You are given the following dataset:

**LoyaltyMemberId:** The user’s ID

**InvoiceId:** The invoice ID

**GrandTotal:** Total invoice amount due

**InvoiceDate:** The date of the transaction

You should segment the customers, by performing an RFM analysis on the given dataset. An RFM analysis consists of the following variables:

* **R**ecency: the number of days that have passed since the customer last purchase
* **F**requency: the number of purchases by the customer in the last 12 months
* **M**onetary Value: the sum of all purchases by the customer

Using RFM analysis, customers are assigned a ranking number of 1,2,3,4, or 5 (with 5 being highest) for each RFM parameter. The three scores together are referred to as an RFM "cell". The dataset is sorted to determine which customers were "the best customers" in the past, with a cell ranking of "555" being ideal.

After assigning each customer his/her ranking number, you should sort them according to their RFM “cell” value, by calculating the weighted average given by the following method:

For instance, for a customer with an RFM “cell” equal to “435”:

Next, given that the customers are sorted using the formula above, you should segment the customers into six classes: **STAR, GOLD, SILVER, BRONZE, YELLOW, WHITE.** The STAR class should consist of the best and most loyal customers, while the weaker ones should fall into the WHITE category. To achieve this, you should use the following percentages:

**STAR** the top 0.8% of the dataset

**GOLD** the next 15% of the dataset

**SILVER** the next 25% of the dataset

**BRONZE** the next 30% of the dataset

**YELLOW** the next 14% of the dataset

**WHITE** the rest of the dataset

Finally, you should fill the following table, which summarizes the analysis results:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Count | Sum(GrandTotal) | | Sum(Transactions) | Avg. Basket |
| Star |  |  |  | |  |
| Gold |  |  |  | |  |
| Silver |  |  |  | |  |
| Bronze |  |  |  | |  |
| Yellow |  |  |  | |  |
| White |  |  |  | |  |

For each segment, you should fill:

**Count:** How many customers there are per segment (e.g. 1750 STAR members)

**Sum(GrandTotal):** The sum of GrandTotal per segment (e.g. €2.000.000 total for STAR members)

**Sum(Transactions):** The sum of transactions per segment (Each invoice counts as one transaction)

**Avg. Basket:** The average basket value per segment (€50.00 average basket for STAR members)

For completing the task, you can use either Python or R, and upload the final script, as well as the document presenting your solutions, on a private repo on Bitbucket. The document can be in any form you prefer (word document, excel file etc.). You should add me, “mdagis@gmail.com” as a collaborator, in your Bitbucket repo.