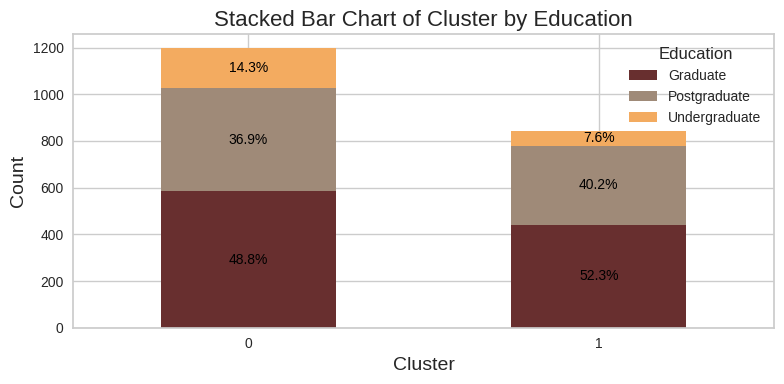
# Cluster 0 is 58.6%

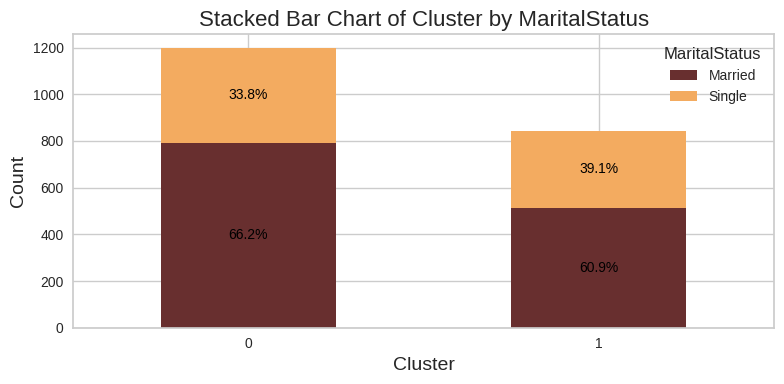
# Cluster 1 is 41.4%



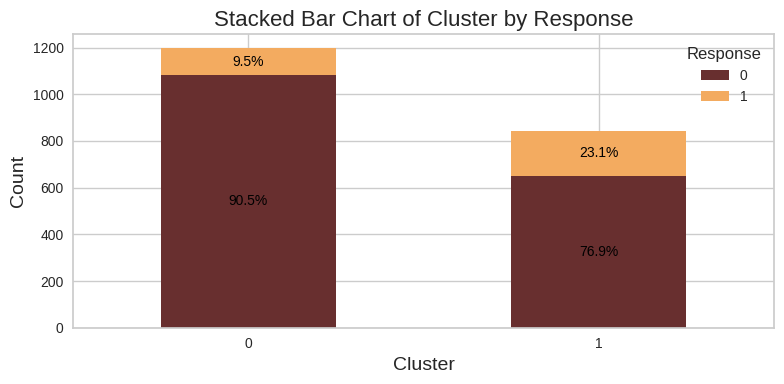
The plot shows the pattern of clusters based on income and spending. The clusters are defined as follows:

* **Group 0**: Low Income to Avg income & Low Spending
* **Group 1**: Avg income To High Income & Avg Spending to High Spending

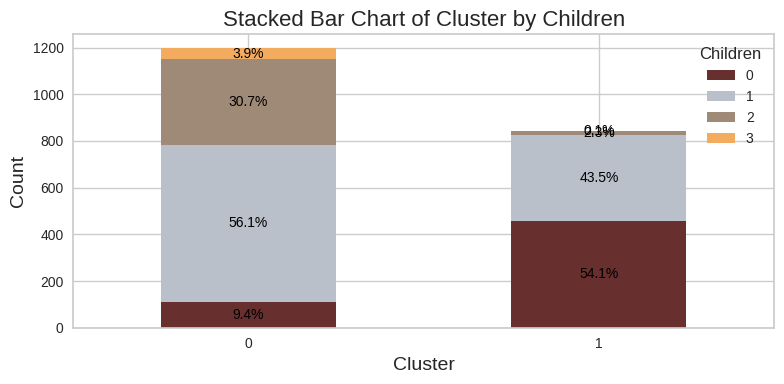


# Cluster 0 and Cluster 1 largest Education level is Graduate and Postgraduate and Undergraduate is more in cluster 0 14,4%  


# The largest percentage of people ‘s MaritalStatus in cluster 0 and 1 is married > 60%



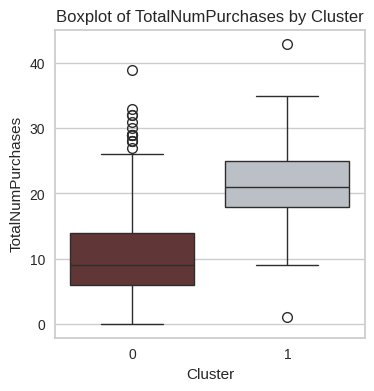
# It seems that percentage of people ‘s Response in cluster 1 larger than cluster 0 by 23.1% but just 9.5 % in cluster 0



# Cluster 0: the largest percent of number of children is 56,1% that has one child

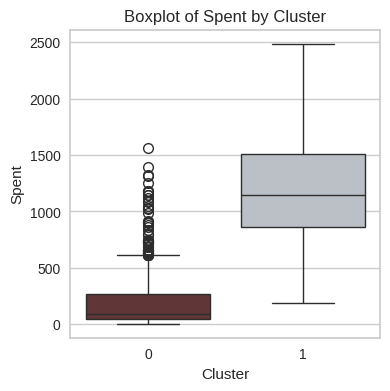
# Cluster 1: the largest percent of number of children is 54.1% that has no child

And the second largest percent is 43.5% that has one child



# in cluster 0 : the median of people ‘s TotalNumPurchases is 9 and ranges from 0 to 26

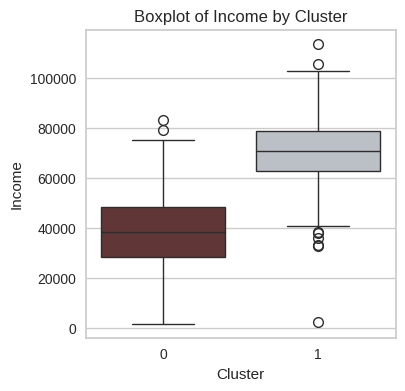
# in cluster 1: the mean of people ‘s TotalNumPurchases is 21 and ranges from 9 to 35



# Cluster 0: spent median is 88.0

# Cluster 1: spent m median is 1148.5

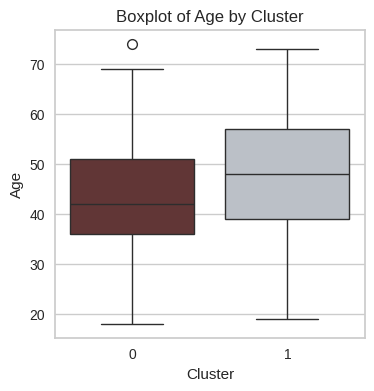
# It seems that Cluster 0 spent less than Cluster 1



# Cluster 0 : income 38241.0

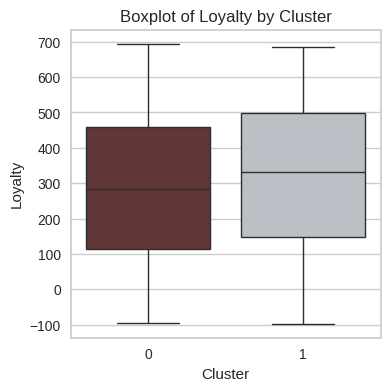
# Cluster 1 : income 70660.6

# It seems that Cluster 1 income is larger than cluster 0 and almost double of it



# Cluster 0: age median 42.0

Cluster 1: age median 48,0



# most customers in both clusters are loyal

**Analysis Conclusion:**

After thoroughly analyzing the clustering results, we observed significant similarities between the characteristics of Cluster 1 and Cluster 2 when the number of clusters (n\_clusters) was set to 3. This overlap indicates that distinguishing these two clusters does not provide additional insights into the customer segmentation.

Additionally, the silhouette score, which measures the cohesion and separation of clusters, was found to be the highest when n\_clusters = 2. This further supports the decision that two clusters provide the most meaningful segmentation.

As a result, the optimal number of clusters for this segmentation problem is 2 (n\_clusters = 2). Below, we provide the detailed profiling of these two distinct clusters, highlighting the key characteristics that differentiate them.

**Profiling customers**

|  |  |  |
| --- | --- | --- |
| **Feature** | **Cluster 0** | **Cluster 1** |
| **Income & Spending** | individuals with low to average income and low spending | individuals with average to high income, and moderate to high spending |
| **Education Level** | |  | | --- | | Majority are Graduate/Postgraduate, with very little percent of Undergraduate | | Majority are Graduate/Postgraduate |
| **Response Rate** | |  | | --- | | Low response rate | | Have more response rate   |  | | --- | |  | |
| |  | | --- | |  |   **Number of Children** | have more than 1 child | have no children or have 1 child |
| **Total Purchases** | Less purchases | Have more number of purchases |
| **Amount Spent** | Lowest spent | Highest spent |
| **Income** | Lowest income | Highest income |
| **Age** | Median: 42.0 | Median: 48.0 |

|  |
| --- |
|  |