



Customer Segmentation & Its Impact on Business Strategy Analysis

Presented By

MD. Ifti Khairul Alam ; ID : 18101014

Tanveer Ahamed Rabby ; ID : 18101025

Sharmin Akter ; ID : 18101065

Supervisor

Shammi Akhtar
Assistant Professor,CSE

External

Molla Rashied Hussein
Assistant Professor,CSE

Contents



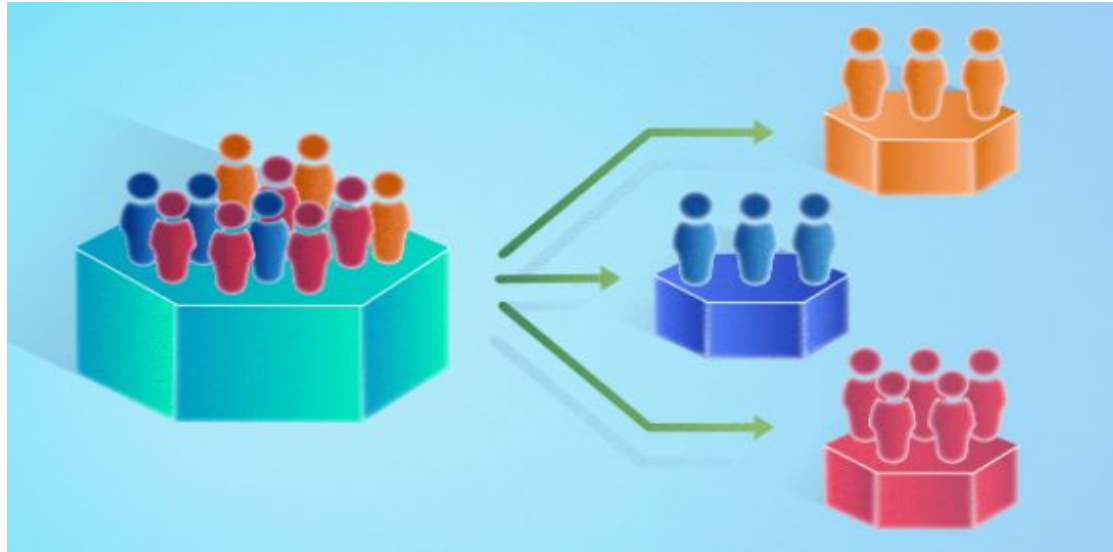
- Introduction
- Motivation
- Types of customer segmentation
- Model for customer segmentation
- RFM Segmentation Model
- Previous Work
- Our Objective
- Workflow
- Data Collection & Understanding
- Data Exploration
- Data Pre Processing
- Feature Engineering
 - Visual of RFM Data
 - Visual of RFM for Clean Data
 - RFM Score Calculation
 - RFM Score Calculation for Clear Data
 - Model Evaluation
 - Cluster Analysis Using K-means
 - Cluster Analysis Using Hierarchical
- Result Analysis
 - Customer Segmentation
 - Product Analysis
 - Recommendation for Business Decision
- Future Work
- CEP Mapping

Introduction



What is customer segmentation?

Customer segmentation involves grouping customers into specific marketing groups, perhaps narrowing them down by gender, interests, buying habits or demographic etc.



Motivation



- People's buying perspective is changing and also buying capability.
- To fulfil their demand and serve them spontaneously we need to segmenting them with catagoriously.
- Best customer satisfaction .
- Make more profit without lose.
- Greater company focus
- Better serve a customer's needs and wants
- Market competitiveness
- Market expansion
- Targeted communication

Types of customer segmentation



GEOGRAPHIC

Based on predefined geographic boundaries and regional needs like type of cities, climate, elevation or region

DEMOGRAPHIC

Based on variable such as age, gender, marital status, education level, family size, occupation, income or religion

PSYCHOGRAPHIC

Based on values, personalities, interests, lifestyle, opinions, temperament or motivators

BEHAVIOURAL

Based on decision-making and buying process with aspects like brand perception, knowledge of the product, the usage and more

Uses of Customer Segmentation



Supershop

E commerce

Hospital

Bank

Real estate

Shopping Mall



RFM Segmentation Model



RECENCY

The *freshness* of customer activity.

e.g. time since last activity



FREQUENCY

The *frequency* of customer transactions.

e.g. the total number of recorded transactions



MONETARY

The *willingness* to spend.

e.g. the total transaction value



Model for Customer Segmentation

- Demographic
- Recency, frequency, monetary (RFM)
- High-value customer (HVCs)
- Customer status

Previous work



Title	Author	Description
CUSTOMER SEGMENTATION USING MACHINE LEARNING	V.Vijilesh , A.Harini, M.Hari Dharshini, R.Priyadharshini	Basic K-means,most selling product analysis missing
RFM ranking – An effective approach to customer segmentation	A.Joy Christy, A.Umamaheswari, L.Priyadarshini, A.Neyaa	Only comparison between 3 algorithms(k-means,fuzzy c means,rm-k means)
e-CLV: a modeling approach for client lifetime assessment in e-commerce spaces, with an application and case considered for online barbers. In innovation, e- Commerce, and e-Service	Etzion, O., Fisher, A., & Wasserkrug, S.	grouped profitable clients and calculated the price for their lifetime with the company
Improving personalization solutions through optimal segmentation of customer bases.	Jiang, T., Tuzhilin, A	proposed in order to increase performance in marketing, both segmentation of customers and targeting of buyers are necessary
CUSTOMER SEGMENTATION USING MACHINE LEARNING	AMAN BANDUNI, Prof AALAVANDHAN A	Without RFm analysis segmentation

Our Objective

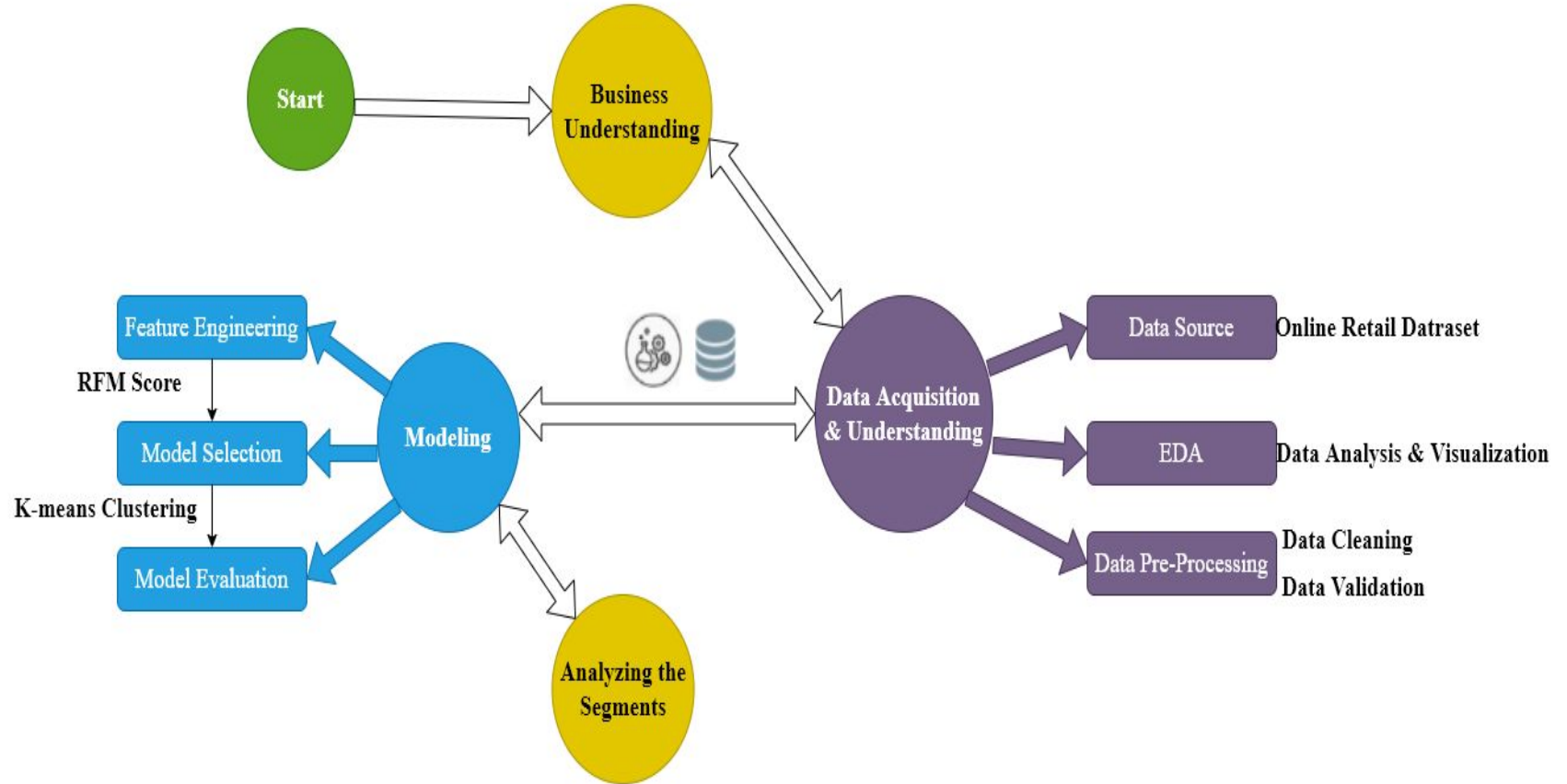


- **Collect Dataset**
- **Data Preprocessing**
- **RFM Analysis**
- **RFM Score Calculation & Segment Customer in Category**
- **K-means & Hierarchical Clustering Model Evaluation (regrading to Silhouette Coefficient & Davies-Bouldin index)**
- **Analysis Customer Segment & Product**
- **Recommendation Business Decision.**



- Silhouette Coefficient score is a metric used to compute the goodness of a clustering technique. Its score varies from -1 to 1
- Davies-Bouldin index, which is a validation metric that is sometimes used in order to evaluate the minimum number of clusters to use.

Workflow



Data Collection & Understanding



The dataset is extracted from UCI Machine Learning Repository

Link : <https://data.world/uci/online-retail>

- The dataset contains transactions occurring between 01/12/2010 and 09/12/2011

- No. of Observation : 541,909
- No. of features: 8
- No. of Transactions: 22,190
- No. of Customers: 4,372
- No. of Products: 3,684

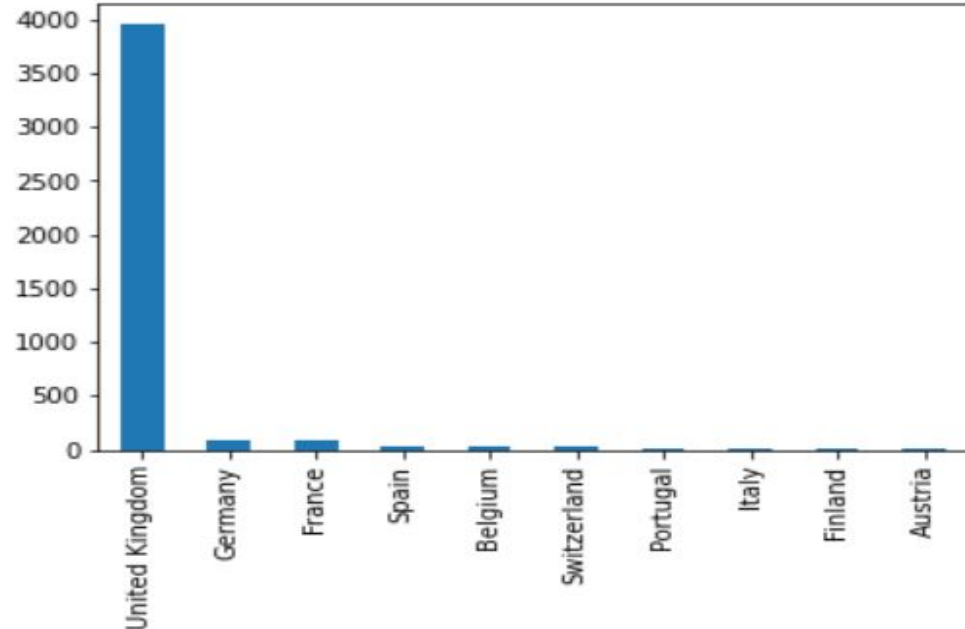
Features	Description
InvoiceNo	6-digit integral number uniquely assigned to each transaction
StockCode	5-digit integral number uniquely assigned to each
Description	Product name
Quantity	The quantities of each product (item) per transaction
InvoiceDate	The day and time when each transaction was generated
UnitPrice	Product price per unit in sterling £
CustomerID	5-digit integral number uniquely assigned to each customer
Country	The name of the country where each customer resides

Data Exploration

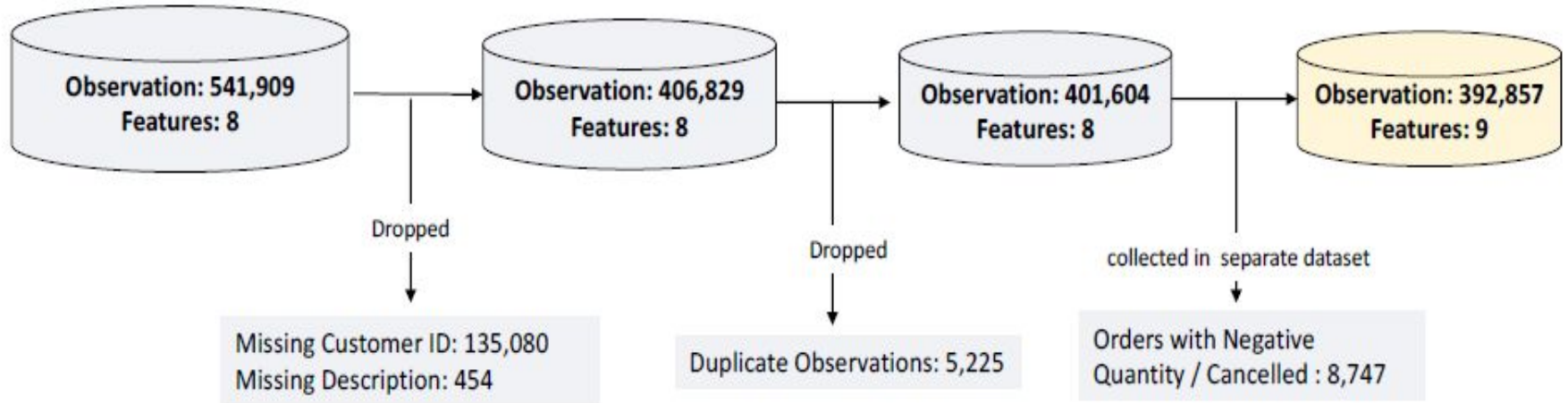


- Checking Null Values
- Checking Duplicate Values
- Checking Negative Quantities

- Largest customer base is United Kingdom with 3,950 customers which is 90% of the dataset



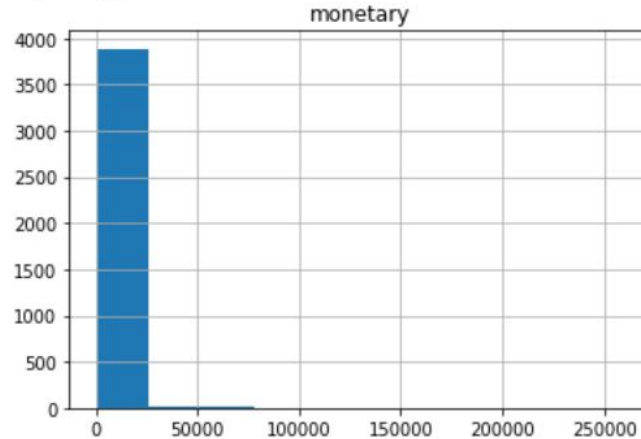
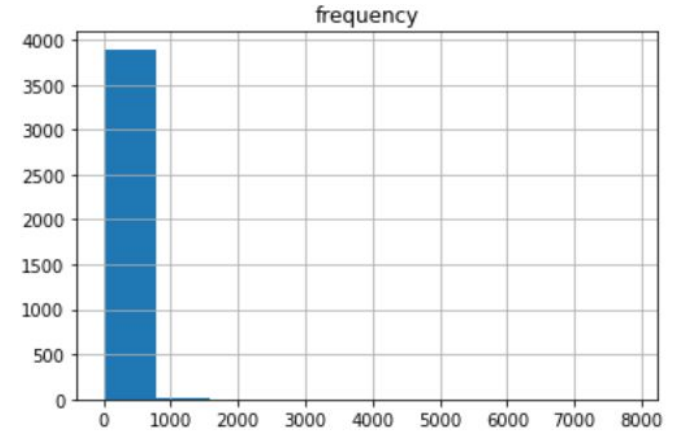
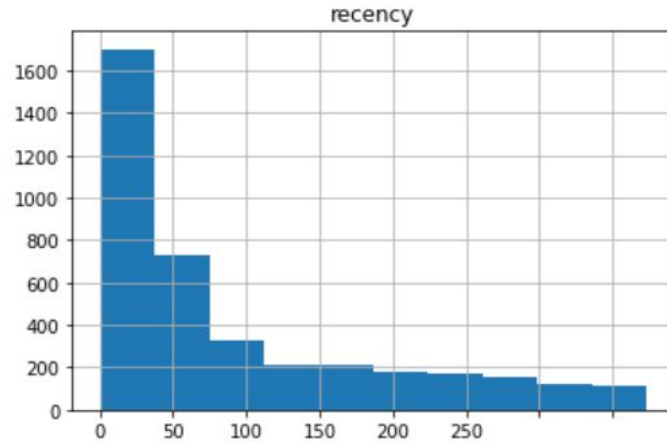
Data Preprocessing





Feature Engineering

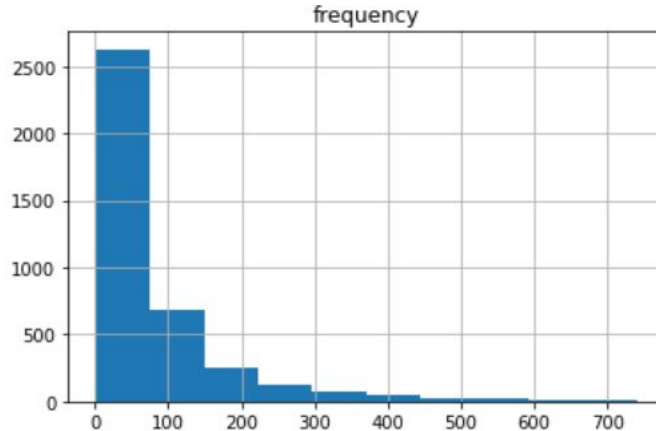
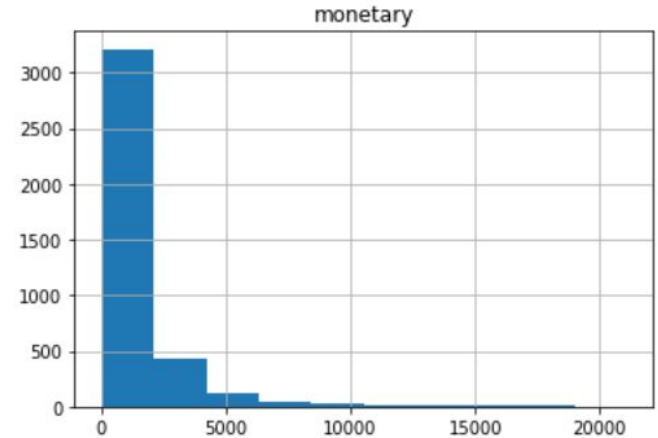
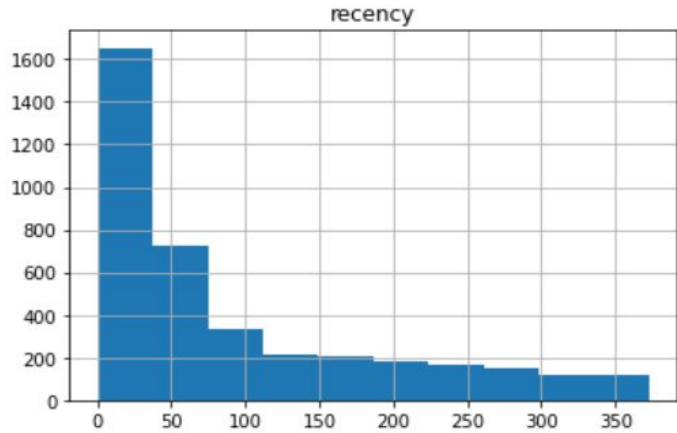
Visuals of RFM Data



Visuals of RFM for Clean Data



Eliminate the outliers in the dataset then the histogram diagrams are looks like



RFM Score Calculation



	recency	frequency	monetary	R	F	M	RFM Score
CustomerID							
12346.0	325	1	77183	1	1	5	115
12747.0	2	103	4196	5	4	5	545
12748.0	0	4596	33719	5	5	5	555
12749.0	3	199	4090	5	5	5	555
12820.0	3	59	942	5	4	4	544

Segment With RFM Score



CustomerID	recency	frequency	monetary	R	F	M	RFM Score	Segment
12346.0	325	1	77183	1	1	5	115	at risk
12747.0	2	103	4196	5	4	5	545	champions
12748.0	0	4596	33719	5	5	5	555	champions
12749.0	3	199	4090	5	5	5	555	champions
12820.0	3	59	942	5	4	4	544	champions

'[1-2] [1-4]': 'at risk',
'[1-2]5': 'can't lose',
'3[1-3]': 'needs attention',
'[3-4] [4-5]': 'loyal customers',
'[4-5]1': 'new customers',
'[4-5] [2-5]': 'champions'.

Model Evaluation(K-means vs Hierarchical)

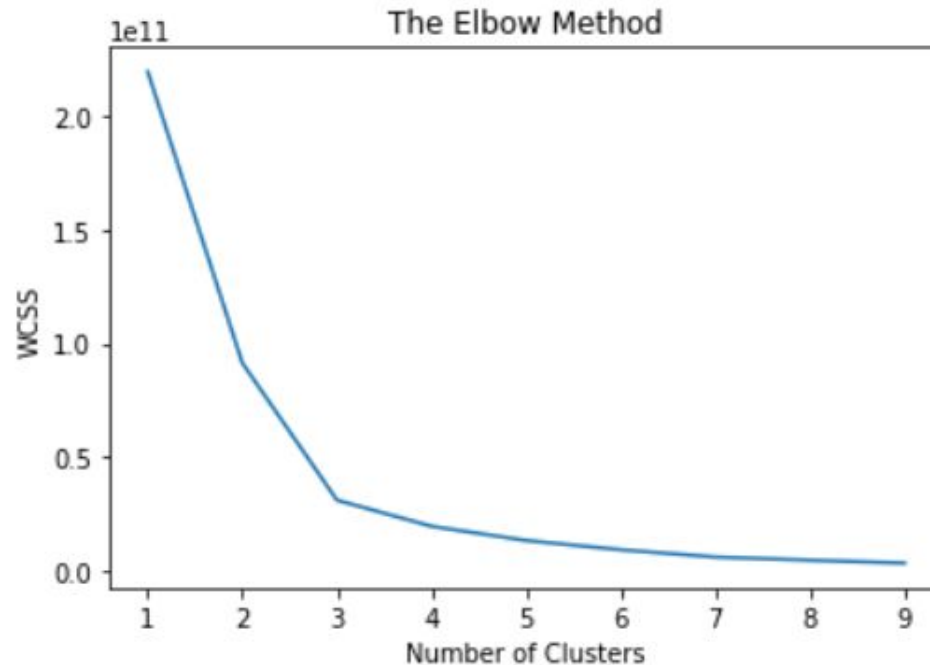


	model	n_clusters	s_score	db_score
0	KMeans	2	0.594541	0.582639
0	Hier	2	0.581698	0.559888
1	KMeans	3	0.588156	0.649014
1	Hier	3	0.575753	0.636482
2	KMeans	4	0.551103	0.642604
2	Hier	4	0.466708	0.738229
3	KMeans	5	0.482530	0.728177
3	Hier	5	0.464246	0.690028
4	KMeans	6	0.499492	0.669170
4	Hier	6	0.421601	0.707550
5	KMeans	7	0.476276	0.733422
5	Hier	7	0.425663	0.752059

- Silhouette Coefficient score is a metric used to compute the goodness of a clustering technique. Its score varies from -1 to 1
- Davies-Bouldin index, which is a validation metric that is sometimes used in order to evaluate the minimum number of clusters to use.
- from n_clusters = 6 to 7, K-means outperforms at n_clusters = 6.
- K-means has a linear time complexity $O(n)$
- Hierarchical which has a quadratic complexity - $O(n^2)$

**With a large number of variables,
K-means compute faster!!!!**

Cluster Analysis Using K-means Clustering

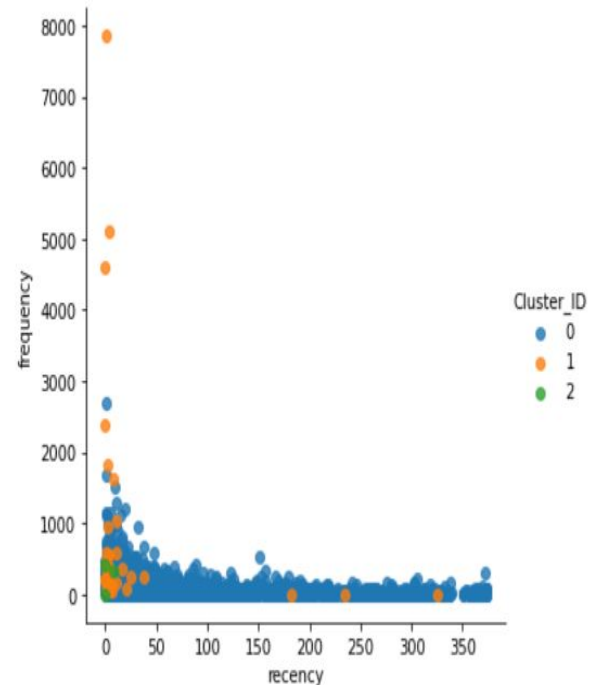
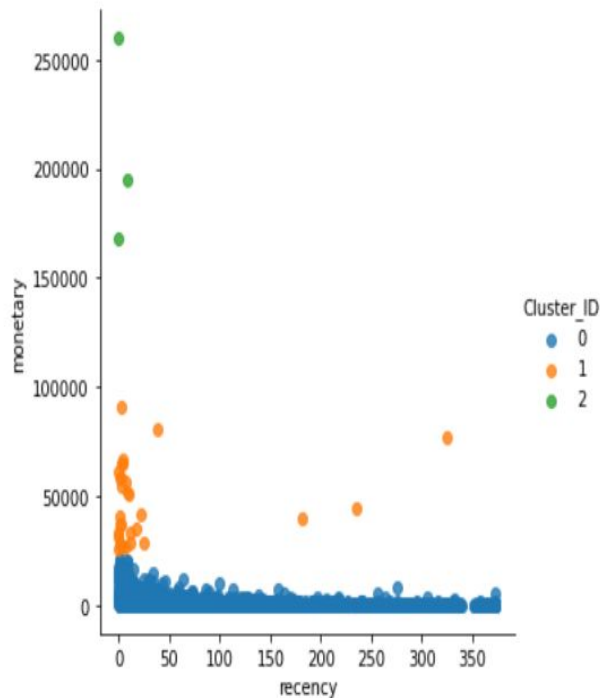
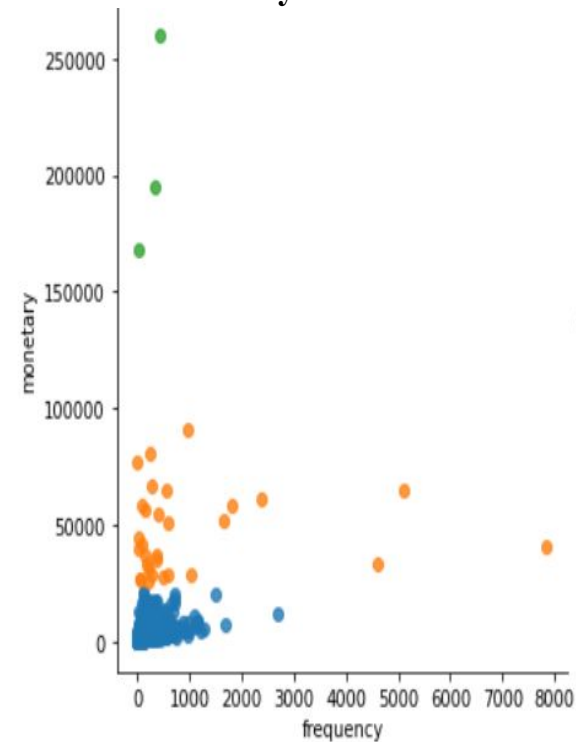


In the following figure we fit K means to the dataset using $K=3$, that means the dataset is segmenting by 3 cluster

Cluster Analysis Using K-means Clustering(Cont.)



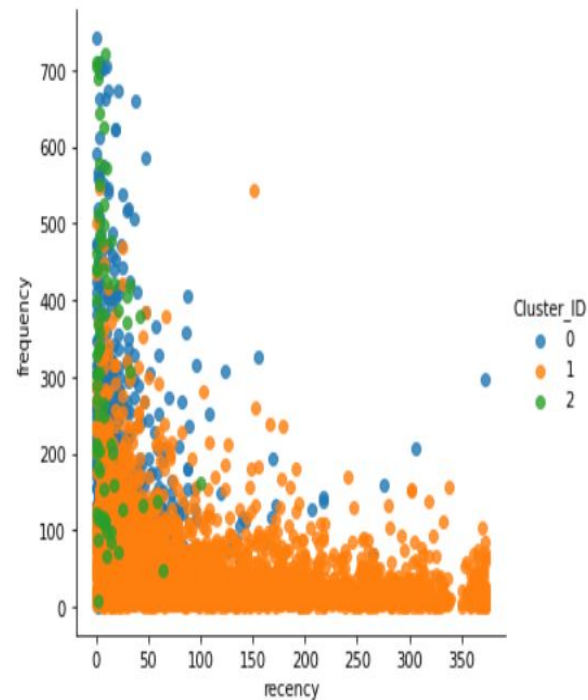
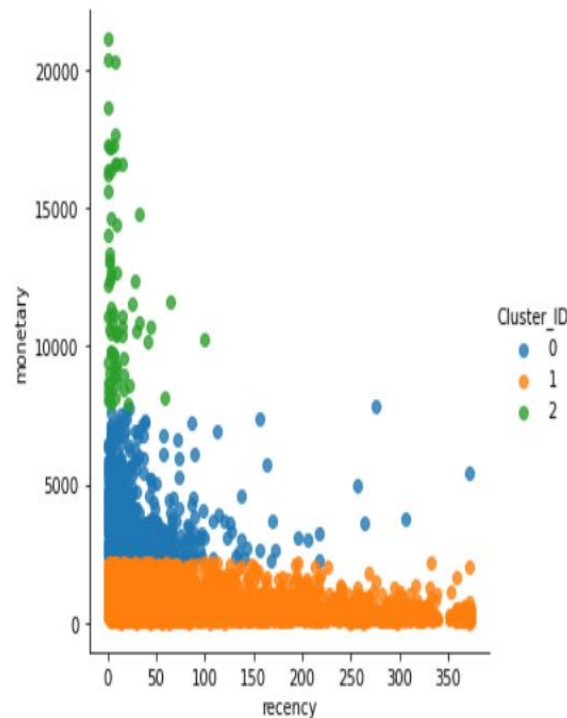
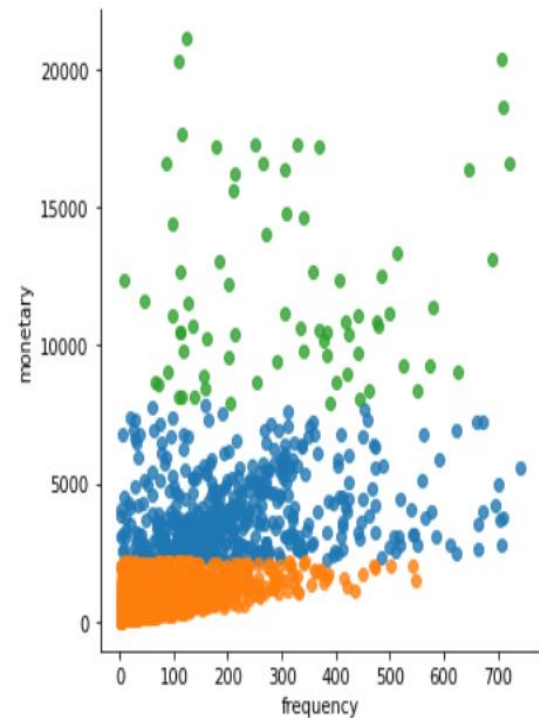
Cluster analysis with outliers



Cluster Analysis Using K-means Clustering(Cont.)



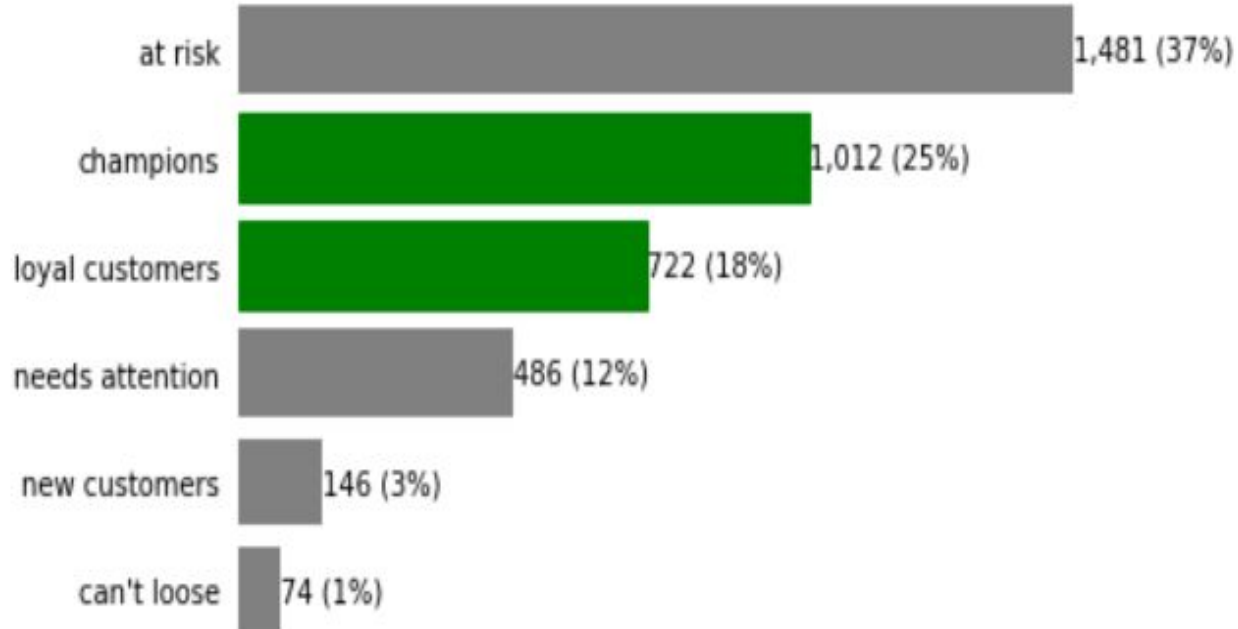
Cluster analysis without outliers





Result Analysis

Customer Segmentation Analysis



Customer Segmentation Analysis



Best Customers

	recency	frequency	monetary	R	F	M	RFM Score	Segment
CustomerID								
18102.0	0	431	259657	5	5	5	555	champions
17450.0	8	337	194550	5	5	5	555	champions
17511.0	2	963	91062	5	5	5	555	champions
16684.0	4	277	66653	5	5	5	555	champions
14096.0	4	5111	65164	5	5	5	555	champions

Churn Customers

	recency	frequency	monetary	R	F	M	RFM Score	Segment
CustomerID								
12346.0	325	1	77183	1	1	5	115	at risk
15749.0	235	10	44534	1	1	5	115	at risk
15098.0	182	3	39916	1	1	5	115	at risk
13093.0	275	159	7832	1	5	5	155	can't loose
17850.0	372	297	5391	1	5	5	155	can't loose

Lose Customers

	recency	frequency	monetary	R	F	M	RFM Score	Segment
CustomerID								
13747.0	373	1	79	1	1	1	111	at risk
14237.0	373	9	161	1	1	1	111	at risk
17643.0	373	8	101	1	1	1	111	at risk
15350.0	373	5	115	1	1	1	111	at risk
13011.0	372	3	50	1	1	1	111	at risk

Loyal Customer

	recency	frequency	monetary	R	F	M	RFM Score	Segment
CustomerID								
18102.0	0	431	259657	5	5	5	555	champions
17450.0	8	337	194550	5	5	5	555	champions
17511.0	2	963	91062	5	5	5	555	champions
16029.0	38	242	81024	3	5	5	355	loyal customers
16684.0	4	277	66653	5	5	5	555	champions

Product Analysis



Most Selling Product

WHITE HANGING HEART T-LIGHT HOLDER	2369
REGENCY CAKESTAND 3 TIER	2200
JUMBO BAG RED RETROSPOT	2159
PARTY BUNTING	1727
LUNCH BAG RED RETROSPOT	1638
...	
damages/showroom etc	1
PINK POLKADOT KIDS BAG	1
Sale error	1
CAMOUFLAGE DOG COLLAR	1
72 CAKE CASES VINTAGE CHRISTMAS	1

Recommendation for Business Decision

R	F	Category	Activity	Recommendation
1-2	1-4	At Risk	Customers who have not bought any product for a long time	Revive interest with a reach out campaign, ignore otherwise.
1-2	5	Can't Lose	Precious customers who don't continue transacting for a long term in the past	Offer other relevant products and special discounts. Recreate brand value It's needed to communicate with them on the idea in their specific choices, as acknowledged from earlier transaction data.
3	1-3	Needs Attention	Above average recency, frequency and monetary values. May not have bought it very recently though.	Make limited time offers, Recommend based on past purchases. Reactivate them.
3-4	4-5	Loyal Customers	Loyal Customers are those customers whose frequency is high. Spend good money with us often. Responsive to promotions.	Upsell higher value products. Ask for reviews. Engage them. Reward them & more promotional offer for them
4-5	1	New Customers	Customers who spent plenty on their first buy	Like with the best clients group, it's essential to lead them to experience valued and liked – and to present them incredible incentives to maintain interacting with the brand
4-5	2-5	Champions	The best customers are the ones with the more rfm value. Bought recently, buy often and spend the most!	Reading their individual choices and affinities will offer extra benefits for even greater personalized messaging Reward them. Can be early adopters for new products. Will promote your brand.

Future Work



- Besides customer segmentation there is a huge opportunity in product segmentation as well.
- Product segmentation can help an organisation in their business strategy and accelerate their goals one step further similarly as customer segmentation.
- It will identify the organization which product is better selling and which we need to focus more.
- Make Smarter business system.



Mapping

Ks are addressed through the project and mapping among Ks,COS and POS

Ks	Attribute	How ks are addressed through the project	CO	PO
K2	Mathematics	Details and basic of Knowledge Statistics	CO-1	PO-a
K3	Engineering Fundamental	Knowledge about Machine Learning, Different types of Learning, Machine Learning Algorithms, Data Analysis Tool, and Python Language, and different types of Framework.	CO-1 CO-2	PO-a PO-b PO-c
K4	Specialist Knowledge	Developing Dataset, Data Cleaning, Data Processing, Data Normalization, Feature Extraction, Design, Train & Test.	CO-1 CO-2	PO-a PO-b PO-c PO-e
K6	Engineering Practice	Knowledge of programming language python, knowledge of using library, Machine learning based model, Understanding data, idea of large data set, classes of machine learning problem	CO-1 CO-2	PO-a PO-c PO-e
K8	Research Literature	The research requires a detailed study of the related research field and other sources and documentation	CO-1 CO-5	PO-a PO-d PO-h


Ps are addressed through the project and mapping among Ps, COs, and POs

Ps	Attribute	How Ps are addressed through the project	CO	PO
P1	Depth of Knowledge Requirement	Basic & advance statistics knowledge(K2) Project requires study of research on Data Science, Data Analysis & Machine Learning Algorithms (K8) Data collection from Online shop, super shop, e-commerce site (Facebook page) (K3, K4) knowledge of using Library, Machine learning Based Model, Understanding Data, Idea of Large Data Set, Classes of Machine Learning Problem (K6)	CO-1 CO-2 CO-8	PO-a PO-b PO-c PO-d PO-j PO-l
P2	Range of Conflicting Requirements	Developing a practical machine learning model with proper regularization with low variance while limited given data will be used. If the learning process is unsupervised or reinforcement rather than supervised learning, then accurate segmentation will be commuted	CO-2 CO-4	PO-a PO-c PO-g
P3	Depth of Analysis Required	A huge algorithm can be adopted but choice of the selected algorithm requires in-detail and depth analysis	CO-1 CO-2	PO-a PO-b PO-d PO-l

Ps are addressed through the project and mapping among Ps, COs, and POs

Ps	Attribute	How Ps are addressed through the project	CO	PO
P4	Familiarity of Issues	CSE graduates are not typically familiar with customer management, business analytics and business policy.	CO-5	PO-f
P5	Extent of Applicable Codes	We maintained user privacy carefully as well as took other ethical approaches and used open-licensed tools to develop the system.	CO-5 CO-6	PO-f PO-h PO-i
P6	Diverse Groups	People of all ages and classes are involved specially the more loyal people	CO-6	PO-i
P7	Interdependence	Research involves a number of sub-system like Data Collection, Training Dataset, Machine Learning Algorithms, Data Analysis, Data Processing.	CO-3 CO-6 CO-8	PO-c PO-i PO-j PO-k

Addressing Complex Activities (As) through the work

As	Attribute	How As are addressed through the project
A1 	Range of Resources	The project requires the use of diverse resources including different types of materials , Information's : dataset (test & training), Dataset (Link: https://archive.ics.uci.edu/ml/datasets/Online+Retail+II) People : (Members: Tanveer Ahamed Rabby,Md.Efti Khirul Alam,Sharmin Akter)
A2	Level of interaction	The level of interaction between the group members has been varied when it comes to making the dataset in our model. By using data analysis & Machine learning algorithms to segment customers from a large dataset for a particular company.
A4	Consequences for society and the environment	By segmenting customers, it can be easy to understand the divergence between loyal customers and so-called customers. If we can segment the customer which contains some features, it can be easier for a businessman or a company to realize the loyal and targeted customer which will be a more efficient way for business.
A5	Familiarity	The project deals with data analysis and machine learning algorithms for segmenting customer and market basket analysis for learners.

ANY
QUESTIONS



THANK
YOU