

## Y42 CASE STUDY RFM ANALYSIS

### Solution:

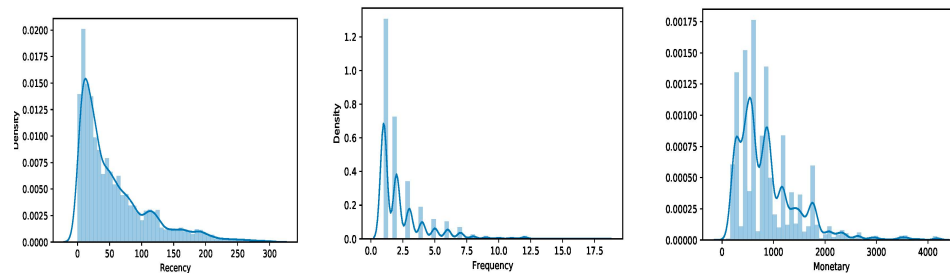
With Python Jupiter notebook I summarized the dataset and checked for missing data. After which, I used the steps highlighted below to solve the problem:

#### Step 1: Calculating RFM values:

To do this, I grouped the dataset by customer ID and then aggregated the columns “order\_date”, “total\_amount”, “orders\_id”.

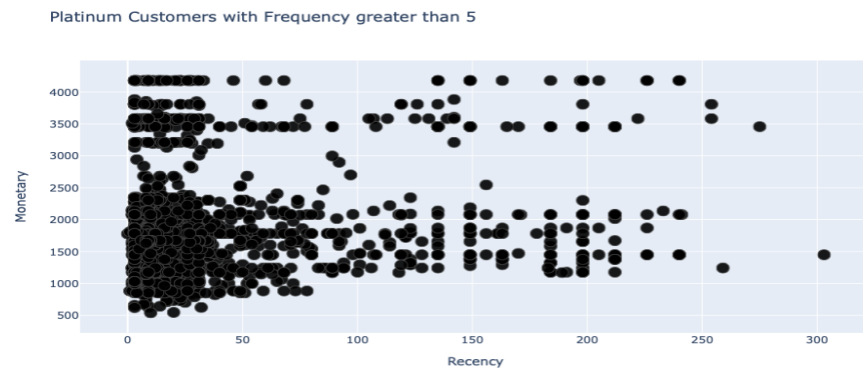
- Recency: calculated by aggregating the “order date” column by subtracting each customers last day of purchase from the latest date in the entire dataset.
- Frequency: calculated by aggregating the “orders\_id” column and counting all “order\_id” for each customer.
- Monetary: calculated by aggregating the and summing the “total\_amount” column for each customer.

**Step 2: Creating an RFM score:** Because the distribution of the obtained RFM values are right skewed as shown below, I used the quantiles 25%, 50%, 75% to attribute RFM scores to each customer. Where 1 is the most valuable point and 4 as the least valuable point



**Distribution of Recency, Frequency and Monetary**

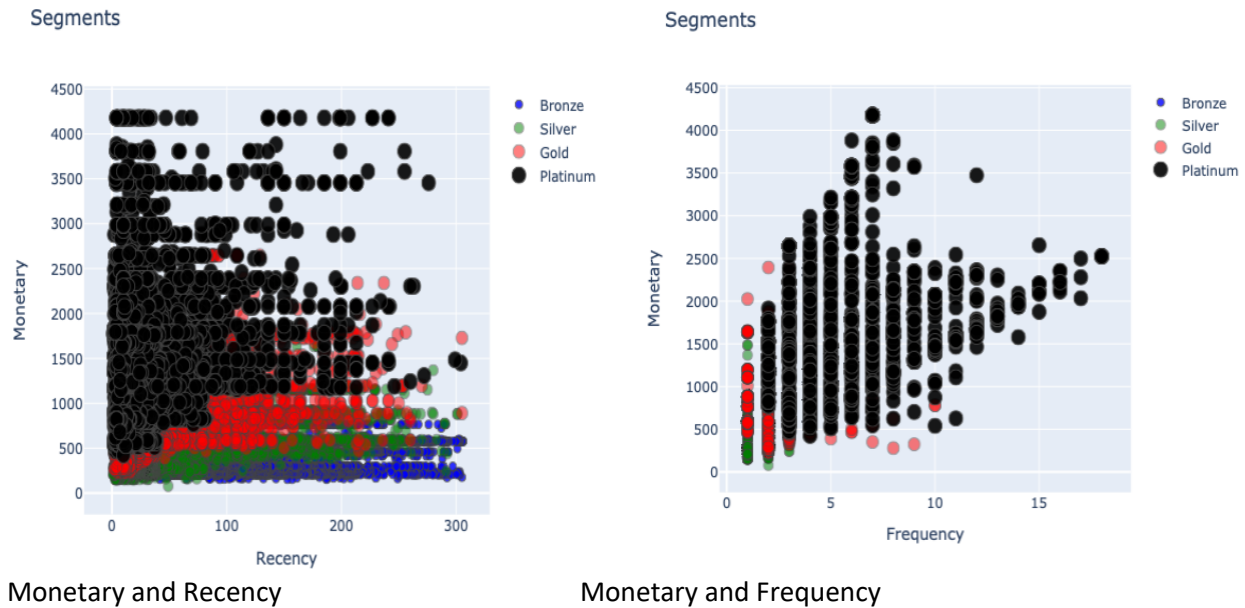
**Step 3: Categorization:** I summed the RFM score for each customer together and grouped all customer into Loyalty groups (Platinum, Gold, Silver, and Bronze). The lower the score the greater the loyalty of the customer. Because of the largeness of the dataset, I further filtered the platinum loyalty group to those whose frequency was greater than 5 as shown below.



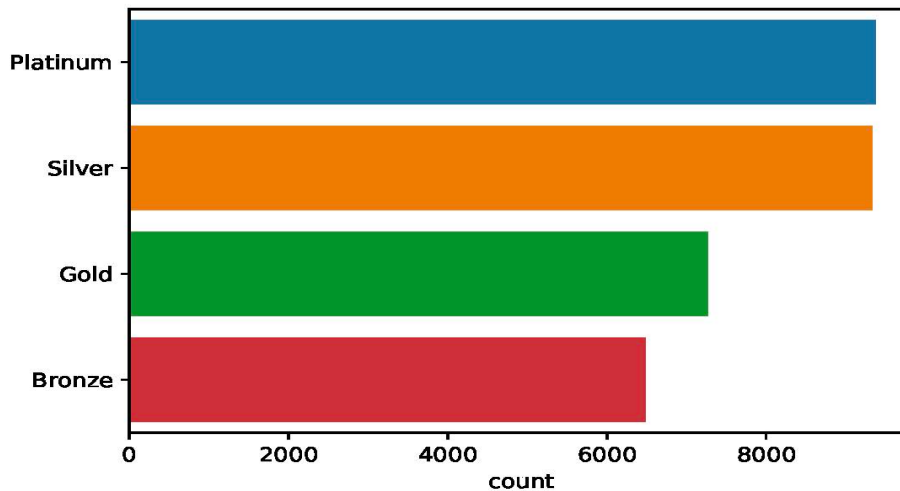
**Most Valuable customers amongst the Platinum Group**

**Other Visualization**

## Other Visualization for all loyalty group



## Conclusion:



The Platinum Group totals 9374. Which makes up 28.9% of the dataset. This figure is consistent with the Pareto principle that about 80% of company sale comes from 20% of its customers. The conclusion is that these customers have sufficiently been captured by the model and belongs to the platinum group. Targeted programs such as loyalty rewards, discounts vouchers e.t.c may be used to ensure the retention of this group of customers.

## Usage of Model on another dataset

My model has been compiled into a python module (RFM\_module) and can easily downloaded, imported and used directly on any data in any preferred python environment.