**Cluster Analysis of the India consumer market – A focus on durable goods**

This report provides insights into the India consumer expenditure market. A range of analytical techniques including socio-economic and cultural factors are used to uncover spending behaviours. Key findings reveal that Gold Ornaments and Motor cycles account for significant expenditure due the cultural value and local necessity to use a motorcycle.

I intend to uncover how household spending patterns vary across different socio-economic strata using the **2014-2015 Indian Household Consumer Expenditure Survey (Microdata.gov.in, 2015)** focusing on durable goods - which can be understood as goods that intended for regular usage for several years without replacement i.e., jewellery and washing machines. **(Lakshmanasamy, 2024)**. I hypothesise that spending shares on durable goods are dependent on MPCE (monthly per capita expenditure) levels, with significant variations in prioritisation across income groups. MPCE is a key variable as it illustrates month-month consumer expenditure behaviour.

Previous research suggests that gold ornaments - understood as objects that are making in gold and worn for personal adornment, such as bracelets - hold particular cultural value in India **(Kumar and Kumar 2014).** **Mouly Potluri, Ansari and Challagundla (2013)** identify spending is emphasised in marriage. **Chellaboina, Gembali and Prabha (2022)** identity that it is crucial to understand customer behaviour and categorise customers according to their demographic and purchasing habits. Customer segmentation is identified as the most effective way of reaching customers therefore a K-means clustering technique is used.

The data **(Microdata.gov.in, 2015)** is a survey that covers the whole of the Indian union except villages in Andaman and Nicobar Islands. It is in the form of sample survey data involving randomly selected households based on sampling procedure. After stratification, it involved **47,535 (rural households)** and **36065 (urban households).** My spending variables are MPCE, gold ornaments, mobile phones and motorcycles. I have chosen these due to my background research on the durable goods industry that shows gold ornaments are sacred. Additionally, I want to analyse spending differences between luxury and necessity goods when demographic variables are present. My demographic variables are rural/urban, whether household is Hindu and household size. These variables aim to capture a range of socio-economic strata that may affect insights into **A graph of a group of bars

Description automatically generated**consumer spending behaviour within India.

Figure 1- Mean spending of variables by MPCE categories

A graph of a number of boxes

Description automatically generated with medium confidenceFigure 1 illustrates the mean spending of 5 variables in “High”, “Medium” and “Low” MPCE categories produced by mutating the MPCE variable using 1/3 quartiles. In all three, motorcycle spending is the highest with a significant increase in the mean spending in the high MPCE category suggesting a few outliers with extreme spending. Gold ornaments also present higher mean spending across all categories in contrast to toys and laptops with minimal spending. We may conclude, high levels of motor cycle spending is due to the necessity of needing transport particularly in rural areas where infrastructure may be underdeveloped. Moreover, high gold spending across all categories may reflect the high cultural value and sacredness of gold especially in weddings as discussed by **Mouly Potluri, Ansari and Challagundla (2013).** Low spending on toys may indicate a lower household size with a lack of children therefore toys are not of interest. Similarly, low spending on laptops may reflect low necessity as jobs may not require them i.e., construction workers.

Figure 2 - Boxplots of the distributions of spending variables

**A diagram of a cluster plot

Description automatically generated**Figure 2 illustrates the distribution of spending variables discussed above. The Y-axis scale is using the log algorithm due to the large amounts of data and variability between variables. Consequently, when analysing the result we must consider small data points may account for large differences in spending. We can infer that gold ornaments have large amounts of outliers both extreme and minimal spenders whilst the median is around 20,000. Motorcycle spending has the highest amount of variability between scores and few outliers reinforcing the necessity to spend on motorcycles. Mobile phones have low a lower median spending due to the significant amount of low spenders skewing the results. The distribution of toys reinforces the low mean spending across all categories in Figure 1 with extreme and minimal spending with the median equating to >1000. Laptop spending has little variability compared to motorcycles with few outliers with minimal spending.

Figure 3 - K means cluster plot of spending and demographic variables

Figure 3 illustrates the result of k-means clustering. I included spending variables MPCE, Gold ornaments, Motorcycles and Mobile Phones with demographic variables Hindu (whether household is Hindu), Rural (Urban or rural habitat) and household size. We can infer that 38% point variability is relatively low however due to the size of the data set this is expected. At first glance, the clusters provide little information as they overlap a significant amount. However, to compliment the k-means plot I decided to create a centroid table which illustrates the values behind the clusters.

**A screenshot of a data

Description automatically generated**

Figure 4 - Cluster Centroid Values Table

**Cluster 1**

Gold spending is highest at 16,790 Rupees which indicates cultural practises that prioritise gold i.e., weddings. Mobile phone spending is moderate at 4428 Rupees reflecting the importance of mobile technology but without extreme spending. MPCE is the highest compared to other clusters at 2708 Rupees suggesting a Middle-High income group. Rural at 1.52 suggests a more urban household which reflects the low necessity of a motor cycle (represented at 1,088 Rupees). A figure of 0.81 Hindu suggests a predominant religious affiliation which might reinforce the high spending on Gold ornaments. The household size of 6.61 suggests multi-generational households.

**Cluster 2**

Motorcycle spending is very high at 48,282 and significantly higher than cluster 1. This suggests a necessity for transportation, possibly in areas of inadequate infrastructure. Gold spending is still high even though far less significant than cluster 1, suggesting some cultural value of gold. Mobile phone spending at 3294 Rupees suggests moderate investment in mobile technology. MPCE of 2318 suggests a middle income household. The rural value of 1.41 suggests a more rural household reflective in the high spending on motorcycles. Moreover, at 6.81 indicating a large household size.

**Cluster 3**

Motorcycle spending is low at 2029 Rupees similar to cluster 2. Gold spending is also the lowest across all clusters suggesting that cultural or social factors that drive gold demands may be less influential here Mobile phone spending is similar to other clusters at 3509 Rupees. MPCE value of 2327 Rupees suggests middle income group. The household size in this cluster is still large (6.23)

**A screenshot of a graph

Description automatically generated**Figure 5 presents my silhouette score results which can be understood as a measure of cluster cohesion and separation, which allow for assessment of the quality of our segmentation. Scores that are closer to 1 indicate high separation and cohesive clusters. Cluster 2 presents the highest score 0.66 suggesting that it is well separated and the data points within the cluster are highly cohesive. Cluster 3 has a score of 0.58 which also demonstrates good cohesion and separation albeit slightly lower than Cluster 2. Cluster 1 has the lowest score 0.43 suggesting less separation and likelihood of overlap with other clusters.

Figure 5 - Silhouette Scores Table

In conclusion, this report aims to provide insights for market research analysts and what areas to focus on. A key issue with the data set is that due to it being a household survey, in-depth analysis into gender and age was not applicable. Both demographic strata are provide key insights which can further provide insight such as, targeted promotions for women.

In relation to Figure 4, **Cluster 1** provides insights that retailers producing gold ornaments particularly within contexts of cultural events such as weddings should target this cluster. There is a significant association between cultural sacredness of gold therefore, emphasising wedding-related promotions during Diwali may be effective. **Cluster 2** provides insights that rural households favour motorcycles illustrated with the significant mean spending. Therefore, I recommend premium motorcycle brands such as Ducati should target this cluster promoting products to rural/urban households where this transportation type is a necessity. Moreover, gold spending is still moderate therefore promotion of mid-range gold products could be valuable to businesses. **Cluster 3** provides insights into low spending of gold and motorcycles. Therefore this cluster may be ideal for budget friendly mobile devices as the spending on mobile phones is still moderate. Research could focus on value for money products that appeal to customers conscious of spending high amounts.

Lastly, in relation to Figure 5, the high silhouette score of **Cluster 2** suggests that market research analysts should focus this group for in-depth analysis and tailored strategies, as it displays distinct characteristics it makes it easier to target effectively. **Cluster 3**’s large size and good silhouette score suggest it is a key segment for broad strategies, although due to its moderate cohesion further analysis may be required to account for the lack of variability. **Cluster 1**’s low scores suggests closer examination into the overlap issues may be beneficial i.e., Re-clustering to improve interpretability.

**References**

* KUMAR, D.V. and KUMAR, D.V. (2014). GOLD ORNAMENT : MEANING, NATURE AND MARKETING STRATEGIES. JETIR, [online] 10(1), pp.a296–a299a296–a299. Available at: https://www.jetir.org/view?paper=JETIR2301037

[Accessed 30 Dec. 2024]

* Lakshmanasamy, T. (2024). Demand for Consumer Durables in India: Quantile Regression Estimation of Household Expenditure on Durable Goods. INDIAN DEVELOPMENT POLICY REVIEW, 5(1), pp.117–130.
* Microdata.gov.in. (2015). India - Household expenditure on Services and Durable goods - JULY 2014 - JUNE 2015, 72 round - variable - V241. [online] Available at: https://microdata.gov.in/NADA43/index.php/catalog/137/datafile/F8/V241 [Accessed 24 Jan. 2025].
* Srinivas Chellaboina, Manoj Gembali and Sathya Prabha S (2022). Product Recommendation based on Customer Segmentation Engine. Product Recommendation based on Customer Segmentation Engine. doi:https://doi.org/10.1109/conit55038.2022.9847990.
* Mouly Potluri, R., Ansari, R. and Challagundla, S. (2013). A Study on Spending Patterns and Buying Motives of Indian Gold Consumers. Journal of Distribution Science, 11(7), pp.31–37. doi:https://doi.org/10.13106/jds.2013.vol11.no7.31.