

Customer Segmentation

Online Retail Store

Dataset Description

Name: Online Retail II Dataset

Dataset was retrieved from UCI ML Repository. Data contains transaction information of products of an online retail company from Dec 2009 to Dec 2010.

Source: Dr. Daqing Chen, Course Director: MSc Data Science. chend '@' lsbu.ac.uk, School of Engineering, London South Bank University, London SE1 0AA, UK.

Fields' Description		
No	Field Name	Description
1	Invoice	Invoice number for the transaction
2	StockCode	Code of the product purchased
3	Description	Text description of the product
4	Quantity	Number of units purchased
5	InvoiceDate	Date of transaction
6	Price	Price of product
7	Customer ID	Unique Id, given to each customer
8	Country	Country of Customer

Data cleaning and preparation

First I checked all the fields to analyze each field.

	Invoice	StockCode	Description	Quantity	InvoiceDate	Price	Customer ID	Country
0	489434	85048	15CM CHRISTMAS GLASS BALL 20 LIGHTS	12	2009-12-01 07:45:00	6.95	13085.0	United Kingdom
1	489434	79323P	PINK CHERRY LIGHTS	12	2009-12-01 07:45:00	6.75	13085.0	United Kingdom
2	489434	79323W	WHITE CHERRY LIGHTS	12	2009-12-01 07:45:00	6.75	13085.0	United Kingdom
3	489434	22041	RECORD FRAME 7" SINGLE SIZE	48	2009-12-01 07:45:00	2.10	13085.0	United Kingdom
4	489434	21232	STRAWBERRY CERAMIC TRINKET BOX	24	2009-12-01 07:45:00	1.25	13085.0	United Kingdom

Then I decided to check how many customers are in each country. Based on the result, I chose to target The United Kingdom, since it has the most number of customers, therefore is a prime location for new products.

Filtering by country, I checked for missing values for customer ID and negative values in price and Quantity, since Price and Quantity would be the key fields for RFM Analysis based on Customer IDs.

	Country	Customer ID
37	United Kingdom	4035
13	Germany	68
12	France	47
31	Spain	25
24	Netherlands	23
28	Portugal	18

Customer Segmentation Online Retail II

Filtering out all negative and missing values, I calculated the Total Amount which is Price*Quantity of each transaction.

	Invoice	StockCode	Description	Quantity	InvoiceDate	Price	Customer ID	Country	TotalAmount
485847	538171	22271	FELTCRAFT DOLL ROSIE	2	2010-12-09 20:01:00	2.95	17530.0	United Kingdom	5.90
485848	538171	22750	FELTCRAFT PRINCESS LOLA DOLL	1	2010-12-09 20:01:00	3.75	17530.0	United Kingdom	3.75
485849	538171	22751	FELTCRAFT PRINCESS OLIVIA DOLL	1	2010-12-09 20:01:00	3.75	17530.0	United Kingdom	3.75
485850	538171	20970	PINK FLORAL FELTCRAFT SHOULDER BAG	2	2010-12-09 20:01:00	3.75	17530.0	United Kingdom	7.50
485851	538171	21931	JUMBO STORAGE BAG SUKI	2	2010-12-09 20:01:00	1.95	17530.0	United Kingdom	3.90

Now the file is ready for RFM Modelling. I chose the last invoice date as the recent date and the calculated Recency field as (recent date - last invoice date of a customer) in the number of days.

Frequency is calculated as the total number of transactions of a customer and Monetary as the total amount spent. This field gives us an idea of how much money a customer has spent in the business.

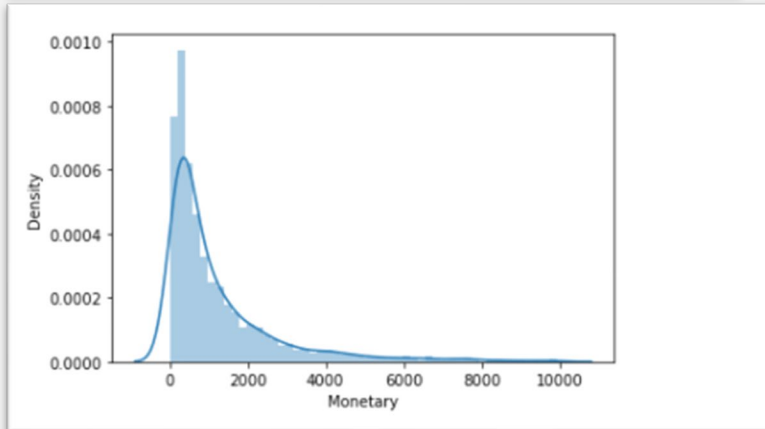
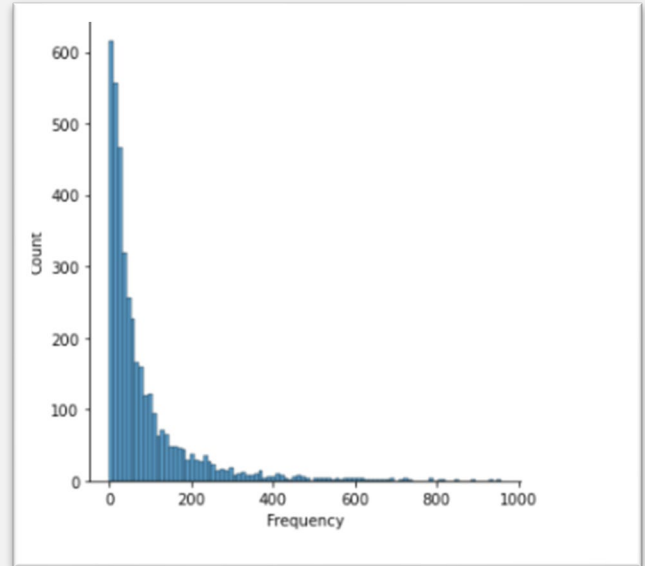
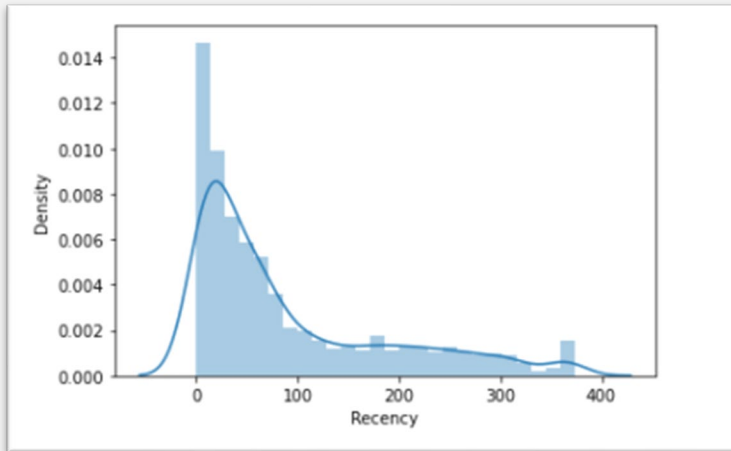
	Customer ID	Recency	Frequency	Monetary
0	12346.0	164	33	372.86
1	12608.0	39	16	415.79
2	12745.0	121	22	723.85
3	12746.0	175	17	254.55
4	12747.0	4	154	5080.53

Then I checked for descriptive stats of each of the fields of Recency, Frequency, and Monetary.

1	#Descriptive Statistics (Monetary)	1	#Descriptive Statistics (Frequency)	1	#Descriptive Statistics (Recency)
2	RFMScores.Monetary.describe()	2	RFMScores.Frequency.describe()	2	RFMScores.Recency.describe()
count	3971.000000	count	3971.000000	count	3971.000000
mean	1867.226382	mean	93.415009	mean	90.201461
std	7379.089721	std	183.278691	std	97.162435
min	0.000000	min	1.000000	min	0.000000
25%	304.110000	25%	18.000000	25%	17.000000
50%	669.850000	50%	43.000000	50%	52.000000
75%	1655.135000	75%	102.000000	75%	135.000000
max	349164.350000	max	5043.000000	max	373.000000
Name: Monetary, dtype: float64		Name: Frequency, dtype: float64		Name: Recency, dtype: float64	

I also created distribution plots for each of the R, F & M fields.

Customer Segmentation Online Retail II



Dividing each of the R, F&M in 4 quantiles and scoring them based on their values I created the RFM Group, and then adding the values I calculated the RFM Score. Labeling the scores as 'Platinum', 'Gold', 'Silver', 'Bronze' loyalty levels.

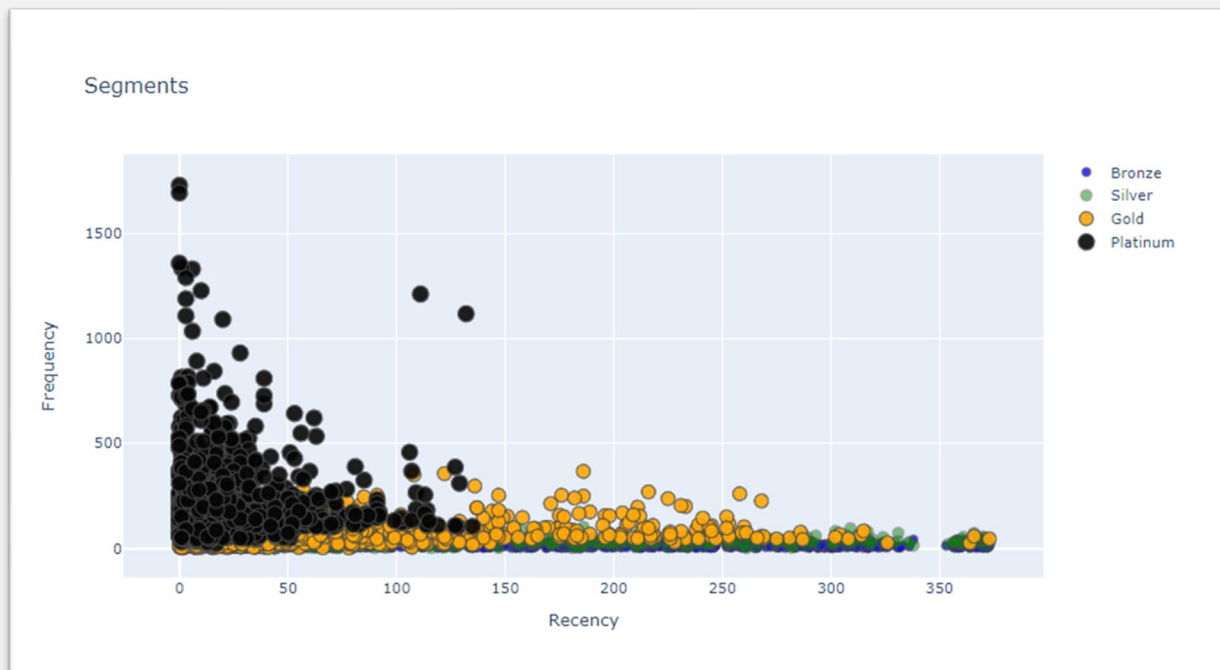
These segments give us an idea of how much value a customer provides.

	Customer ID	Recency	Frequency	Monetary	R	F	M	RFMGroup	RFMScore	RFM_Loyalty_Level
0	12346.0	164	33	372.86	4	3	3	433	10	Silver
1	12608.0	39	16	415.79	2	4	3	243	9	Silver
2	12745.0	121	22	723.85	3	3	2	332	8	Gold
3	12746.0	175	17	254.55	4	4	4	444	12	Bronze
4	12747.0	4	154	5080.53	1	1	1	111	3	Platinum

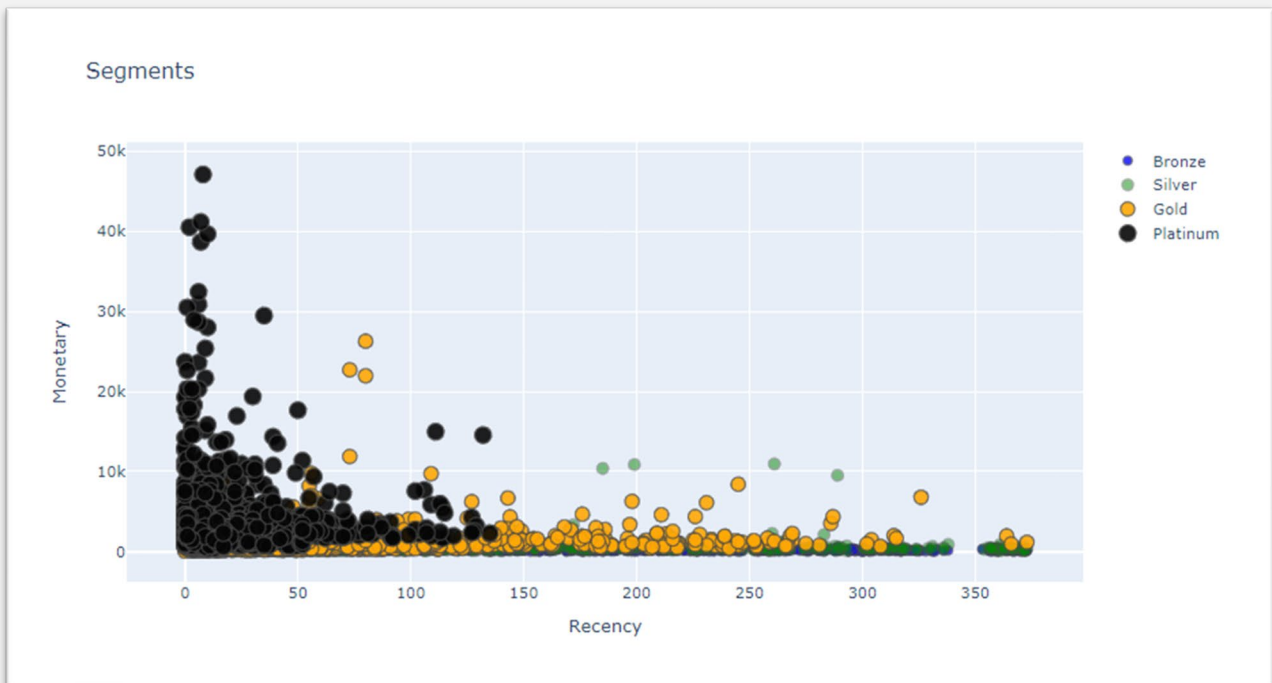
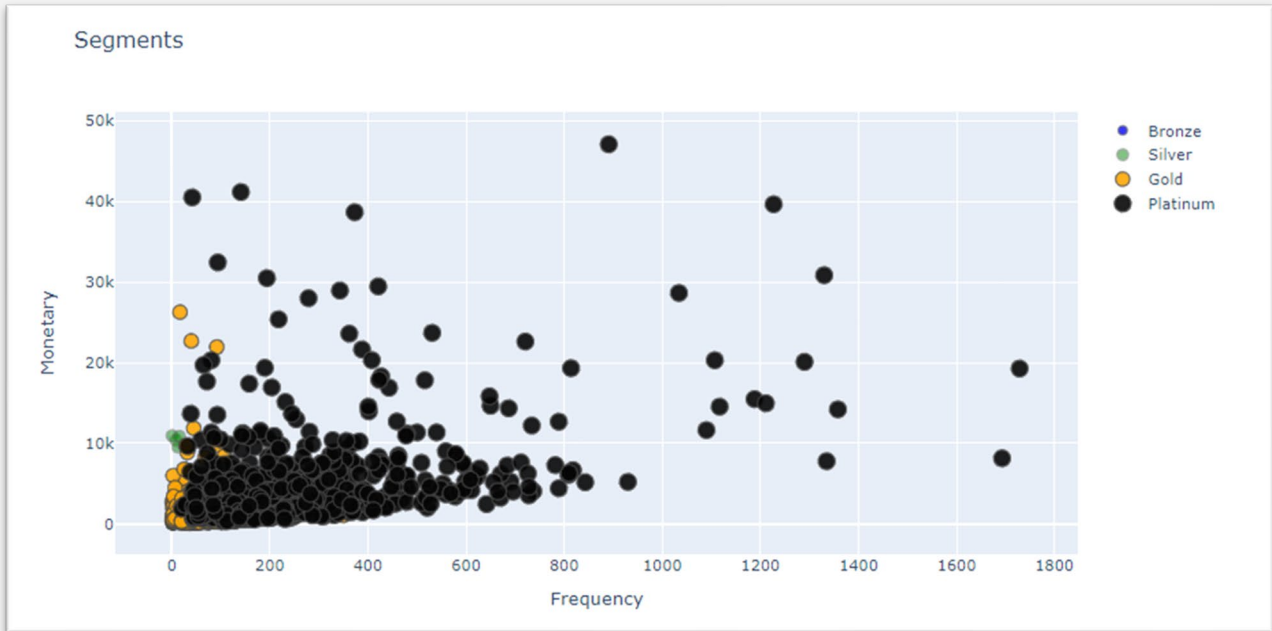
Customer Segmentation Online Retail II

```
1 RFMScores['RFM_Loyalty_Level'].value_counts()
Gold      1260
Platinum   1139
Silver     853
Bronze     719
Name: RFM_Loyalty_Level, dtype: int64
```

Total values in each segment combined with individual scores also give us an idea of which segment can be targeted more based on the total potential to move up a segment. Here, there is an almost similar number of Gold and Platinum members. The scatter plots further give an understanding of how the customer segments are related w.r.t each of the fields of Recency, Frequency, and Monetary.



Customer Segmentation Online Retail II



Recommendations

The analysis of quantity and price reveals that the products in question are not high-priced products, 75% of the products have a price lower than 3.75 pounds. Also, the segmentation analysis reveals the relationship that as frequency increases, a customer moves up the segmentation profile, and also the monetary value of the customer increase.

Therefore, in industries where products are low priced and transaction quantities are high, almost all the customer segments are important.

- The marketing towards platinum and Gold customers has to be designed to keep them buying. Here loyalty points, special privileges such as access to pre-sale items, referral bonuses may work very well. Most of the marketing should be targeted towards the exclusivity of products and special treatment.
- Since these products are low margin, it is important to analyze the cost of reaching a customer through marketing channels and converting them along with the lifetime value to make sure that it is worth targeting bronze customers. If it is, then they should be given and future purchase discounts. Packaged deals, free shipping, etc work well with bronze customers.
- For Silver customers, it is worth targeting them more than bronze customers since they are closer to move to gold customers. Therefore, the ROI of marketing is higher than that of bronze customers. Silver customers should also be provided with loyalty program deals, free shipping, and next purchase discounts, time-bound coupons, etc.