BLACKWELL ELECTRONICS REPORT

Customer purchasing behavior

1. Overview of data

Blackwell Electronics demographic data was analyzed using data mining methods to explore customer purchasing behavior. The given dataset didn't have missing values. The dataset contains 5 variables (attributes):

- "in-store" indicates whether the transaction made place in store (1) or online (0)
- "age" indicates the age of the customer
- "items" tracks the number of items the customer purchased
- "amount" indicates the amount of money spent per the transaction
- "region" indicates in which of the four regions the purchase was made (1 – East; 2 – West; 3 – South; 4 - Central)

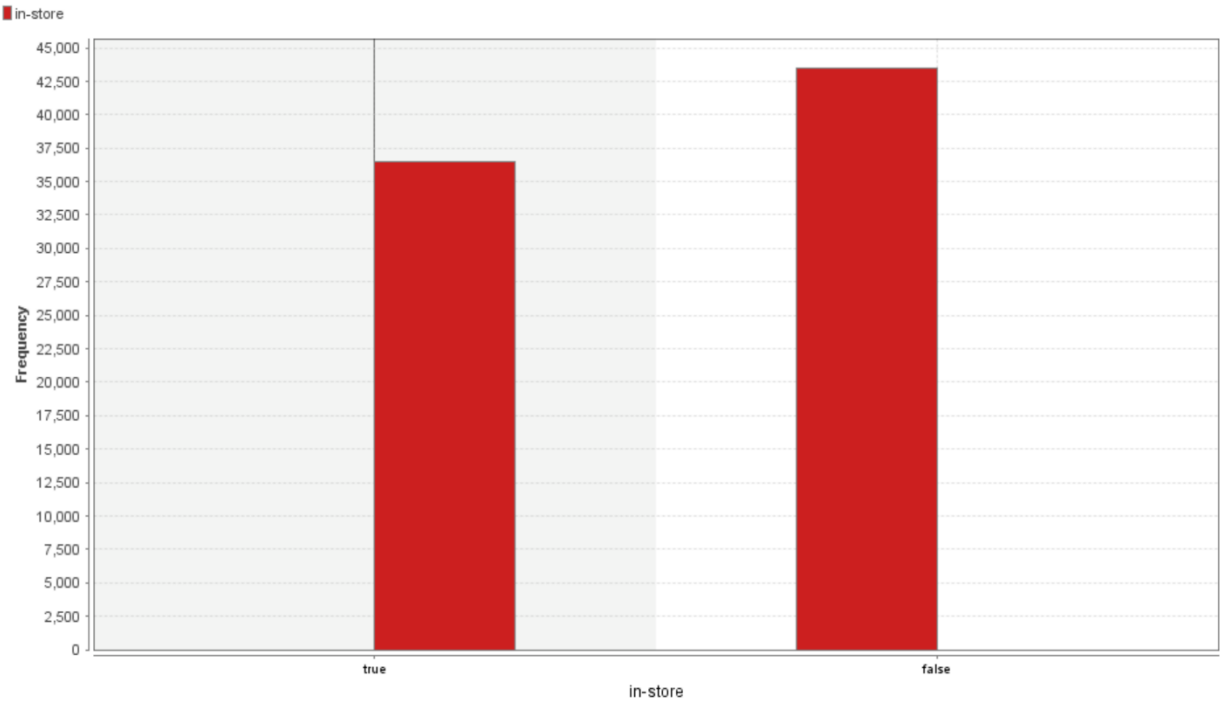

Figure 1.1: In store attribute distribution

In store attribute shows (Figure 1.1) that the data is equally distributed, meaning half of all recorded transactions were made in store and the other half online, with slight advantage of online purchases.

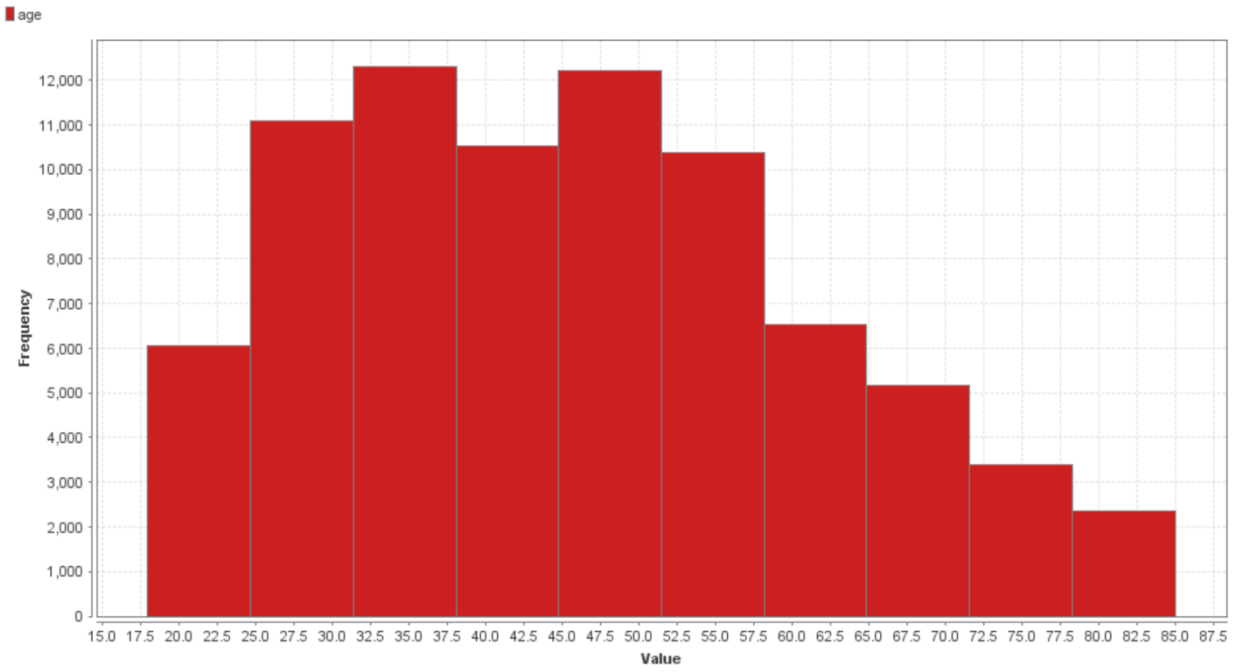

Figure 1.2: Age attribute distribution

On the first glance the data shows that the recorded age range of customers is from 18 to 85 with the average of 45.7, but most customers fall in range between 25 and 55 (Figure 1.2).

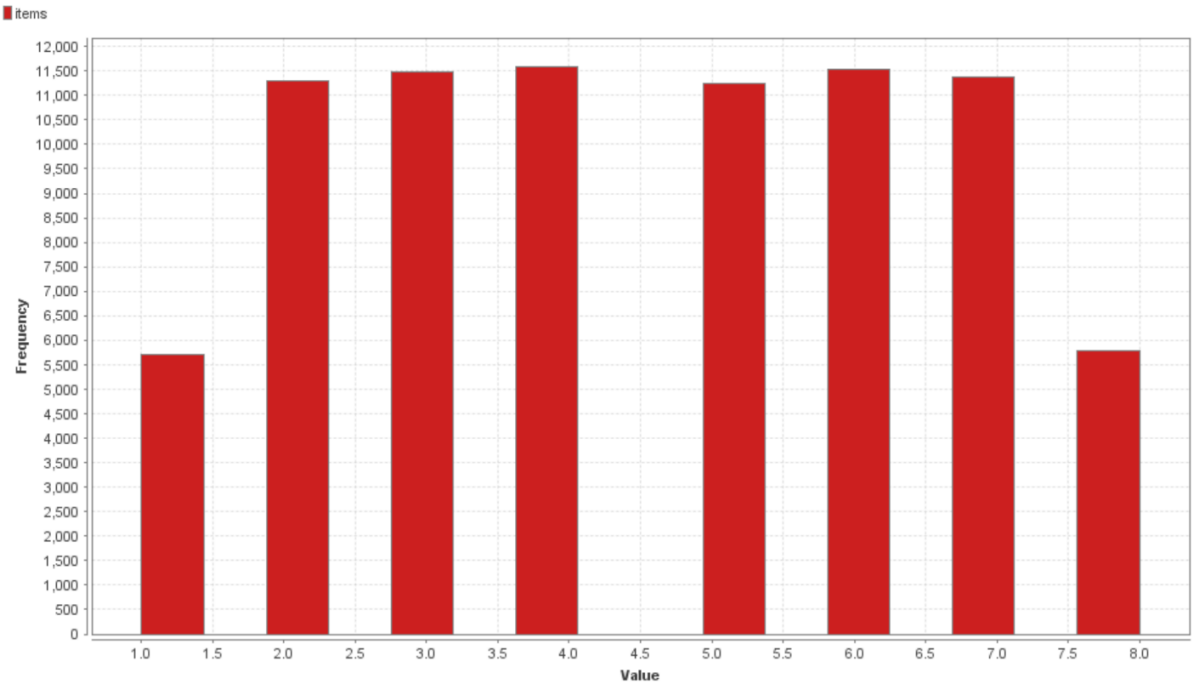

Figure 1.3: Items attribute distribution

Number of items they purchased ranges from 1 to 8 items per transaction with average of 4.5 (Figure 1.3).

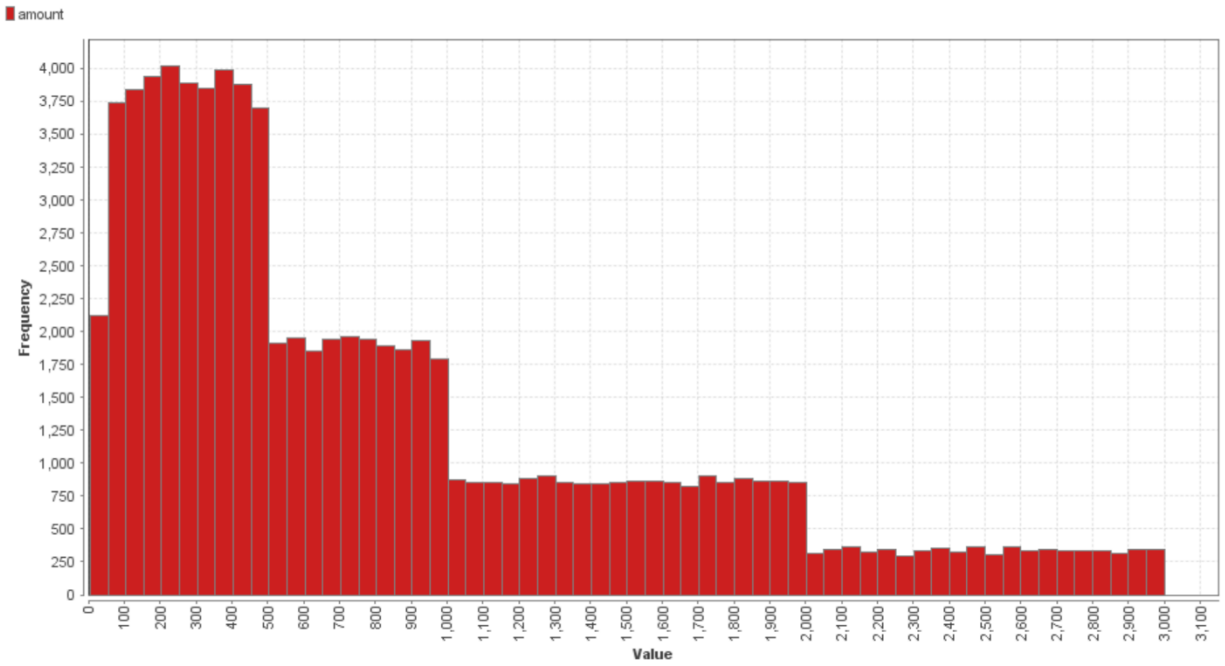

Figure 1.4: Amount attribute distribution

The amount spent per transaction ranges from 5 to 3000, with the average of 835. Almost half of all transactions had the amount between 5 and 500 (Figure 1.4).

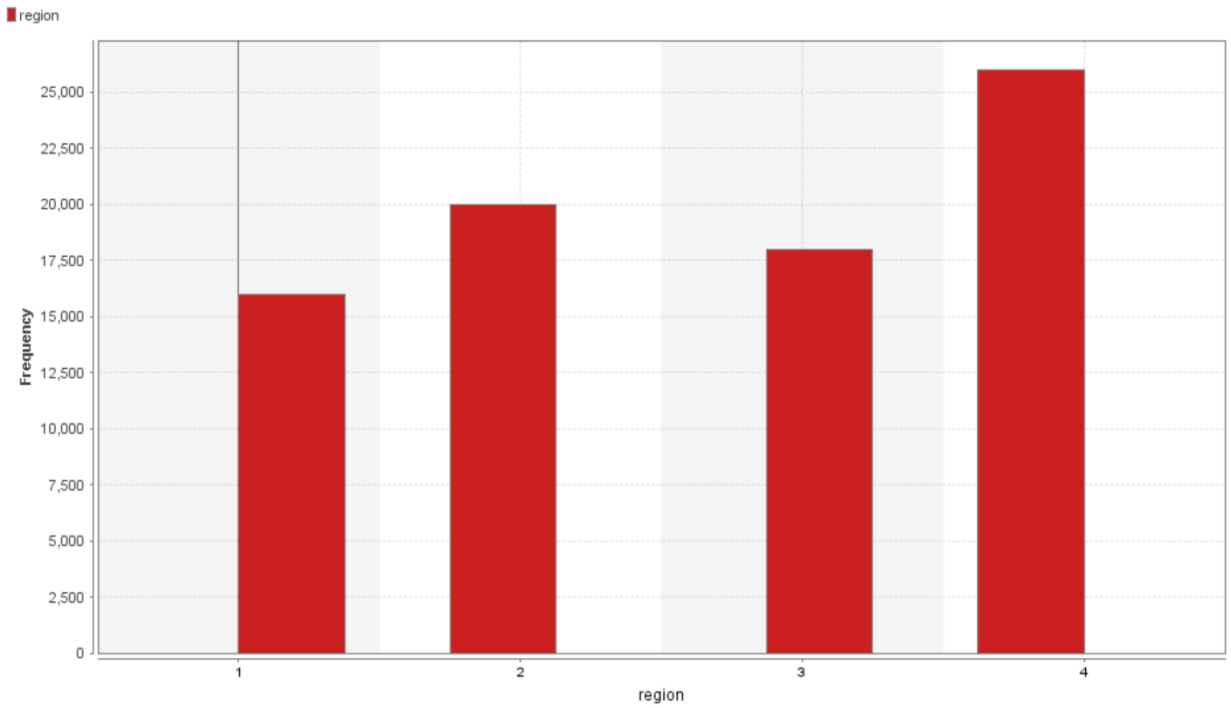

Figure 1.5: Region attribute distribution

The region attribute shows that the recorded transactions are mostly equally distributed by regions, with a slight advantage of region 4 (central region).

2. Region and amount connection

Analysis of the dataset and the connection between the region and the amount spent per transaction shows that customers in first and second region (east and west) spend less than customers in other two regions (south and central). Customers in east region (first) spend up to 2000 per one transaction, while customers in second region spend the least amount of money per transaction. Data shows that they spend up to 600 per transaction. Customers in south and central region (3. and 4.) spend the most per transaction.

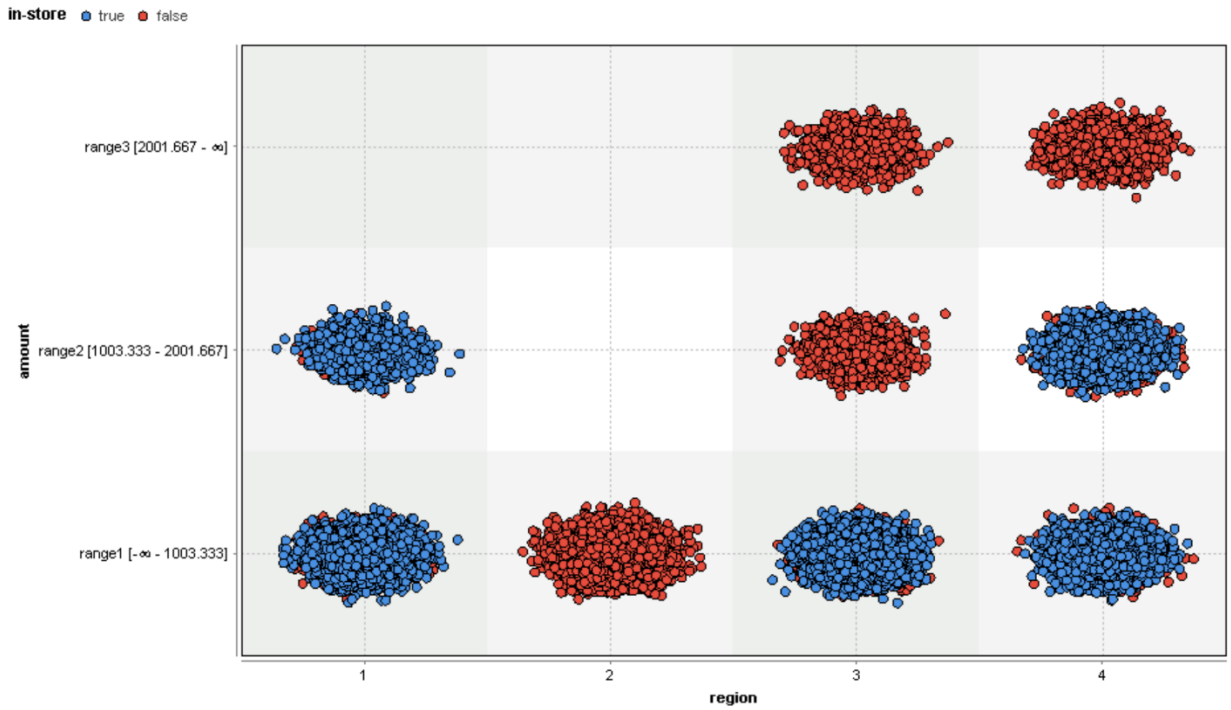

Figure 2.1: Region and amount scatterplot with segmentation

There are also few other observed differences between regions. Data from the east region (first) shows that purchases were dominantly made in store. West region (second), besides having the lowest amount per transaction, has only online transactions recorded. Third region (south), besides being one of two regions with highest amounts per transaction, has no in store purchases recorded for amounts above 1100. Amounts below that amount are dominantly from in store purchases. The picture changes with larger amounts. For transactions ranging from 1100 to 3000 there are only online purchases observed. Central region (fourth) is also one of the regions with highest amount per transaction. Besides that, central region has mostly in store purchases for the amounts ranging up to 2000. Everything above that threshold is purchase made online.

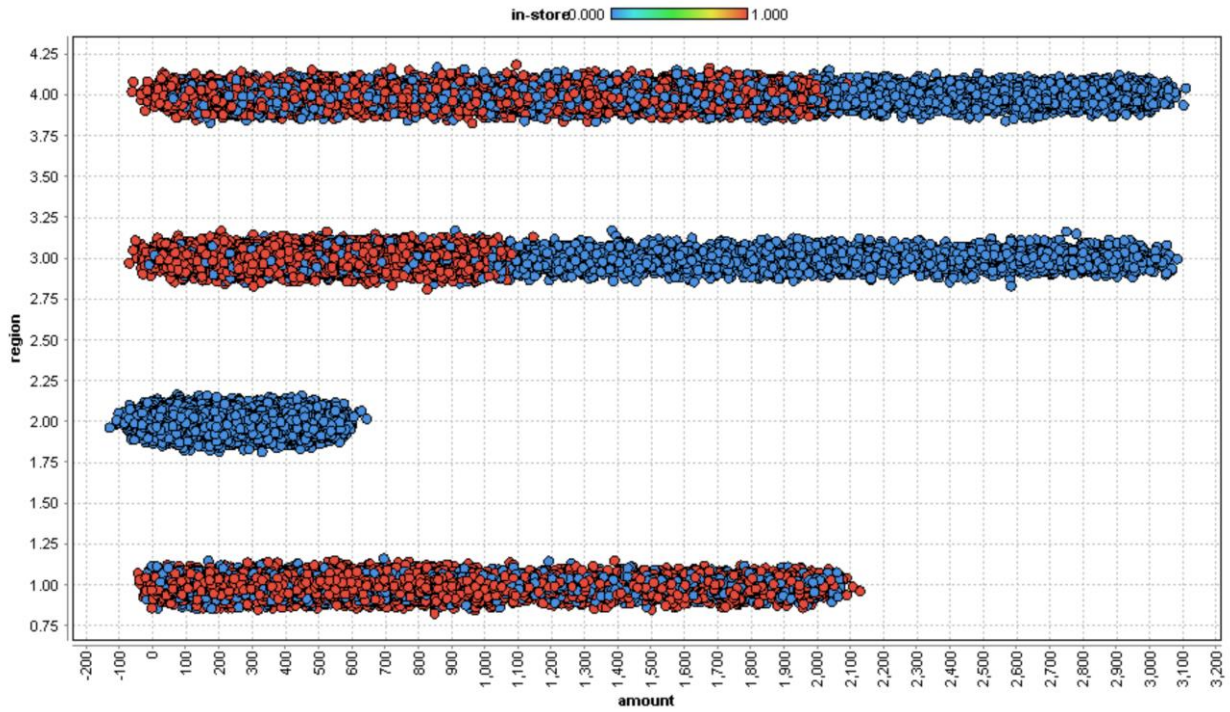

Figure 2.2: Amount and region scatterplot without segmentation

3. Age and region connection

Blackwell dataset exploration gave few insights regarding connection of customers age and region of purchase.

First region (east) has equal distribution of customers ranging from 18 to 68 with a slight advantage of age group of customers between age 34 to 51. Age group above 65 is visibly less present among Blackwell customers in this region. Also, there are no customers older than 75 in this region.

Second region (west) has similar distribution of customers regarding their age. Opposite from first region, they have predominantly customers from age groups from 34 to 85, with less customers from 25 to 34. Also, it should be noted that this region doesn't have customers younger than 25.

Third region (south) has similar distribution of customers as the first region. Customers ranging from 18 to 68 are almost equally represented in this region. There are clearly less customers older than 68, and there are no customers older than 75.

Age groups from 18 to 51 are dominantly represented in fourth region (central). Age group from 51 to 65 is present but not dominant. This region doesn't have customers older than 65.

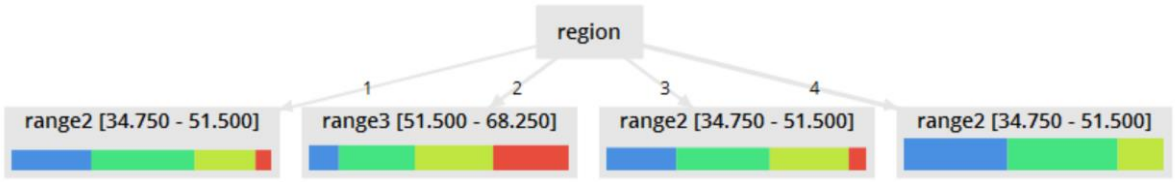

Figure 3.1: Decision tree on age attribute

From the given data, we can see that region is mostly connected with in store purchases and the amount spent per transaction, and a bit less with age. We can´t properly predict the customers age in region based on other data.

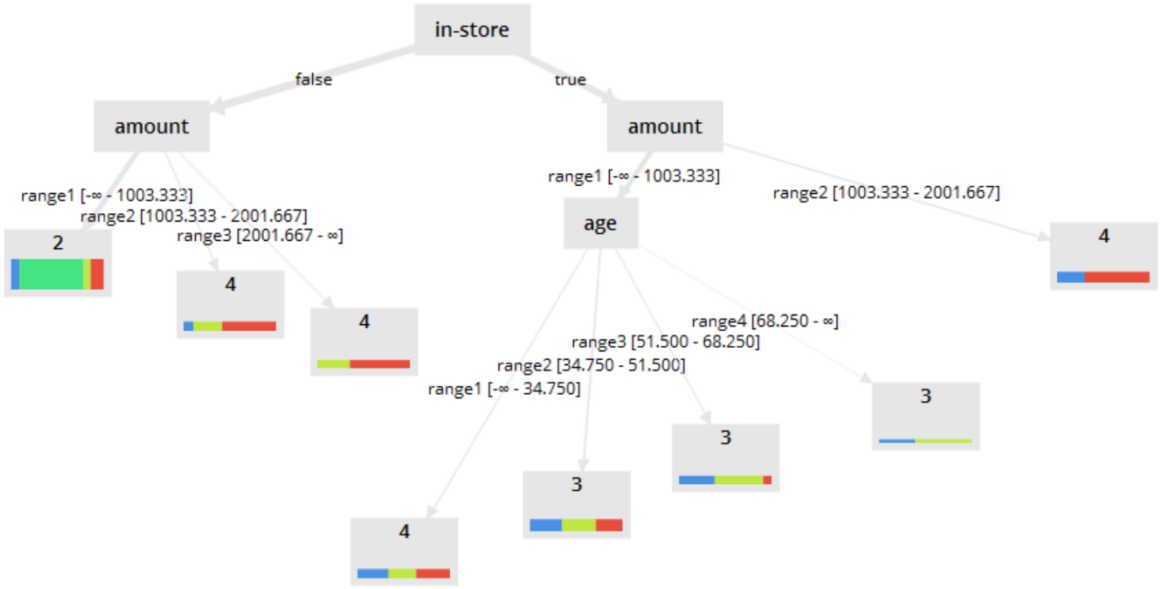

Figure 3.2: Decision tree of region attribute

4. Location of a transaction

Looking at Blackwell data we can see that with the age rising, the online purchases rise. Younger people have no preference between online and in store purchases. The same goes for the age group from 34 to 51. The difference in buying habits starts to be seen in the age range from 51 to 68. This age group shows tendency of online purchases. The biggest difference in buying habits are seen in the age group from 68 to 85. The data shows that online purchases double in store purchases for this age group.

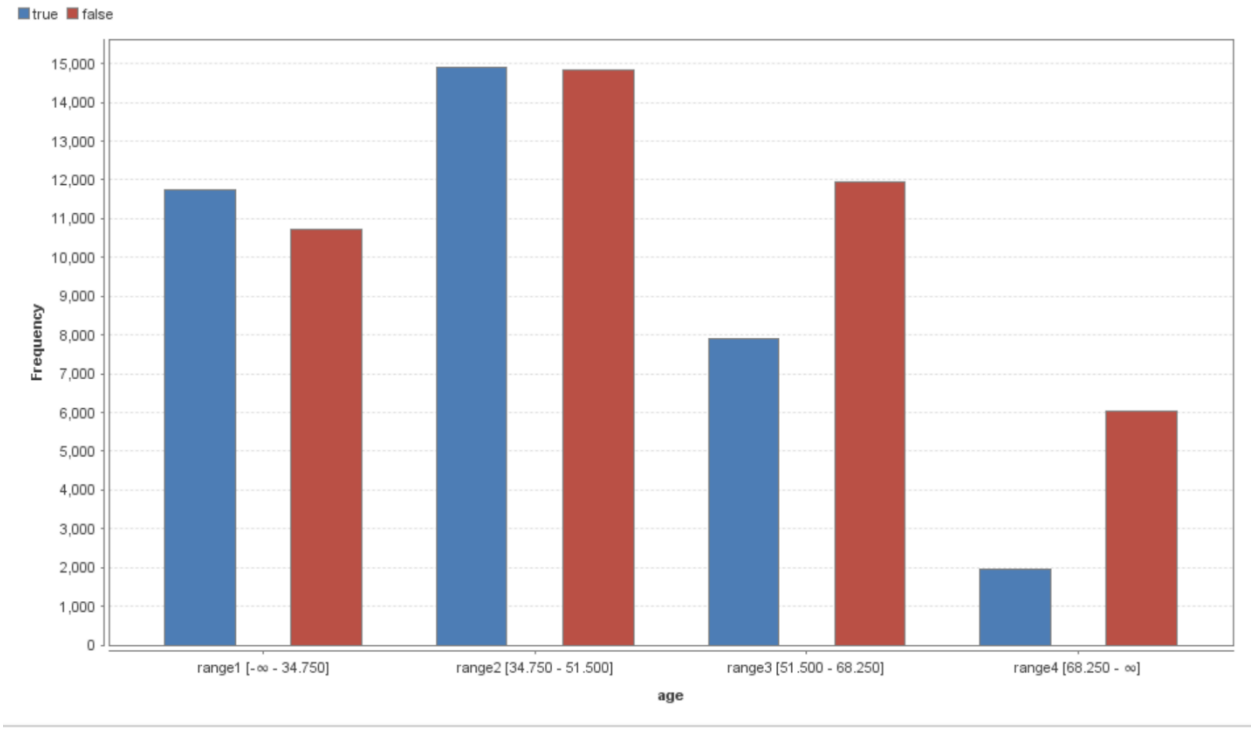

Figure 4.1: Age and in store attribute correlation

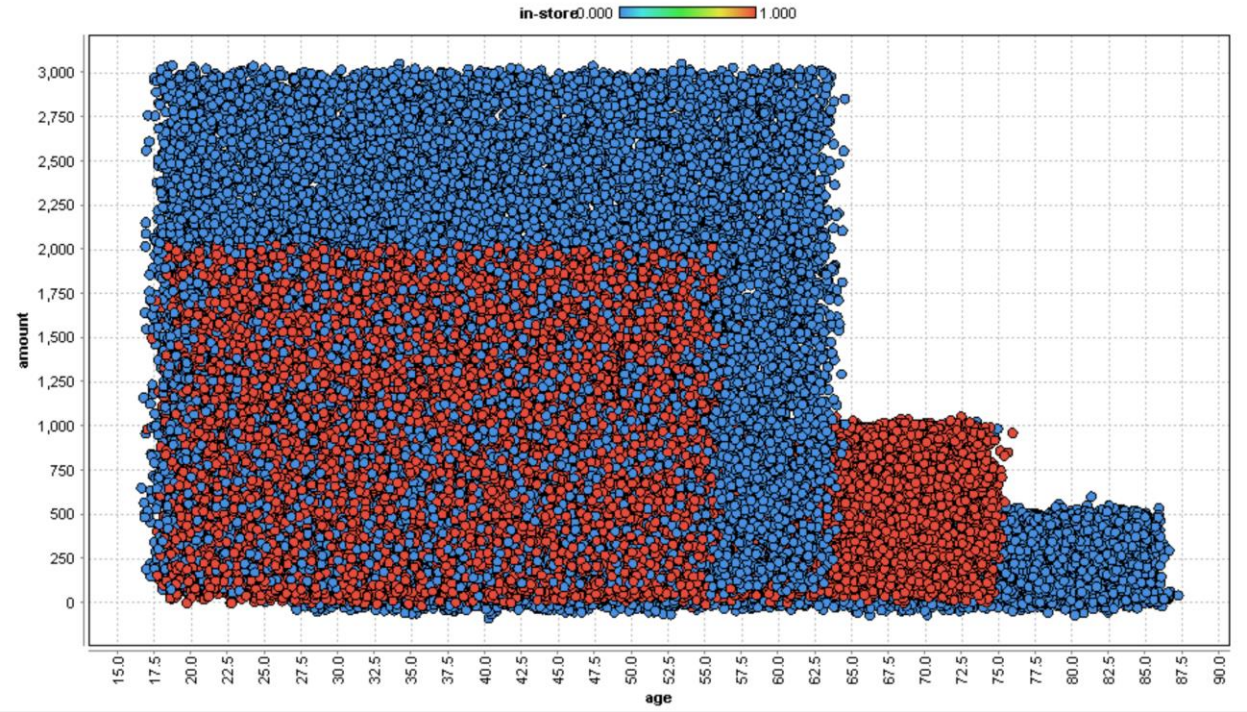

Figure 4.2. Age and amount correlation with in store data

Data also shows the connection between the location of purchase and the amount of the transaction (Figure 4.2). As it can be seen, online purchases have the higher amount per transaction. In store purchases only go to 2000, while online purchases go to 3000 per transaction.

Age groups from 18 to 62 spend more online than in a store. Age group from 55 to 62 buys mostly online, unlike the population from 18 to 55 who buys almost equally online and in the store with only difference in the amount spent; they spend less in a store (<=2000) and more online (<=3000). Age group from 62 to 75 spends less (up to a 1000) and mostly in store, while the age group from 75 to 85 spends the least amount of all groups but they do their purchases only online.

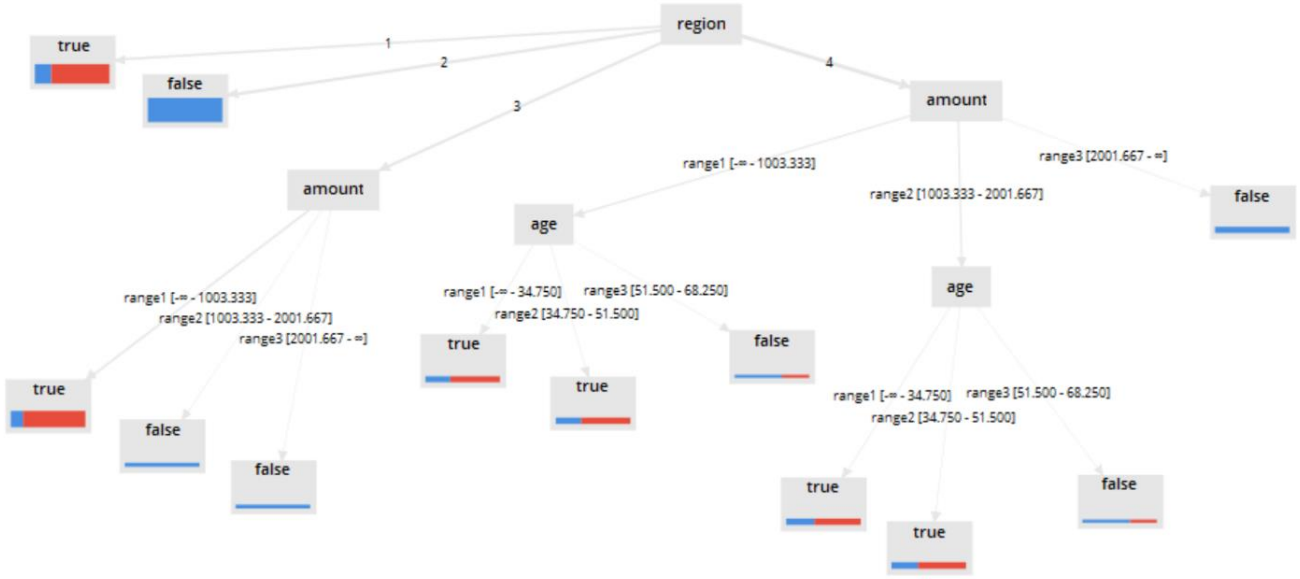

Figure 4.3: Decision tree for in store attribute

The data showed that the location of the transaction is mostly connected with the region and the amount of the transaction, and after that with age. We can see (Figure 4.3) that in first region there are dominantly in store purchases unlike in the second region, where we only have online transactions. The location of the transaction for the third region depends on the amount spent. For the transaction up to 1000 the purchases were made in store, but for the more expensive transactions purchases were solely made online. Transaction location for the fourth region depends on the amount. For the amounts higher than 2000 the transaction was made online, but for the lower amounts it depended on age. Age group up to 51 mostly made purchases in the store while the age group from 51 to 68 mostly purchased products online.

5. Items and amount connection

The Blackwell customer data shows no connection between the number of purchased items and the amount spent. Analysis showed that the amount attribute has no direct connection with other recorded attributes. The data shows that customers who purchased more than average number of items (>4.5) didn't spend more in a transaction than customers who purchased less than average number of items (<4.5).

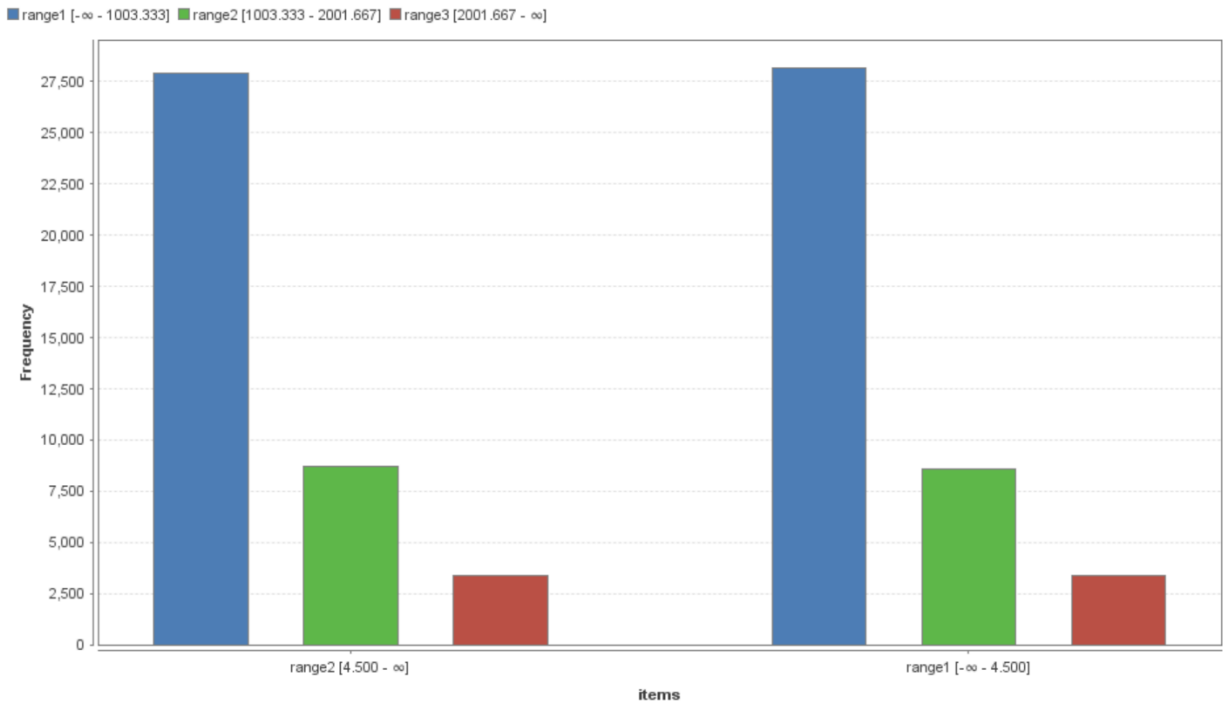

Figure 5.1: Items and amount attribute correlation