**INTRODUCTION**

This report presents a detailed analysis of transaction data using clustering techniques to uncover patterns and anomalies. Our dataset includes various features such as transaction dates, amounts, merchant names, and categorical codes. The primary goal is to identify significant trends and patterns across different clusters of transactions, thereby gaining insights into spending behaviours and potential anomalies.

**DEVELOPMENT.**

* **Distribution of Transactions by Month**

The bar graph showing the distribution of transactions by month highlights clear seasonal trends. June and November stand out as peak months, each with over 50,000 transactions, likely due to mid-year sales promotions and holiday shopping. Conversely, August has the lowest transaction count, possibly due to a seasonal dip in business activity. These trends emphasize the importance of targeted promotions and resource planning during peak periods and strategies to boost sales during slower months.

* **Top 20 Merchants by Transaction Count**

The top 20 merchants by transaction count reveal Amazon and ASDA as dominant players. "amznmktplace" leads with over 35,000 transactions, followed by "amazon uk marketplace" and "amazon uk retail," indicating Amazon's significant market presence. ASDA appears multiple times, reflecting its strong foothold in grocery and home shopping sectors. This data can guide strategic partnerships and promotional efforts with high-volume merchants to leverage their popularity.

* **Transaction Amounts by Month**

The box plot of transaction amounts by month identifies significant outliers, particularly in May, where amounts spike up to 4 million units. This anomaly suggests unusual high-value transactions during this period, warranting further investigation. Overall, transaction amounts are relatively low and consistent across other months, highlighting the need to monitor and analyze high-value transactions for potential anomalies or significant events.

* **Transaction Amounts by Day of the Month**

The scatter plot of transaction amounts by day of the month reveals significant outliers on the 4th and 9th days, with amounts reaching up to 4 million units. These high-value transactions, sporadic in nature, suggest the need for regular monitoring and investigation to detect potential fraud or errors. Most transactions cluster around lower values, emphasizing routine financial activities.

* **Transaction Amounts by Top 10 Merchants**

The box plot of transaction amounts for the top 10 merchants highlights variability in spending patterns. "Travelodge gb0000" and "Travelodge website" exhibit higher median transaction amounts with significant outliers, suggesting occasional high-value transactions. In contrast, "Amazon UK Marketplace" and "Amazon UK Retail" show lower, more frequent transactions. Understanding these patterns can aid in monitoring high-value transactions and managing merchant relationships effectively.

* **Transaction Amounts by Top 10 Transaction Descriptions**

The box plot for the top 10 transaction descriptions indicates that most categories have low-value transactions clustered around zero. However, "Equip Operational" stands out with several high-value transactions, suggesting significant spending in this category. This highlights the need to monitor operational expenditures closely and investigate high-value transactions for potential optimization.

* **Transaction Amounts by Top Directorate**

The box plot for transaction amounts across top directorates shows most categories with low-value transactions centered around zero. However, the "CYP&F SCHOOLS" directorate exhibits high-value transactions and notable negative outliers, indicating potential anomalies or errors. Investigating these outliers can provide insights into spending behaviors and help identify areas for cost optimization.

* **Scatter Plot of Transaction Amounts by Day and Month**

The scatter plot of transaction amounts by day and month illustrates continuous transaction activity with significant peaks in February, April, and June. These high-value transactions indicate specific periods with increased financial activity, which could be seasonal or event-driven. Analyzing these patterns can enhance financial planning and anomaly detection.

* **Monthly Transaction Amounts by Year**

The line plot of monthly transaction amounts over several years highlights a significant spike in May 2014, suggesting an unusual event or high transaction volume. Apart from this anomaly, transaction amounts remain relatively stable across other years, indicating consistent business activities. This long-term stability can inform strategic planning and resource allocation.

* **Decomposition of Time Series**

The time series decomposition into observed, trend, seasonal, and residual components reveal an increasing trend in transaction amounts, particularly towards the end of the series. Consistent seasonal patterns and significant residuals indicate regular periodic fluctuations and occasional anomalies. Understanding these components can improve forecasting and anomaly detection.

* **Clustering Results**

The clustering analysis categorizes transactions into five clusters, each with distinct transaction patterns:

* Cluster 0: Routine low-value transactions with minimal variability.
* Cluster 1: High-value transactions concentrated in May, indicating significant events or seasonal spikes.
* Cluster 2: Mixed transactions with moderate values and occasional high amounts.
* Cluster 3: Routine low-value transactions with some high-value outliers.
* Cluster 4: Similar to Cluster 0, with low-value transactions.
* **Yearly Transaction Amounts**

The line graph showing total transaction amounts from 2013 to 2024 reveals significant fluctuations. Key observations include:

* Growth Periods: 2013 to 2017 shows steady growth peaking in 2017, and 2021 to 2024 shows another notable increase, peaking in 2024.
* Decline Periods: A noticeable decline occurs from 2017 to 2020, and a slight dip in 2022 before rising again in 2024.
* Spikes: 2017 and 2024 are peak years, suggesting significant business activities during these periods.

**CONCLUSION**

The analysis reveals significant seasonal trends, high-value transactions, and consistent low-value transactions across different clusters. Notable findings include significant peaks in transaction amounts in May, particularly in 2014, and steady growth in transaction volumes from 2021 to 2024. The clustering results highlight routine low-value transactions and high-value transactions in specific periods.

* **Hypothesis Validation**

A t-test conducted to compare the mean transaction amounts for 2024 with those of previous years provided a t-statistic of 2.3668 and a p-value of 0.0179, leading to the rejection of the null hypothesis. This indicates that transaction amounts in 2024 are significantly higher than in previous years.

* **Recommendations**

1. Investigate Peaks: Analyze factors contributing to high transaction amounts in 2017 and 2024 to leverage these peaks.
2. Understand Decline: Examine reasons behind the decline from 2017 to 2020 to identify areas for improvement.
3. Sustain Growth: Develop strategies to sustain the recent growth trend, especially the substantial increase in 2024.

Overall, this report provides a statistical foundation for understanding transaction patterns, detecting anomalies, and informing future business strategies.