*SYED MUHAMMAD IRTIZA ZAIDI - 23973*

*[Company Name] | [Company Address]*

*REport on Transaction data*

Table of Contents

[Report on Bank Transaction Data Analysis 2](#_Toc137636577)

[1. Introduction 2](#_Toc137636578)

[RFM 2](#_Toc137636579)

[2. Data Cleaning and Transformation 3](#_Toc137636580)

[3. RFM Score Calculation 4](#_Toc137636581)

[4.Statistical Analysis 4](#_Toc137636582)

[5. Correlation Analysis 4](#_Toc137636583)

[6. RFM Visualization 5](#_Toc137636584)

[7.Python Libraries Used: 6](#_Toc137636585)

[8. Conclusion 6](#_Toc137636586)

# Report on Bank Transaction Data Analysis

## 1. Introduction

This report provides a detailed analysis of bank transaction data using a code snippet. The code processes a dummy dataset, originally containing 116,201 rows and 9 columns, which was obtained from the following source: Bank Transaction Data. The dataset includes information such as :

* account numbers
* transaction dates
* cheque numbers
* value dates
* withdrawal amounts
* deposit amounts
* and balance amounts.

The code performs various data processing tasks and calculates the RFM (Recency, Frequency, Monetary) scores for each transaction. The RFM scores are then analyzed and visualized to gain insights into customer transaction behavior.

### RFM

RFM stands for Recency, Frequency, and Monetary Value. It is a widely used framework for analyzing and segmenting customers based on their transactional behavior. The RFM model assesses three key aspects of customer behavior:

* Recency (R): Recency measures the time elapsed since the customer's last transaction. It indicates how recently a customer has interacted with the business. Customers who have made a purchase or transaction more recently are generally considered more engaged or active.
* Frequency (F): Frequency refers to the number of transactions or interactions made by a customer within a specific period. It quantifies how often a customer engages with the business. Customers with higher transaction frequency are typically more loyal and valuable.
* Monetary Value (M): Monetary Value represents the total amount of money spent by a customer over a certain period. It indicates the value or profitability of a customer. Customers who have spent more money are generally considered more valuable to the business.

By analyzing and combining these three factors (R, F, and M), the RFM model provides a framework for segmenting customers into distinct groups based on their transactional behavior. This segmentation allows businesses to identify different customer segments, understand their characteristics, and tailor marketing strategies accordingly.

To calculate RFM scores, each factor (Recency, Frequency, and Monetary) is divided into quantiles or score ranges. The customers' actual values for each factor are then mapped to the corresponding score ranges. Typically, higher scores indicate more desirable behavior. The RFM scores can be combined using weighted averages or other methods to obtain a composite score that represents the overall value or engagement level of a customer.

The RFM Score was used to categorize them into four groups:

* Best Value: Customers with the highest RFM Scores, indicating recent, frequent, and high-value transactions.
* High Value: Customers with relatively high RFM Scores, demonstrating good engagement and value.
* Mid Value: Customers with moderate RFM Scores, reflecting average engagement and value.
* Low Value: Customers with lower RFM Scores, indicating less recent, infrequent, or lower-value transactions.

These segments provide insights into customer behavior, preferences, and potential opportunities for targeted marketing campaigns, personalized offerings, and customer retention strategies.

Overall, RFM analysis helps businesses understand and segment their customer base, identify high-value customers, optimize marketing efforts, and improve customer satisfaction and loyalty.

## 2. Data Cleaning and Transformation

The code begins by reading the bank transaction data from an Excel file and performs necessary data cleaning operations. These operations include :

* filling null values with zeros
* removing duplicates
* correcting formats and types of certain columns
* and dropping unnecessary columns.

The resulting dataset is transformed into a more suitable format for RFM analysis.

## 3. RFM Score Calculation

The code calculates the RFM scores for each transaction based on the Recency, Frequency,

and Monetary factors. The RFM scores are calculated using different weights for each factor. The Recency weight, Frequency weight, and Monetary weight are set to 2/10, 3/10, and 5/10, respectively. The RFM scores are then normalized by dividing them by 2 to obtain a final RFM score ranging from 0 to 5. Additionally, the code assigns RFM ranks to each transaction based on specific score ranges.

## 4.Statistical Analysis

The code provides statistical analysis of the RFM scores by calculating descriptive statistics, such as mean, count, minimum, and maximum values, for the Recency, Frequency, and Monetary factors. These statistics offer insights into the distribution and central tendencies of the RFM scores.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Recency | Frequency | Monetary |
| COUNT | 10 | 10 | 10 |
| mean | 540.556534 | 11616.2 | 47856.98376 |
| std | 283.117278 | 15882.18 | 72358.91389 |
| min | 142.640672 | 51 | 191.036906 |
| 25% | 369.963542 | 1056.25 | 321.328303 |
| 50% | 465.456827 | 5301 | 5030.276072 |
| 75% | 732.587868 | 12720 | 78055.42467 |
| max | 1041.028523 | 48758 | 203656.053 |

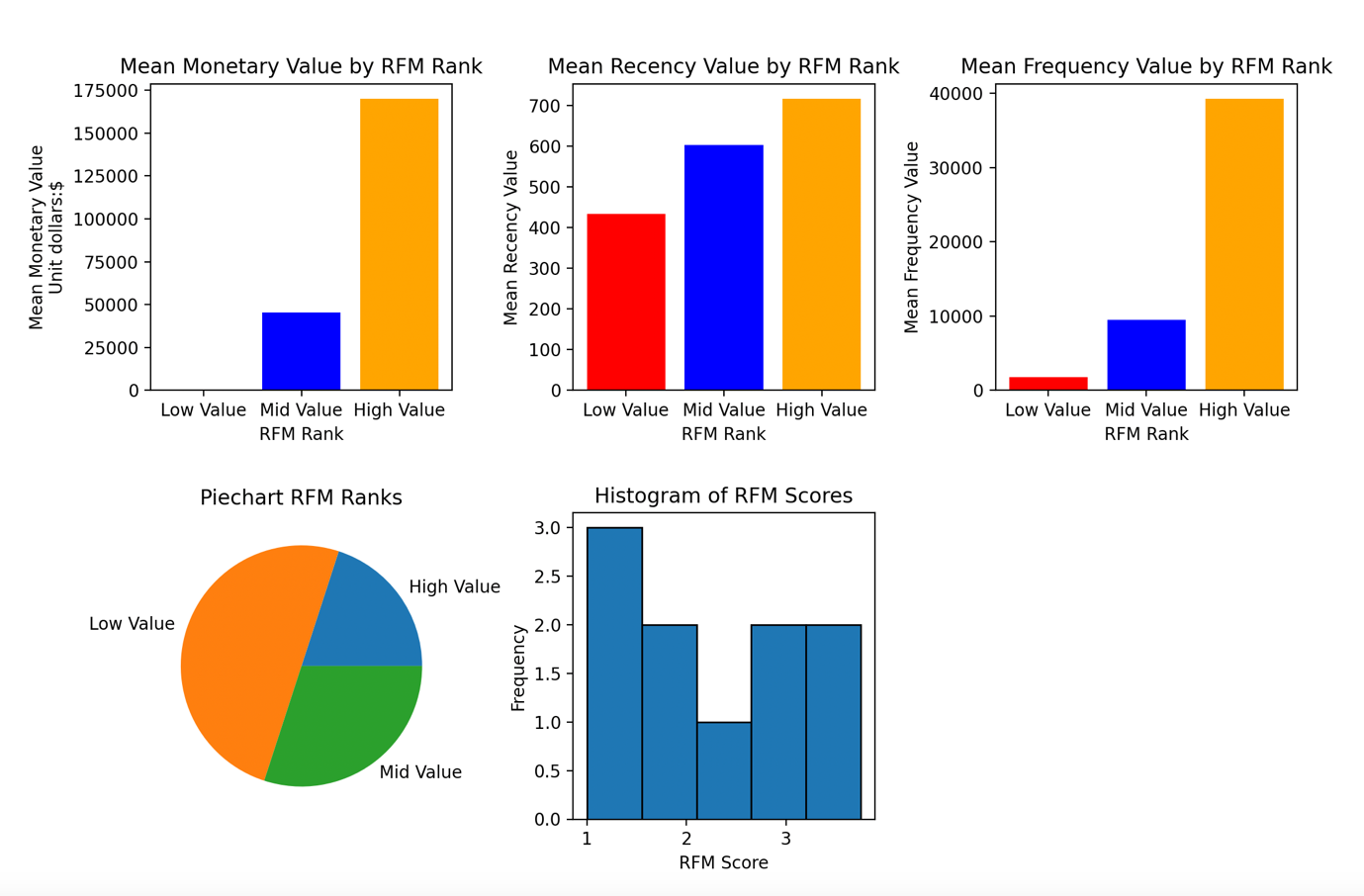
## 5. Correlation Analysis

The code calculates the correlation matrix for the Recency, Frequency, and Monetary factors. The correlation matrix indicates the degree of linear relationship between these factors, providing insights into any potential patterns or associations between them.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Recency | Frequency | Monetary |
| Recency | 1 | 0.325635 | 0.552396 |
| Frequency | 0.325635 | 1 | 0.838032 |
| Monetary | 0.552396 | 0.838032 | 1 |

## 6. RFM Visualization

The code includes visualizations to aid in understanding the RFM scores and their relationship with other variables. Initially, the code plots a bar graph showing the mean monetary value for each RFM rank. This visualization helps identify variations in monetary value based on the RFM ranks. Additionally, a pie chart is created to represent the distribution of RFM ranks in the dataset, providing an overview of the customer segments. Furthermore, a histogram is plotted to visualize the distribution of RFM scores and their frequency in the dataset.



## 7.Python Libraries Used:

This analysis utilized several Python libraries to manipulate, analyze, and visualize the bank transaction data. The following libraries were employed:

1. pandas: pandas is a powerful library for data manipulation and analysis. It provides data structures and functions for efficiently handling structured data, such as data frames. In this analysis, pandas was used for reading and cleaning the transaction data, performing group-wise operations, and merging data frames.

2. datetime: The datetime module is a built-in Python library that supplies classes for manipulating dates and times. It was used in this analysis to calculate the recency of transactions by finding the difference between the transaction date and the reference date.

3. numpy: numpy is a fundamental library for scientific computing in Python. It provides support for large, multi-dimensional arrays and mathematical functions to operate on these arrays efficiently. In this analysis, numpy was used to perform numerical computations and transformations, such as calculating the RFM scores and selecting conditions for segmentation.

4. matplotlib: matplotlib is a widely-used library for creating visualizations in Python. It provides a flexible and comprehensive set of functions for creating various types of plots and charts. In this analysis, matplotlib was utilized to generate bar graphs, pie charts, histograms, and other visualizations to visualize the RFM analysis results.

These libraries greatly facilitated the data processing, analysis, and visualization tasks in this analysis, allowing for efficient and effective exploration of the bank transaction data and extraction of valuable insights for decision-making purposes.

## 8. Conclusion

In conclusion, the code provides a comprehensive analysis of bank transaction data by calculating RFM scores and visualizing the results. The RFM scores offer valuable insights into customer transaction behavior, enabling businesses to identify high-value customers, segment their customer base, and design targeted marketing strategies. The code also performs statistical and correlation analysis, providing additional understanding of the dataset. The visualizations enhance the interpretation of RFM scores and facilitate the communication of key findings. The analysis presented in this report serves as a foundation for further exploration and decision-making based on the bank transaction data.