

RFM Analysis. Report

Presented by: Dusukov Aidar
dusukov.aidar@gmail.com

ETL of data

	email		phone	first_name		last_name	zip	city
0	5b31baf92171234adacc3aaf0ee5a14cf0b1adb35194	3567550f9cf51d6b9fe94cbbf17f8cf5da83677b20a6	be34ad33b31774df67565b3db888fa2361331e6566f1	1086b76dc02f395636156d9ca711d54a043f6640c690	92027	licata		
1	b47a23dfa8a71dd52879e32b46b95d3b0673e6ea38b0	0ff8479f1516d3de1091624ded4b0eeb2a0144854801	e6d5c8867b7a46dc7b81cdc689e47fe84647d49022d6	bd3dae5fb91f88a4f0978222dfd58f59a124257cb081	36100	vicenza		
2	6929f69b848bb6ea80905aa169a80c962644f6e1ea89	09a7a47a69b67e1d65c4ce385ec430bf49397261f38c	208ad55e4c4200ca42a73ab260efd6df57efccc95aad	3d67c6abd93267d5c5cb698b79401f76615b8f47f9a9	71016	san severo		
3	c5c2d17bbf20b772ff72abfcb2f5021cf02cebe422d3	c15aa8b05de5ccd5816dc3f69e8babff8b8998b0e64	f805bcb3efc982ec60744b3cc23f11ed1a4a0c4b783f	0f46ffbc599ddb2ebbd4b73dad90f240f502798a7986	20038	busto garolfo		
4	e60a1afac6a840b383ad8172de7304eb404348e72f66	f09585362157ce6813345220cc5de55cc72e01c9905f	09c12d01508bb6b857cf2f84199fb3134b58985806b4	cd165a891948a7ffae249980f8ca64e963f73f4116bc	10048	vinovo		

state	country	birthday_year	gender	age	conversion_name	conversion_date	conversion_id	conversion_value	conversion_value_margin	handling_cost	conversion_currency	ad_click_id_value
NaN	IT	NaN	NaN	NaN	Purchase	2020-08-08	918183	377.0	154.22	18.0	EUR	NaN
NaN	IT	NaN	NaN	NaN	Cancel	2020-08-07	917656	-643.5	-48.29	-5.0	EUR	NaN
NaN	IT	NaN	NaN	NaN	ReturnToSender	2020-08-07	905587	-318.5	-198.33	-183.1	EUR	NaN
NaN	IT	NaN	NaN	NaN	Purchase	2020-08-08	918391	462.8	307.12	44.0	EUR	NaN
NaN	IT	NaN	NaN	NaN	Cancel	2020-08-07	917776	-513.5	5.17	-5.0	EUR	NaN

First of all. Receiving data, we need to make sure it does not contain any useless and incompetent values to improve data quality and establish consistency.

ETL of data

	City	Conversion_Value
0	(cr) gussola	1203.80
1	- nerviano	819.00
2	- seleziona	1677.00
3	- seleziona il comune	0.00
4	- venafro	1287.00
...
73	??s????? ??????s	0.00
74	??ßade?a	-708.50
75	a coruña	18221.84
76	a coruña- españa	871.00
77	a guarda, pontevedra	1950.00

	Zip	Conversion_Value
15579	9990	591.50
15580	9991	2028.00
15581	a45d950	344.50
15582	a63 pd79	2373.80
15583	a67 cp28	670.80
...
19679	yo42 2pp	1287.00
19680	yo7 3ha	1153.75
19681	yo8 9xn	809.90
19682	ze1 0hn	1756.30
19683	ze2 9ep	809.90

Sometimes several columns can affect our analysis and even lead to inadequate results.

RFM Analysis

is a method used to determine customer value by looking at three dimensions:



01

Recency

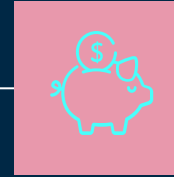
When is the last time the user takes an action?



02

Frequency

How many times does the user take this action?

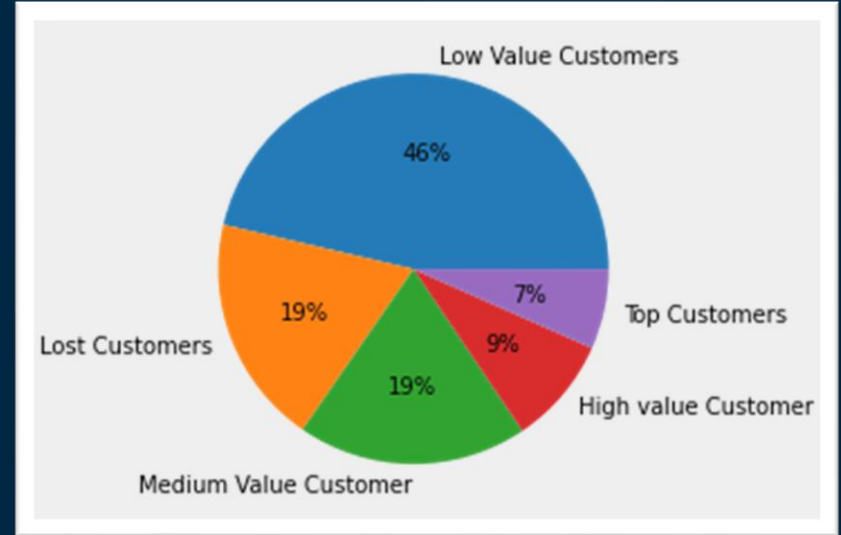
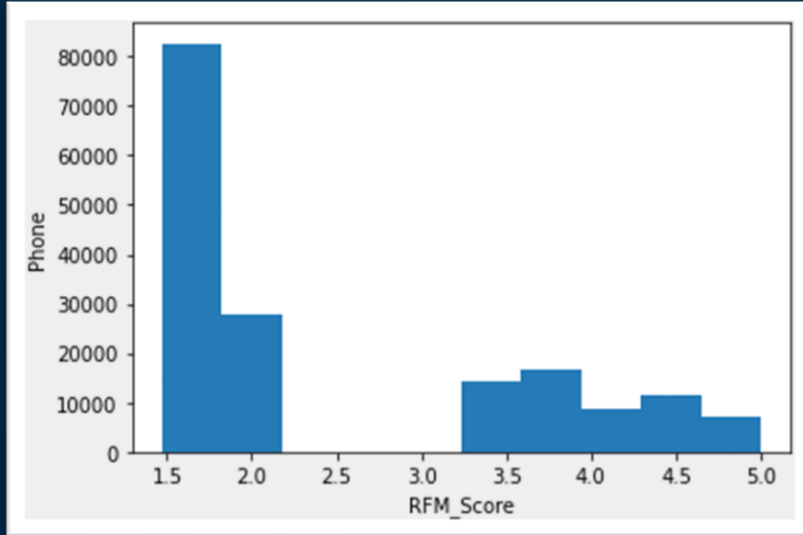


03

Monetary

What is the sum of monetary value from this user throughout his/her lifetime?

RFM Analysis



Following graphs were obtained during RFM Analysis.

Clearly, almost half of all the customers are in the Low-value category, while the amount of Medium and Lost customers is surprisingly equal. Minor categories are represented by High-value and Top customers, which contribute to 16 % of all buyers and have a difference of 2 %.

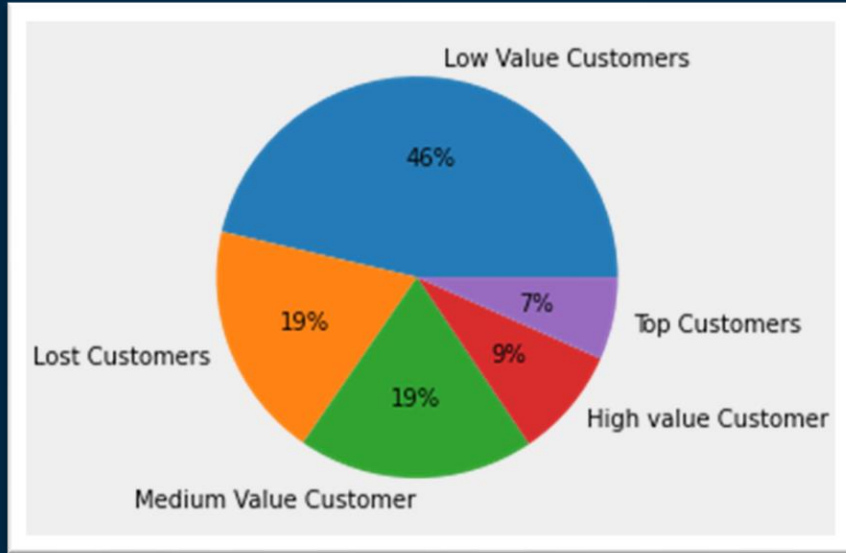
RFM Analysis

	Country	Frequency
15	IT	280766
10	GB	4914
4	DE	3007
9	FR	2867
19	NL	2217
7	ES	898
22	RO	554
8	FI	496
1	BE	473
21	PT	448
11	GR	388
12	HR	356
14	IE	320

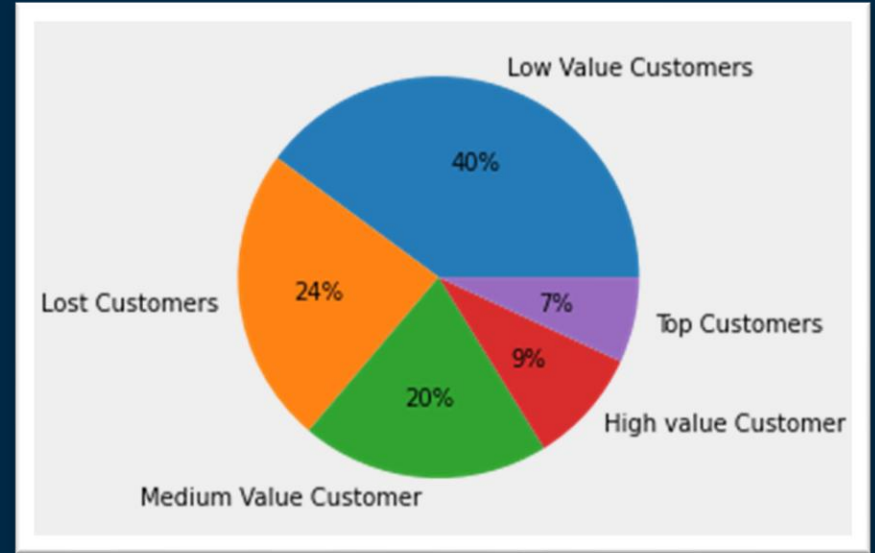
5	DK	301
24	SI	246
25	SK	227
13	HU	202
0	AT	187
23	SE	184
20	PL	173
18	LV	172
2	BG	161
16	LT	126
3	CZ	126
6	EE	125
17	LU	66

Let's analyze the countries distribution.

RFM Analysis



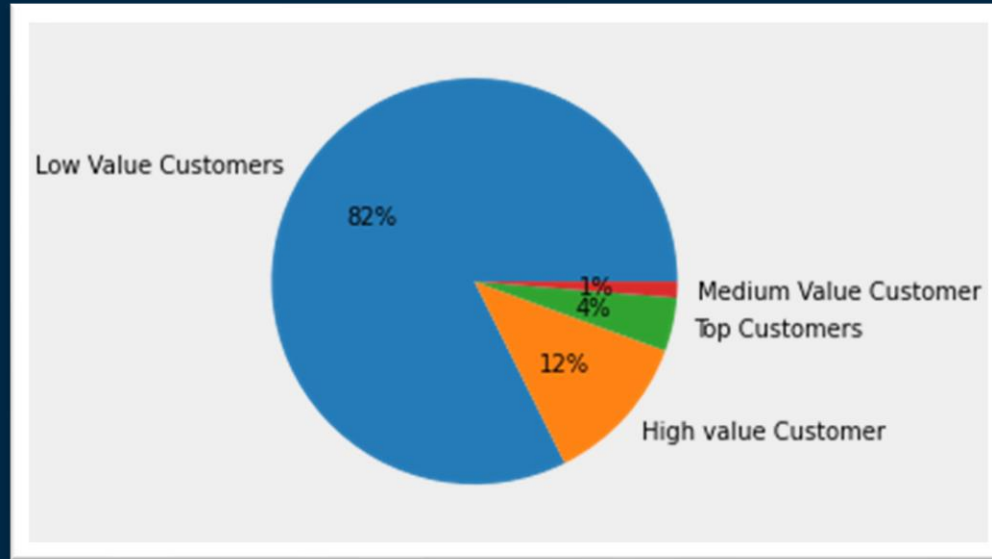
International Market



Italian Market

Obviously, two graphs do not differ much, since Italian market contributes to 93,6 % of the whole market.

RFM Analysis



Another situation with non-Italian market. Here, 82 % percent goes to Low-value customers. Moreover we don't see Lost customers category. And number of High-value buyers is the largest among top 3 categories by 12 and 3 respectively.

Conclusion

- Data cleansing and organizing are significant stages in data analysis, since they help to process data correctly and prevent future errors
- RFM Analysis is a very useful method for identifying customers categories, which is incredibly important for market and business
- During RFM Analysis some patterns were identified. Thus, Low-value customers contribute to about half of the market. Some actions should be performed to remove them from current category to the highest ones, as well as decrease the amount of Lost customers
- International outside Italy market showed an importance of decreasing the huge percentage of Low-value customers and potential opportunities for High-value category

Thank you for attention!