Market Segmentation Analysis

Step 4: Exploring Data

Prepared by Ankit Kumawat

Introduction

After data collection, the next critical step is to *explore the data*. This ensures the data is clean, valid, and ready for segmentation. Good exploration helps identify errors, outliers, and key patterns — all of which impact the quality of the final segments.

1) A First Look at the Data

- Begin with simple checks: inspect raw data for obvious issues.
- Use frequency tables, plots, or basic summaries.
- Look for missing values, outliers, implausible entries (e.g., negative ages).
- Understand how variables are coded and distributed.

2) Data Cleaning

- Correct obvious data errors: typos, impossible values, duplicates.
- Decide whether to remove, adjust, or impute missing or extreme data points.
- Document all cleaning steps to keep the process transparent.
- Cleaning reduces noise and bias.

3) Descriptive Analysis

- Use descriptive statistics (mean, median, mode) to get an overview.
- Visualize data: histograms, box plots, scatterplots help spot trends.
- Identify patterns, groupings, or relationships that could inform segmentation.
- Check correlations between variables.

4) Pre-Processing

- Standardize or normalize numerical variables if needed.
- Convert categorical variables into usable forms (e.g., dummy coding).
- Reduce variables if too many: use factor analysis or merge similar categories.
- Ensure variables are suitable for clustering or other segmentation methods.

5) Principal Components Analysis (Optional)

- If many numeric variables exist, PCA helps reduce dimensions.
- PCA finds new variables that explain most of the variation.
- Useful for simplifying complex data sets before segment extraction.
- Helps detect hidden structures.

Step 4 Checklist

- Did you check for errors and missing values?
- Is the data cleaned and ready?
- Did you describe and visualize key variables?
- Have you transformed variables appropriately?
- Did you reduce complexity where needed?

Key Takeaway

Exploring data is not just about cleaning — it's about *understanding* what you have. Good exploration uncovers hidden problems and insights that make segmentation more accurate and meaningful.