For our dataset with 4,502 rows, we can consider various factors to determine the appropriate sample size for random sampling. While a sample size of 30 is often cited as a minimal threshold for the Central Limit Theorem to apply, we should consider a larger sample size to enhance the reliability of our results, especially given the diversity of categories and segments in our data.

Sample Size Calculation

Using a sample size calculator can provide a more tailored recommendation based on our dataset's specific characteristics. Here are the factors we can consider in the calculation:

1. **Population Size**: 4,502 (the total number of transactions).

2. **Margin of Error (Confidence Interval)**: Common choices are ±5% or ±3%. A smaller margin of error requires a larger sample size.

3. **Confidence Level**: Typically set at 90%, 95%, or 99%. A higher confidence level requires a larger sample size.

4. **Variability**: If we know the variability of our data, we can use that information. If not, a conservative estimate (50% variability) is often used.

Example Calculation

Assuming we aim for a 95% confidence level and a ±5% margin of error:

Using the formula for sample size:

$$n = \frac{N \cdot Z^2 \cdot p \cdot (1-p)}{(E^2 \cdot (N-1)) + (Z^2 \cdot p \cdot (1-p))}$$

Where:

- n = required sample size
- N = population size (4,502)
- Z = Z-value (1.96 for 95% confidence)
- p = estimated proportion (0.5 for maximum variability)
- E = margin of error (0.05)