E Commerce Retail Store

Online buying /selling store

Break down of e commerce retail store

- First word come into mind is **trade**
- commerce is related to trading
- Adding **e** into it makes it electronic
- So e-commerce retail store stands for doing electronically trading of goods, daily use things etc.

Understanding Problem at store

- To attain New customer
- Holding on to the old customers
- Meeting to customer demand
- Good at servicing part

Some other factors

- Age
- Education
- Society
- Occasion
- Salary

Customer Attrition

Customer attrition - also known as customer churn, turnover, or defection - is when clients or customers end their relationship with a company.

- **Voluntary**
  when the customer decides to stop purchasing your product or service.
- **Involuntary**
  is when customers leave due to factors beyond your control (e.g., they go out of business.

Factors for Customer Churn

1. Poor Onboarding
2. Weak Relationship Building
3. Poor Customer Service

Ways for reducing Customer churn

1. Segment customer Base
2. Collect Consumer data
3. Improving response time
4. Personalize the customer journey

why, how and what part regarding this retail store's problem?


Solution for this Problem

Customer Segmentation using RFM analysis

a score from 1 to 5 is given, with 5 being the highest. The collection of three values for each customer is called an RFM cell. , then sort customers from highest to lowest to find the most valuable customers

- **Recency**

  How recent was the customer's last purchase?
- **Frequency**

  How often did this customer make a purchase in a given period?
- **Monetary**

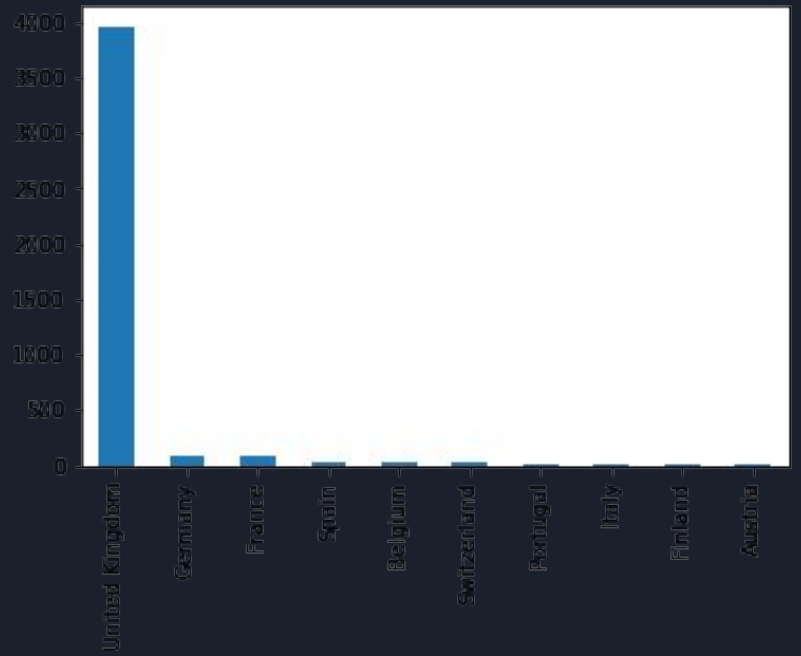
  How much money did the customer spend in a given period?

RFM Facts:

- the more recent the purchase, the more responsive the customer is to promotions
- the more frequently the customer buys, the more engaged and satisfied they are
- monetary value differentiates heavy spenders from low-value purchasers

Dataset Description

- InvoiceNo
- StockCode
- Description
- Quantity
- InvoiceDate
- UnitPrice
- CustomerID
- Country



Data cleaning

1. Converting Invoice date to Date
2. Examine the Descriptive Statistics of Data
3. Are there any missing observations in the dataset? If yes, how many missing observations in which variable?
4. *Subtract Missing Observations from the Data Set.*
5. *How Many of Which Product Are There?*
6. The 'C' in the invoices shows the canceled transactions. Remove the canceled transactions from the dataset.

	count	mean	std	min	25%	50%	75%	max
Quantity	406829.0	12.061303	248.693370	-80995.0	2.00	5.00	12.00	80995.0
UnitPrice	406829.0	3.460471	69.315162	0.0	1.25	1.95	3.75	38970.0
CustomerID	406829.0	15287.690570	1713.600303	12346.0	13953.00	15152.00	16791.00	18287.0

Descriptive Statics

Missing Data

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

Product Count

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

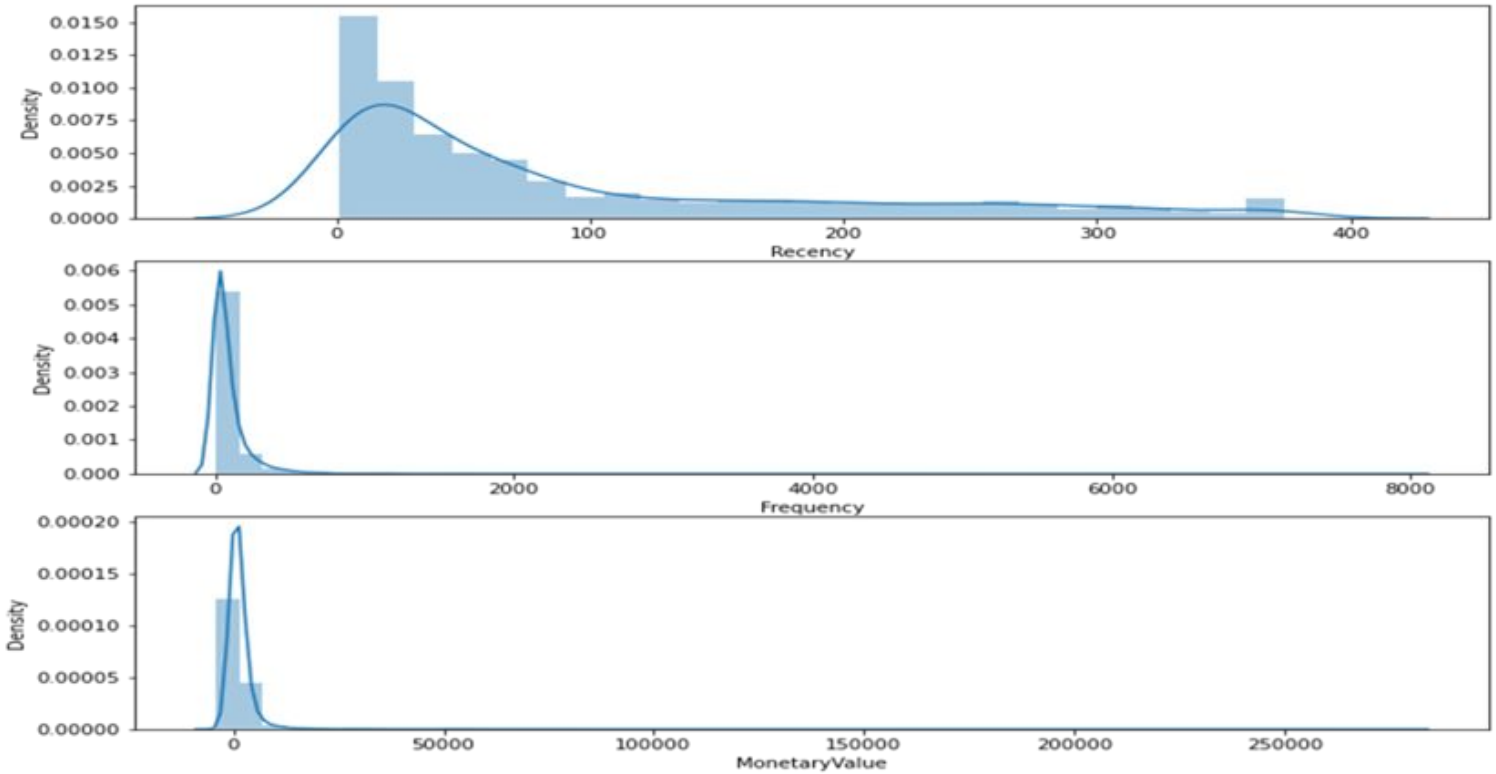
153	536382	22783	SET 3 WICKER OVAL BASKETS W LIDS	4	12/1/2010 9:45	16.95	16098.0	United Kingdom
154	C536383	35004C	SET OF 3 COLOURED FLYING DUCKS	-1	12/1/2010 9:49	4.65	15311.0	United Kingdom

Before removing the the “C” invoiceNo

After removing the the “c” InvoiceNo

150	536382	22926	IVORY GIANT GARDEN THERMOMETER	12	2010-12-01 09:45:00	5.95	16098.0	United Kingdom
157	536384	22464	HANGING METAL HEART LANTERN	12	2010-12-01 09:53:00	1.65	18074.0	United Kingdom

Analysis of the Dataset:
1.Distribution of RFM



Calculating RFM metrics

CustomerID	Recency	Frequency	Monetary
12346.0	325	1	77183.60
12747.0	2	103	4196.01
12748.0	0	4596	33719.73
12749.0	3	199	4090.88
12820.0	3	59	942.34

Generating RFM scores and defining rfm level

CustomerID	Recency	Frequency	MonetaryValue	R	F	M	RFM_Segment_Concat	RFM_Score	RFM_Level
12346.0	326	2	0.00	1	1	1	1.01.01.0	3	Require Activation
12347.0	2	182	4310.00	4	4	4	4.04.04.0	12	Can't Loose Them
12348.0	75	31	1797.24	2	2	4	2.02.04.0	8	Champions
12349.0	19	73	1757.55	3	3	4	3.03.04.0	10	Can't Loose Them
12350.0	310	17	334.40	1	1	2	1.01.02.0	4	Needs Attention

Analysing the data segments

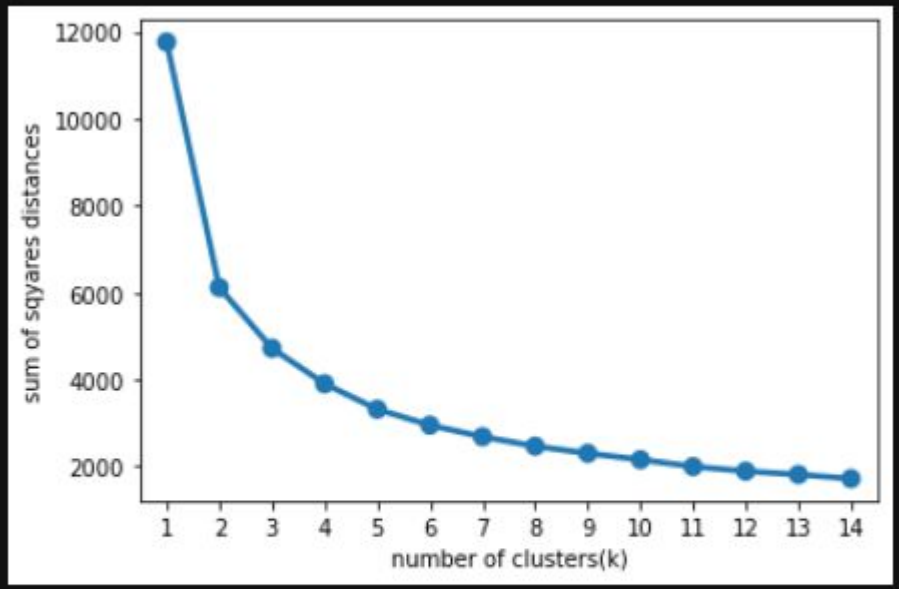
	recency			frequency			monetary		
	mean	count	max	mean	count	max	mean	count	max
segment									
about_to_sleep	52.250000	360	71	1.300000	360	2	439.892778	360	6207.67
at_Risk	155.695868	605	373	3.348760	605	7	969.751904	605	21535.90
cant_loose	132.300000	70	313	9.771429	70	35	2383.257714	70	10217.48
champions	6.088012	659	12	14.687405	659	248	6552.265372	659	279489.02
hibernating	213.615236	1037	374	1.209257	1037	2	399.946703	1037	7829.89
loyal_customers	32.701031	776	71	8.068299	776	76	2732.938455	776	123725.45
need_attention	49.196629	178	71	2.674157	178	4	821.471966	178	3545.69
new_customers	7.238095	42	12	1.000000	42	1	377.234286	42	3861.00
potential_loyalists	16.484909	497	32	2.253521	497	4	717.294567	497	12393.70
promising	23.187500	96	32	1.000000	96	1	306.213646	96	1757.55

K-Means clustering

Unsupervised learning

We use unsupervised learning when we don’t have target variable in data set

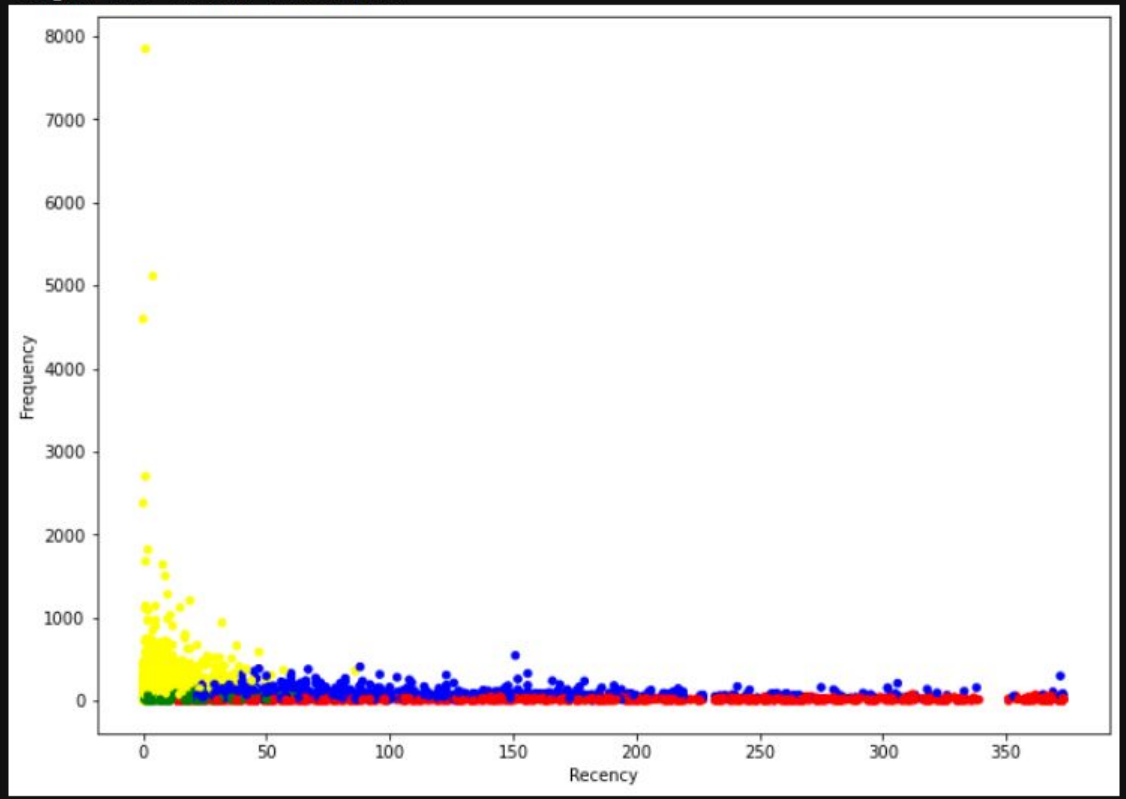
How to decide value of k in k–means




```
sse(2) 6115.101188497703
sse(3) 4718.958308381104
sse(4) 3902.8667646470117
sse(5) 3316.9174154916404
```

K -Means Scatter Plot

```
: <AxesSubplot:xlabel='Recency', ylabel='Frequency'>
  <Figure size 504x504 with 0 Axes>
```



Solutions

```
RFMScores.head()
```

CustomerID	Recency	Frequency	Monetary	R	F	M	RFMGroup	RFMScore	RFM_Loyalty_Level	Cluster	Color
12346.0	325	1	77183.60	4	1	4	414	9	Bronze	2	blue
12747.0	2	103	4196.01	1	4	4	144	9	Bronze	0	red
12748.0	0	4596	33719.73	1	4	4	144	9	Bronze	0	red
12749.0	3	199	4090.88	1	4	4	144	9	Bronze	0	red
12820.0	3	59	942.34	1	3	3	133	7	Gold	0	red