

SEGMENTASI CUSTOMER BERDASARKAN ANALISIS RFM

Yanuar Eka Putera

Fullstack Data Science - Batch 4



Analysis Flow

1. Introduction
2. Data Understanding
3. Exploratory Data Analysis
4. Data Preprocessing
5. Insight of Segmentation Customer
6. Conclusion & Suggestion

Introduction

Data yang digunakan merupakan **data publik** salah satu perusahaan **e-commerce** dari brazil, yang dibuat dari pesanan yang terjadi di **Olist Store**.

Dataset memiliki informasi 100k pesanan **dari 2016 hingga 2018**.

Data Understanding

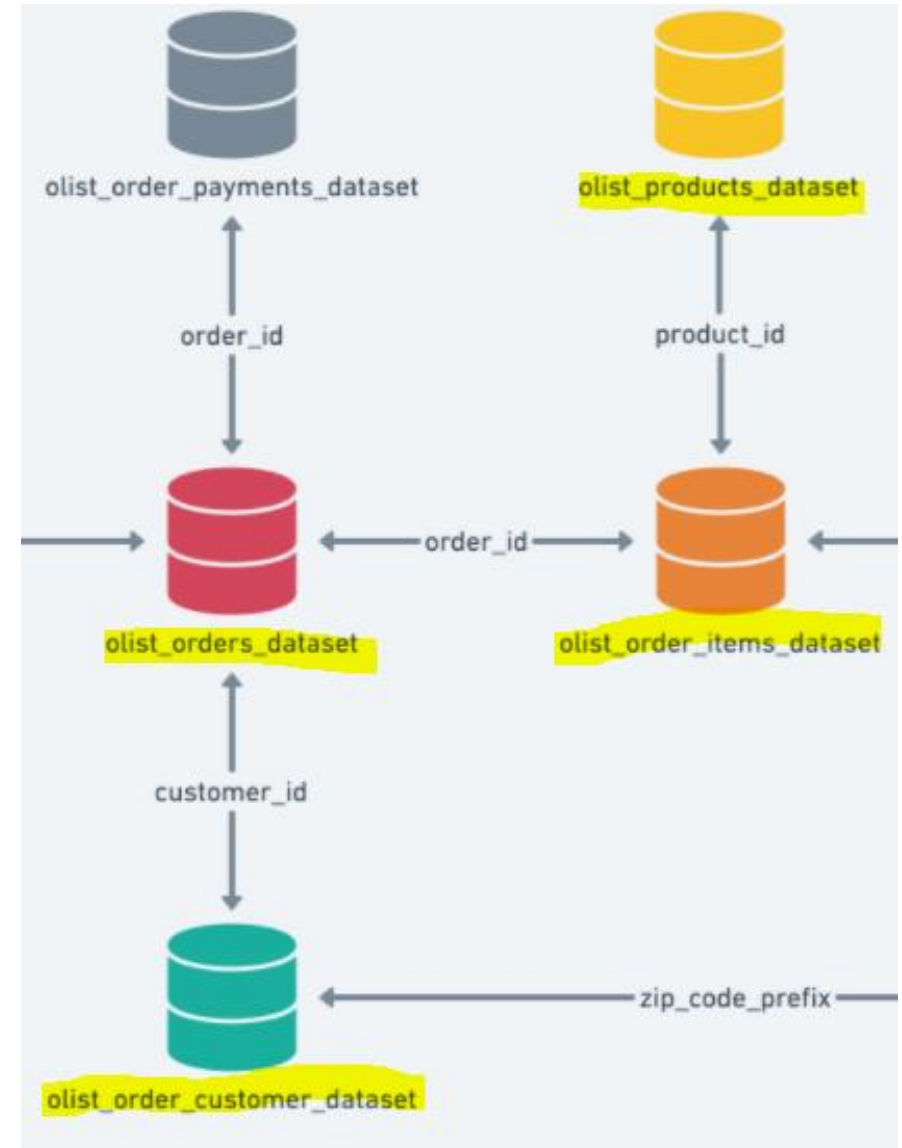
Column Name	Description
order_id	Unique identifier of the order.
order_item_id	Sequential number identifying number of items included in the same order.
product_id	Product unique identifier
price	Item price
product_category_name	Category of product, in Portuguese.
customer_id	key to the orders dataset. Each order has a unique customer_id
order_purchase_timestamp	Shows the purchase timestamp.
customer_state	Code customer state
state	Name customer state

Exploratory Data Analysis

Data Scheme

Berikut data yang diperlukan untuk segmentasi pelanggan berdasarkan RFM:

- Olist_product_dataset
- Olist_order_items_dataset
- Olist_order_dataset
- Olist_order_customer_dataset



Data Overview

	order_id	order_item_id	product_id	price	product_category_name	customer_id	order_purchase_timestamp	customer_state	state
0	00010242fe8c5a6d1ba2dd792cb16214	1	4244733e06e7ecb4970a6e2683c13e61	58.90	cool_stuff	3ce436f183e68e07877b285a838db11a	2017-09-13 08:59:02	RJ	Rio de Janeiro
1	00018f77f2f0320c557190d7a144bdd3	1	e5f2d52b802189ee658865ca93d83a8f	239.90	pet_shop	f6dd3ec061db4e3987629fe6b26e5cce	2017-04-26 10:53:06	SP	Sao Paulo
2	000229ec398224ef6ca0657da4fc703e	1	c777355d18b72b67abbeef9df44fd0fd	199.00	moveis_decoracao	6489ae5e4333f3693df5ad4372dab6d3	2018-01-14 14:33:31	MG	Minas Gerais
3	00024acbcd0a6daa1e931b038114c75	1	7634da152a4610f1595efa32f14722fc	12.99	perfumaria	d4eb9395c8c0431ee92fce09860c5a06	2018-08-08 10:00:35	SP	Sao Paulo
4	00042b26cf59d7ce69dfabb4e55b4fd9	1	ac6c3623068f30de03045865e4e10089	199.90	ferramentas_jardim	58dbd0b2d70206bf40e62cd34e84d795	2017-02-04 13:57:51	SP	Sao Paulo
...
112645	ffff94f6ce00a00581880bf54a75a037	1	4aa6014eceb682077f9dc4bffe5bc05b0	299.99	utilidades_domesticas	b51593916b4b8e0d6f66f2ae24f2673d	2018-04-23 13:57:06	MA	Maranhao
112646	fffc46ef2263f404302a634eb57f7eb	1	32e07fd915822b0765e448c4dd74c828	350.00	informatica_acessorios	84c5d4fbaf120aae381fad077416eaa0	2018-07-14 10:26:46	PR	Parana
112647	fffc4705a9662cd70adb13d4a31832d	1	72a30483855e2eafc67aee5dc2560482	99.90	esporte_lazer	29309aa813182aaddc9b259e31b870e6	2017-10-23 17:07:56	SP	Sao Paulo
112648	fffe18544ffabc95dfada21779c9644f	1	9c422a519119dcad7575db5af1ba540e	55.99	informatica_acessorios	b5e6afd5a41800fdf401e0272ca74655	2017-08-14 23:02:59	SP	Sao Paulo
112649	fffe41c64501cc87c801fd61db3f6244	1	350688d9dc1e75ff97be326363655e01	43.00	cama_mesa_banho	96d649da0cc4ff33bb408b199d4c7dcf	2018-06-09 17:00:18	SP	Sao Paulo

112650 rows x 9 columns

- No Duplicated
- Missing Values 1603 on product_category_name

Handle Missing Values

```
[168] df['product_category_name'].fillna('others', inplace=True)
```

```
df[(df['product_category_name'].isin(["others"]))]
```

order_item_id	product_id	price	product_category_name	customer_id
1	ff6caf9340512b8bf6d2a2a6df032cfa	7.79	others	ae55e228ded796581eccc67045516023
1	a9c404971d1a5b1cbc2e4070e02731fd	7.60	others	baa4ed05f22ffb6c4a41b71b699a492c
1	5a848e4ab52fd5445cdc07aab1c40e48	122.99	others	a69c820c317aa34
1	41eee23c25f7a574dfaf8d5c151dbb12	20.30	others	c664dc5de5b5f4
1	e10758160da97891c2fdcbbc35f0f031d	56.00	others	97dc92f1aae3a7d
...

```
[169] df.isnull().sum()
```

order_id	0
order_item_id	0
product_id	0
price	0
product_category_name	0
customer_id	0
order_purchase_timestamp	0
customer_state	0
state	0
dtype: int64	

Data Preprocessing

What is RFM?

RFM (Recency, Frequency, Monetary) adalah Model pemasaran untuk segmentasi pelanggan berbasis perilaku. Dengan mengelompokkan pelanggan berdasarkan riwayat transaksi mereka, yaitu **Seberapa baru mereka belanja, Seberapa sering mereka belanja, dan Berapa banyak yang mereka beli.**

Recency – Keterkinian Pembelian

- Hasil selisih order_purchase_date dengan order_purchase _date terbaru.

	CustomerID	recency
0	00012a2ce6f8dcda20d059ce98491703	292 days 16:58:31
1	000161a058600d5901f007fab4c27140	413 days 23:26:25
2	0001fd6190edaaf884bcaf3d49edf079	551 days 22:00:14
3	0002414f95344307404f0ace7a26f1d5	382 days 19:57:37
4	000379cdec625522490c315e70c7a9fb	153 days 19:24:40

Frequency – Frekuensi Pembelian

- Menghitung berapa kali customer membeli produk di Olist

CustomerID	frequency
fc3d1daec319d62d49bfb5e1f83123e9	21
be1b70680b9f9694d8c70f41fa3dc92b	20
bd5d39761aa56689a265d95d8d32b8be	20
adb32467ecc74b53576d9d13a5a55891	15
10de381f8a8d23fff822753305f71cae	15

Monetary – Berapa banyak customer menghabiskan uang

- Metotalkan nominal yang sudah dikeluarkan oleh customer.

CustomerID	monetary
1617b1357756262bfa56ab541c47bc16	13440.00
ec5b2ba62e574342386871631fafd3fc	7160.00
c6e2731c5b391845f6800c97401a43a9	6735.00
f48d464a0baaea338cb25f816991ab1f	6729.00
3fd6777bbce08a352fddd04e4a7cc8f6	6499.00

Data RFM

- Menggabungkan ketiga data recency, frequency, dan monetary

CustomerID	recency	frequency	monetary
00012a2ce6f8dcda20d059ce98491703	292	1	89.80
000161a058600d5901f007fab4c27140	413	1	54.90
0001fd6190edaaf884bc3af3d49edf079	551	1	179.99
0002414f95344307404f0ace7a26f1d5	382	1	149.90
000379cdec625522490c315e70c7a9fb	153	1	93.00

Data RFM terstandarisasi dan skor

- Standarisasi didapat berdasarkan quantile pada masing-masing RFM

CustomerID	recency	frequency	monetary	r	f	m	rfm_segment_concat	rfm_score
00012a2ce6f8dcda20d059ce98491703	292	1	89.80	2	1	3	213	6
000161a058600d5901f007fab4c27140	413	1	54.90	1	1	2	112	4
0001fd6190edaaf884bc3d49edf079	551	1	179.99	1	1	4	114	6
0002414f95344307404f0ace7a26f1d5	382	1	149.90	1	1	4	114	6
000379cdec625522490c315e70c7a9fb	153	1	93.00	3	1	3	313	7

Pelabelan Customer

- Memberi label kepada Customer berdasarkan skor

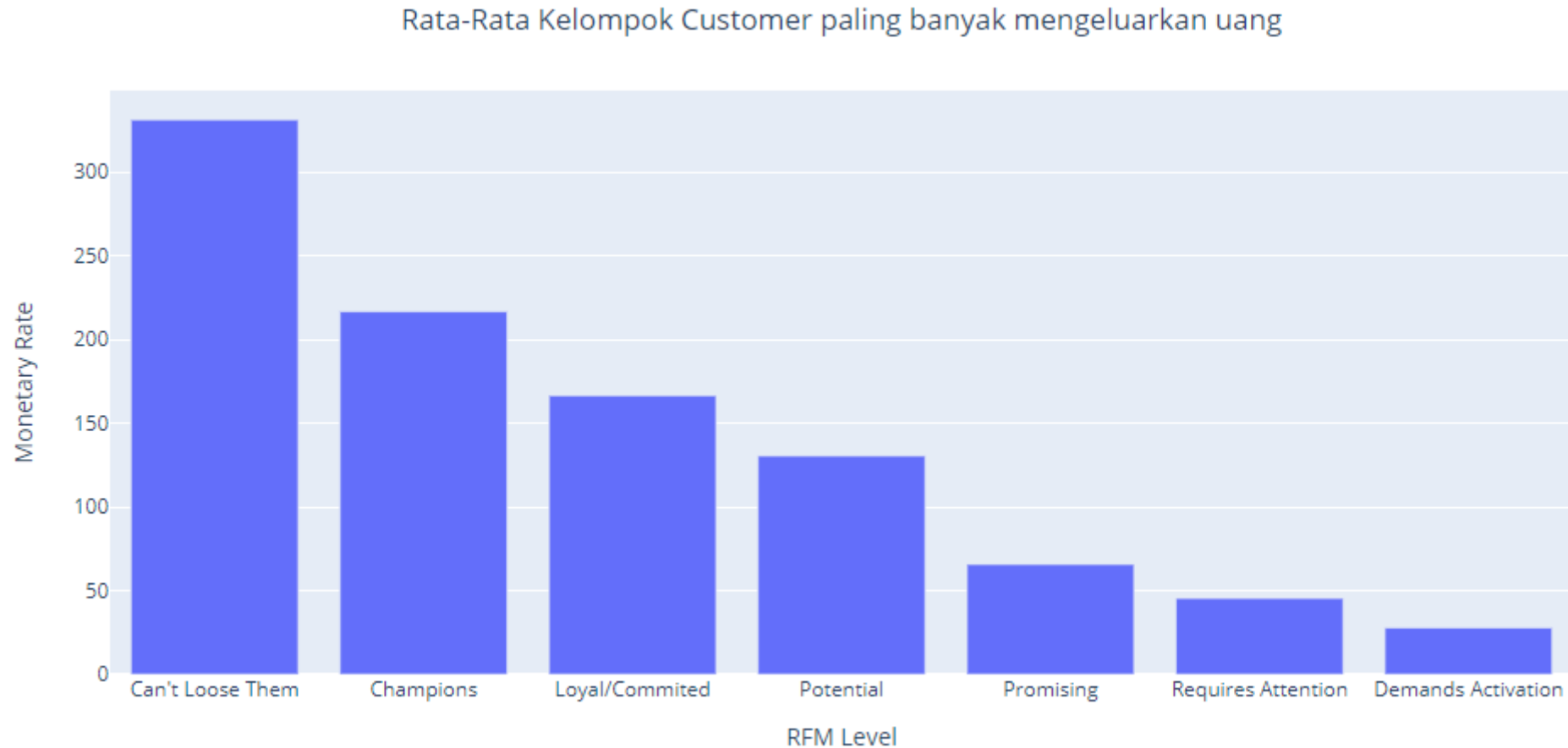
CustomerID	recency	frequency	monetary	r	f	m	rfm_segment_concat	rfm_score	rfm_level
00012a2ce6f8dcda20d059ce98491703	292	1	89.80	2	1	3	213	6	Potential
000161a058600d5901f007fab4c27140	413	1	54.90	1	1	2	112	4	Requires Attention
0001fd6190edaaf884bcaf3d49edf079	551	1	179.99	1	1	4	114	6	Potential
0002414f95344307404f0ace7a26f1d5	382	1	149.90	1	1	4	114	6	Potential
000379cdec625522490c315e70c7a9fb	153	1	93.00	3	1	3	313	7	Loyal/Committed

Insight Segmentasi Customer

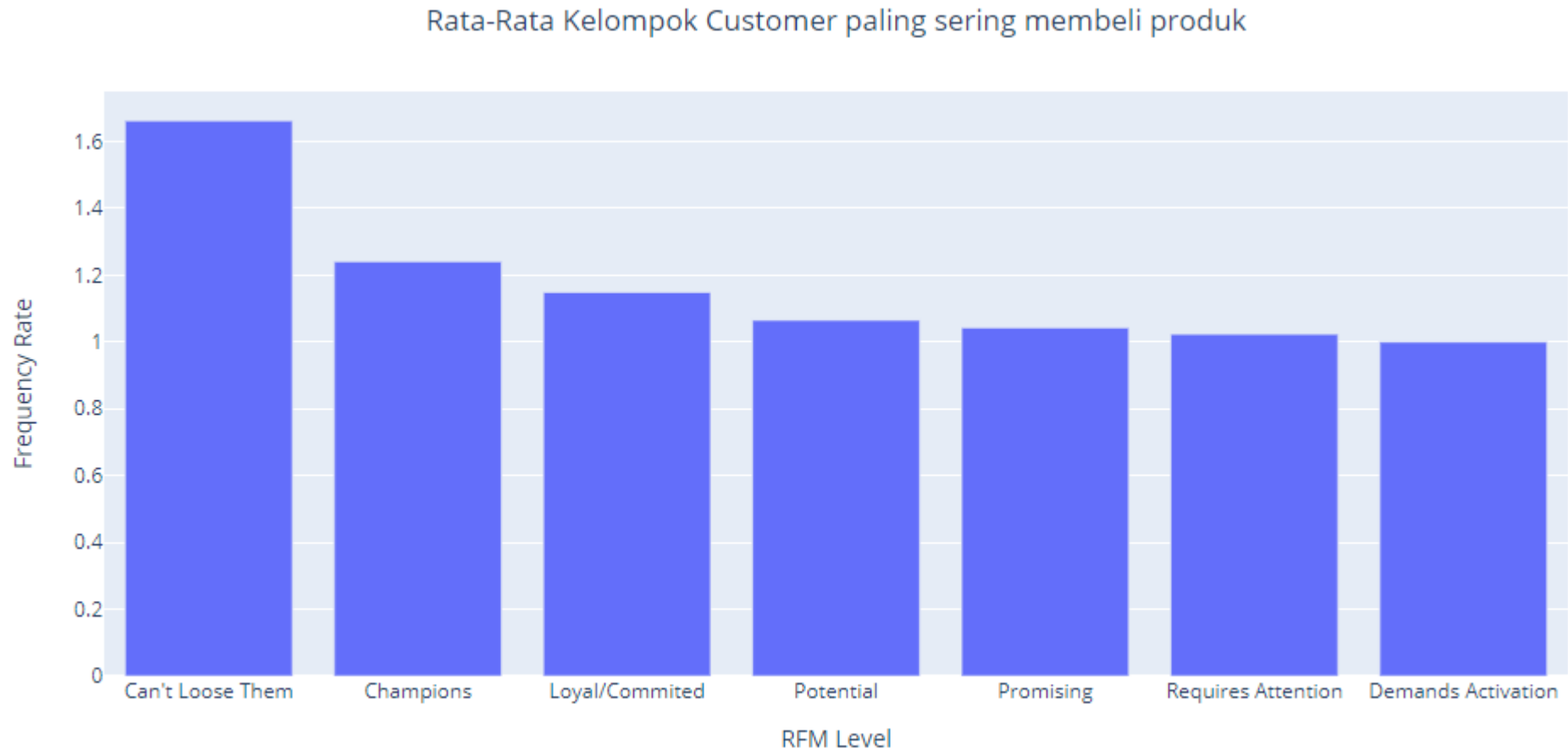
Jumlah Customer tiap Kelompok



Rata-Rata kelompok customer yang paling banyak mengeluarkan uang



Rata-Rata kelompok customer yang paling sering membeli produk



Rata-Rata kelompok customer yang baru-baru ini membeli produk



Conclusion & Suggestion

Conclusion

- Fokus melakukan pemasaran kepada customer yang berada di kelompok potensial karena jumlah customernya lebih banyak dibandingkan dengan yang lain.
- Selain itu melakukan diskon atau promo terhadap customer yang memiliki daya beli rendah yaitu customer yang berada di kelompok Requires Attention dan Demands Activation.

Suggestion

- Para pembaca bisa melakukan CLV(Customer Lifetime Value) dapat digunakan untuk mendapatkan insight lain dari customer.
- Para pembaca juga bisa menambahkan churn rate untuk mengetahui churn yang terjadi pada customer
- Pembaca dapat melakukan market basket analysis untuk mencocokkan produk dengan customer.
- 3 hal tersebut dapat dilakukan dari dataframe (df) yang sudah dibuat oleh penulis.