Cusommer segmentation using RFM

November 8, 2023

1 Marketing Data Science/ Customer Segmentation

2 Importing Libreries

[ ]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
```

3 Data Importing

[21]:
```
import warnings
warnings.filterwarnings("ignore")
```

[22]:
```
retail = pd.read_excel('Online Retail.xlsx')
retail.head()
```

[22]:

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
0	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	2010-12-01 08:26:00	2.55	17850.0	United Kingdom
1	536365	71053	WHITE METAL LANTERN	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	2010-12-01 08:26:00	2.75	17850.0	United Kingdom
3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom
4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom

[23]:
```
retail_df=retail.copy()
retail_df
```

[23]:
```
        InvoiceNo StockCode                          Description  Quantity  \
0          536365    85123A   WHITE HANGING HEART T-LIGHT HOLDER         6
1          536365     71053                  WHITE METAL LANTERN         6
2          536365    84406B       CREAM CUPID HEARTS COAT HANGER         8
3          536365    84029G  KNITTED UNION FLAG HOT WATER BOTTLE         6
4          536365    84029E       RED WOOLLY HOTTIE WHITE HEART.         6
…             …       …                                     …       …
541904     581587     22613          PACK OF 20 SPACEBOY NAPKINS        12
541905     581587     22899         CHILDREN'S APRON DOLLY GIRL          6
541906     581587     23254        CHILDRENS CUTLERY DOLLY GIRL          4
541907     581587     23255      CHILDRENS CUTLERY CIRCUS PARADE         4
541908     581587     22138        BAKING SET 9 PIECE RETROSPOT          3

                InvoiceDate  UnitPrice  CustomerID         Country
0       2010-12-01 08:26:00       2.55     17850.0  United Kingdom
1       2010-12-01 08:26:00       3.39     17850.0  United Kingdom
2       2010-12-01 08:26:00       2.75     17850.0  United Kingdom
3       2010-12-01 08:26:00       3.39     17850.0  United Kingdom
4       2010-12-01 08:26:00       3.39     17850.0  United Kingdom
…                       …        …         …              …
541904 2011-12-09 12:50:00       0.85     12680.0          France
541905 2011-12-09 12:50:00       2.10     12680.0          France
541906 2011-12-09 12:50:00       4.15     12680.0          France
541907 2011-12-09 12:50:00       4.15     12680.0          France
541908 2011-12-09 12:50:00       4.95     12680.0          France

[541909 rows x 8 columns]
```

[24]:
```
retail_df.info
```

[24]:
```
<bound method DataFrame.info of        InvoiceNo StockCode
Description  Quantity  \
0          536365    85123A   WHITE HANGING HEART T-LIGHT HOLDER         6
1          536365     71053                  WHITE METAL LANTERN         6
2          536365    84406B       CREAM CUPID HEARTS COAT HANGER         8
3          536365    84029G  KNITTED UNION FLAG HOT WATER BOTTLE         6
4          536365    84029E       RED WOOLLY HOTTIE WHITE HEART.         6
…             …       …                                     …       …
541904     581587     22613          PACK OF 20 SPACEBOY NAPKINS        12
541905     581587     22899         CHILDREN'S APRON DOLLY GIRL          6
541906     581587     23254        CHILDRENS CUTLERY DOLLY GIRL          4
541907     581587     23255      CHILDRENS CUTLERY CIRCUS PARADE         4
541908     581587     22138        BAKING SET 9 PIECE RETROSPOT          3

                InvoiceDate  UnitPrice  CustomerID         Country
0       2010-12-01 08:26:00       2.55     17850.0  United Kingdom
1       2010-12-01 08:26:00       3.39     17850.0  United Kingdom
```

```
2      2010-12-01 08:26:00       2.75     17850.0  United Kingdom
3      2010-12-01 08:26:00       3.39     17850.0  United Kingdom
4      2010-12-01 08:26:00       3.39     17850.0  United Kingdom
…                      …          …           …              …
541904 2011-12-09 12:50:00       0.85     12680.0          France
541905 2011-12-09 12:50:00       2.10     12680.0          France
541906 2011-12-09 12:50:00       4.15     12680.0          France
541907 2011-12-09 12:50:00       4.15     12680.0          France
541908 2011-12-09 12:50:00       4.95     12680.0          France

[541909 rows x 8 columns]>
```

[25]:
```
retail_df.describe().round(2)
```

[25]:
```
         Quantity                    InvoiceDate  UnitPrice  CustomerID
count  541909.00                         541909  541909.00   406829.00
mean        9.55  2011-07-04 13:34:57.156386048       4.61    15287.69
min    -80995.00            2010-12-01 08:26:00  -11062.06    12346.00
25%         1.00            2011-03-28 11:34:00       1.25    13953.00
50%         3.00            2011-07-19 17:17:00       2.08    15152.00
75%        10.00            2011-10-19 11:27:00       4.13    16791.00
max     80995.00            2011-12-09 12:50:00   38970.00    18287.00
std       218.08                            NaN      96.76     1713.60
```

4 Data Preparation & Exploration

[26]:
```
retail_df.isna().sum()
```

[26]:
```
InvoiceNo          0
StockCode          0
Description     1454
Quantity           0
InvoiceDate        0
UnitPrice          0
CustomerID     135080
Country            0
dtype: int64
```

[27]:
```
retail_df.dropna(inplace=True)
```

[28]:
```
retail_df.shape
```

[28]:
```
(406829, 8)
```

5 Data Description

[29]:
```
retail_df['Description']
```

[29]:
```
0          WHITE HANGING HEART T-LIGHT HOLDER
1                         WHITE METAL LANTERN
2              CREAM CUPID HEARTS COAT HANGER
3         KNITTED UNION FLAG HOT WATER BOTTLE
4              RED WOOLLY HOTTIE WHITE HEART.
                          …
541904            PACK OF 20 SPACEBOY NAPKINS
541905           CHILDREN'S APRON DOLLY GIRL
541906          CHILDRENS CUTLERY DOLLY GIRL
541907        CHILDRENS CUTLERY CIRCUS PARADE
541908          BAKING SET 9 PIECE RETROSPOT
Name: Description, Length: 406829, dtype: object
```

[30]:
```
retail_df.groupby('Description').agg({"Quantity": "sum"}).
 ↪sort_values("Quantity", ascending=False).head()
```

[30]:

Description	Quantity
WORLD WAR 2 GLIDERS ASSTD DESIGNS	53215
JUMBO BAG RED RETROSPOT	45066
ASSORTED COLOUR BIRD ORNAMENT	35314
WHITE HANGING HEART T-LIGHT HOLDER	34147
PACK OF 72 RETROSPOT CAKE CASES	33409

[31]:
```
retail_df['InvoiceNo'].str.contains('c').count()
```

[31]:
```
8905
```

6 Product conceled

[32]:
```
retail_df['InvoiceNo'].str.contains("c", na=False)
```

[32]:
```
0         False
1         False
2         False
3         False
4         False
          …
541904    False
541905    False
541906    False
541907    False
541908    False
```

```
Name: InvoiceNo, Length: 406829, dtype: bool
```

[33]:
```
retail_df['TotalPrice']=retail_df["Quantity"]*retail_df["UnitPrice"]
```

[34]:
```
retail_df
```

[34]:

	InvoiceNo	StockCode	Description	Quantity	\
0	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	
1	536365	71053	WHITE METAL LANTERN	6	
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	
3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	
4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	
…	…	…	…	…	
541904	581587	22613	PACK OF 20 SPACEBOY NAPKINS	12	
541905	581587	22899	CHILDREN'S APRON DOLLY GIRL	6	
541906	581587	23254	CHILDRENS CUTLERY DOLLY GIRL	4	
541907	581587	23255	CHILDRENS CUTLERY CIRCUS PARADE	4	
541908	581587	22138	BAKING SET 9 PIECE RETROSPOT	3	

	InvoiceDate	UnitPrice	CustomerID	Country	TotalPrice
0	2010-12-01 08:26:00	2.55	17850.0	United Kingdom	15.30
1	2010-12-01 08:26:00	3.39	17850.0	United Kingdom	20.34
2	2010-12-01 08:26:00	2.75	17850.0	United Kingdom	22.00
3	2010-12-01 08:26:00	3.39	17850.0	United Kingdom	20.34
4	2010-12-01 08:26:00	3.39	17850.0	United Kingdom	20.34
…	…	…	…	…	…
541904	2011-12-09 12:50:00	0.85	12680.0	France	10.20
541905	2011-12-09 12:50:00	2.10	12680.0	France	12.60
541906	2011-12-09 12:50:00	4.15	12680.0	France	16.60
541907	2011-12-09 12:50:00	4.15	12680.0	France	16.60
541908	2011-12-09 12:50:00	4.95	12680.0	France	14.85

```
[406829 rows x 9 columns]
```

7 RFM Analysis

[35]:
```
import datetime as dt
```

[36]:
```
retail_df['InvoiceDate'].max()
```

[36]:
```
Timestamp('2011-12-09 12:50:00')
```

[37]:
```
today_date=dt.datetime(2011,12,11)
today_date
```

[37]:
```
datetime.datetime(2011, 12, 11, 0, 0)
```

[38]:
```
rfm=retail_df.groupby('CustomerID').agg({'InvoiceDate': lambda date:(today_date␣
 ↪- date.max()).days,
                                         'InvoiceNo': lambda num: num.nunique(),
                                         'TotalPrice': lambda TotalPrice:␣
 ↪TotalPrice.sum()})
rfm
```

[38]:

CustomerID	InvoiceDate	InvoiceNo	TotalPrice
12346.0	326	2	0.00
12347.0	3	7	4310.00
12348.0	76	4	1797.24
12349.0	19	1	1757.55
12350.0	311	1	334.40
…	…	…	…
18280.0	278	1	180.60
18281.0	181	1	80.82
18282.0	8	3	176.60
18283.0	4	16	2094.88
18287.0	43	3	1837.28

[4372 rows x 3 columns]

[39]:
```
rfm.columns=["Recency","Frequency","Monetary"]
rfm
```

[39]:

CustomerID	Recency	Frequency	Monetary
12346.0	326	2	0.00
12347.0	3	7	4310.00
12348.0	76	4	1797.24
12349.0	19	1	1757.55
12350.0	311	1	334.40
…	…	…	…
18280.0	278	1	180.60
18281.0	181	1	80.82
18282.0	8	3	176.60
18283.0	4	16	2094.88
18287.0	43	3	1837.28

[4372 rows x 3 columns]

[40]:
```
rfm=rfm[rfm["Monetary"]>0]
rfm
```

[40]:

CustomerID	Recency	Frequency	Monetary

```
12347.0           3           7   4310.00
12348.0          76           4   1797.24
12349.0          19           1   1757.55
12350.0         311           1    334.40
12352.0          37          11   1545.41
…               …           …       …
18280.0         278           1    180.60
18281.0         181           1     80.82
18282.0           8           3    176.60
18283.0           4          16   2094.88
18287.0          43           3   1837.28

[4320 rows x 3 columns]
```

[41]:
```
rfm.describe().T
```

[41]:
```
             count         mean          std           min      25%     50%  \
Recency     4320.0    90.892130    99.142113  1.000000e+00   17.000   50.00
Frequency   4320.0     5.117130     9.386392  1.000000e+00    1.000    3.00
Monetary    4320.0  1924.373832  8264.936833  7.105427e-15  302.435  657.85

                75%        max
Recency      139.00     374.00
Frequency      6.00     248.00
Monetary    1626.26  279489.02
```

[42]:
```
rfm["recency_score"]= pd.qcut(rfm["Recency"],5, labels=[5,4,3,2,1])
```

[43]:
```
 rfm["frequency_score"]=pd.qcut(rfm['Frequency'].rank(method='first'),5,␣
 ↪labels=[1,2,3,4,5])
```

[44]:
```
rfm["monetary_score"]=pd.qcut(rfm['Monetary'],5, labels=[1,2,3,4,5])
```

rfm.head()

[45]:
```
rfm["RFM_SCORE"]=(rfm["recency_score"].astype(str)+rfm["frequency_score"].
 ↪astype(str))
```

rfm.head()

8 Segmenting Customers using RFM score

[46]:
```
seg_map= {
    r'[1-2][1-2]' : 'hebernating',
    r'[1-2][3-4]' : 'at_risk',
    r'[1-2]5' : 'cant_loose',
    r'3[1-2]' : 'about-tosleep',
    r'33': ' need_Attention',
```

```
    r'[3-4][4-5]' :'loyal_customer',
    r'41' : 'promissing',
    r'51' : 'new_costumers',
    r'[4-5][2-3]': 'potential_loyaliste',
    r'5[4-5]' :'championes'
}
rfm['segment']=rfm['RFM_SCORE'].replace(seg_map,regex=True)
rfm.head()
```

[46]:

CustomerID	Recency	Frequency	Monetary	recency_score	frequency_score	monetary_score	RFM_SCORE	segment
12347.0	3	7	4310.00	5	4	5	54	championes
12348.0	76	4	1797.24	2	3	4	23	at_risk
12349.0	19	1	1757.55	4	1	4	41	promissing
12350.0	311	1	334.40	1	1	2	11	hebernating
12352.0	37	11	1545.41	3	5	4	35	loyal_customer

[47]:
```
rfm[["segment","Recency","Frequency","Monetary"]].groupby("segment").
 ↪agg(["mean","count","max"]).round()
```

[47]:

	Recency			Frequency			Monetary		
	mean	count	max	mean	count	max	mean	count	max
segment									
need_Attention	49.0	178	71	3.0	178	4	821.0	178	3546.0
about-tosleep	52.0	360	71	1.0	360	2	440.0	360	6208.0
at_risk	156.0	605	373	3.0	605	7	970.0	605	
cant_loose	132.0	70	313	10.0	70	35	2383.0	70	
championes	6.0	659	12	15.0	659	248	6552.0	659	
hebernating	214.0	1037	374	1.0	1037	2	400.0	1037	
loyal_customer	33.0	776	71	8.0	776	76	2733.0	776	
new_costumers	7.0	42	12	1.0	42	1	377.0	42	
potential_loyaliste	16.0	497	32	2.0	497	4	717.0	497	
promissing	23.0	96	32	1.0	96	1	306.0	96	

```
at_risk               21536.0
cant_loose            10217.0
championes           279489.0
hebernating            7830.0
loyal_customer       123725.0
new_costumers          3861.0
potential_loyaliste   12394.0
promissing             1758.0
```

[ ]: