



Customer Segmentation

Dilla Arta Febriyani

Prologue



There are 541909 rows data in dataset. After cleansing and handling some missing value, we have total transaction 398003 rows, with 389196 transaction above 0 and 8807 below 0.

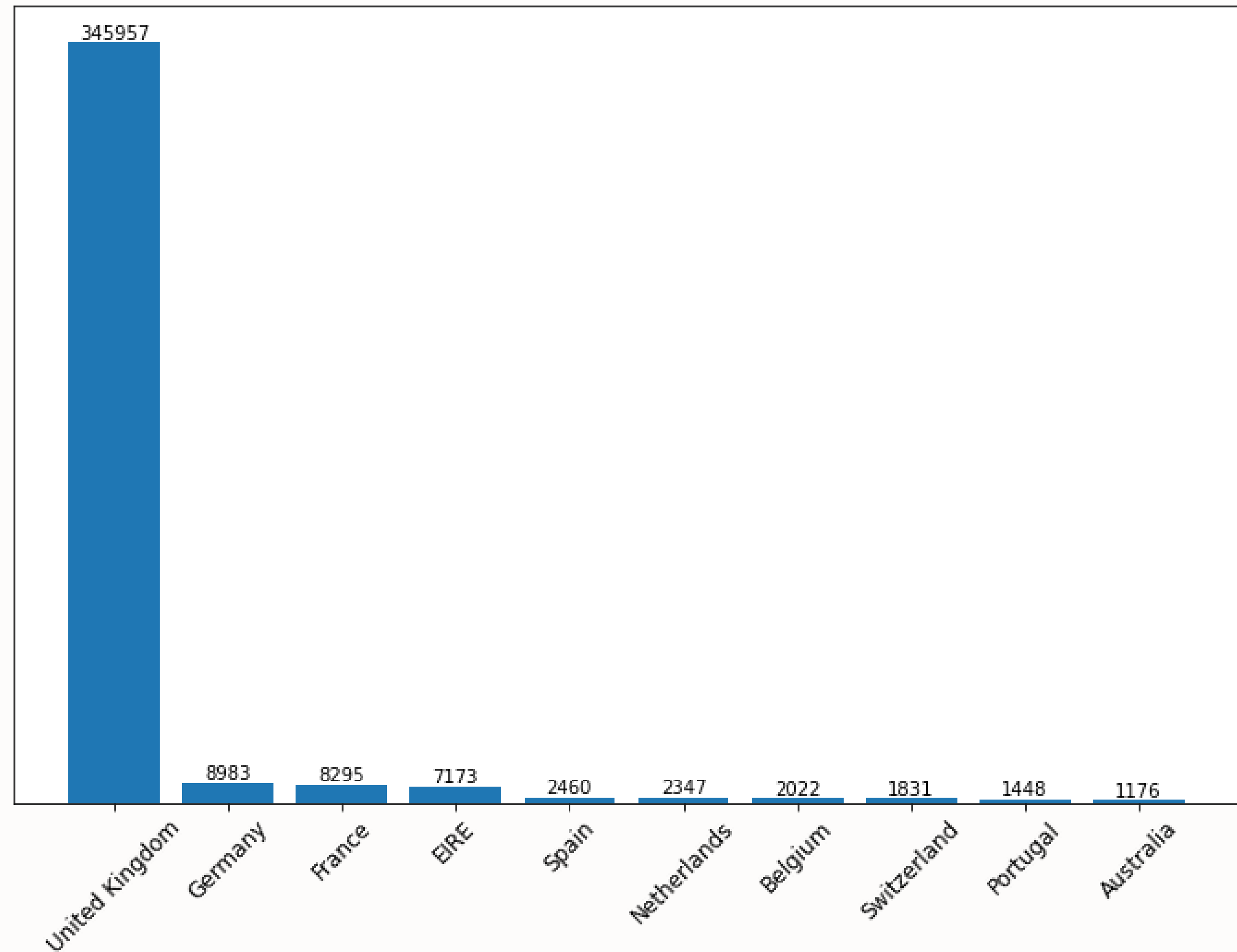
This data consist of 8 columns (Kode_Bayar, Kode_Barang, Barang, Jumlah, Tanggal_Transaksi, Harga_Satuan, Kode_Pelanggan, Negara)



United Kingdom

Country with The Highest Transactions Done

Top 10 Country with Big Transaction

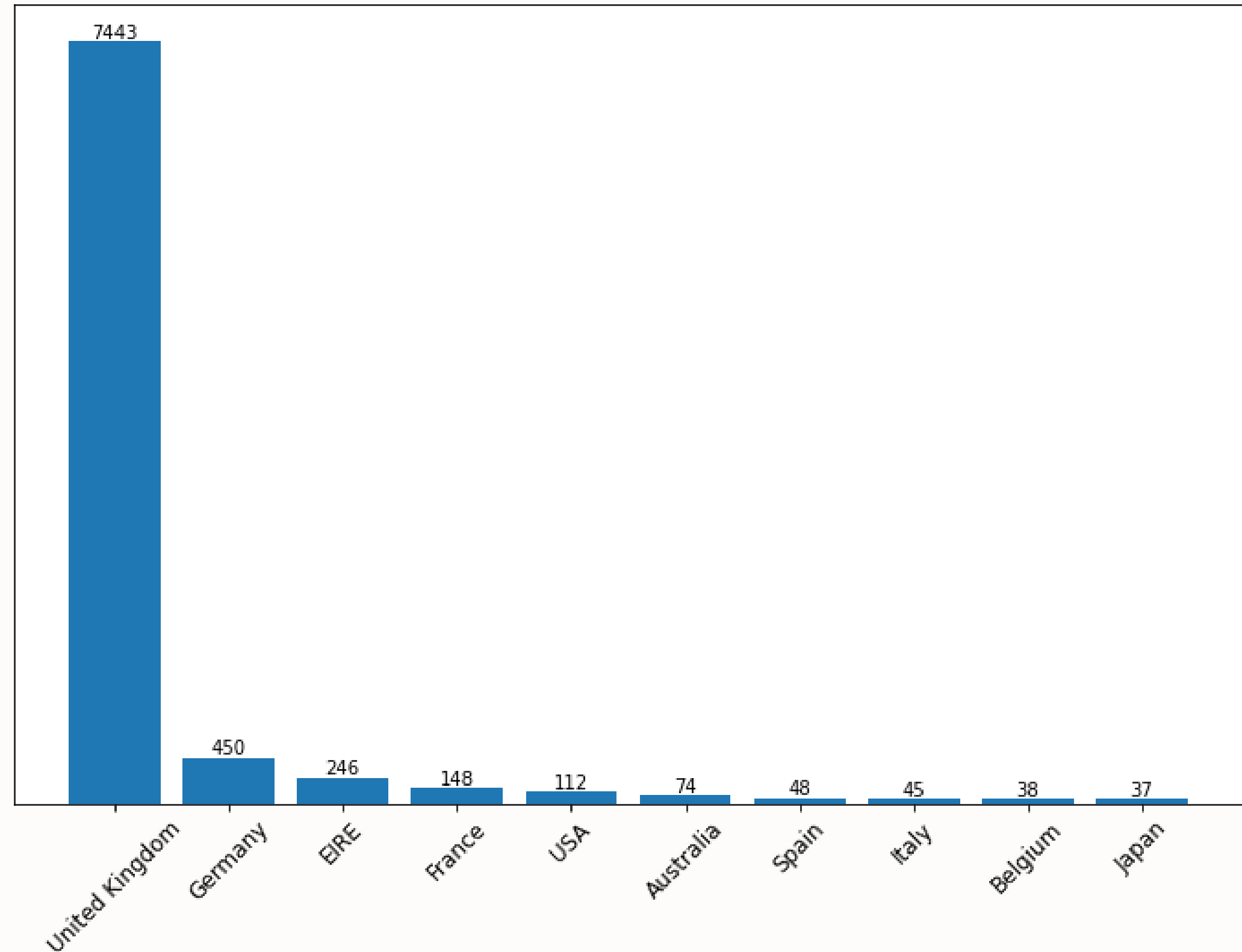




United Kingdom

Country with The Highest Cancelation Transactions Done

Top 10 Country with Big Cancelation Transaction



Best Seller Product



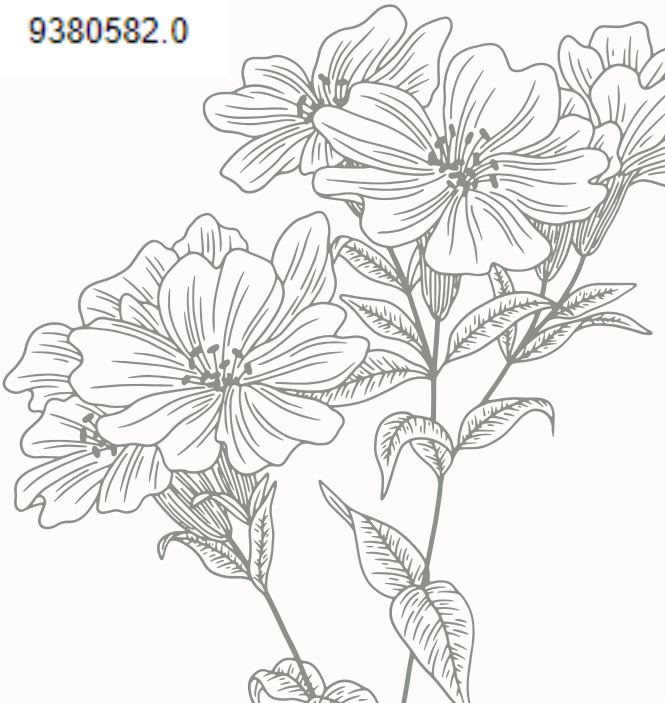
WORLDWIDE

	Jumlah	Harga_Satuan	Kode_Pelanggan
Barang			
WORLD WAR 2 GLIDERS ASSTD DESIGNS	53119.0	139.30	7285096.0
JUMBO BAG RED RETROSPOT	44963.0	3339.80	25348589.0
ASSORTED COLOUR BIRD ORNAMENT	35215.0	2361.49	21400078.0
WHITE HANGING HEART T-LIGHT HOLDER	34128.0	5949.72	32027292.0
PACK OF 72 RETROSPOT CAKE CASES	33386.0	581.93	16022432.0
POPCORN HOLDER	30492.0	579.54	10626441.0
RABBIT NIGHT LIGHT	27045.0	1686.84	12495412.0
MINI PAINT SET VINTAGE	25880.0	219.87	4919126.0
PACK OF 12 LONDON TISSUES	25305.0	128.40	6179923.0
PACK OF 60 PINK PAISLEY CAKE CASES	24129.0	376.06	10463457.0

UNITED KINGDOM

	Jumlah	Harga_Satuan	Kode_Pelanggan
Barang			
WORLD WAR 2 GLIDERS ASSTD DESIGNS	47886.0	125.54	6623614.0
JUMBO BAG RED RETROSPOT	40777.0	3023.72	23289934.0
ASSORTED COLOUR BIRD ORNAMENT	32580.0	2236.42	20412053.0
WHITE HANGING HEART T-LIGHT HOLDER	32079.0	5694.77	30777283.0
POPCORN HOLDER	28550.0	529.32	9805404.0
PACK OF 12 LONDON TISSUES	24297.0	121.80	5927090.0
BROCADE RING PURSE	22672.0	237.63	3225133.0
PACK OF 72 RETROSPOT CAKE CASES	22182.0	482.75	13560076.0
VICTORIAN GLASS HANGING T-LIGHT	21427.0	1161.53	12362915.0
PACK OF 60 PINK PAISLEY CAKE CASES	19882.0	332.94	9380582.0

Top 4 best selling product in the world are mostly purchased from United Kingdom.



Unpopular Product



CANCELATION PRODUCT

	Jumlah	Harga_Satuan	Kode_Pelanggan
Barang			
MEDIUM CERAMIC TOP STORAGE JAR	-74494.0	11.66	142297.0
ROTATING SILVER ANGELS T-LIGHT HLDR	-9367.0	5.13	48937.0
Manual	-3995.0	105725.11	2645143.0
FAIRY CAKE FLANNEL ASSORTED COLOUR	-3150.0	4.99	51599.0
WHITE HANGING HEART T-LIGHT HOLDER	-2578.0	117.90	650949.0
GIN + TONIC DIET METAL SIGN	-2030.0	15.98	101011.0
HERB MARKER BASIL	-1527.0	1.20	31500.0
FELTCRAFT DOLL MOLLY	-1447.0	16.65	89896.0
TEA TIME PARTY BUNTING	-1424.0	25.30	105945.0
PAPER POCKET TRAVELING FAN	-1385.0	0.42	58120.0

LOWEST SELLING PRODUCT

		Kode_Bayar
Barang	Negara	
4 PURPLE FLOCK DINNER CANDLES	EIRE	1
ORANGE VOTIVE CANDLE	France	1
	Sweden	1
ORANGE/WHT/FUSCHIA STONES NECKLACE	United Kingdom	1
ORGANISER WOOD ANTIQUE WHITE	EIRE	1
	Finland	1
	Lebanon	1
ORIGAMI JASMINE INCENSE/CANDLE SET	France	1
ORIGAMI OPIUM INCENSE/CANDLE SET	EIRE	1
ORANGE VOTIVE CANDLE	Belgium	1

These are 10 products that we should put more attention because they are canceled by customer and have lowest number of selling.

To increase sales, we should expand our channel and products to some countries in Europe such as Germany, Eire, France, Spain. We can adding more best selling products in each country and subtract some products that have no buyer.





— "How customer
segmentation in United
Kingdom?" —

Recency, Frequency, Monetary, (RFM)



This analysis is to identify potential customer, based on their behavior on spending money.

- Recency = to know the number of days since last purchases.
- Frequency = to know the number of purchases.
- Monetary = to know how much customers spent their money.

	Kode_Pelanggan	Tanggal_Transaksi	Kode_Bayar	Jumlah	Harga_Satuan	Total_Pembelian
0	17850.0	2010-12-01 08:26:00	536365	6.0	2.55	15.30
1	17850.0	2010-12-01 08:26:00	536365	6.0	3.39	20.34
2	17850.0	2010-12-01 08:26:00	536365	8.0	2.75	22.00
3	17850.0	2010-12-01 08:26:00	536365	6.0	3.39	20.34
4	17850.0	2010-12-01 08:26:00	536365	6.0	3.39	20.34

There are 345957 transactions that we'll identify to know how customer behavior in United Kingdom



Recency, Frequency, Monetary, (RFM)



	recency	frequency	monetary
Kode_Pelanggan			
12346.0	325	1	77183.60
12747.0	2	103	4196.01
12748.0	0	4373	32691.09
12749.0	3	197	4045.58
12820.0	3	59	942.34

Calculate for each customer.

- Recency (Tanggal_Transaksi) = present last purchase date
- Frequency (Kode_Bayar) = transaction's frequency according to transaction ID.
- Monetary (Total_Pembelian) = total purchased amount



Recency, Frequency, Monetary, (RFM)



	recency	frequency	monetary	r_quartile	f_quartile	m_quartile	Nilai_RFM
Kode_Pelanggan							
12346.0	325	1	77183.60	4	4	1	441
12747.0	2	103	4196.01	1	1	1	111
12748.0	0	4373	32691.09	1	1	1	111
12749.0	3	197	4045.58	1	1	1	111
12820.0	3	59	942.34	1	2	2	122

Comput quantile of RFM values.

Top customer is when they have low recency, high frequency, and high monetary (111)

- Recency value has quantile 1 (low) - 4 (high)
- Frequency value has quantile 4 (low) - 1 (high)
- Monetary value has quantile 4 (low) - 1 (high)



Recency, Frequency, Monetary, (RFM)



There are 410 customers that have low recency, high frequency, and high monetary (111). This customer class need promo or discount to keep them coming back again.

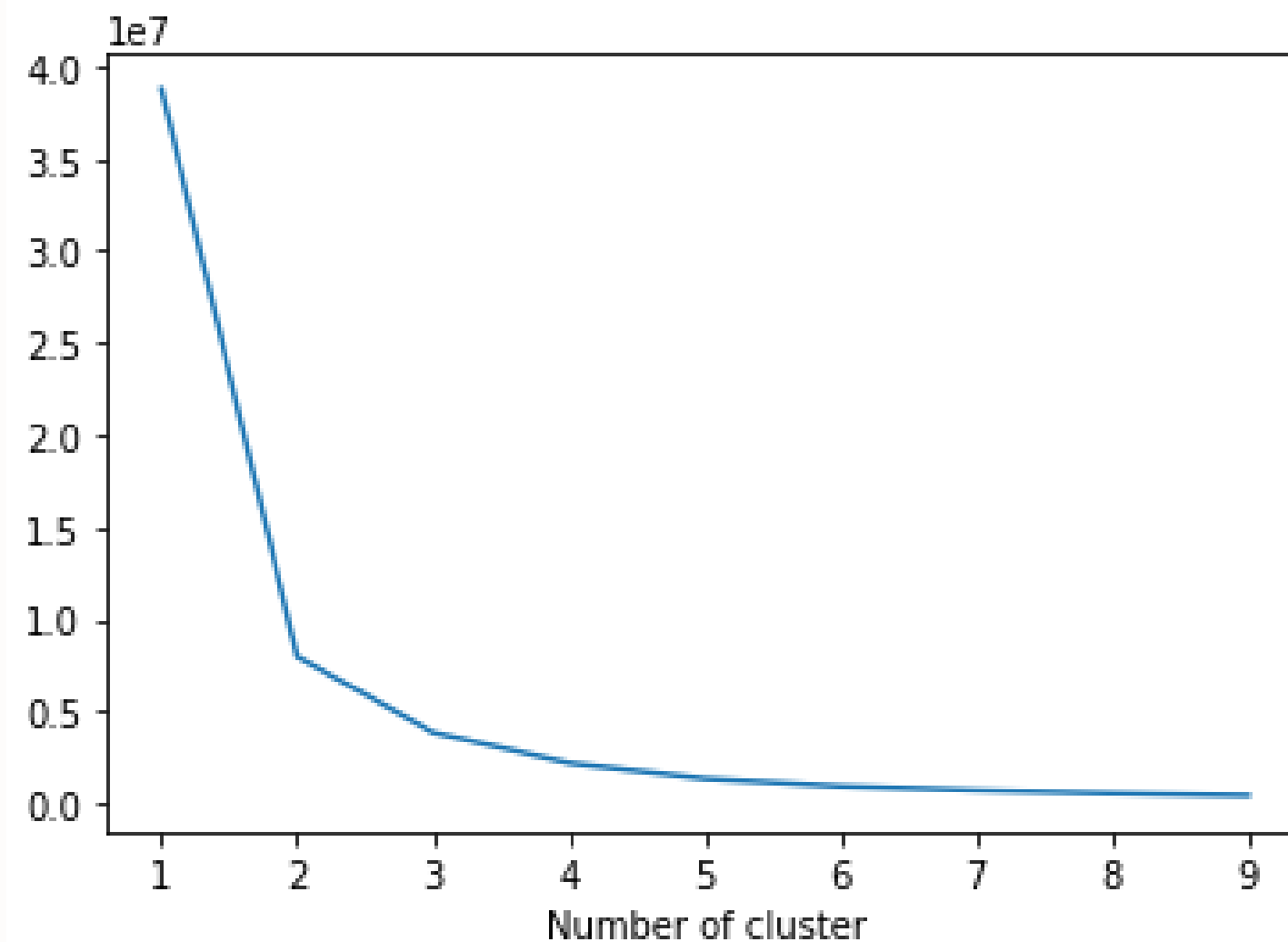
In other note, there are 329 customers that have high recency, low frequency, and low monetary (444). They need promo, discount, or more products variation suggestion to increase their transaction volume.



K-Means



K-Means is one of unsupervised algorithm machine learning clustering. Similar with RFM, but we should define how many cluster that we need to assign. We can identify it with Elbow Method, that will show the optimal cluster.



In this case we assign recency value into Elbow Method and will use 4 cluster. Actually we get optimal cluster is 3/4 (take the point in the end of steep line)



K-Means



We assign recency, frequency, and monetary or revenue value into 4 cluster and here are some statistic values of them.

RECENCY

	count	mean	std	min	25%	50%	75%	max
RecencyCluster								
0	441.0	307.560091	39.478206	252.0	273.0	303.0	336.0	373.0
1	552.0	193.090580	31.397096	140.0	168.0	191.0	218.0	250.0
2	930.0	83.658065	24.391304	52.0	64.0	77.0	100.0	138.0
3	1996.0	19.848697	14.371622	0.0	8.0	17.0	30.0	51.0

FREQUENCY

	count	mean	std	min	25%	50%	75%	max
FrequencyCluster								
0	3464.0	47.759238	42.869037	1.0	14.0	32.0	70.0	180.0
1	429.0	315.174825	127.102393	183.0	214.0	274.0	379.0	733.0
2	23.0	1231.695652	469.800193	780.0	932.0	1086.0	1359.0	2564.0
3	3.0	5660.000000	1699.141842	4373.0	4697.0	5021.0	6303.5	7586.0

MONETARY

	count	mean	std	min	25%	50%	75%	max
RevenueCluster								
0	3743.0	1027.883346	1112.371459	0.00	283.2800	602.510	1371.9900	5727.89
1	152.0	10508.622368	5429.470721	5756.89	6745.1925	8550.855	11791.5425	30400.88
2	22.0	53339.602727	16258.897510	32691.09	39906.7725	52162.995	63523.7175	90927.50
3	2.0	227024.045000	46150.391805	194390.79	210707.4175	227024.045	243340.6725	259657.30

K-Means



We sum up all score into overall score table and divide it into 3 segments, low value, mid value, and high value.

MEAN

	recency	frequency	monetary
OverallScore			
0	307.450801	21.915332	344.414577
1	194.124542	31.450549	522.167749
2	84.826816	44.549721	941.670315
3	23.525552	66.088959	1209.088840
4	15.073846	276.507692	3919.550738
5	9.854369	415.766990	11994.026117
6	7.631579	919.578947	27229.688421
7	3.333333	1240.666667	118937.656667
8	1.666667	5660.000000	45718.370000

Low value

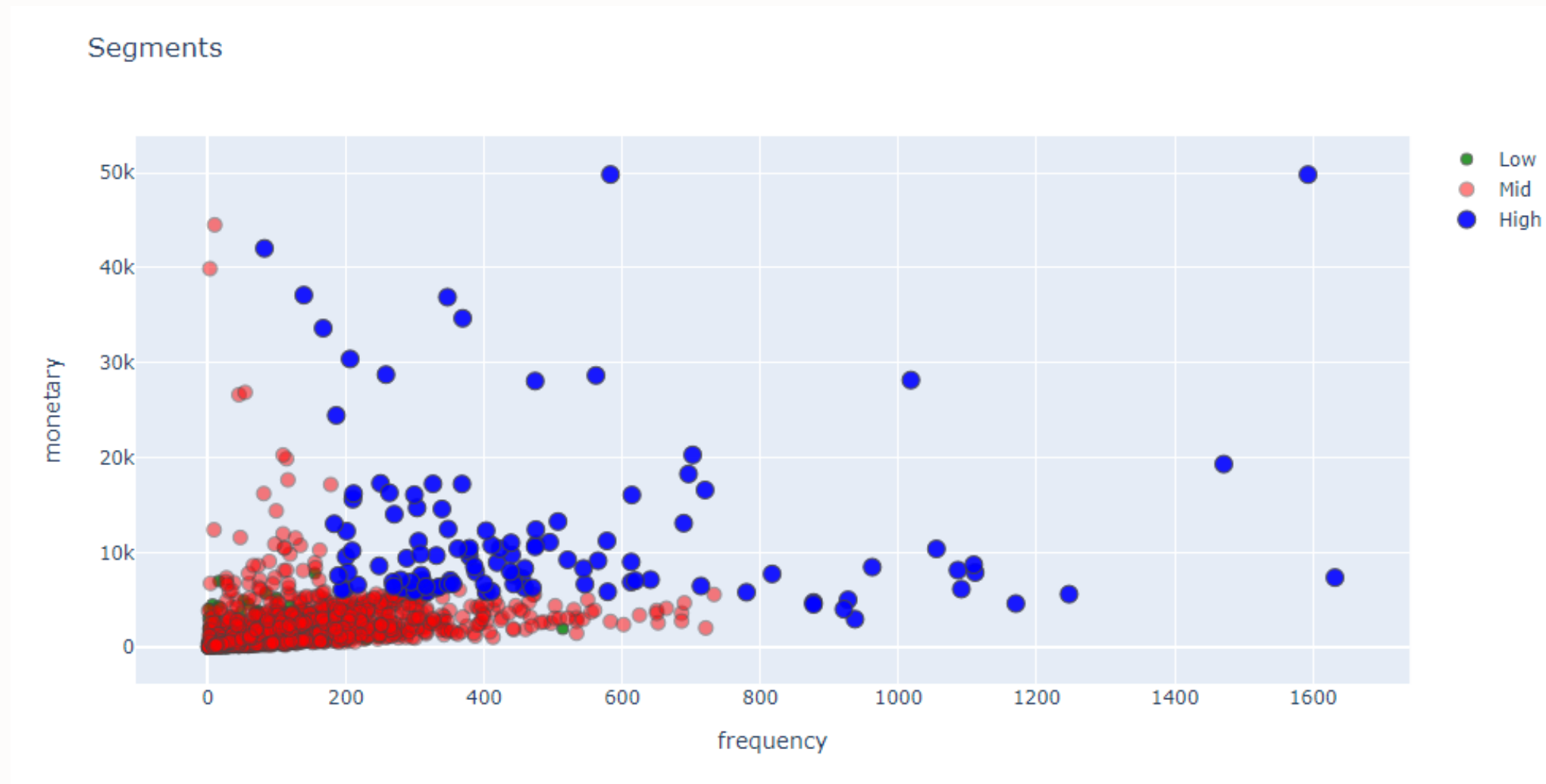
Mid value

High value

COUNT

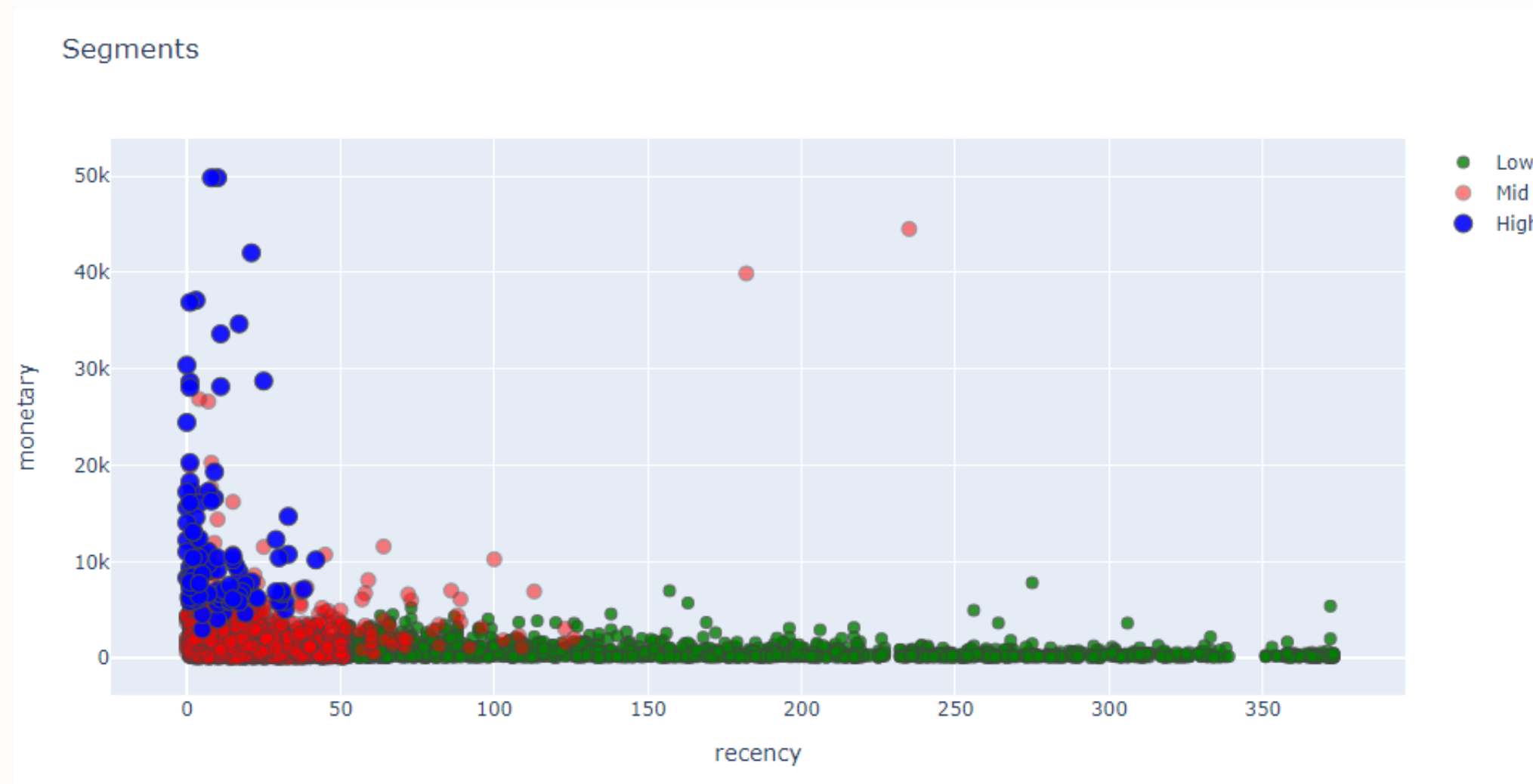
	recency	frequency	monetary
OverallScore			
0	437	437	437
1	546	546	546
2	895	895	895
3	1585	1585	1585
4	325	325	325
5	103	103	103
6	19	19	19
7	6	6	6
8	3	3	3

K-Means



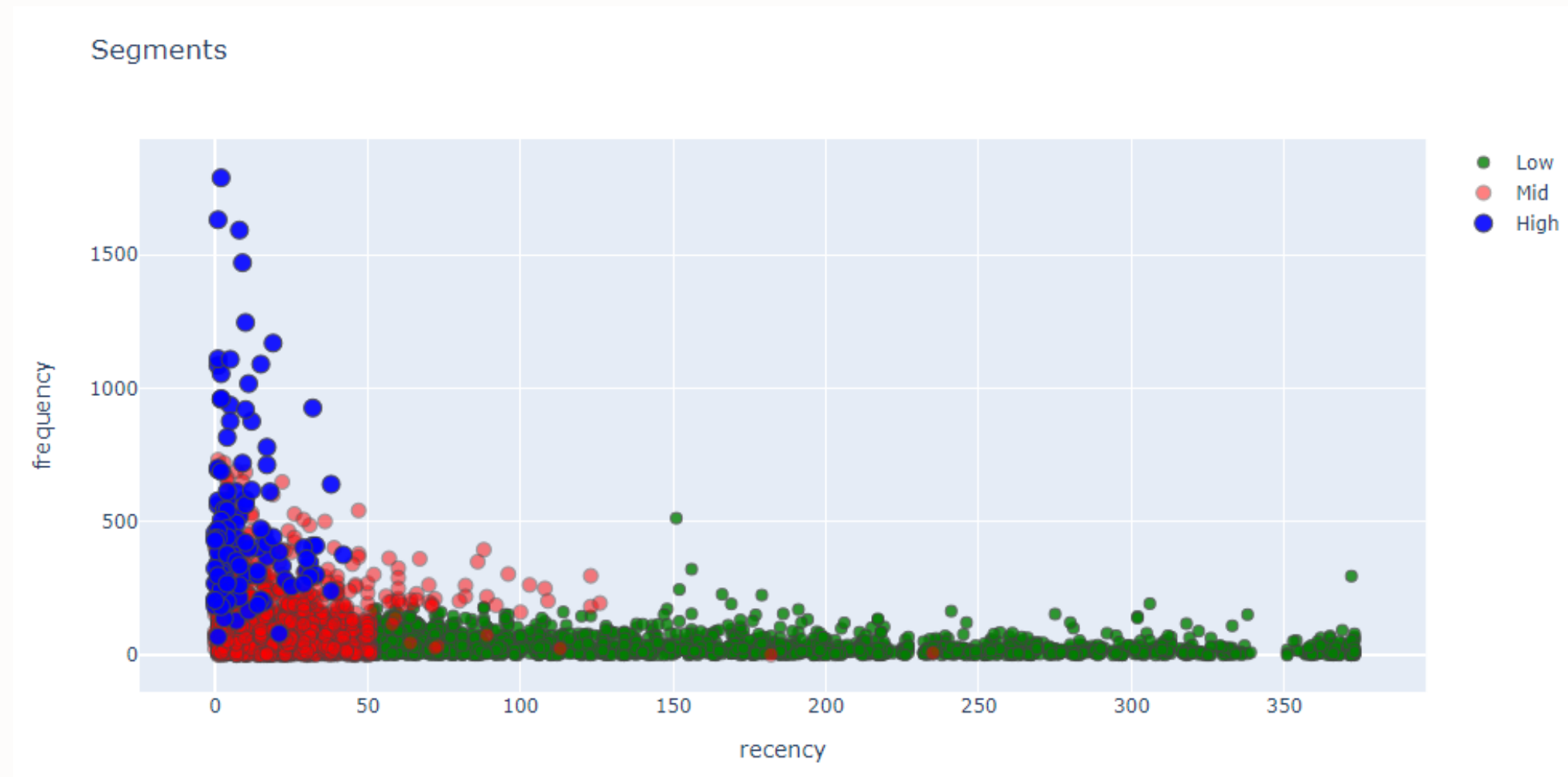
Distribution each segments according to frequency and monetary value. High value customer shows high amount of purchases and frequency purchases made.

K-Means



Distribution each segments according to recency and monetary value. High value customer shows low recency and high frequency purchases made.

K-Means



Distribution each segments according to recency and frequency value. High value customer shows low recency and high frequency purchases made.

Conclusion



To improve our sales, we should targeting high value customer. We can maintain with cashback/ discount in period time, to keep them coming back again.

while for mid value customer, we can maintain with promo buy one get one or buy two or more cheaper, discount, and cashback. This treatment to increase their spending habit and keep them coming back again.

For low value customer, we can maintain with promo that will increase their spending habit such as buy one get one, flash sale, discount, add some money to get 2 and so on.

According to K-Means and RFM analysis, most customers in United Kingdom are in mid value class. They often using our service lately, fairly frequent and generates moderate revenue.



—

Thank You

—

https://github.com/dillaarta/cust_segmentation.git