Customers Segmentation

By Nina Aulia N. Azhary

TABLE OF CONTENTS
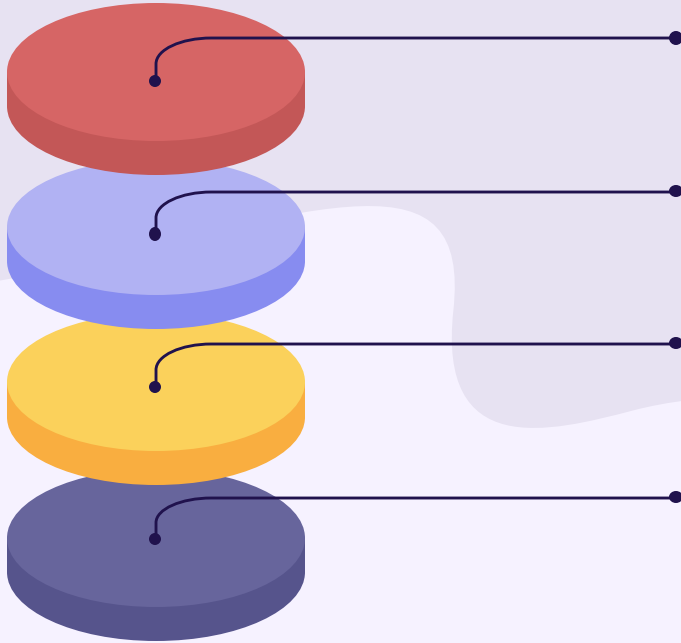
What is Customer Segmentation?

Customer segmentation divides a company's customers into groups that reflect similarities among customers in each group. The goal of segmenting customers is to decide how to relate to customers' segments to maximize the value of each customer to the business.

01

Business Problem

The company wants to make customer segmentation practically. So we try to use machine learning to allow advanced algorithms to surface insights and identify discrete groups of customers with a high degree of accuracy based on demographic, behavioral, and other indicators.

02

Data Understanding

Understand general information dataset, check for missing values, duplicate data, etc.

General Information

Our data has 70864 rows and 9 columns with no missing values

Column	Description
Invoiceno	Number uniquely assigned to each transaction
Stockcode	Number uniquely assigned to each distinct product.
Description	Product (item) name.
Quantity	The quantities of each product (item) per transaction
Invoicedate	Purchase date
Unitprice	Product price per unit
Customerid	Number uniquely assigned to each customer.
Country	The name of the country where a customer resides.
Totalprice	Total price (unitPrice * quantity)

```python
In [4]: df.drop('Unnamed: 0', inplace = True, axis = 1)
        df.describe()
```

Out[4]:

	InvoiceNo	Quantity	UnitPrice	CustomerID	totalPrice
count	70864.000000	70864.000000	70864.000000	70864.00000	70864.000000
mean	560639.843136	11.511233	3.063033	15552.19790	19.874129
std	13176.494003	41.017582	31.894970	1595.50593	78.123349
min	536365.000000	1.000000	0.001000	12747.00000	0.001000
25%	549128.000000	2.000000	1.250000	14194.00000	4.160000
50%	562012.500000	4.000000	1.950000	15525.00000	10.200000
75%	572283.000000	12.000000	3.750000	16931.00000	17.700000
max	581586.000000	4300.000000	8142.750000	18287.00000	8142.750000

```python
In [6]: df['InvoiceDate'] = df['InvoiceDate'].astype('datetime64[ns]') # convert datatype
        df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 70864 entries, 0 to 70863
Data columns (total 9 columns):
InvoiceNo      70864 non-null int64
StockCode      70864 non-null object
Description    70864 non-null object
Quantity       70864 non-null int64
InvoiceDate    70864 non-null datetime64[ns]
UnitPrice      70864 non-null float64
CustomerID     70864 non-null int64
Country        70864 non-null object
totalPrice     70864 non-null float64
dtypes: datetime64[ns](1), float64(2), int64(3), object(3)
memory usage: 4.9+ MB
```

```python
In [3]: # create new column to calculate total price
        df['totalPrice'] = df['Quantity'] * df['UnitPrice']
        df.head()
```

Out[3]:

	Unnamed: 0	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country	totalPrice
0	416792	572558	22745	POPPY'S PLAYHOUSE BEDROOM	6	2011-10-25 08:26:00	2.10	14286	United Kingdom	12.60
1	482904	577485	23196	VINTAGE LEAF MAGNETIC NOTEPAD	1	2011-11-20 11:56:00	1.45	16360	United Kingdom	1.45
2	263743	560034	23299	FOOD COVER WITH BEADS SET 2	6	2011-07-14 13:35:00	3.75	13933	United Kingdom	22.50
3	495549	578307	72349B	SET/6 PURPLE BUTTERFLY T-LIGHTS	1	2011-11-23 15:53:00	2.10	17290	United Kingdom	2.10
4	204384	554656	21756	BATH BUILDING BLOCK WORD	3	2011-05-25 13:36:00	5.95	17663	United Kingdom	17.85

```python
In [5]: df.describe(include = 'object')
```

Out[5]:

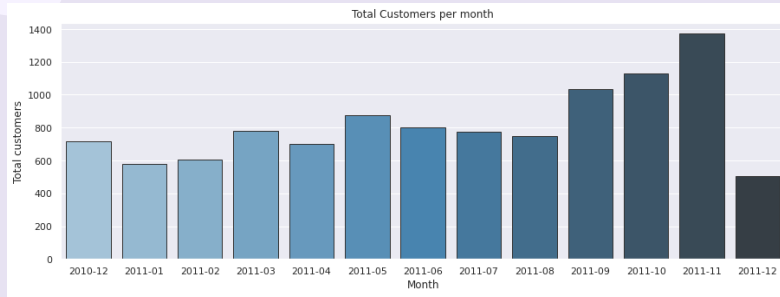|	StockCode	Description	InvoiceDate	Country
count	70864	70864	70864	70864
unique	3227	3352	13140	1
top	85123A	WHITE HANGING HEART T-LIGHT HOLDER	2011-11-14 15:27:00	United Kingdom
freq	381	381	128	70864

01. Total customers per month

03. Total Transaction per month

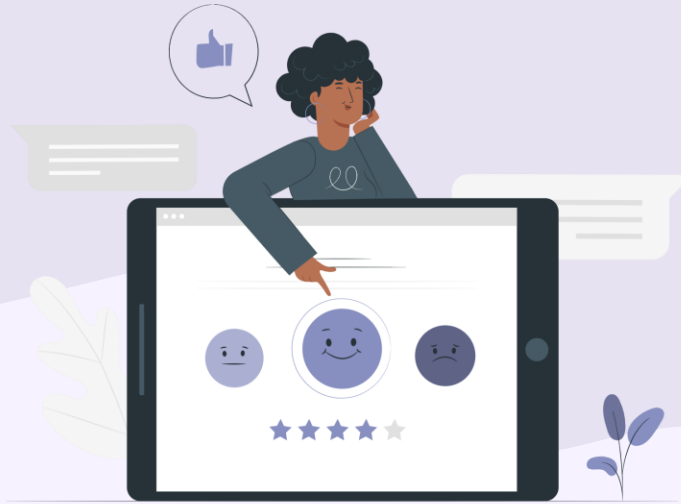# 02. Total orders per month

Findings :

our sales increased in the last three months at the end of the year but there was a clear drop off in December 2011.
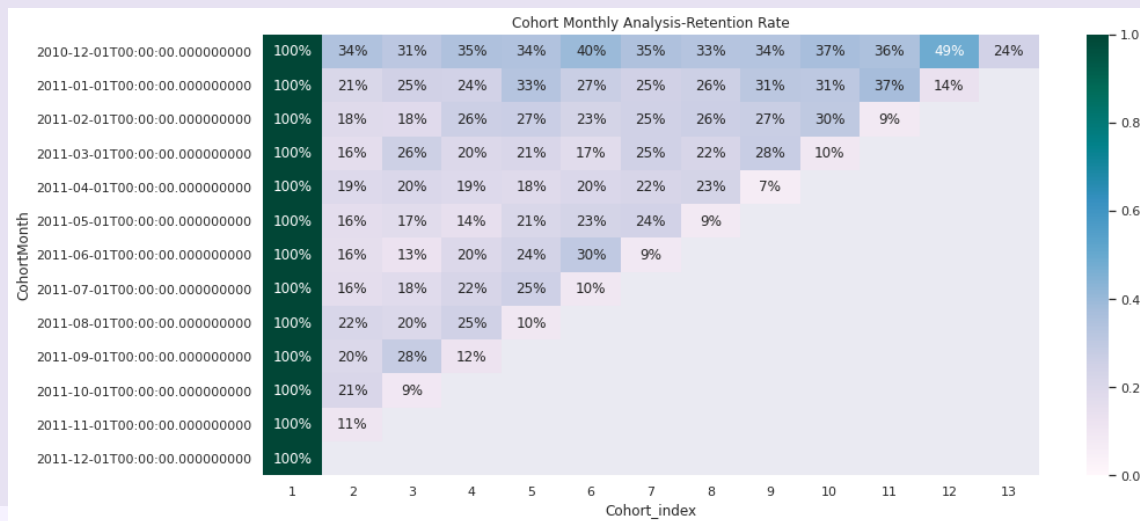
03

Cohort Analysis & RFM

grouping customers based on existing data

Cohort Analysis

Cohort Monthly Analysis-Retention Rate

From the above cohort retention rate heatmap, notice that there is average retention of 34% for the Cohort Month 2010–12–01, with the highest retention rate occurring after 11 months 49% (despite the drop off in Month 12, 30-40% of visitors were still coming back until then). The highest retention rate was in the last three months (August– Oktober2011). For all the other Cohort Months, the average retention rates are around 18–30%.

RMF Analysis


```
In [16]:  # create recency, frequency & monetary
          rfm = df.groupby(['CustomerID']).agg({'InvoiceDate' : lambda date : (PRESENT - date.max()).days,
                                                 'InvoiceNo' : 'count',
                                                 'totalPrice' : 'sum'})

          # Rename the columns
          rfm.rename(columns={'InvoiceDate': 'Recency',
                              'InvoiceNo': 'Frequency',
                              'totalPrice': 'Monetary'},
                              inplace=True)

          # Print head data
          rfm.head(5)
```

Out[16]:

CustomerID	Recency	Frequency	Monetary
12747	2	27	992.82
12748	1	967	7522.06
12749	4	37	813.45
12820	3	17	268.02
12822	71	9	146.15

```
In [22]:  rfm.tail(10)
```

Out[22]:

| CustomerID	Recency	Frequency	Monetary	R	F	M	RFM_Segment	RFM_Score	Level	segmentation
18270	38	3	50.30	2	1	1	211	4.0	Low	Lost Potential
18272	3	38	642.44	3	3	3	333	9.0	Top	Champions
18274	30	2	30.15	2	1	1	211	4.0	Low	Lost Potential
18276	44	4	67.68	2	1	1	211	4.0	Low	Lost Potential
18277	58	4	48.78	2	1	1	211	4.0	Low	Lost Potential
18280	278	2	38.70	1	1	1	111	3.0	Low	Lost
18281	181	2	31.80	1	1	1	111	3.0	Low	Lost
18282	8	2	30.70	3	1	1	311	5.0	Middle	New
18283	4	152	432.93	3	3	3	333	9.0	Top	Champions
18287	43	15	395.76	2	2	3	223	7.0	Middle	Big Spender
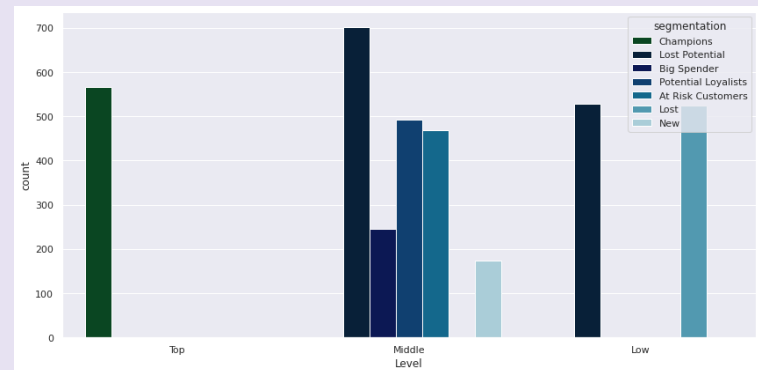
Data Pre processing

Implement log transformation and standard scaling to reduce skewness and make the data a better approximation of the normal distribution.
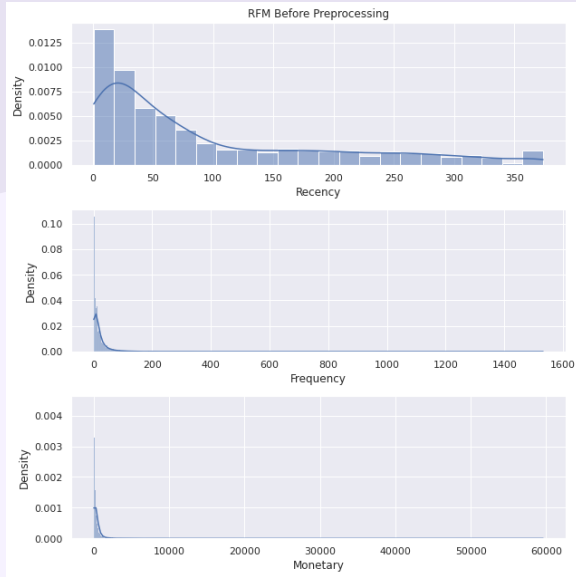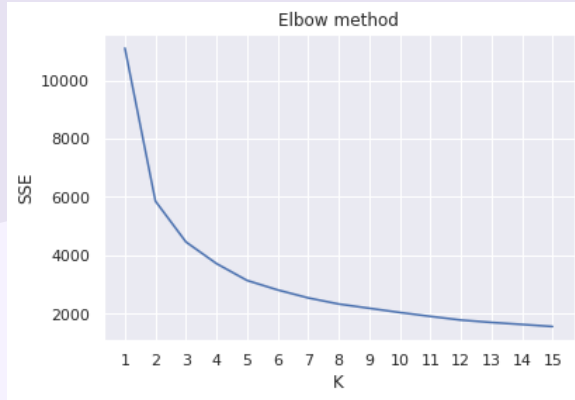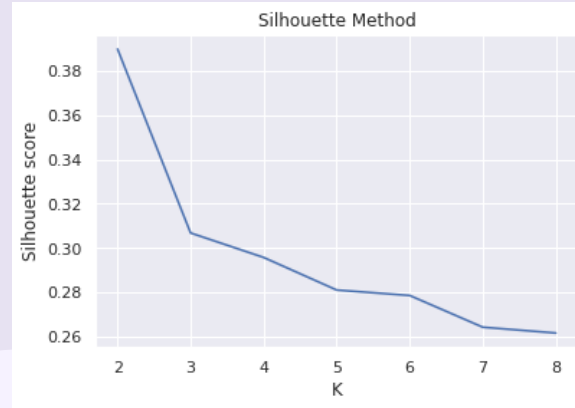

Before
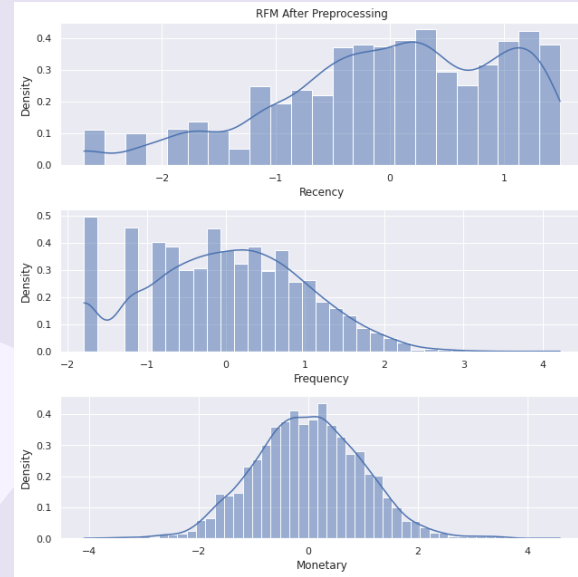
After

K-means for Clustering

01.Find the best number of K for our cluster using Elbow method and Silhouette analysis.
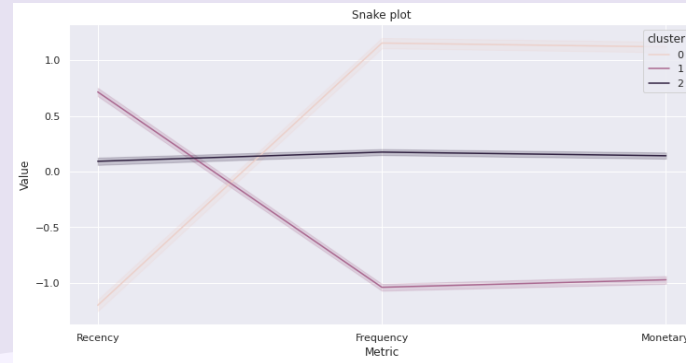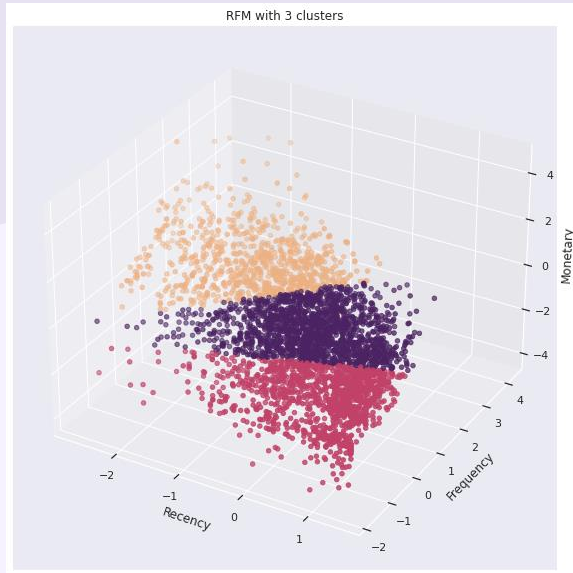

Elbow method

Silhouette method

Visualizing 3D clustering

02.apply k-means with 3 clusters and plotting the data

RFM with 3 clusters

Snake plot

Snake plot is market research technique to compare different segments, it describes more than we use the summarized table. We infer that cluster 0 is frequent, spend more, and they buy the product recently. Then, the cluster 1 is less frequent, less to spend, but they buy the product recently. Finally, the cluster 2 is less frequent, less to spend, and they buy the product at the old time.

04

Business Strategy

After analyzing and grouping customer segmentation based on the dataset, we will create a business strategy to improve customer satisfaction and mobilize our business operations.

Business Strategy

Segmentation	Description	Strategy
Champions	Best customers, who bought most recently, most often, and are heavy spenders.	Reward customers by offer free shipping or deliver a free gift with a purchase. Create customer loyalty programs.
Big Spander	Customers who spend big amount.	Keep releasing products that are consistently high-caliber and unmatched. Throwing them in personalized services to make them feel extra special.
Potential Loyalists	Our recent customers with average frequency and who spent a good amount.	Create loyalty tiers, create a point system on purchases or other customer features. Offer discounts.
New	Customers who have a high overall RFM score but are not frequent shoppers.	Offer discounts, encourage referrals, ask for feedback and reviews.
At Risk Customers	Customers who purchased often and spent big amounts, but haven't purchased recently.	Send them personalized reactivation campaigns to reconnect, ask for feedback and offer renewals and helpful products to encourage another purchase.
Lost Potential	Customers who used to visit and purchase quite often, but haven't been visiting recently.	Run surveys to find out what went wrong and avoid losing them to a competitor then create personalized campaigns.
Lost	Customers who haven't made a purchase for a long time, only make a one-time transaction and spend less.	Run surveys to find out what went wrong and avoid losing them to a competitor then create personalized campaigns (create a new and more profitable offer for your customers).

Another Business Goals

Based on the cohort and data analysis, we found that the company's sales increased at the end of the year. So we made a promotional campaign strategy to boost sales and set our business up for a successful upcoming year. such us;

- spooktacular sales promotions for Halloween.
- Christmas gift shoppers.
- Flash sales after Christmas.