10/14/2024

**OMOREDE IGUMA**

Customer Behavior Analysis in E-Commerce

**TABLE OF CONTENTS**

**RESULTS AND DISCUSSIONS**

1. Characteristics of the Customers ---------------------------------------------------- 1

2.0 K-Means Cluster Analysis ----------------------------------------------------------- 3

3.0 Chi-Square Test for Associations --------------------------------------------------- 5

4.0 Analysis of Variance ----------------------------------------------------------------- 6

5.0 Chi-Square Test for Associations --------------------------------------------------- 9

6.0 Independent Sample t-test ------------------------------------------------------------ 11

7.0 Correlation Analysis ------------------------------------------------------------------ 12

**LIST OF TABLES**

**TABLES PAGES**

Table 1.0: Characteristics of the Customers 1

Table 2.1: Final Cluster 3

Table 2.2: Number of Cases in each Cluster 3

Table 3.1: Gender \* Membership Type Crosstabulation 5

Table 4.1.1: Descriptives 6

Table 4.1.2: Robust Tests of Equality of Means 6

Table 4.2.1: Multiple Comparisons 7

Table 5.1: Satisfaction level \* Membership Type Crosstabulation 9

Table 6.0: Independent Sample t-test 11

Table 7.0: Correlations 12

**RESULTS AND DISCUSSIONS**

1. **Descriptive Statistics**

The characteristics of the customers are presented in Table 1.1.

**Table 1.1**

***Characteristics of the Customers***

**Parameters Frequency Percentage**

**Age**

Under 30 117 33.6

31 – 39 173 49.7

40 -46 58 16.7

Total 348 100

Mean 33.6 Years

**Gender**

Female 173 49.7

Male 175 50.3

Total 348 100

**Membership Type**

Bronze 114 32.8

Silver 117 33.6

Gold 117 33.6

Total 348 100

**Satisfaction Level**

Satisfied 125 35.9

Neutral 107 30.7

Unsatisfied 116 33.3

Total 348 100

The demographics of the sample, as outlined in Table 1.1, reveal that the majority of customers (49.7%) are aged between 31 and 39, with a mean age of 33.6 years. The gender distribution was close with (50.3%) male and (49.7%) female participants (*N* = 348). Regarding membership type, customers were evenly split among Bronze (32.8%), Silver (33.6%), and Gold (33.6%). Satisfaction levels are distributed as follows: (35.9%) of customers reported being satisfied, (30.7%) were neutral, and (33.3%) were unsatisfied.

**2.0 K-Means Cluster Analysis**

The results from a k-mean cluster analysis are presented in Table 2.1 and Table 2.2.

**Table 2.1**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *Final Cluster Centre* | | | | |
|  | Cluster | | | |
| 1 | 2 | 3 | 4 |
| Total Spent | 1459.77 | 474.22 | 1165.04 | 748.43 |
| Items Purchased | 20 | 9 | 15 | 12 |
| Age | 29 | 39 | 31 | 32 |

*Note:* Hierarchical cluster analysis was conducted to determine the number of clusters via a dendrogram to use for the k-means cluster analysis (see SPSS output on the GitHub page)

**Table 2.2**

***Number of Cases in each Cluster***

|  |  |  |
| --- | --- | --- |
|  | | |
| Cluster | 1 | 58 |
|  | 2 | 114 |
|  | 3 | 59 |
|  | 4 | 117 |
| Valid |  | 348 |

A K-means cluster analysis was performed to group customers based on age, items purchased, and total spent. A four-cluster solution was selected, and the final cluster centres are reported in Table 2.1.

Cluster 1 (*n* = 58) is characterized by the youngest customers (*M* = 29) with the highest number of items purchased (*M* = 20). Cluster 2 (*n* = 114) includes older customers (*M* = 39) with the lowest number of items purchased (*M* = 9), while Cluster 3 (*n* = 59) represents the slightly younger customers (*M* = 31) with moderate item purchased (*M* = 15). Finally, Cluster 4 (*n* =117) represents the slightly older customers (*M* = 32) with a low number of items purchased (*M* = 15).

The frequency of total money spent was significantly different across all clusters, younger customers tend to spend more while as their age increases, customers tend to spend less as well as buy fewer items.

The four clusters represent distinct customer groups, primarily differing in age, number of items purchased, and total spent. Cluster 1 represents younger, high-spending customers, Cluster 2 includes older, low-spending customers, Cluster 3 consists of slightly younger, moderate-spending customers and Cluster 4 consists of middle-aged customers with low spending. Since younger and slightly younger customers tend to spend more and buy more, the company should treat this cluster of customers properly by doing promos. For slightly older and older customers, marketing research should be conducted to find ways to incentivise them to spend more and buy more.

**3.0: Chi-Square Test for Association**

The results from a Chi-Square Test for Association between gender and membership applied are presented in Table 3.1.

**Table 3.1**

*Gender \* Membership Type Crosstabulation*

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Gender \* Membership Type Crosstabulation | | | | |  |  |  |  |  |
|  |  |  | *Membership Type* | |  |  |  |  |  |
|  |  |  | *Bronze* | *Gold* | *Silver* | *Total* | *χ²* | *df* | *φ*c |
| Gender | Female | Count | 114 | 58 | 1 | 173 | 227.04\* | 2 | 0.808\* |
|  |  | Expected Count | 56.7 | 58.2 | 58.2 | 173.0 |  |  |  |
|  |  | Std. Residual | 7.6 | 0 | -7.5 |  |  |  |  |
|  | Male | Count | 0 | 59 | 116 | 175 |  |  |  |
|  |  | Expected Count | 57.3 | 58.8 | 58.8 | 175.0 |  |  |  |
|  |  | Std. Residual | -7.6 | .0 | 7.5 |  |  |  |  |
| Total |  | Count | 114 | 117 | 117 | 348 |  |  |  |
|  |  | Expected Count | 114.0 | 117.0 | 117.0 | 348.0 |  |  |  |

*Note*. 2x3 contingency table so Cramer’s V coefficient was used for the effect size.

\* = *p* < .05

A chi-square test of independence was performed to examine the relationship between gender and membership type. The relation between these variables was significant, χ² (2, *N* = 348) = 227.04, *p* < .001. The Cramer’s V coefficient (**φ*c*** = 0.808, *p* < .001) indicates a strong association between gender and the membership type.

From the crosstabulation, it was found that females were majorly bronze membership subscribers, while males were majority gold membership subscribers.

**4.0 Analysis of Variance (ANOVA)**

The results from an ANOVA between total spending and age category are presented in Table 4.1, table 4.2, and Table 4.3.

**Table 4.1.1**

*Descriptives*

Total Spent

|  |  |  |  |
| --- | --- | --- | --- |
| *Age Category* | *N* | *Mean* | *SD* |
| Under 30 | 117 | 1154.02 | 272.89 |
| 31-39 | 173 | 724.77 | 318.24 |
| 40-46 | 58 | 595.18 | 57.28 |
| Total | 348 | 847.49 | 354.21 |

**Table 4.1.2**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *Robust Tests of Equality of Means* | | | | |  |
| Total Spend | | | | |  |
|  | *Statistica* | *df1* | *df2* | *p-value* | *ε2* |
| Welch | 230.106\* | 2 | 209.795 | <.001 | 0.394 |
| *Note*: The assumption of homogeneity of variance was violated, so the alternative Welch statistics was used for the main analysis and post hoc test. | | | | |  |
| a. Asymptotically F distributed. | | | | |  |

\* = *p* < .05

**4.2 Post Hoc Test**

**Table 4.2.1**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| *Multiple Comparisons* | | | | | | |
| Dependent Variable: Total Spending | | | | | | |
| Tamhane | | | | | | |
| *(I) Age Category* | *(J) Age Category* | *Mean Difference (I-J)* | *Std. Error* | *p-value* |
|  |
| Under 30 | 31 - 39 | 429.25\* | 34.96 | <.001 |  |  |
| 40 - 46 | 558.84\* | 26.33 | <.001 |  |  |
| 31 - 39 | Under 30 | -429.25\* | 34.96 | <.001 |  |  |
| 40 - 46 | 129.59\* | 25.34 | <.001 |  |  |
| 40 - 46 | Under 30 | -558.84\* | 26.33 | <.001 |  |  |
| 31 - 39 | -129.59\* | 25.34 | <.001 |  |  |
| \* The mean difference is significant at the 0.05 level. | | | | | | |

A one-way ANOVA was conducted to determine whether there were significant differences in total spending based on age category (see Table 4.0). There was a statistically significant difference in total spending between the three age groups, as determined by Welch’s *F* (2, 209.795) = 230.106, *p* < .001. The effect size Epsilon-squared was 0.394 (39.4%); indicating that the independent variable age category explained 39.4% of the variation in the dependent variable total spent. Post hoc analyses using the Tamhane test revealed that customers under 30 spent significantly more ($1154.02) than those aged 31–39 ($724.77), with a mean difference of $429.25, *p* < .001. Similarly, those under 30 spent significantly more than those aged 40–46 ($595.18), with a mean difference of $558.84, *p* < .001. Significant differences were also between the 31–39 and 40–46 age groups (mean difference = $129.59, *p* < .001).

5.0 **Chi-Square Test for Association**

The results from a Chi-Square Test for Association between Satisfaction Level and Membership Type are presented in Table 5.1.

**Table 5.1**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| *Satisfaction level \* Membership Type Crosstabulation* | | | | | | |
|  | | | *Membership Type* | | | *Total* |
| *Bronze* | *Gold* | *Silver* |
| Satisfaction level | Satisfied | Count | 0 | 117 | 8 | 125 |
| Expected Count | 40.9 | 42.0 | 42.0 | 125.0 |
| Standardized Residual | -6.4 | 11.6 | -5.2 |  |
| Neutral | Count | 56 | 0 | 51 | 107 |
| Expected Count | 35.1 | 36.0 | 36.0 | 107.0 |
| Standardized Residual | 3.5 | -6.0 | 2.5 |  |
| Unsatisfied | Count | 58 | 0 | 58 | 116 |
| Expected Count | 38.0 | 39.0 | 39.0 | 116.0 |
| Standardized Residual | 3.2 | -6.2 | 3.0 |  |
| Total | | Count | 114 | 117 | 117 | 348 |
| Expected Count | 114.0 | 117.0 | 117.0 | 348.0 |

*Note*: Satisfaction Level and Membership Type are ordinal variables, so the Gamma coefficient was used to determine the effect size.

*χ²* (4) = 315.80, *p <* .05, *γ**=* -.804

A chi-square test of independence was conducted to examine the relationship between satisfaction level and membership type (see Table 5.0). The association was significant, χ²(4, *N* = 348) = 315.80, *p* < .001, with a negatively weak and non-significant effect size (Gamma = -.036, *p>*.05). Gold members were more likely to report being satisfied (std. residual = 11.6), while bronze and silver members were less likely to report satisfaction (std. residual = 3.2 and -6.2, respectively).

6.0 **Independent Sample T-test**

The result from an independent sample t-test between total spending and the discount applied is presented in Table 6.1.

**Table 6.1**

*Independent Sample t-test*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | *Discount applied?* | *N* | *M* | *SD* | *t* | *df* | *p-value* | *𝑑* |
| Total Spent |  |  |  |  |  |  |  |  |
|  | TRUE | 175 | 787.27 | 281.87 | -3.17 | 300.56 | 0.002\* | -0.34 |
|  | FALSE | 173 | 909.01 | 419.59 |  |  |  |  |

*Note:* The assumption of homogeneity of variance was violated, hence, heterogeneity of variance results was used for the t-statistics, df and p-value

\* = *p* < .05

An independent samples t-test was conducted to compare the total spending between customers who had a discount applied and those who did not (see Table 6.1). There was a significant difference in total spending for customers with a discount (*M* = 787.27, *SD* = 281.87) and without a discount (*M* = 909.01, *SD* = 419.59), *t*(300.56) = -3.17, *p* = .002. The effect size, as measured by Cohen’s d, was d = -0.34, indicating a negatively small effect size.

7.0 **Correlation Analysis**

The results from spearman’s rank correlation analysis between days since the last purchase, membership type, and satisfaction level are presented in Table 7.1.

**Table 7.1**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *Correlations* | | | | |
|  | | *Days Since the Last Purchase* | *Membership type* | *Satisfaction Level* |
| Days Since Last the Purchase | Pearson Correlation | 1 | -.480\*\* | .751\*\* |
| Sig. (2-tailed) |  | <.001 | <.001 |
| N | 348 | 348 | 348 |
| Membership type | Pearson Correlation | -.480\*\* | 1 | -.748\*\* |
| Sig. (2-tailed) | <.001 |  | <.001 |
| N | 348 | 348 | 348 |
| Satisfaction Level | Pearson Correlation | .751\*\* | -.748\*\* | 1 |
| Sig. (2-tailed) | <.001 | <.001 |  |
| N | 348 | 348 | 348 |
| \*. Correlation is significant at the 0.05 level (2-tailed). | | | | |

A Spearman’s rank correlation analysis was performed to assess the relationship between days since the last purchase, membership type, and satisfaction level (Table 7.1). A positively significant correlation was found between days since the last purchase and satisfaction level (rs = .751, *p* < .001) it could be that satisfied customers don’t feel pressured to buy frequently, and thus, they may only return when they really need something but still report high satisfaction., while a correlation between days since the last purchase and membership type was negative and significant (rs = -.480, *p*>.001) indicating that higher-tier members (Gold) purchase more often. Membership type and satisfaction level were negative and significant (rs = -.748, *p* < .001. A possible explanation could be that gold members may expect VIP treatment, so even minor inconveniences lower their satisfaction more than bronze members.