**Project Documentation: Unlocking Insights from Merged Transactional Data**

**Objective**

The primary objective of this project is to analyze transaction records, particularly those involving merged categories and amounts, in order to gain insights into customer behavior, preferences, and spending patterns. This project demonstrates the transformative power of clean and well-organized data.

**Techniques Used**

* Data Cleaning
* Data Transformation
* Data Visualization
* Descriptive Statistics
* Advanced Analysis (Clustering)

**Data Source**

The data used in this project was sourced from Wikipedia and consists of transaction records with order IDs, categories, and amounts.

#### ****Step-by-Step Process****

**Step 1: Import Libraries and Load Data**

**Explanation:** Imported necessary libraries and loaded the dataset.

**Step 2: Data Transformation and Reshaping**

**Explanation:** Transformed and reshaped the data by splitting and exploding the 'Category' and 'Amount' columns, ensuring amounts are converted to float.

**Step 3: Data Export**

**Explanation:** Exported the cleaned data to a CSV file for future use.

**Step 4: Data Transformation and Summarization**

**Explanation:** Summarized the data by grouping it by 'Category', calculating the total amount spent and the count of transactions.

**Step 5: Category Summary Export**

**Explanation:** Exported the summarized data to a CSV file for reporting.

**Step 6: Data Visualization**

**Explanation:** Created a horizontal bar chart to visualize the total amount spent by category.

**Step 7: Statistics Summary**

**Summary statistics for 'Total Amount Spent'**

**count:** The number of non-missing (non-null) values in the "Total Amount Spent" column is 10. In other words, there are 10 data points in this column

**unique:** The number of unique values in the "Total Amount Spent" column is 10, indicating that all values in the column are unique.\*\*top:\*\* The most frequent (mode) value in the "Total Amount Spent" column is 1256.22

**freq:** The frequency of the most frequent value, 1256.22, is 1. This means it appears only once in the column.

**dtype:** The data type of the "Total Amount Spent" column is indicated as 'float64', which means it contains floating-point numbers

**Summary statistics for 'Transaction Count'**

**count:** The number of non-missing values in the "Transaction Count" column is 10, which matches the number of data points in this column.

**mean:** The mean (average) value of the "Transaction Count" column is approximately 1.3. This means that, on average, there are 1.3 transactions per category.

**std:** The standard deviation of the "Transaction Count" column is approximately 0.483. It measures the dispersion or spread of the data around the mean.

**min:** The minimum value in the "Transaction Count" column is 1, which is the lowest number of transactions.

**25%:** The first quartile (25th percentile) value is 1, indicating that 25% of the categories have one transaction.

**50%:** The median (50th percentile) value is 1, suggesting that half of the categories have one transaction.

**75%:** The third quartile (75th percentile) value is 1.75, implying that 75% of the categories have 1.75 transactions or fewer.

**max:** The maximum value in the "Transaction Count" column is 2, which is the highest number of transactions.

**Step 8: Advanced Analysis: Customer Segmentation**

**Explanation:** Applied KMeans clustering for customer segmentation based on spending and transaction count.

**Conclusion**

The concluded project demonstrates the transformative power of clean and well-organized data by analyzing transaction records to uncover valuable insights into customer behavior, preferences, and spending patterns. By employing a meticulous step-by-step approach to data cleaning, transformation, and visualization, I was able to successfully converted raw, unstructured data into a structured format that facilitated meaningful analysis.

Through the summarization and visualization of transaction data, i identified key spending trends across various product categories. The integration of advanced techniques such as KMeans clustering further enhanced our understanding by segmenting customers based on their spending habits and transaction frequency. This segmentation provides a foundation for targeted marketing strategies and personalized customer experiences.

In conclusion, this project not only highlights the critical role of data cleaning and transformation but also demonstrates how advanced analytical methods and interactive visualizations can unlock new dimensions of understanding in customer behavior and business strategy. The insights gained from this analysis are instrumental in driving informed decision-making and fostering a data-driven culture within the organization.