

Title: Exploratory Data Analysis

Introduction

In this assignment we will define some questions and find interesting patterns answering for these questions. I will describe with plots and explanation of this and at the end sum up and conclude it. As a major dataset there was taken a finance_11.csv file.

Methods

Only methods which we will use is just generating questions and trying to answer them with plots and bars.

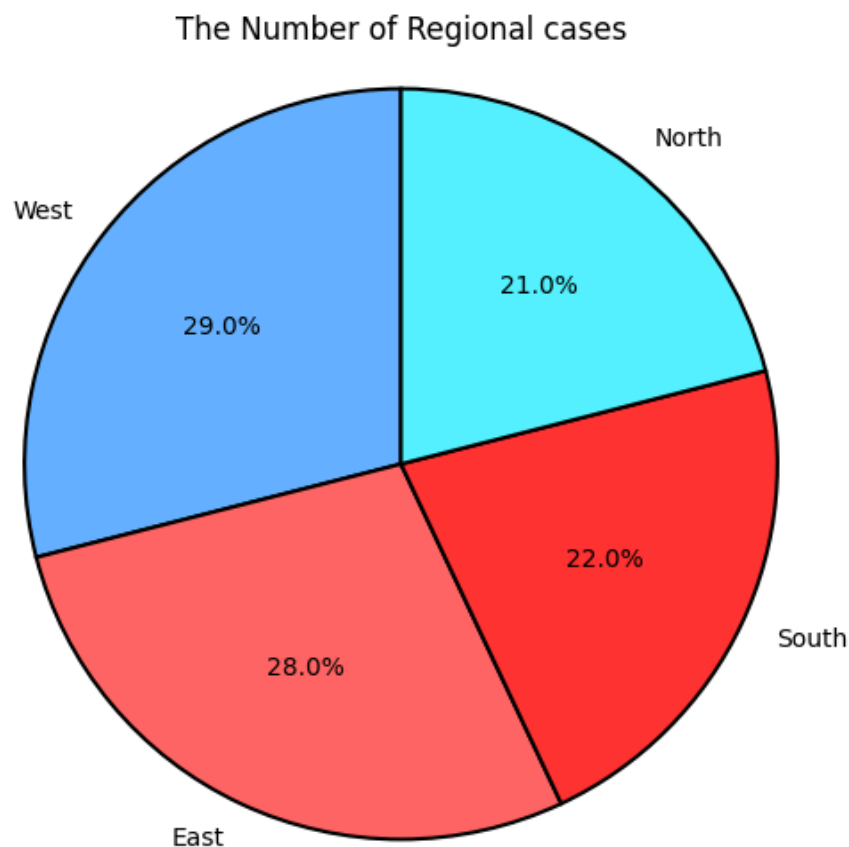
Question 1: Analyze the total and average transaction amounts across different regions?

Question 2: Investigate how different transaction categories affect the account balance?

Question 3: Determine the proportion of transactions for each category across regions?

Question 4: Identify transactions with amounts significantly higher than the average. Are these high-value transactions concentrated in specific categories or regions?

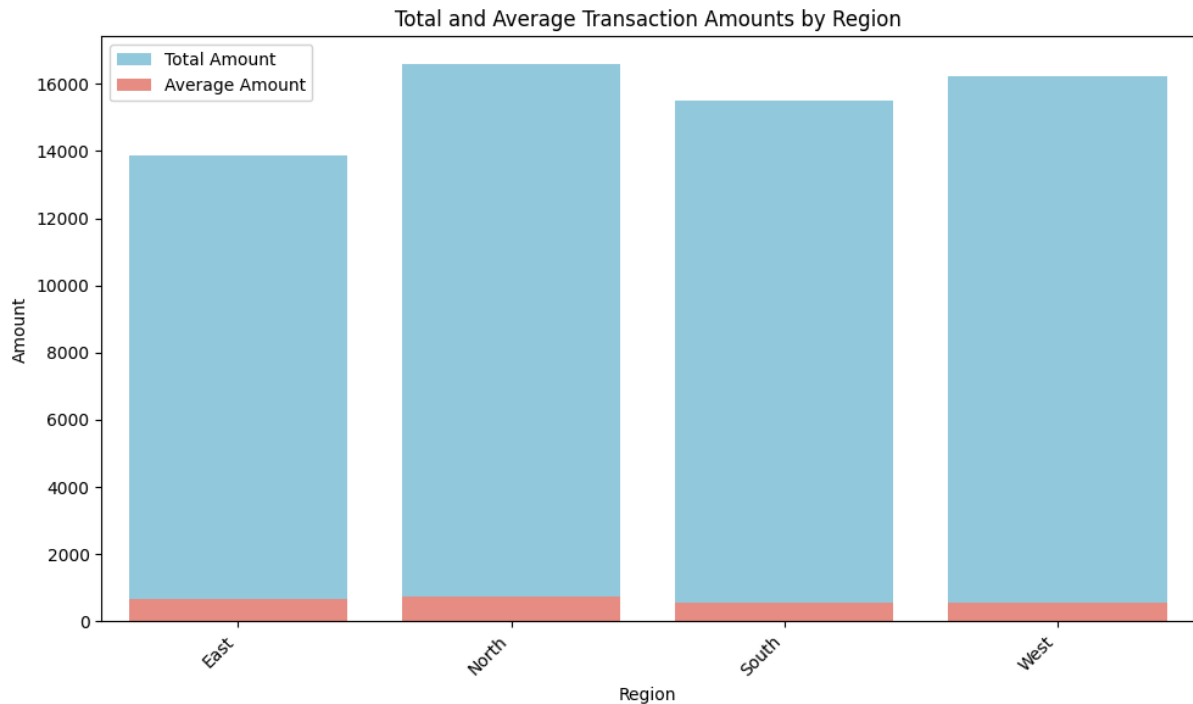
Findings



Firstly, I defined the proportion of transactions by region. And the most transactions were made on the West side with almost 30% when the minimum was on the North side with a bit more than 20%.

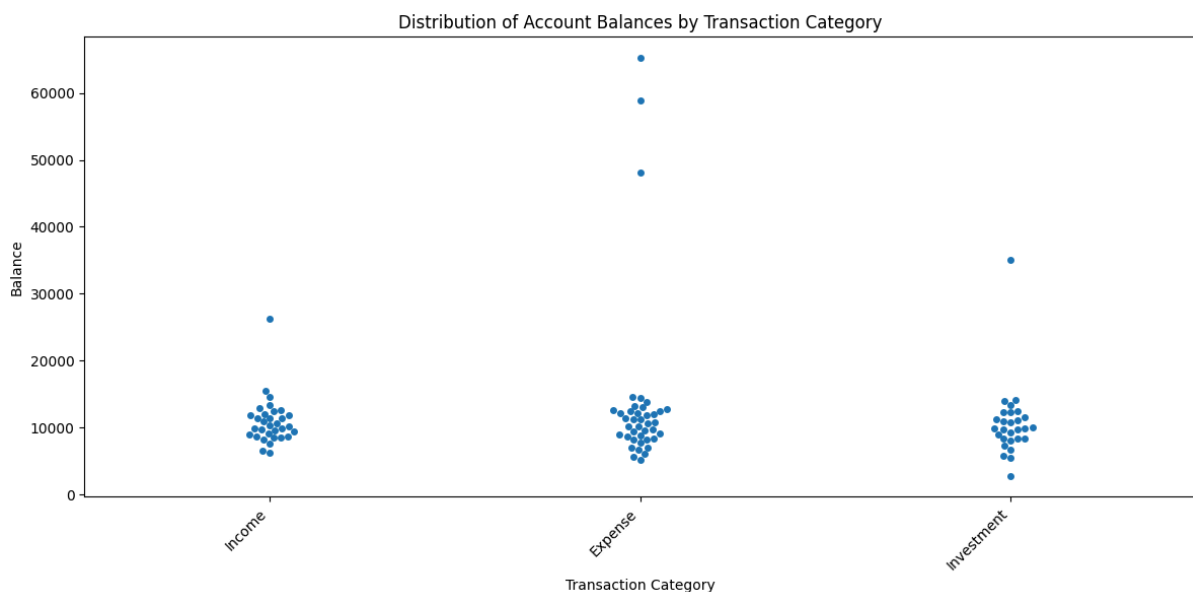
Answer 1

Now, answering for the first question, let's draw the bar chart with total and avg bars



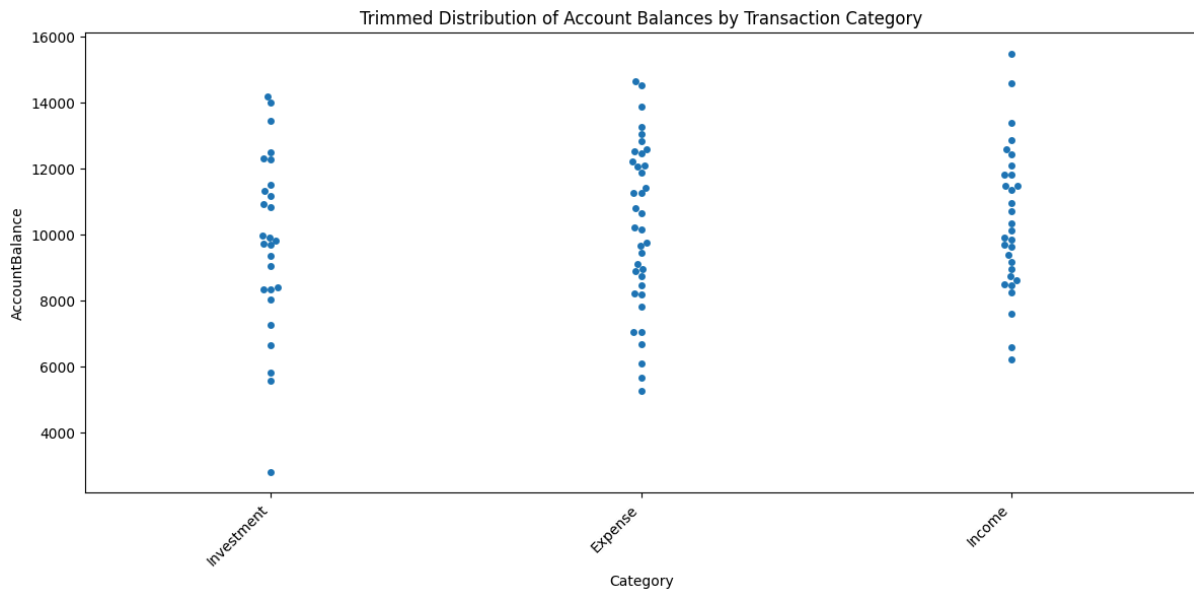
So, looking at this barplot it is clear that even the most transactions made on the West side but the biggest total amount on the North side, considering that on this exact region the least transactions were made. Naturally the most avg amount in the North region. Furthermore, even the East side was one of tops with the number of transactions the amount of it the least one compared with others.

Answer 2



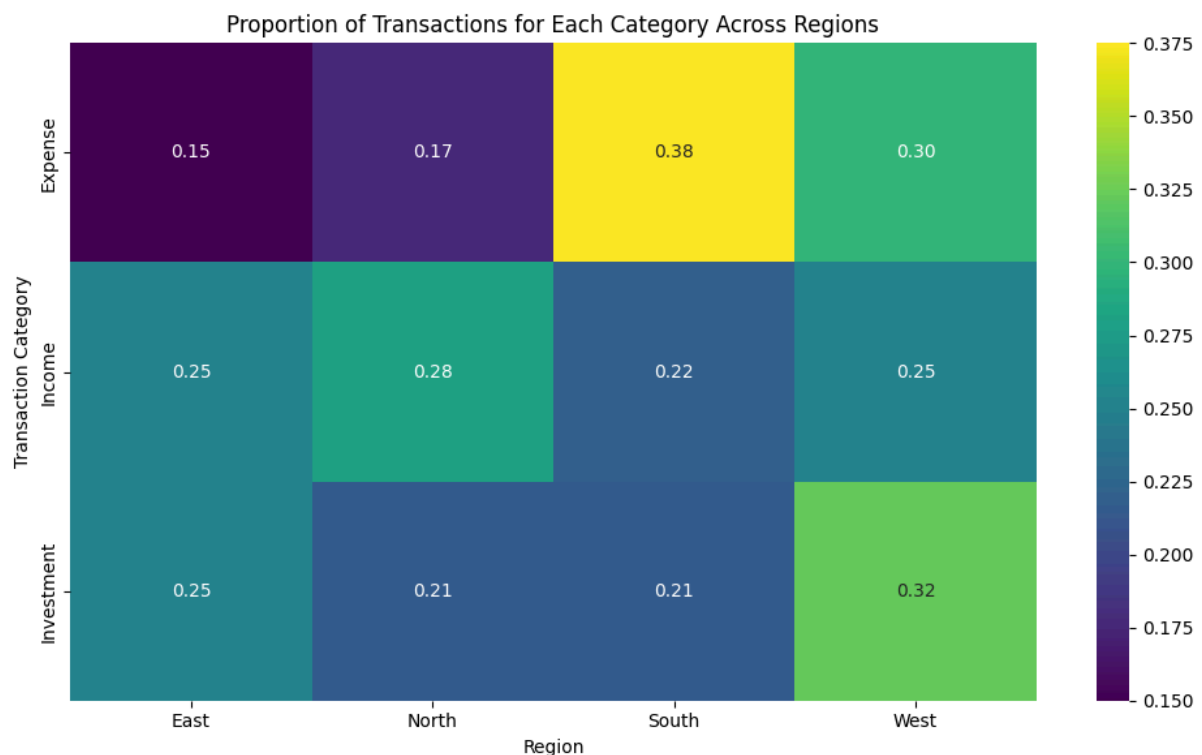
Here is a swarm plot showing distribution of account balance by transaction category. We can say that each category has at least one outlier. From them the

most one and the fewest outliers in the Expense category. So, by trimming the df we get this kind of distribution



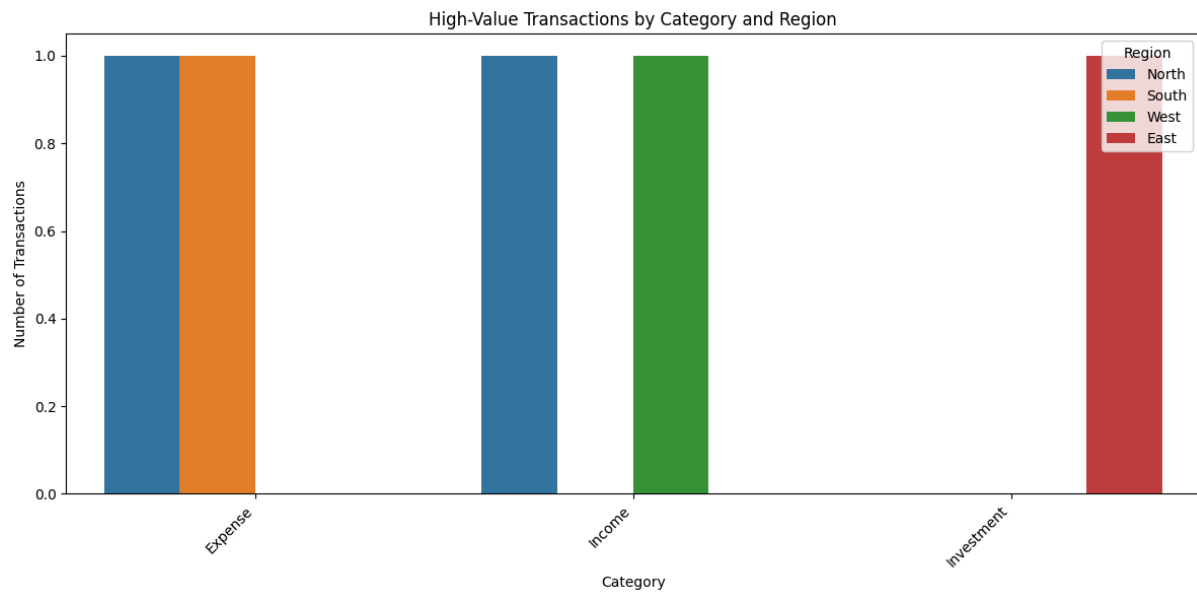
There are no interesting patterns, data distributed separately. And answering the question, the category almost did not affect the account total.

Answer 3



This visualization provides a comprehensive overview of the proportion of transactions for each category across regions. And it is clear that the South region shows the highest proportion of Expense transactions (38%). Furthermore the East and North have the lowest proportions of Expense transactions with 15% and 17%, respectively.

Answer 4



The last question bar chart illustrates the distribution of high-value transactions across different categories and regions. The South stands out for high-value Expense transactions, aligning with its spending dominance seen earlier. The West has significant high-value Income transactions, supporting its investment focus. Furthermore, the East is prominent in high-value Investment transactions, reflecting a balanced but strategic approach.

Conclusion

Concluding the findings and answers it is clear that without additional plot we can summarize incorrectly and our findings will be not clear. So, with numbers there right to see the percentages also and overall views.