Online Retail - RFM & KMeans Clustering

Dzung Do

August 11, 2023

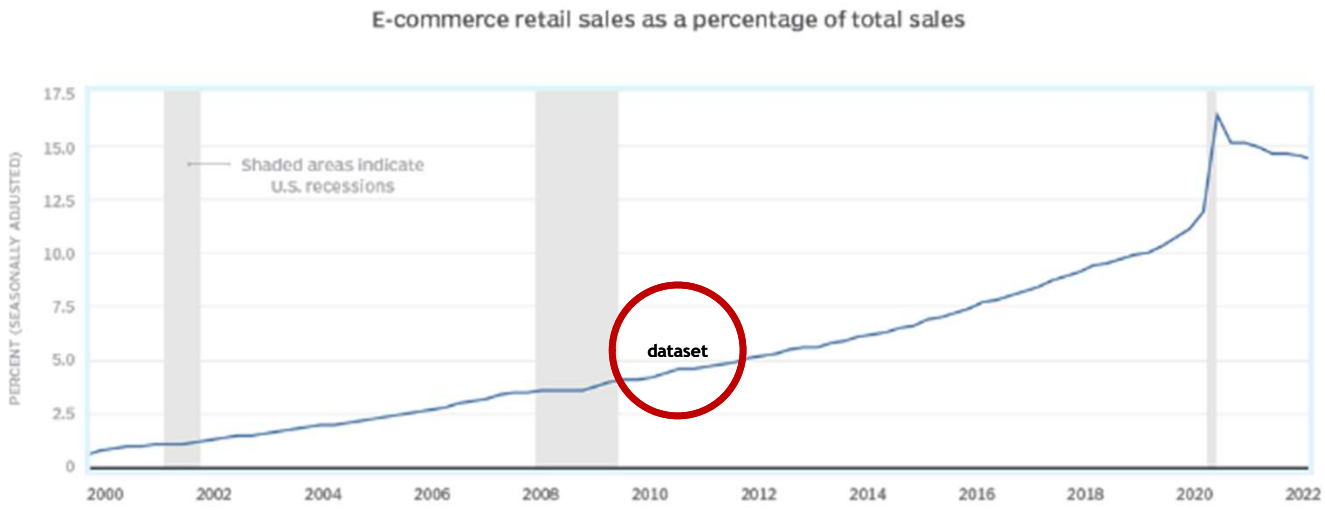

The online retail
RFM analysis
Further takes
1
2
3
4
5
Dataset
KMeans Clustering

Content

1 The online retail (E-Commerce)

- E-commerce retail sales on the rise
- Online shopping can track process, activities, essential information, etc.

  → Customer-centric business intelligence



1 The online retail (E-Commerce)

Online retailers have the following common business concerns:

- Who are the most/least valuable or loyal customers? What are the distinct characteristics of them?
- What are customers' purchase behavior patterns?
- What are the sales patterns in terms of various perspectives such as products, regions, and time?
- Which types of customers are more likely to respond to a certain promotion mailing and so on?

1 The online retail (E-Commerce)

2 Dataset

- Dataset has the transactions occurring from Dec 2010 to Dec 2011 for a **non-store** online retail based in UK.
- Items are unique all-occasion **gifts**.
- Many customers of the company are wholesalers in **37** countries.

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

```
df.shape
```

```
(541909, 8)
```

2 Dataset

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

For applying RFM method

	CustomerID	Recency	Frequency	Monetary
0	12346	326	1	77183.60
1	12347	3	7	4310.00
2	12348	76	4	1437.24
3	12349	19	1	1457.55
4	12350	311	1	294.40

3 RFM analysis

The RFM model is based on three quantitative factors:

- **Recency**: How recently a customer has made a purchase
- **Frequency**: How often a customer makes a purchase
- **Monetary**: How much money a customer spends on purchases

Then ranking a customer in each factor on a scale of 1 to 5 (the higher the number, the better the result)

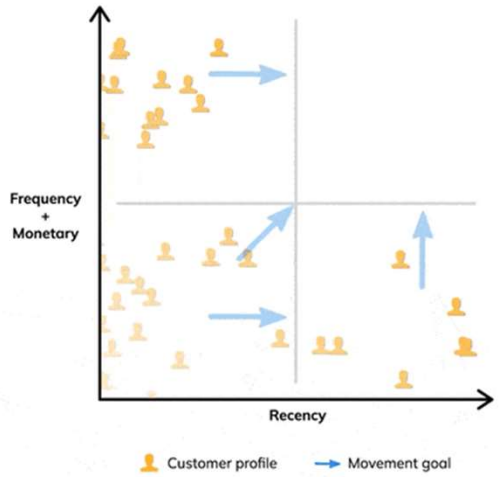
3 RFM analysis

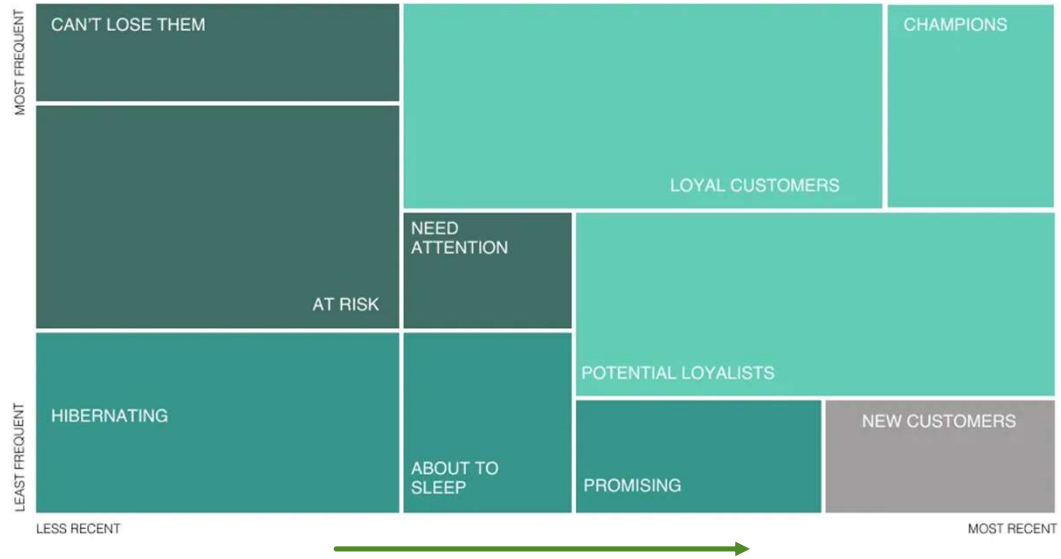
	CustomerID	Recency	Frequency	Monetary
0	12346	326	1	77183.60
1	12347	3	7	4310.00
2	12348	76	4	1437.24
3	12349	19	1	1457.55
4	12350	311	1	294.40

create 3 new columns
of scores by 'qcut'

	CustomerID	Recency	Frequency	Monetary	Recency_score	Frequency_score	Monetary_score	RFM_score
0	12346	326	1	77183.60	1	1	5	115
1	12347	3	7	4310.00	5	5	5	555
2	12348	76	4	1437.24	2	4	4	244
3	12349	19	1	1457.55	4	1	4	414
4	12350	311	1	294.40	1	1	2	112

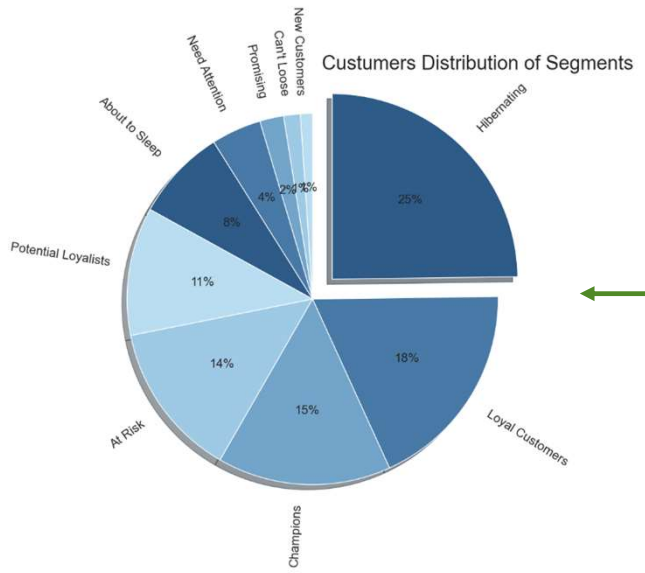
3 RFM analysis





3 RFM analysis

	CustomerID	Recency	Frequency	Monetary	Recency_score	Frequency_score	Monetary_score	RFM_score	Segment
0	12346	326	1	77183.60	1	1	5	115	Hibernating
1	12347	3	7	4310.00	5	5	5	555	Champions
2	12348	76	4	1437.24	2	4	4	244	At Risk
3	12349	19	1	1457.55	4	1	4	414	Promising
4	12350	311	1	294.40	1	1	2	112	Hibernating

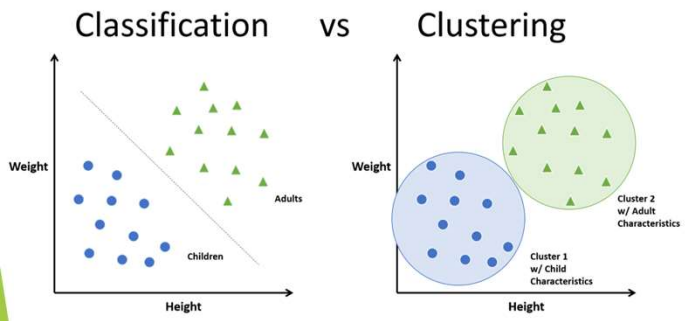


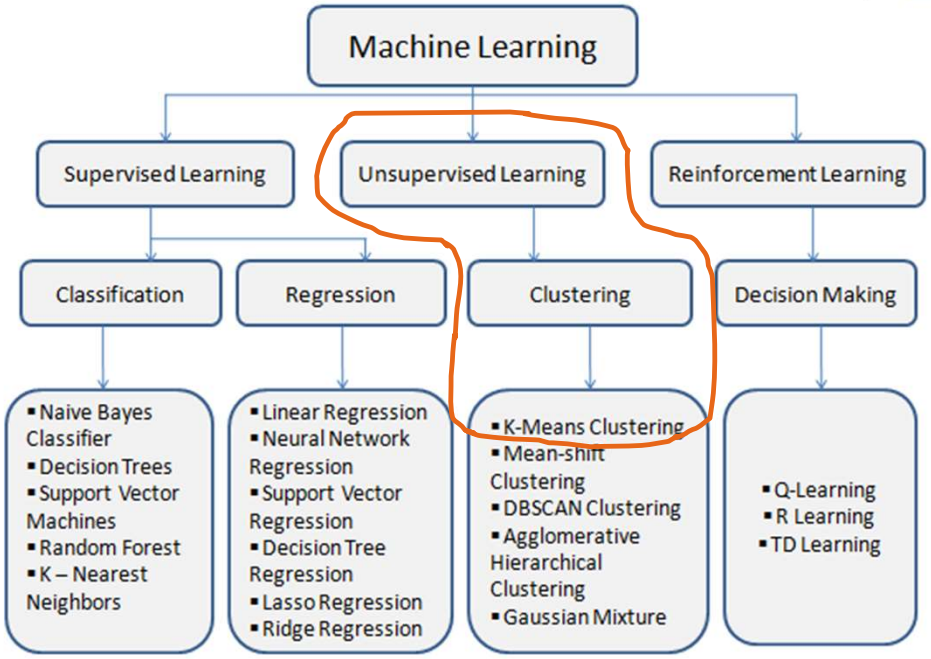
	Segment	Count
7	New Customers	46
2	Can't Loose	63
9	Promising	90
6	Need Attention	189
0	About to Sleep	348
8	Potential Loyalists	484
1	At Risk	587
3	Champions	655
5	Loyal Customers	798
4	Hibernating	1074

4 KMeans Clustering

Three popular types of **clustering** algorithms:

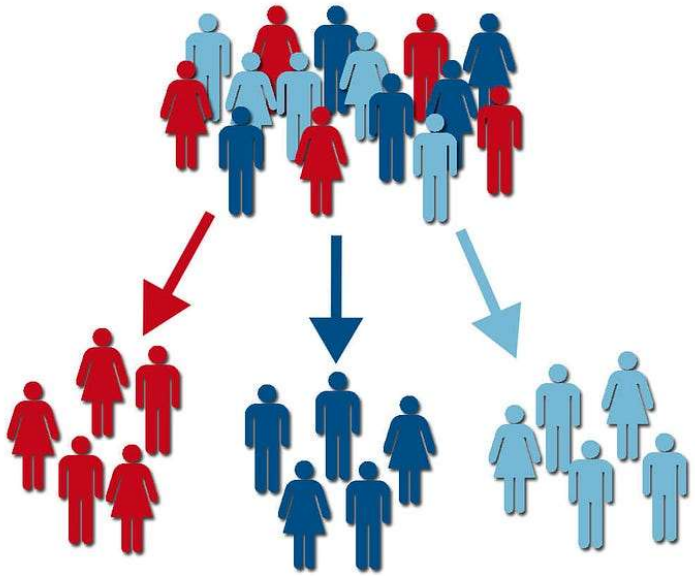
1. Partitional clustering -> **KMeans**
2. Hierarchical clustering
3. Density-based clustering



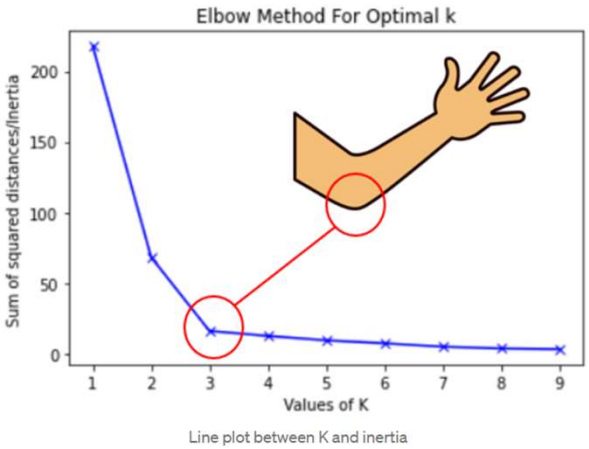


4 KMeans Clustering

E-commerce Dataset → Customer Segmentation → RFM & KMeans



4 KMeans Clustering

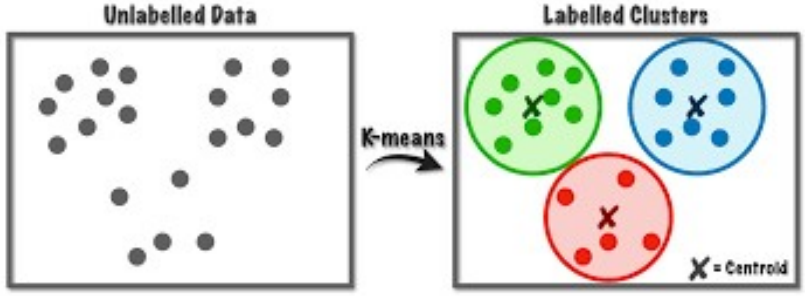


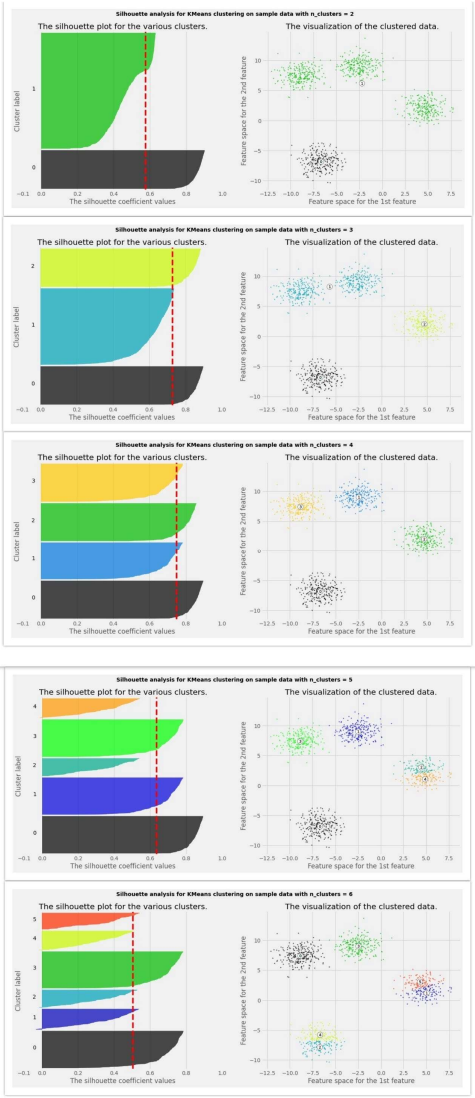
Line plot between K and inertia

Elbow method or/and
Silhouette score

to get **k**

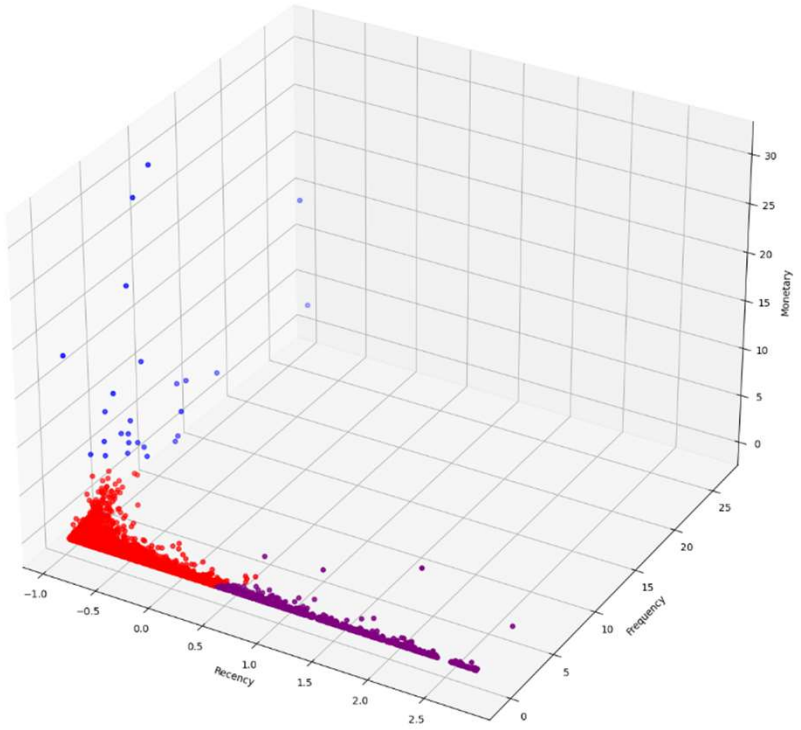
model = KMeans(n_clusters = 3)





4 KMeans Clustering

RFM & KMeans → Apply 2 model layers → Deep Learning



5 Further takes

- Answer more questions in Exploratory Data Analysis (EDA) such as sales trend, new customers, & transactions by time
- Apply Time Series models for forecasting the trend of purchases
- Study more 3D visualizations for Clustering models

Thank you!