RECENCY, FREQUENCY AND MONETARY

RFM Analysis

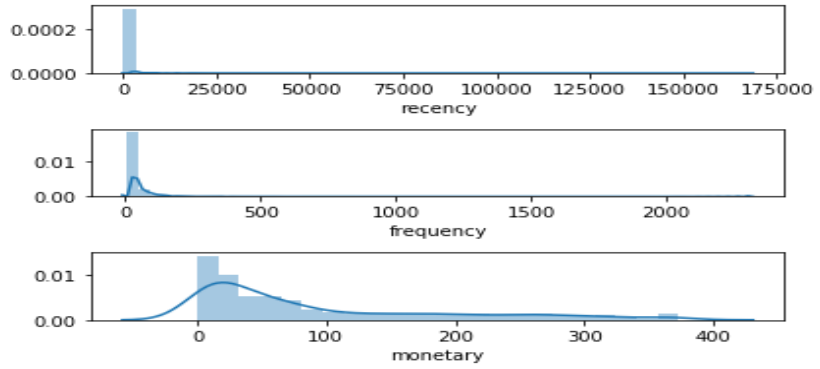
The customer transaction dataset held by the U.K. merchant has 5 variables as shown in Table 1 , and it contains all the transactions occurring in years 2010 and 2011.

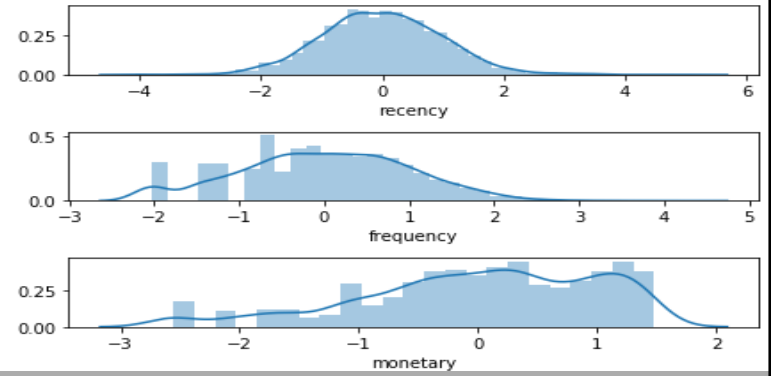
It makes each individual consumer, and therefore it makes some in-depth analyses in the present study.

As Kmeans clustering require data to be normalized and has a symmetric distribution, preprocessing process in scale is needed.

Table 1 Variables in the target dataset

Variable Names	Data Types	Description
Customer ID	Nominal	Corresponding to each distinct product category
Recency	Numeric	Recency in month
InvoiceDay	Numeric	Time in month since the first purchase in 2011
Frequency	Numeric	Frequency of purchase per product category
Monetary	Numeric	Total amount spent per product category





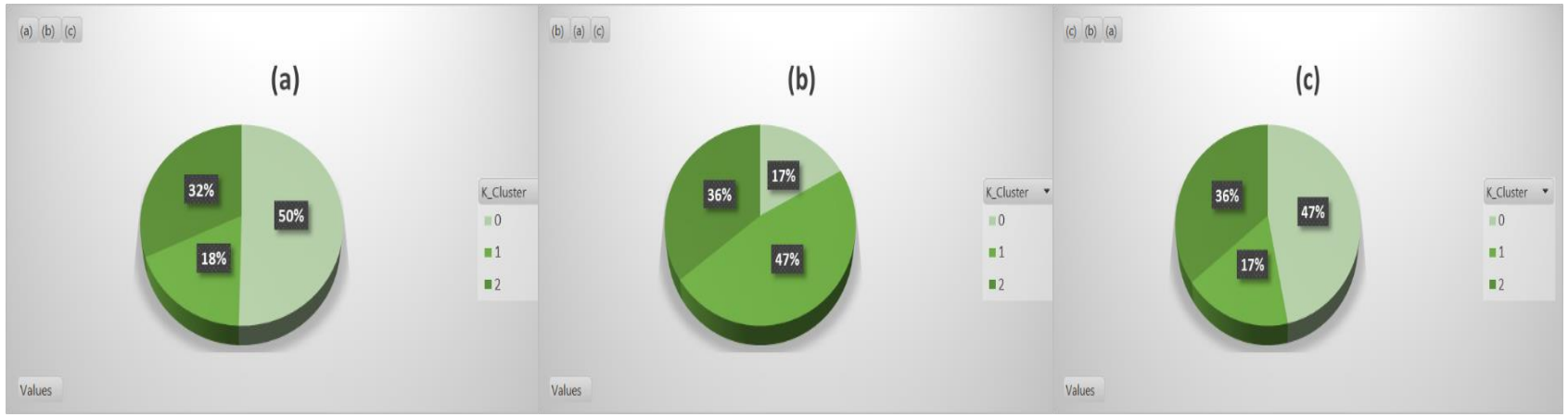
Before normalized distribution

After normalized distribution

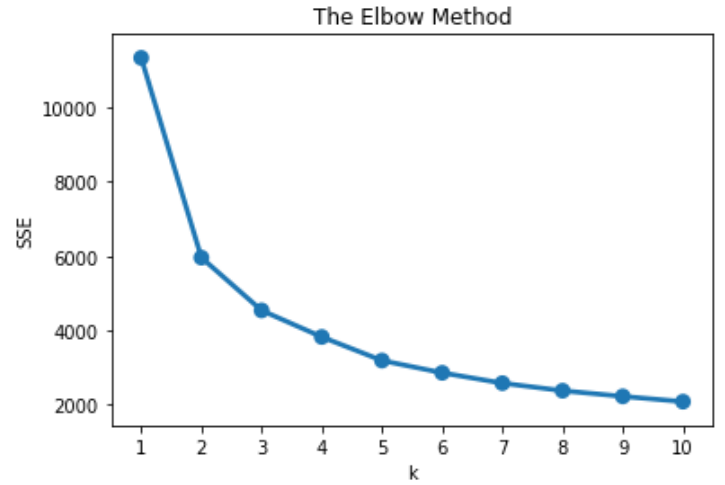
CUSTOMER LEVEL ON ONLINE STORE

Understanding the cluster

We ran k-means segmentation for several k values around the recommended value.
Cluster 0 relates to some 1900 consumers, composed of 50.0 per cent of the whole population
Cluster 1 relates to some 627 consumers, composed of 17.0 per cent of the whole population
Cluster 2 relates to some 1368 consumers, composed of 36.0 per cent of the whole population



Customer Segmentation and Instances (a) 'R', (b) 'F' and (c) 'M' in each cluster



Plot of the WSS indices using k clusters, on RFM data.

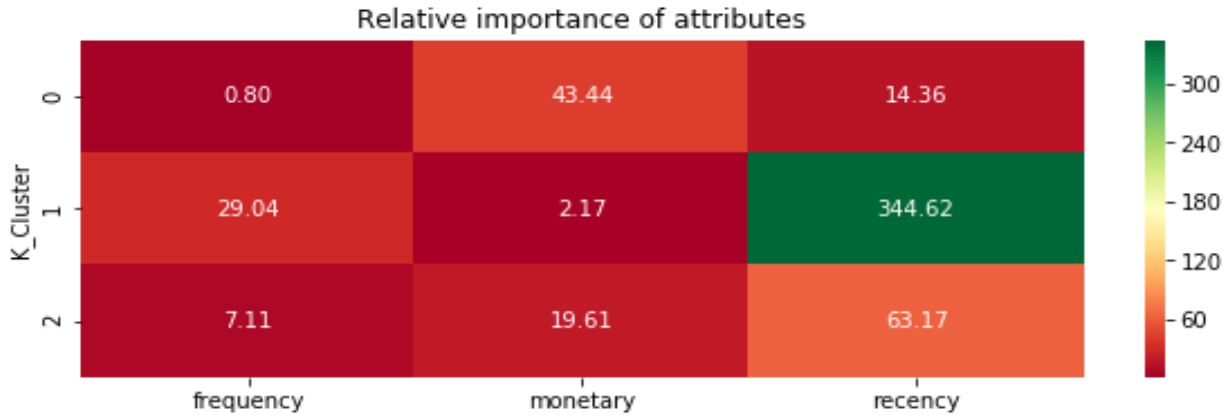
FILTER TOP/ BEST CUSTOMER

RFM Segment in 3 clusters

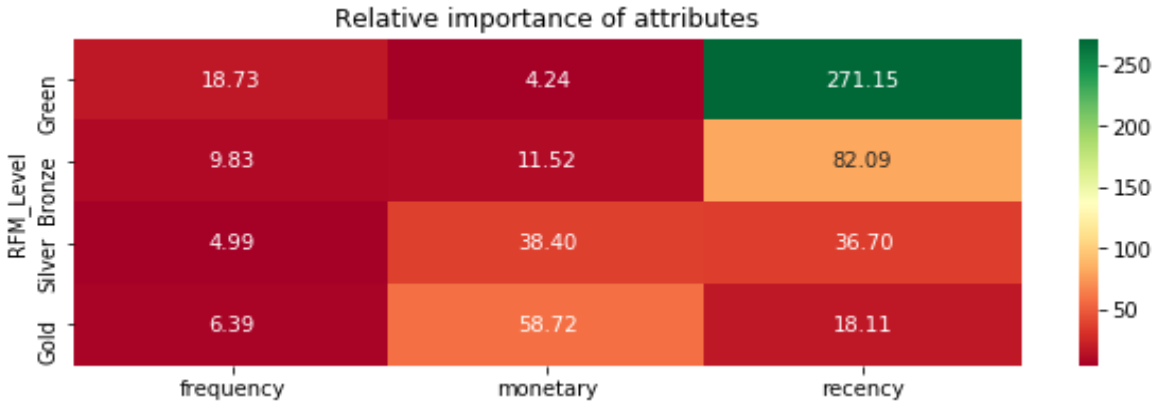
out of the total 10 instances, there was only two instances in customer ID such as 17968.0 and 16274.0 taking monetary value that instances were of less than £ 248 on comparable RFM Segment "434".

Table 2 Customer ID for the U.K. merchants by segments

Customer ID	RFM_Segment	RFM_Score	MonetaryValue	RFM_Leve
17968.0	334	10.0	218.8	Gold
16274.0	334	10.0	218.8	Gold
17908.0	434	11.0	248.6	Gold
14729.0	434	11.0	248.6	Gold
18011.0	424	10.0	218.8	Gold
17976.0	434	11.0	248.6	Gold
15923.0	424	10.0	218.8	Gold
17967.0	434	11.0	248.6	Gold
15880.0	444	12.0	369.0	Gold
16402.0	424	10.0	218.8	Gold



Relative importance of attributes by cluster



Relative importance of attributes by RFM level

CUSTOMER-CENTRIC ONLINE RETAIL / RECOMMENDATION

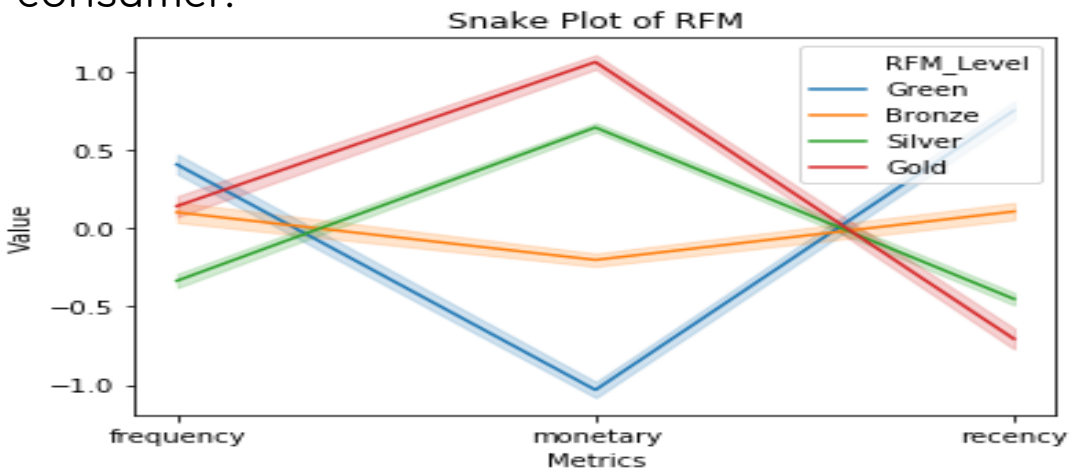
overview: parametric...

Snake plot (from market research)

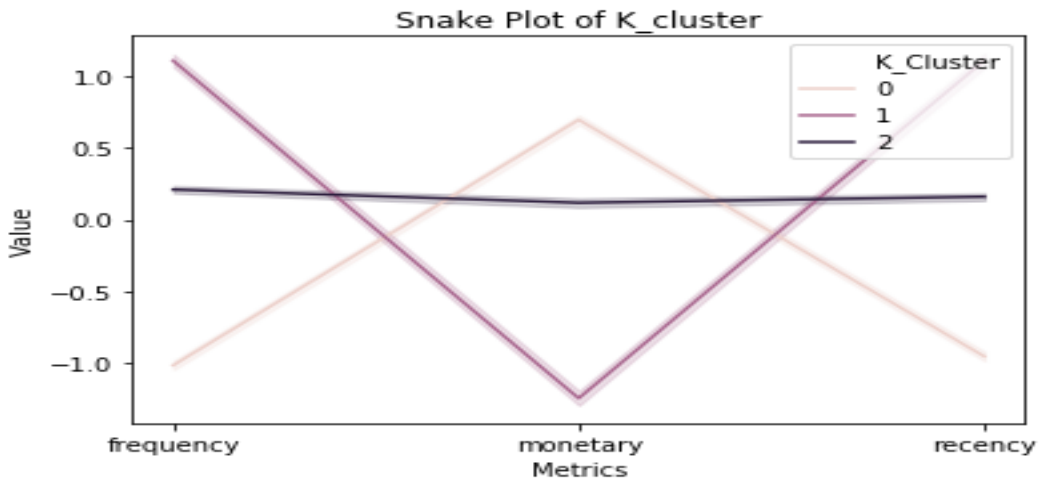
Overall, Gold and as a result, spent on product with word occurrence “pink”, “heart” and “candle” respectively can be categorized as low recency, high frequency and high monetary with a purchased quite often spending per consumer.

Silver can be categorized as medium recency, low frequency and medium monetary with a medium spending per consumer.

Bronze can be categorized as high recency, high frequency and low monetary with a regular spending per consumer.



Market research technique to compare different segment



Visual representation of each segment’s attributes