31/03/2025

**OMOREDE IGUMA**

Consumer Behaviour 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 ----------------------------------------------------------------- 7

5.0 Chi-Square Test for Associations --------------------------------------------------- 10

6.0 Independent Sample t-test ------------------------------------------------------------ 12

7.0 Correlation Analysis ------------------------------------------------------------------ 13

**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 10

Table 6.0: Independent Sample t-test 12

Table 7.0: Correlations 13

**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**

20 - 29 117 33.4

31 – 39 175 50

40 -46 58 16.6

Total 350 100

Mean 33.6 Years

**Gender**

Male 175 50

Female 175 50

Total 350 100

**Membership Type**

Bronze 116 33.1

Silver 117 33.4

Gold 117 33.4

Total 350 100

**Satisfaction Level**

Unsatisfied 116 33.1

Neutral 107 30.6

Satisfied 127 36.3

Total 350 100

The demographics of the consumer, as outlined in Table 1.1, reveal that the majority of customers (50%) are aged between 31 and 39, with a mean age of 33.6 years. The gender distribution was similar, 50% male and 50% female participants (*N* = 350). Regarding membership type, customers were evenly split among Bronze (33.1%), Silver (33.4%), and Gold (33.4%). Satisfaction levels are distributed as follows: 36.3% of customers reported being satisfied, 30.6% were neutral, and 33.1% were unsatisfied.

**2.0 K-Means Cluster Analysis**

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

**Table 2.1**

|  |  |  |
| --- | --- | --- |
| *Final Cluster Centres* | | |
|  | Cluster | |
| 1 | 2 |
| Total Spend | 611.50 | 1311.14 |
| Items Purchased | 10 | 18 |
| Age | 35 | 30 |

*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 | 233.000 |
| 2 | 117.000 |
| Valid | | 350.000 |
| Missing | | .000 |

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 centers are reported in Table 2.1.

Cluster 1 (*n* = 223) is characterized by the older customers (*M* = 35) with the lowest number of items purchased (*M* = 10). Cluster 2 (*n* = 117) includes younger customers (*M* = 30) with the highest number of items purchased (*M* = 18).

The frequency of total money spent was significantly different across both clusters; younger customers tend to spend more ($1311.14), while older customers tend to spend less ($611.50), 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 older, low-spending customers, and Cluster 2 includes younger, higher-spending customers. Since younger customers tend to spend more and buy more, the company should treat this cluster of customers properly by doing promotions. For 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-squared 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* | *Silver* | *Gold* | *Total* | *χ²* | *df* | *φ*c |
| Gender | Male | Count | 0 | 116 | 59 | 175 | 229.04\* | 2 | 0.809\* |
|  |  | Expected Count | 58 | 58.5 | 58.5 | 175.0 |  |  |  |
|  |  | Adjusted Residual | -13.2 | 13.0 | .1 |  |  |  |  |
|  | Female | Count | 116 | 1 | 58 | 175 |  |  |  |
|  |  | Expected Count | 58 | 58.5 | 58.5 | 175.0 |  |  |  |
|  |  | Adjusted Residual | 13.2 | -13.0 | -.1 |  |  |  |  |
| Total |  | Count | 116 | 117 | 117 | 350 |  |  |  |
|  |  | Expected Count | 116 | 117 | 117 | 350 |  |  |  |

*Note*. A 2x3 contingency table was used, so Cramer’s V coefficient was employed to calculate the effect size.

\* = *p* < .05

A chi-square test of independence was performed to examine the relationship between gender and membership type. The relationship between these variables was significant, χ²(2, *N* = 350) = 229.04, *p* < .001. The Cramer’s V coefficient (**φ*c*** = 0.809, *p* < .001) indicates a strong association between gender and the membership type.

From the crosstabulation, it was found that males were highly overrepresented in the silver membership category (adjusted residual = 13.0) and completely absent from the bronze category (adjusted residual = -13.2). Conversely, females were significantly overrepresented in the bronze category (adjusted residual = 13.2) and almost absent from the silver category (adjusted residual = -13.0). No significant deviation was observed in the gold membership category (adjusted residual = 0.1 for males, -0.1 for females), suggesting an equal distribution of gold membership across genders.

**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* |
| 20-29 | 116 | 1156.29 | 272.07 |
| 30-39 | 175 | 720.90 | 318.71 |
| 40-49 | 58 | 597.62 | 56.43 |
| Total | 348 | 845.13 | 354.99 |

**Table 4.1.2**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *Robust Tests of Equality of Means* | | | | |  |
| Total Spend | | | | |  |
|  | *Statistica* | *df1* | *df2* | *p-value* | *ε2* |
| Welch | 229.401\* | 2 | 209.181 | <.001 | 0.395 |
| *Note*: The assumption of homogeneity of variance was violated, so the alternative Welch statistic 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 | | | | | | |
| Games-Howell | | | | | | |
| *(I) Age Category* | *(J) Age Category* | *Mean Difference (I-J)* | *Std. Error* | *p-value* |
|  |
| 20-29 | 30 - 39 | 435.40 | 34.91 | <.001 |  |  |
| 40 - 49 | 558.68 | 26.33 | <.001 |  |  |
| 30 - 39 | 20-29 | -435.40 | 34.91 | <.001 |  |  |
| 40 - 49 | 123.28 | 25.21 | <.001 |  |  |
| 40 - 49 | 20-29 | -558.68 | 26.33 | <.001 |  |  |
| 30 - 39 | -123.28 | 25.21 | <.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 among the three age groups, as determined by Welch’s *F*(2, 209.181) = 229.401, *p* < .001. The effect size Epsilon-squared was 0.395 (39.5%), indicating that the independent variable age category explained 39.5% of the variation in the dependent variable total spent. Post hoc analysis using the Games-Howell test revealed that customers aged 20-29 spent significantly more ($1156.29) than those aged 30–39 ($720.90), with a mean difference of $435.40, *p* < .001. Similarly, those aged 20-29 spent significantly more than those aged 40–49 ($597.62), with a mean difference of $558.68, *p* < .001. Significant differences were also between the 30–39 and 40–49 age groups (mean difference = $123.28, *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* | *Sliver* | *Gold* |
| Satisfaction level | Unsatisfied | Count | 58 | 58 | 0 | 116 |
| Expected Count | 38.4 | 38.8 | 38.8 | 116.0 |
| Adjusted Residual | 4.7 | 4.6 | -9.3 |  |
| Neutral | Count | 56 | 51 | 0 | 107 |
| Expected Count | 35.5 | 35.8 | 35.8 | 107.0 |
| Adjusted Residual | 5.1 | 3.7 | -8.8 |  |
| Satisfied | Count | 2 | 8 | 117 | 127 |
| Expected Count | 42.1 | 42.5 | 42.5 | 127.0 |
| Adjusted Residual | -9.5 | -8.1 | 17.6 |  |
| Total | | Count | 116 | 116 | 117 | 117 |
| Adjusted Count | 116.0 | 116.0 | 117.0 | 117.0 |

*Note*: Satisfaction Level and Membership Type are ordinal variables, so the Gamma coefficient was used to determine the effect size. *χ²* () = 309.44, *p* < .05, *γ =* .793

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* = 350) = 309.44, *p* < .001, with a and positively large effect size (Gamma = -.793).

From the cross tabulation, gold members were significantly more likely to be satisfied (adjusted residual = 17.6) than expected, while Bronze (adjusted residual = -9.5) and Silver (adjusted residual = -8.1) members were significantly underrepresented in the satisfied category. In contrast, Bronze (adjusted residual = 5.1) and Silver (adjusted residual = 3.7) members were overrepresented in the Neutral category, while Gold members were significantly underrepresented (adjusted residual = -8.8). Similarly, Bronze (adjusted residual = 4.7) and Silver (adjusted residual = 4.6) members were overrepresented among unsatisfied customers, while Gold members were absent from this category (-9.3).

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 | 821.40 | 213.61 | -1.251 | 245.596 | 0.212 | -.134 |
|  | FALSE | 174 | 868.99 | 454.38 |  |  |  |  |

*Note:* The assumption of homogeneity of variance was violated, hence, heterogeneity of variance results were 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 non-statistically significant difference in total spending for customers with a discount (*M* = 821.40, *SD* = 213.61) and without a discount (*M* = 868.99, *SD* = 454.38), *t*(245.596) = -1.251, *p* = 0.212. The effect size, as measured by Cohen’s d, was d = -.134, indicating a 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 Last Purchase* | *Membership Type* |
| Spearman's rho | Days Since Last Purchase | Correlation Coefficient | 1.000 | -.476\*\* |
| Sig. (2-tailed) |  | 0.000 |
| N | 350 | 350 |
| Membership Type | Correlation Coefficient | -.476\*\* | 1.000 |
| Sig. (2-tailed) | 0.000 |  |
| N | 350 | 350 |
| \*\*. Correlation is significant at the 0.01 level (2-tailed). | | | | |

A Spearman’s rank correlation analysis was performed to assess the relationship between days since the last purchase and satisfaction level (Table 7.1). A negatively significant correlation was found between days since the last purchase and membership type (rs = -.476, *p* < .001). This indicates that higher-tier members (Gold) make more purchases, while Bronze and Silver members are more likely to experience large gaps in the number of days between purchases.