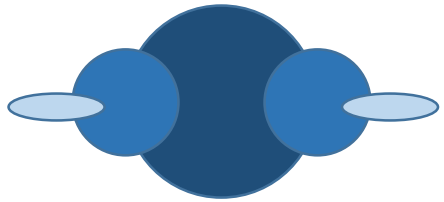


Customer Segmentation (Online Retail)

Moh. Ridwan

Summary

Business Background

A company X mainly sells unique all-occasion gift-ware. Many customers of the company are wholesalers

Problems Statements

How to deal with customers on a group basis to maximize each customer's value to the business?

Objective

Create a consumer segmentation

Proposed Solutions

Company x can use the RFM model to segment consumers

Result:

RFM Model Using Quartiles:

There are 11 customer segments consisting of champion (1151), loyal (658), potential loyal (653), new customers (343), promising (172), about to sleep (304), needs attention (339), can't lose them (104), at risk (516), hibernating (1020), and lost (662)

RFM Model Using Kmeans:

There are 2 customer segments consisting of champion (22) and loyal (5856)

Business Benefit

- Company X can determine consumer groups based on their activities so as to make treatment based on consumer groups to increase business activities

Data Overview

Data

This Online Retail II data set contains all the transactions occurring for a UK-based and registered, non-store online retail between 01/12/2009 and 09/12/2011.The company mainly sells unique all-occasion gift-ware. Many customers of the company are wholesalers

Attribute Information:

InvoiceNo: Invoice number. Nominal. A 6-digit integral number uniquely assigned to each transaction. If this code starts with the letter 'c', it indicates a cancellation.
StockCode: Product (item) code. Nominal. A 5-digit integral number uniquely assigned to each distinct product.
Description: Product (item) name. Nominal.
Quantity: The quantities of each product (item) per transaction. Numeric.
InvoiceDate: Invice date and time. Numeric. The day and time when a transaction was generated.
UnitPrice: Unit price. Numeric. Product price per unit in sterling (Â£).
CustomerID: Customer number. Nominal. A 5-digit integral number uniquely assigned to each customer.
Country: Country name. Nominal. The name of the country where a customer resides.

Data can be downloaded via this link : https://archive.ics.uci.edu/ml/datasets/Online+Retail+II#

Data Overview

Data

In the downloaded folder, there are two dataset files in the form of .xlsx so they need to be merged into one. Here's a look at the top 5 data

	Invoice	StockCode	Description	Quantity	InvoiceDate	Price	Customer ID	Country
0	489434	85048	15CM CHRISTMAS GLASS BALL 20 LIGHTS	12	2009-12-01 07:45:00	6.95	13085.0	United Kingdom
1	489434	79323P	PINK CHERRY LIGHTS	12	2009-12-01 07:45:00	6.75	13085.0	United Kingdom
2	489434	79323W	WHITE CHERRY LIGHTS	12	2009-12-01 07:45:00	6.75	13085.0	United Kingdom
3	489434	22041	RECORD FRAME 7" SINGLE SIZE	48	2009-12-01 07:45:00	2.10	13085.0	United Kingdom
4	489434	21232	STRAWBERRY CERAMIC TRINKET BOX	24	2009-12-01 07:45:00	1.25	13085.0	United Kingdom

```
Int64Index: 1067371 entries, 0 to 541909
Data columns (total 8 columns):
 #   Column       Non-Null Count    Dtype
---  ------       --------------    -----
 0   Invoice      1067371 non-null  object
 1   StockCode    1067371 non-null  object
 2   Description  1062989 non-null  object
 3   Quantity     1067371 non-null  int64
 4   InvoiceDate  1067371 non-null  datetime64[ns]
 5   Price        1067371 non-null  float64
 6   Customer ID  824364 non-null   float64
 7   Country      1067371 non-null  object
```

We will take data with a customer id column that is not empty because it is in line with the main goal, which is to form consumer segmentation

Additional data in the form of regional data for EDA . Here's a look at the top 5 data.
Link: https://www.kaggle.com/tomvebrcz/countriesandcontinents

	Country	Region	Population
0	Angola	AFRICA	12127071
1	Benin	AFRICA	7862944
2	Botswana	AFRICA	1639833
3	Burkina Faso	AFRICA	13902972
4	Burundi	AFRICA	8090068

```
RangeIndex: 227 entries, 0 to 226
Data columns (total 3 columns):
 #   Column      Non-Null Count  Dtype
---  ------      --------------  -----
 0   Country     227 non-null    object
 1   Region      227 non-null    object
 2   Population  227 non-null    int64
dtypes: int64(1), object(2)
```

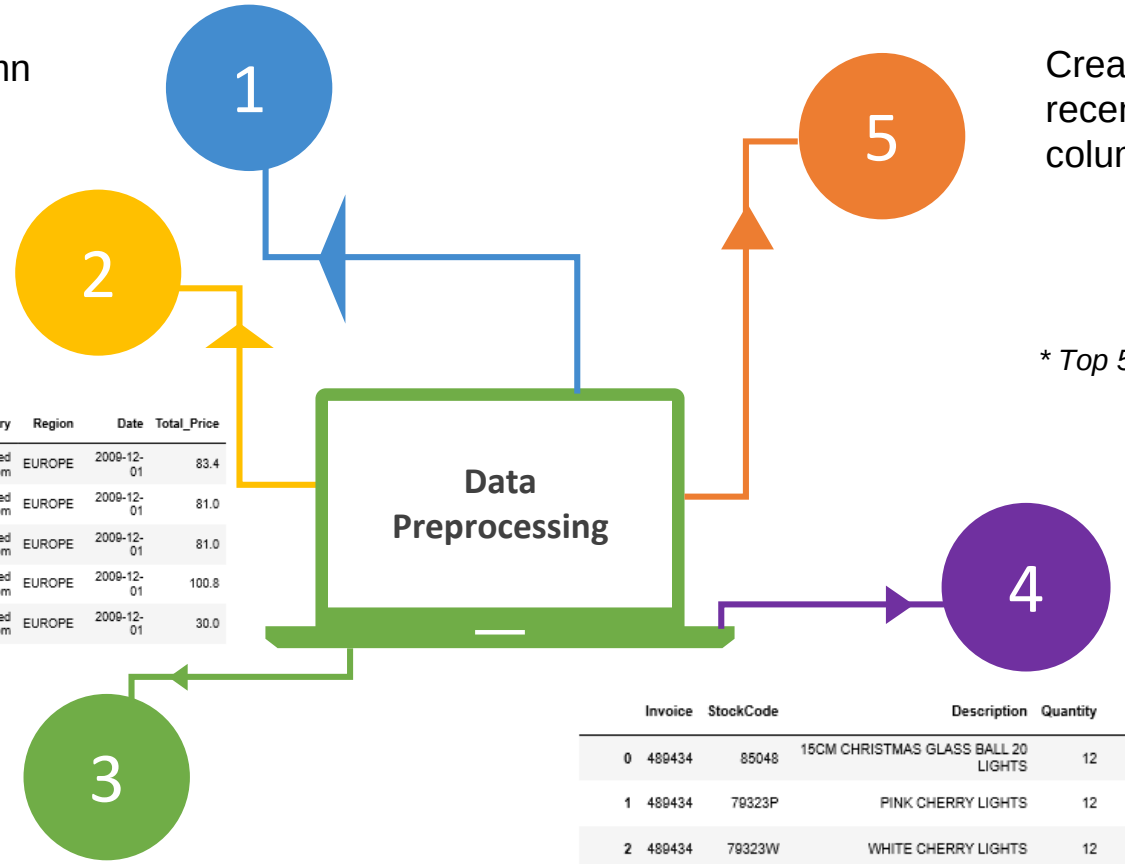
Data Preprocessing

Delete empty data in the customer id column

Create date, region, and total price columns for EDA

	Invoice	StockCode	Description	Quantity	InvoiceDate	Price	Customer ID	Country	Region	Date	Total_Price
0	489434	85048	15CM CHRISTMAS GLASS BALL 20 LIGHTS	12	2009-12-01 07:45:00	6.95	13085.0	United Kingdom	EUROPE	2009-12-01	83.4
1	489434	79323P	PINK CHERRY LIGHTS	12	2009-12-01 07:45:00	6.75	13085.0	United Kingdom	EUROPE	2009-12-01	81.0
2	489434	79323W	WHITE CHERRY LIGHTS	12	2009-12-01 07:45:00	6.75	13085.0	United Kingdom	EUROPE	2009-12-01	81.0
3	489434	22041	RECORD FRAME 7" SINGLE SIZE	48	2009-12-01 07:45:00	2.10	13085.0	United Kingdom	EUROPE	2009-12-01	100.8
4	489434	21232	STRAWBERRY CERAMIC TRINKET BOX	24	2009-12-01 07:45:00	1.25	13085.0	United Kingdom	EUROPE	2009-12-01	30.0

** Top 5 data*



Retrieve data with non-cancelled orders

Create a dataframe (rfm) consisting of recency, frequency and monetary columns

Customer ID	recency	frequency	monetary
12346.0	326	12	77556.46
12347.0	3	8	5633.32
12348.0	76	5	2019.40
12349.0	19	4	4428.69
12350.0	311	1	334.40

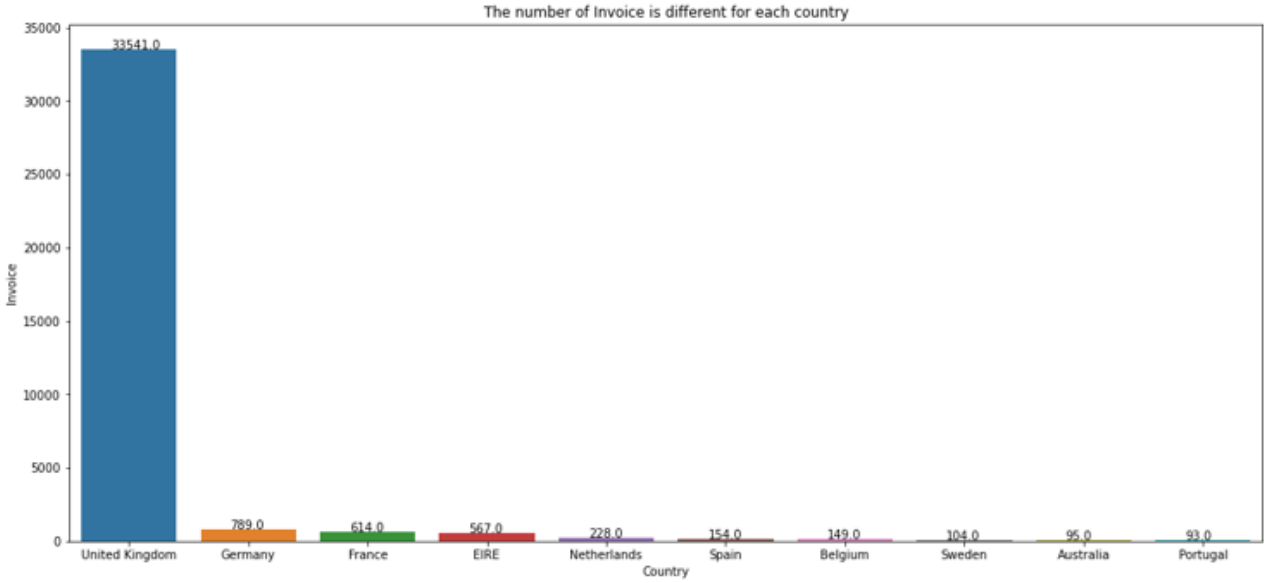
** Top 5 data*

Retrieve data whose total price is more than 0

	Invoice	StockCode	Description	Quantity	InvoiceDate	Price	Customer ID	Country	Region	Date	Total_Price
0	489434	85048	15CM CHRISTMAS GLASS BALL 20 LIGHTS	12	2009-12-01 07:45:00	6.95	13085.0	United Kingdom	EUROPE	2009-12-01	83.40
1	489434	79323P	PINK CHERRY LIGHTS	12	2009-12-01 07:45:00	6.75	13085.0	United Kingdom	EUROPE	2009-12-01	81.00
2	489434	79323W	WHITE CHERRY LIGHTS	12	2009-12-01 07:45:00	6.75	13085.0	United Kingdom	EUROPE	2009-12-01	81.00
3	489434	22041	RECORD FRAME 7" SINGLE SIZE	48	2009-12-01 07:45:00	2.10	13085.0	United Kingdom	EUROPE	2009-12-01	100.80
4	489434	21232	STRAWBERRY CERAMIC TRINKET BOX	24	2009-12-01 07:45:00	1.25	13085.0	United Kingdom	EUROPE	2009-12-01	30.00
...	...	...	...	...	...	...	...	...	...	...	...
541905	581587	22899	CHILDREN'S APRON DOLLY GIRL	6	2011-12-09 12:50:00	2.10	12680.0	France	EUROPE	2011-12-09	12.60
541906	581587	23254	CHILDRENS CUTLERY DOLLY GIRL	4	2011-12-09 12:50:00	4.15	12680.0	France	EUROPE	2011-12-09	16.60
541907	581587	23255	CHILDRENS CUTLERY CIRCUS PARADE	4	2011-12-09 12:50:00	4.15	12680.0	France	EUROPE	2011-12-09	16.60
541908	581587	22138	BAKING SET 9 PIECE RETROSPOT	3	2011-12-09 12:50:00	4.95	12680.0	France	EUROPE	2011-12-09	14.85
541909	581587	POST	POSTAGE	1	2011-12-09 12:50:00	18.00	12680.0	France	EUROPE	2011-12-09	18.00

805549 rows × 11 columns

Exploratory Data Analysis



The highest number of invoices at all times is in Europe (UK), this is appropriate because online retailers are based in the UK. The highest number of invoices in November. This is because there are several special events in November, one of which is the Christmas event

Modeling and Evaluation

RFM Model Using Quartiles

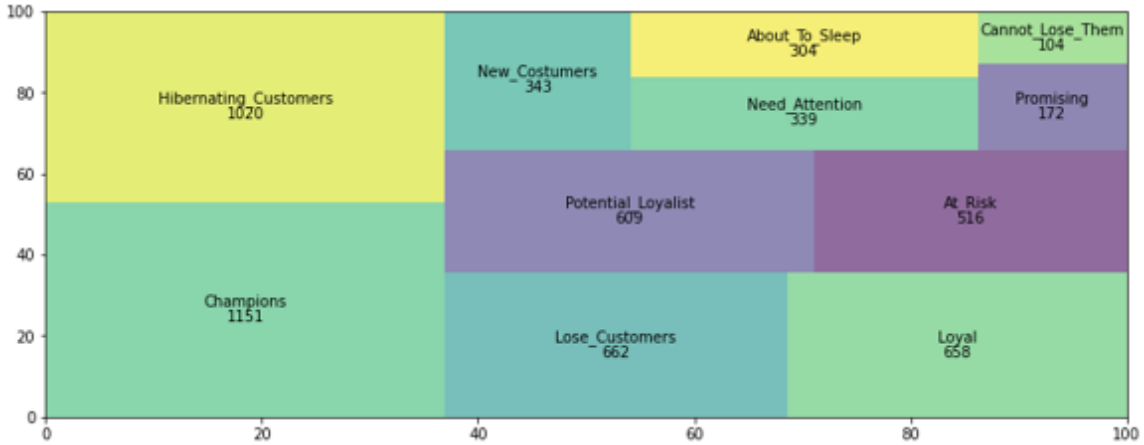
score on recency, frequency, and monetary by quartiles

create 11 segments for consumers

visualization

RFM segmentation categorizes your customers into different segments, according to their interactions with your website, which will allow you to subsequently approach these groups in the most effective way. RFM stands for the three dimensions:

1.Recency – How recently did the customer purchase?
2.Frequency – How often do they purchase?
3.Monetary Value – How much do they spend?



*11 segments on consumers refer to the following sources.
https://docs.exponea.com/docs/rfm-segmentation

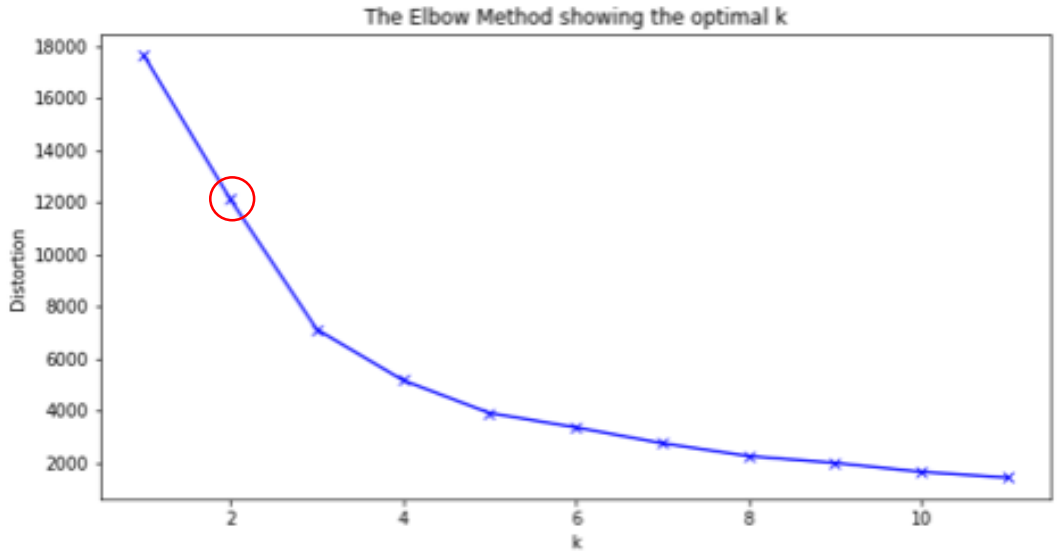
Modeling and Evaluation

RFM Model Using K-Means

Scaling columns

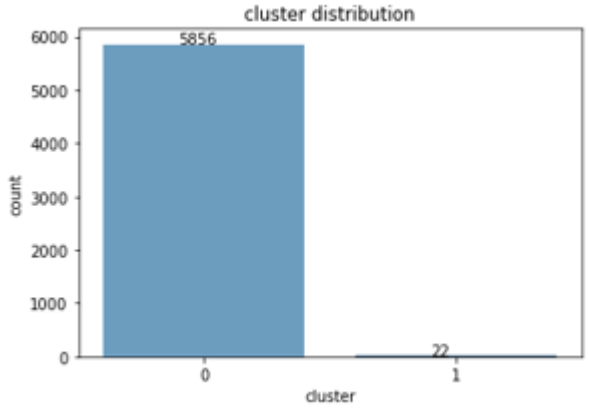
Determine the optimal value of k

Make k means model



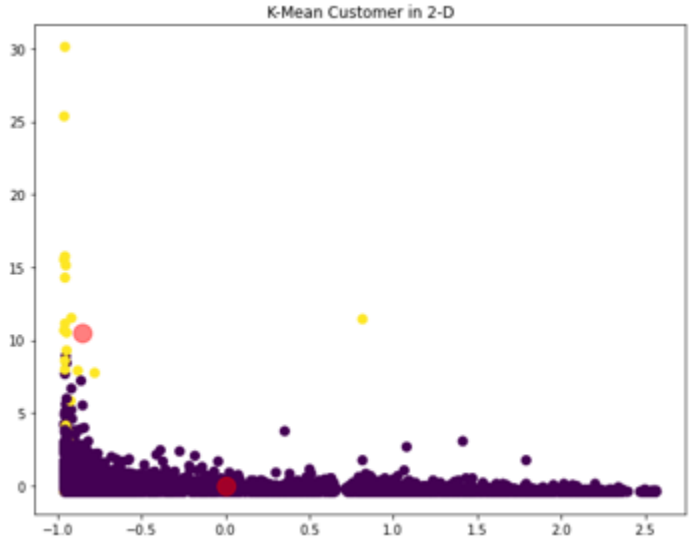
The optimal value of k is:
k=2

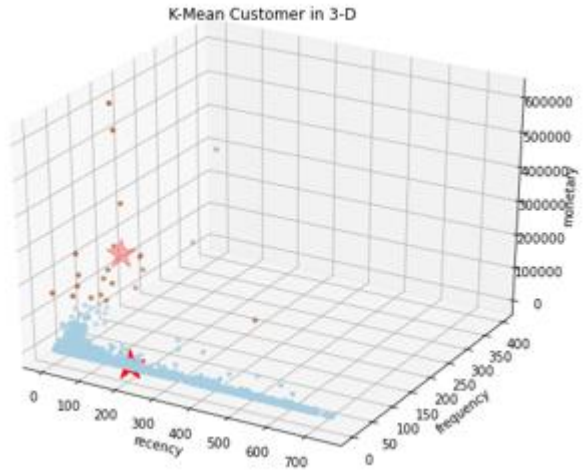
Silhouette Coefficient for kmeans: **0.9163506547916832**



cluster	recency	frequency	monetary
0	202.536714	5.775615	2366.659960
1	23.545455	143.045455	176557.656818

Based on this average, it can be seen that cluster 1 is champions while 0 means not champions





Recommendation

RFM Model Using Quartiles

Segment	Number of Customers	Actionable Tip
Champion	1151	Reward them. Can be early adopters of new products. Will promote your brand. Most likely to send referrals.
Loyal	658	Upsell higher value products. Ask for reviews.
Potential Loyalist	653	Offer membership / loyalty program. Keep them engaged. Offer personalised recommendations.
New Customers	343	Provide on-boarding support, give them early access, start building relationship.
Promising	172	Offer coupons. Bring them back to the platform and keep them engaged. Offer personalised recommendations.
About to sleep	304	Make limited time offers.
Needs attention	339	Make limited time offers. Offer personalised recommendations.
Can’t lose them	104	Win them back via renewals or newer products, don’t lose them to competition. Talk to them if necessary. Spend time on highest possible personalisation.
At Risk	516	Provide helpful resources on the site. Send personalised emails.
Hibernating customers	1020	Make subject lines of emails very personalised. Revive their interest by a specific discount on a specific product.
Lost	662	Revive interest with reach out campaign. Ignore otherwise.

Recommendation

RFM Model Using K-Means

Segment	Number of Customers	Actionable Tip
Champion	22	Reward them. Can be early adopters of new products. Will promote your brand. Most likely to send referrals.
Not Champion	5856	Offer membership, offer coupons, or make limited time offers .

THANK YOU