[BTM539] Project 1: RFM Analysis with UCI online retail data 20223484 이연수

Recency, frequency, and monetary value are the three key concepts in RFM analysis. Recency refers to how long has it been since the customer’s last purchase. It captures the elapsed time since the last purchase and can be measured in days, weeks, months, or years. In the analysis, I defined recency as the number of days between the customer’s last purchase and January 1st, 2012. I used the first day of 2012 as the reference point since the data consists of purchase records during 2009-2011. By taking the first day of 2012 as the reference point, I can safely calculate recency without having negative values. I measured recency in days to make the analysis precise.

Frequency captures how frequently the customer has made a purchase. According to Robert, Byung, & Scott (2008)[[1]](#footnote-1), frequency can be measured as “the number of purchase occasions since the first purchase” or as “the number of purchase occasions divided by the duration of being the customer. Since the data does not include information on the duration of being the customer, I use the first measurement method provided in the book. Specifically, I defined frequency as the number of invoices (purchases) during 2009-2011, i.e., during the whole period when data was collected.

Monetary value refers to how much money the customer previously spent. There are many ways to measure monetary value – the total amount of money spent during a certain period, the total amount spent divided by the duration of being a customer, or the average spending per order. In this analysis, I define monetary value as the average monetary spending per invoice (purchase) during 2009-2011, i.e., during the whole period when data was collected. Similar to frequency, I cannot use the second measure since the data does not include information on the duration of being the customer. I chose the third measure over the first one because the length of the customer relationship is already confounded in the frequency measure I defined above.

The total number of customers included in the dataset is 5,878. Since 5,878 cannot be divided by three, I split the customers in each dimension so that the middle range group has an extra - 1,959 customers in the high score group, 1,960 customers in the medium score group, and 1,960 customers in the low score group. For convenience, I will use the following notation for each group: Group*ijk* for *i,j,k* ∈{1,2,3}where *i* is the recency score (3 for smallest recency group, 1 for biggest recency group, and 2 for middle), *j* is the frequency score (3 for biggest frequency group, 1 for smallest frequency group, and 2 for middle), and *k* is the monetary value score (3 for biggest monetary value group, 1 for smallest monetary value group, and 2 for middle).[[2]](#footnote-2)

Now, I will discuss the features of the 27 groups. All group features discussed here are summarized in Table 1. First, group size is significantly different between groups ranging from 27 to 456. The smallest group was Group 133 and the next smallest group was Group311 with 43 customers. On the other hand, the largest group was Group 113 and the next largest group was Group112 with 359 customers. We can observe that in general group size is big when recency score and frequency score are similar. Note how there are small numbers of customers in groups with (recency, frequency) = (3,1) and (recency, frequency) = (1,3). Monetary value, on the other hand, does not show such patterns. The correlations between recency, frequency, and monetary values support this as well (Figure 2). Possibly, this pattern might be due to the definition of the three concepts. To illustrate, because the length of being a customer is confounded in frequency but not monetary value, this might imply that recency and length of being a customer have some correlation (although further examination is needed for confirmation).

The number of different countries within each group also varied from 2 to 21. Customers in Group331 are either from the United Kingdom or Spain while Group333 and Group223 have customers from 21 and 20 different countries, respectively. In all the groups, there were most customers from the United Kingdom. Next, I examined in which weekday customers made purchases. It turns out that in most groups, the most frequent weekday for invoices is Thursday. Monday and Saturday were not the most frequent weekday for any group. I identified one interesting pattern through weekday analysis – groups with low monetary value scores tend to have the most orders on Sunday. That is, customers who spend less per invoice have higher propensity to order on Sunday compared to other groups.

Lastly, we evaluate how much each group contributed to the aggregated sales. Overall, the contribution of the groups is highly unbalanced with few groups accounting for most of the sales. Specifically, Group333 contributed the most to the aggregated sales with £7,667,092.99, which is approximately 43.21% of the total sales. Group331 follows with £2,166,2,31.38, which is around 10.71% of the total sales. Next comes Group332 with £1,543,246.18 which around 8.70% of the total sales. These top three groups combined account for more than 60% of the total sales. The group with the least contribution is Group311 with £8,162.32, which accounts for approximately 0.05% of the total sales. The contributions of each group are summarized in Table 1 and Figure 3. In Figure3, one can observe that overall spending tend to be big when both recency score and frequency score are big. Moreover, with recency score and frequency score fixed, spending per group tends to increase with monetary value score.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **R** | **F** | **M** | **Number of Customers** | **Number of countries** | **Most Frequent Weekday of Order (Invoice)** | **Contribution to Sales (**£) |
| 1 | 1 | 1 | 266 (4.53%) | 9 | Tuesday, Wednesday | 44,072.51 |
| 1 | 1 | 2 | 359 (6.11%) | 14 | Thursday | 91,822.07 |
| 1 | 1 | 3 | 456 (7.76%) | 17 | Tuesday, Wednesday | 297,082.22 |
| 1 | 2 | 1 | 317 (5.39%) | 9 | Sunday | 128,164.14 |
| 1 | 2 | 2 | 225 (3.83%) | 15 | Thursday | 209,068.87 |
| 1 | 2 | 3 | 134 (2.28%) | 18 | Tuesday | 360,259.20 |
| 1 | 3 | 1 | 132 (2.25%) | 3 | Sunday | 190,263.70 |
| 1 | 3 | 2 | 43 (0.73%) | 3 | Thursday | 176,885.49 |
| 1 | 3 | 3 | 27 (0.46%) | 9 | Thursday, Wednesday | 274,258.04 |
| 2 | 1 | 1 | 105 (1.79%) | 4 | Thursday | 18,088.95 |
| 2 | 1 | 2 | 198 (3.37%) | 9 | Thursday | 58,307.86 |
| 2 | 1 | 3 | 297 (5.05%) | 16 | Thursday | 320,341.09 |
| 2 | 2 | 1 | 188 (3.20%) | 3 | Thursday | 100,644.51 |
| 2 | 2 | 2 | 313 (5.32%) | 18 | Thursday | 315,284.89 |
| 2 | 2 | 3 | 241 (4.10%) | 20 | Thursday | 616,486.64 |
| 2 | 3 | 1 | 271 (4.61%) | 4 | Sunday | 625,216.73 |
| 2 | 3 | 2 | 204 (3.47%) | 10 | Thursday | 721,828.65 |
| 2 | 3 | 3 | 143 (2.43%) | 18 | Thursday | 959,399.93 |
| 3 | 1 | 1 | 43 (0.73%) | 2 | Tuesday | 8,162.32 |
| 3 | 1 | 2 | 103 (1.75%) | 7 | Thursday | 30,437.10 |
| 3 | 1 | 3 | 132 (2.25%) | 10 | Wednesday | 291,856.46 |
| 3 | 2 | 1 | 152 (2.59%) | 3 | Sunday | 85,234.31 |
| 3 | 2 | 2 | 206 (3.50%) | 10 | Thursday | 205,836.75 |
| 3 | 2 | 3 | 184 (3.13%) | 15 | Thursday | 503,807.05 |
| 3 | 3 | 1 | 485 (8.25%) | 6 | Thursday | 1,900,280.51 |
| 3 | 3 | 2 | 309 (5.26%) | 11 | Thursday | 1,543,246.18 |
| 3 | 3 | 3 | 345 (5.87%) | 21 | Thursday | 7,667,092.99 |

Table Summary of RFM group features

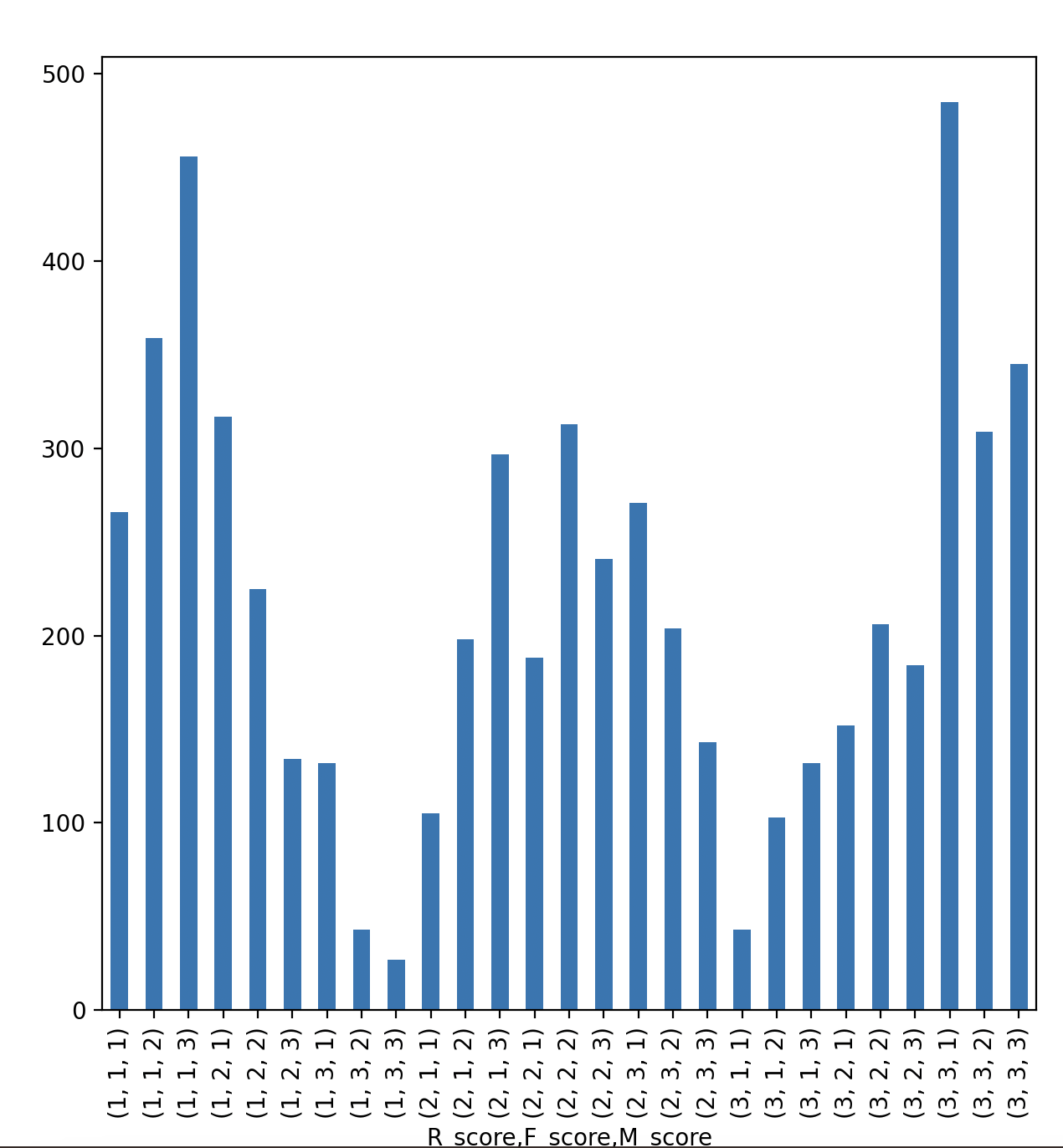


Figure Number of customers in each of 27 groups

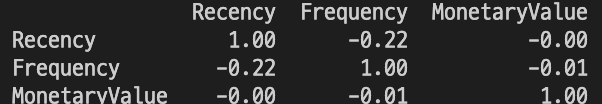


Figure Correlations between recency, frequency, and monetary value. Recall that for recency, the lower the recency value, the higher the recency score

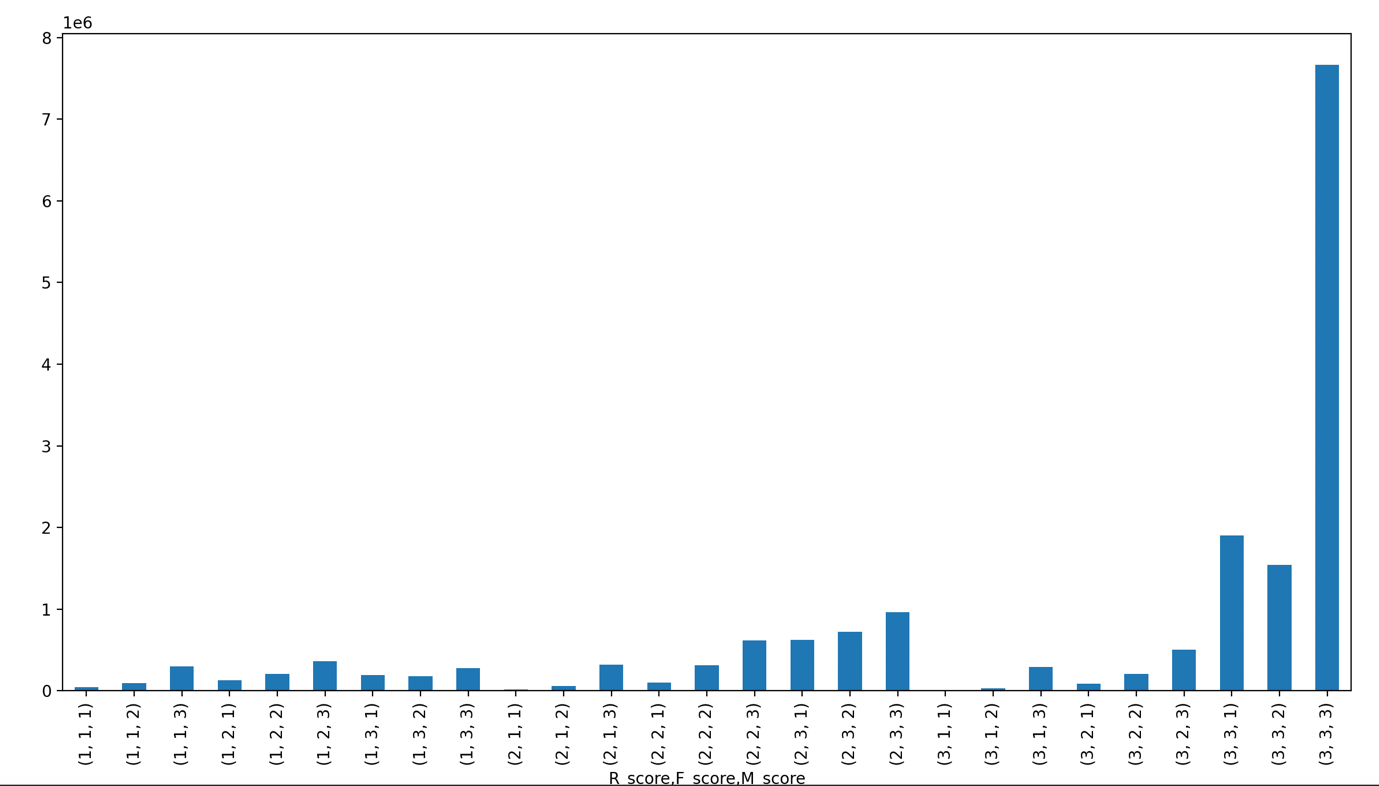


Figure Total spending per group. One can observe that a few groups contribute to most of the total sales

1. Robert, C. B., Byung, D. K., & Scott, A. N. (2008). Database Marketing: Analyzing and Managing Customers. [↑](#footnote-ref-1)
2. For instance, Group312 refers to the group of customers with small recency values (recently visited the store), small frequency values (small number of invoices), and middle range of monetary value. [↑](#footnote-ref-2)