

Customer Segmentation and Recommendation System to Increase Customer Retention Olist (Brazilian E-Commerce Integrator)

by : AlphaGroup_JC_DS_OL_09



OUTLINE

01

Profile Group Intro

02

**Problem
Formulation**

03

Data Understanding

04

Findings

05

Solution

06

**Conclusion &
Recommendation**



01

PROFILE GROUP

ALPHA GROUP MEMBERS



Irfan Zidni

zidniirfan155@gmail.com

JCDSOL-009-047

- EDA Main Dataset
- Data Cleaning
- Recommendation System



Ali Yafi

ir.yafee@gmail.com

JCDSOL-009-036

- Business Problem
- Tableau
- PowerPoint



Noor Kharismawan Akbar

akbar.noorkharismawan@gmail.com

JCDSOL-009-038

- RFM & K-Means Clustering
- EDA Cluster
- Conclusion & Recommendation

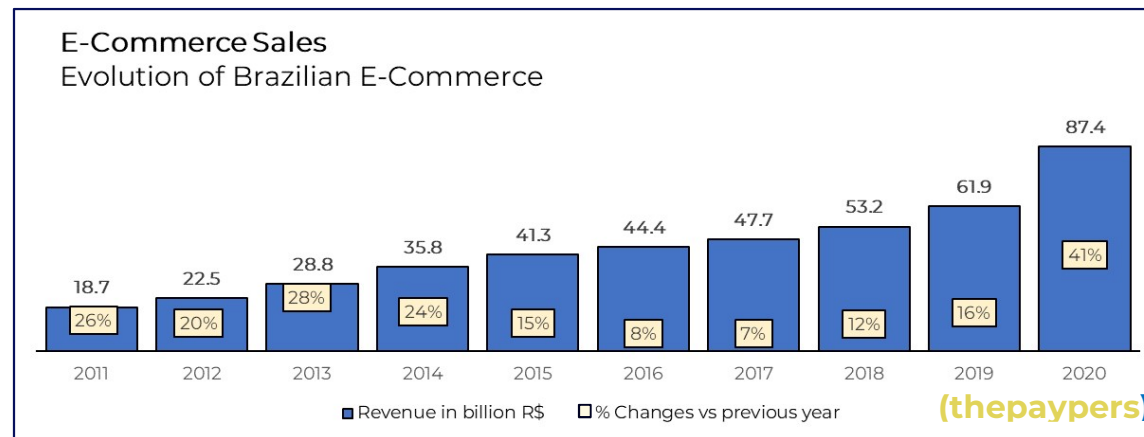
02

PROBLEM FORMULATION



CONTEXT

Semakin terhubungnya dunia dengan dunia digital, **e-commerce** terus tumbuh dengan kecepatan yang belum pernah terjadi sebelumnya dengan proyeksi yang mengindikasikan bahwa jumlah ini akan **terus meningkat**. (Forbes)



Brazil merupakan pasar **terbesar ke-12** untuk eCommerce dengan prediksi pendapatan sebesar R\$ 231 Bill pada tahun 2023. (Ecommercedb)

CONTEXT

Di Brazil terdapat E-Commerce yang cukup terkenal bernama Olist. Olist dibentuk pada 2015 oleh Tiago Dalvi, berkantor pusat di Curitiba, Parana, Brazil.



tiny

business management

Sistem ERP yang dibuat untuk menyederhanakan operasi bisnis Anda



olist store

sell on marketplaces

Marketplace Integrator



sale

own virtual store

Teknologi lengkap dan fitur eksklusif yang terintegrasi ke dalam e-commerce Anda untuk meningkatkan pertumbuhan bisnis Anda.



olist pax

optimized logistics

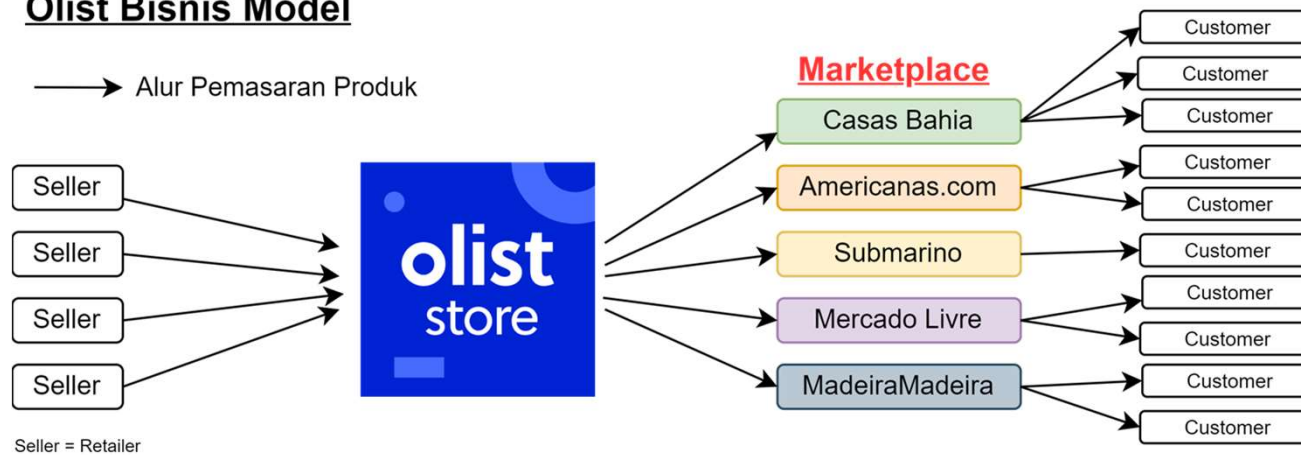
menyatukan semua layanan logistik yang dibutuhkan bisnis Anda untuk menjadi lebih kompetitif di satu tempat.

CONTEXT

Olist Store (Marketplace Integrator), e-commerce platform yang memberikan kemudahan kepada seller untuk menjual produknya secara online dengan cara mempromosikan ke eCommerce marketplace di Brazil.

Olist Bisnis Model

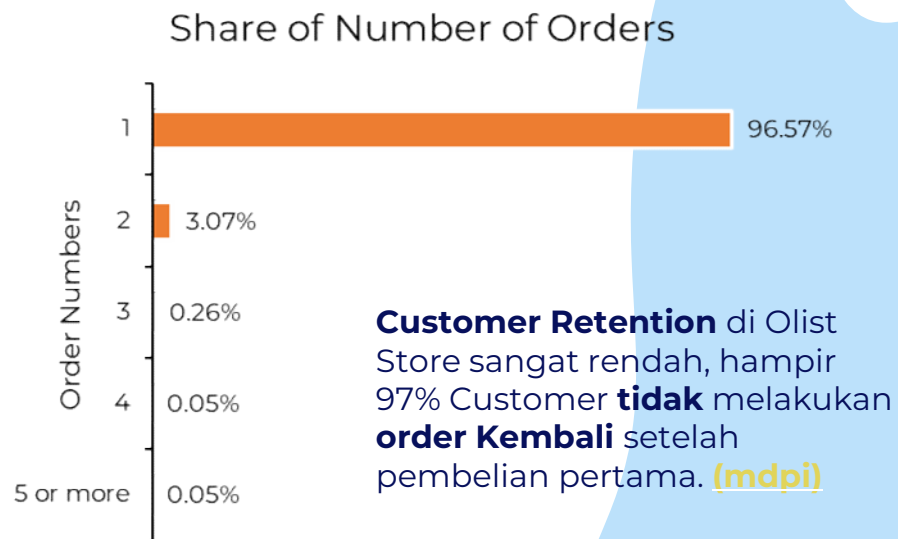
→ Alur Pemasaran Produk



Olist Store mendapatkan keuntungan dari bagi hasil dari **penjualan** dan **monthly fee** dari seller

PROBLEM FORMULATION

Problem



Goals

Menaikkan **customer retention** sehingga dapat **meningkatkan Revenue** Olist Store



METHOD

Analytics Approach

1. Melakukan **segmentasi** customer dengan **RFM & K-Means Clustering**. Dengan memisahkan customer berdasarkan karakteristiknya, strategi marketing jadi lebih efektif (**ada.cx**)
2. Memberikan **Recommendation System** di setiap segmentnya untuk meningkatkan penjualan (**exposebox**)

Limitation

Project ini akan berfokus pada:

1. Layanan **Olist Store**.
2. Masalah rendahnya **Customer retention**
3. Pengembangan **strategi marketing** yang efektif

Metrics Evaluation

Menentukan Jumlah Cluster:

1. **Silhouette Score**, Ukuran kemiripan suatu cluster dibandingkan dengan cluster lain
2. **Elbow Method**, perubahan tingkat kemiripan setiap penambahan satu cluster



03

DATA UNDERSTANDING

DATASET

dataset publik **e-commerce Brazil** dari pesanan yang dibuat di **Olist Store**. Memiliki informasi lebih dari **100 ribu pesanan** dari tahun **2016 hingga 2018** yang dibuat di beberapa Marketplace di Brasil.

Sumber: ([Kaggle](#))

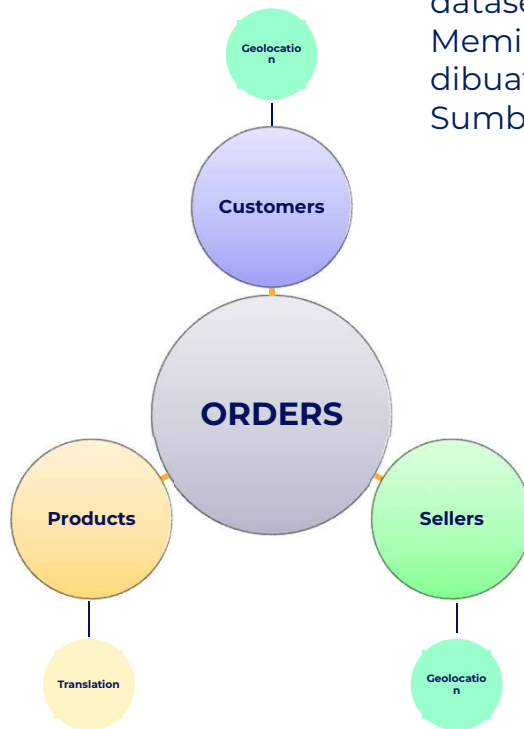
9 Dataset

Seluruh dataset digunakan (digabung) → 112.647 baris data

Deskripsi Entity:

1. **ORDERS**
detail transaksi yang dilakukan, meliputi pembayaran, waktu, produk yang dibeli, review dan harga.
(dataset order, order_payment, order_review, order_item)
2. **CUSTOMERS**
detail profil dari pembeli meliputi ID, Lokasi, dll. (dataset customer)
3. **SELLERS**
detail profil retailer meliputi ID, Lokasi dll. (dataset seller)
4. **PRODUCTS**
detail deskripsi produk, foto, ukuran dll. (dataset product)

Dataset Geolocation & Product_category_translation sebagai penunjang informasi tambahan



ANOMALI DATA

Data yang dalam kondisi yang tidak normal

1. **Incorrect Types**, seller_city berbentuk angka **04482255** dan berbentuk email **vendas@creditparts.com.br**
2. **Missing Value**, product_category_name tidak terisi (NaN)
3. **Inconsistent Data**, misal geolocation_city ada **sao paulo** dan **são paulo**

	seller_id	seller_zip_code_prefix	seller_city	seller_state
517	ceb7b4fb9401cd378de7886317ad1b47	22790	04482255	RJ
2067	c8143b3069f6746a77421b5ce30a450c	75345	abadia de goias	GO

	product_id	product_category_name_english
6	a1804276d9941ac0733cfd409f5206eb	NaN
122	71225f49be70df4297892f6a5fa62171	NaN
212	87285b34884572647811a353c7ac498a	housewares
264	595fac2a385ac33a80bd5114aec74eb8	perfumery

Anomali dapat menyebabkan **bias** pada tahapan Analisis sehingga penanganannya dengan domain knowledge (memanfaatkan kolom lain sebagai informasi) maupun didrop.



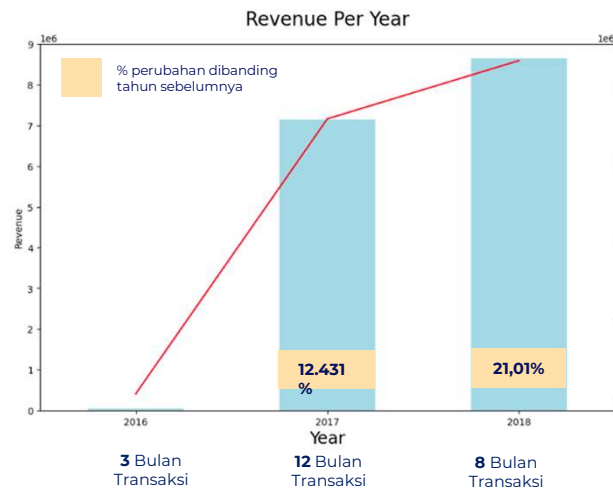
04

FINDINGS

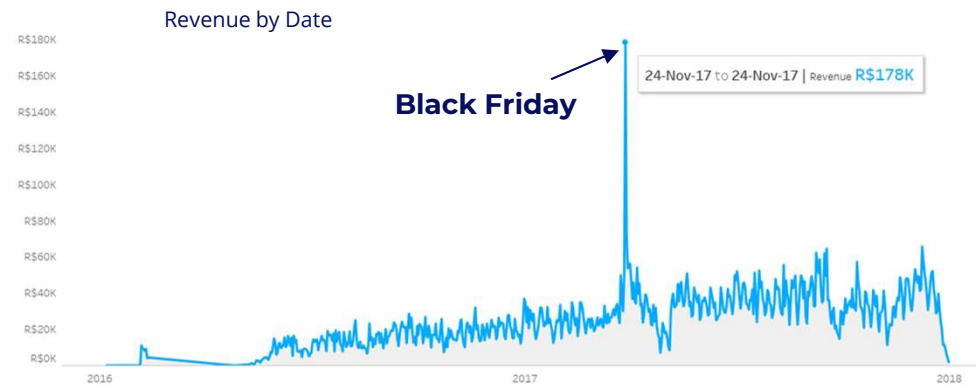
FINDINGS

10/25

REVENUE



Revenue setiap tahun **naik**



Pada **24 Nov 2017** order melonjak **300%** dibandingkan hari sebelumnya, hal ini kemungkinan disebabkan oleh tradisi **black friday** di Brazil. diskon besar-besaran diberikan sehingga banyak terjadi lonjakan order.

Potensi Olist Store untuk terus berkembang cukup tinggi, dimana tren **revenue** cenderung **naik**. Hal ini juga sejalan dengan metrics **jumlah order** maupun **jumlah customer** yang juga semakin **naik**.

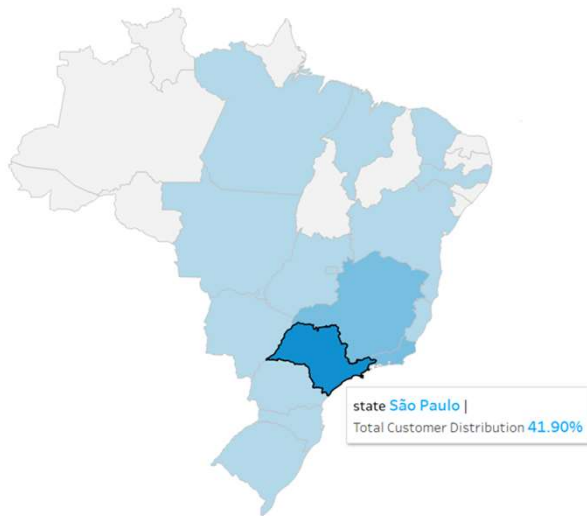
FINDINGS

11/25

CUSTOMER BEHAVIOUR

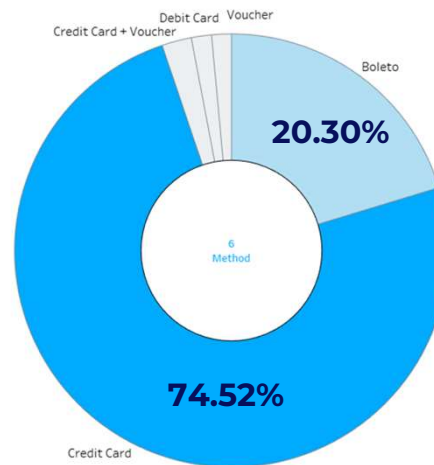
Mengetahui **kebiasaan** yang dilakukan customer dalam berbelanja sangat penting untuk menentukan Langkah strategis yang bisa diambil selanjutnya.

Customer by State



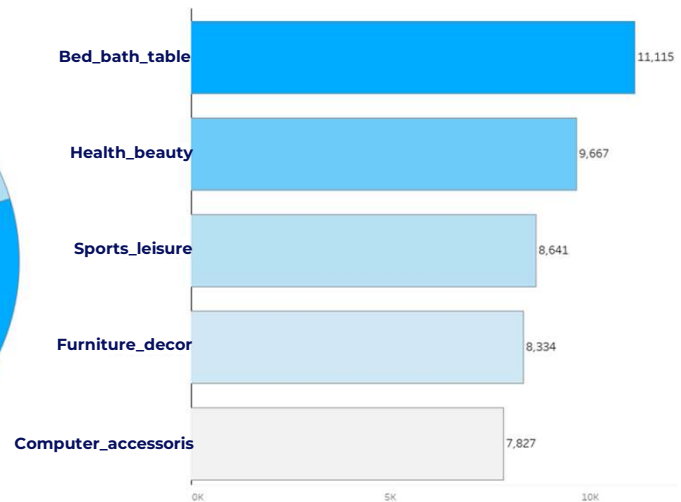
Mayoritas customer dari state
Sao Paulo

Payment Types



94,82% customer menggunakan
Credit Card & Boleto

Top 5 Category by order count



bed_bath_table merupakan
produk yang paling laris

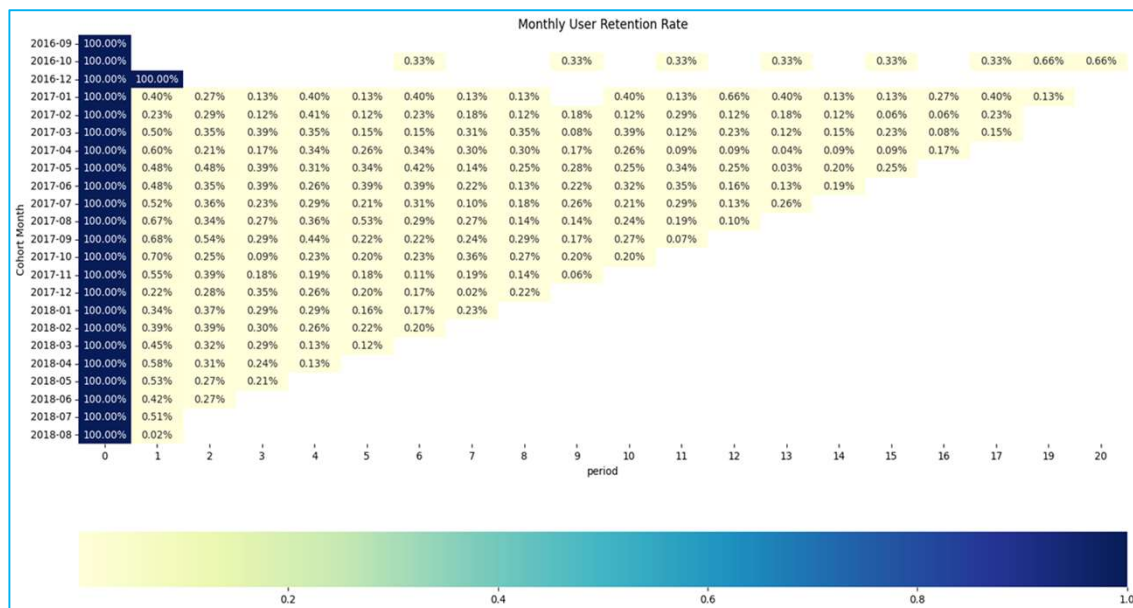
FINDINGS

RETENTION ANALYSIS

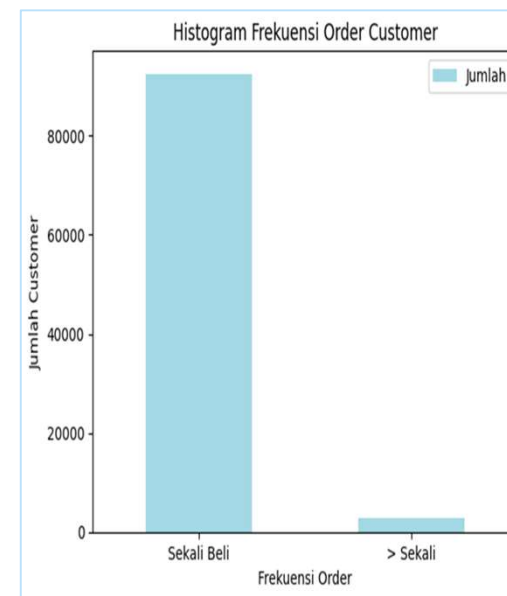
Pada 2016-2018 **retention rate** Olist store sangat rendah, terlihat pada cohort maupun grafik frekuensi order (Mayoritas customer hanya order satu kali)

Cohort Analysis

Tool untuk mengetahui berapa besar tingkat retensi pelanggan



Frekuensi Order





05

SOLUTION

Metode untuk mengetahui perilaku customer dalam berbelanja dengan mengukur kapan terakhir suatu customer berbelanja (**Recency**), seberapa sering customer membeli produk kita (**Frequency**), dan seberapa banyak customer mengeluarkan uang untuk produk kita (**Monetary**).



Data dibagi menjadi **5 bins** yang di berdasarkan **equal frequency** (Recency & Monetary) dan **equal interval** (Frequency). Kemudian dilakukan scoring dengan metode ranking dimana nilai tertinggi dimiliki customer yang:

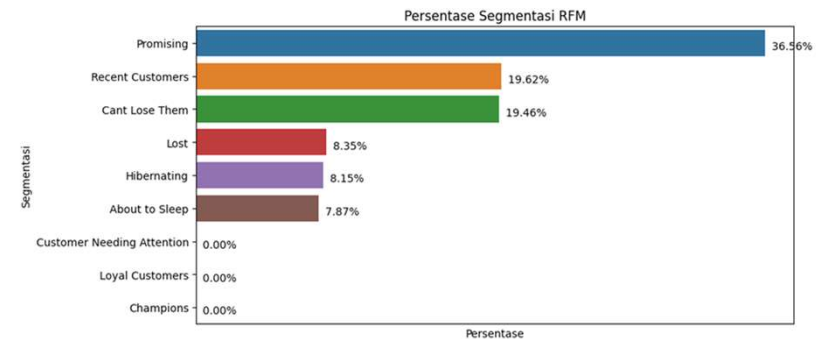
- Recency : recent buy
- Frequency : pembelian terbanyak
- Monetary : Spent terbanyak

<https://www.dase-analytics.com/blog/en/rfm-analysis/>

Reference Cluster (11 Cluster)

No	Nama Segmen	Skor RFM
1	Champions	555, 554, 544, 545, 454, 455, 445
2	Loyal Customers	543, 444, 435, 355, 354, 345, 344, 335
3	Potential Loyalist	553, 551, 552, 541, 542, 533, 532, 531, 452, 451, 442, 441, 431, 453, 433, 432, 423, 353, 352, 351, 342, 341, 333, 323
4	Recent Customers	512, 511, 422, 421, 412, 411, 311
5	Promising	525, 524, 523, 522, 521, 515, 514, 513, 425, 424, 413, 414, 415, 315, 314, 313
6	Customers Needing Attention	535, 534, 443, 434, 343, 334, 325, 324
7	About To Sleep	331, 321, 312, 221, 213
8	At Risk	255, 254, 245, 244, 253, 252, 243, 242, 235, 234, 225, 224, 153, 152, 145, 143, 142, 135, 134, 133, 125, 124
9	Cant Lose Them	155, 154, 144, 214, 215, 115, 114, 113
10	Hibernating	332, 322, 231, 241, 251, 233, 232, 223, 222, 132, 123, 122, 212, 211
11	Lost	111, 112, 121, 131, 141, 151

Dataset Cluster (9 Cluster)



Segmentasi scoring RFM yang terbentuk:

- Cluster yang terbentuk belum bisa memisahkan frequency 1x dan >1x order.
- Dengan adanya, 3 segmen RFM yang anggota nya sangat sedikit (<3 customer)
- Cluster Hibernating dan Lost, jika dilihat dari karakteristiknya masih bisa digabung menjadi satu cluster

Karakteristik Cluster RFM

	Recency	Frequency	Monetary	
	mean	mean	mean	count
segment				
About to Sleep	274.57556	1.01624	89.81615	7511
Cant Lose Them	415.77942	1.04956	271.90131	18565
Champions	14.00000	16.00000	902.04000	1
Customer Needing Attention	217.33333	6.00000	811.20000	3
Hibernating	322.55713	1.00579	56.14261	7772
Lost	480.06467	1.00640	55.50745	7964
Loyal Customers	188.00000	9.00000	1172.67000	1
Promising	138.58321	1.05702	237.15379	34881
Recent Customers	122.21489	1.00433	51.59348	18721

Dengan **keterbatasan Scoring RFM**, Analytic Approach selanjutnya menggunakan K-Means Clustering

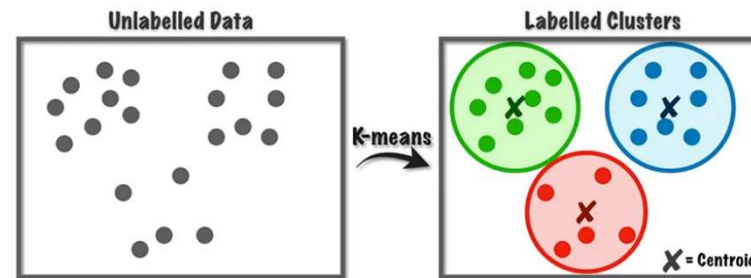
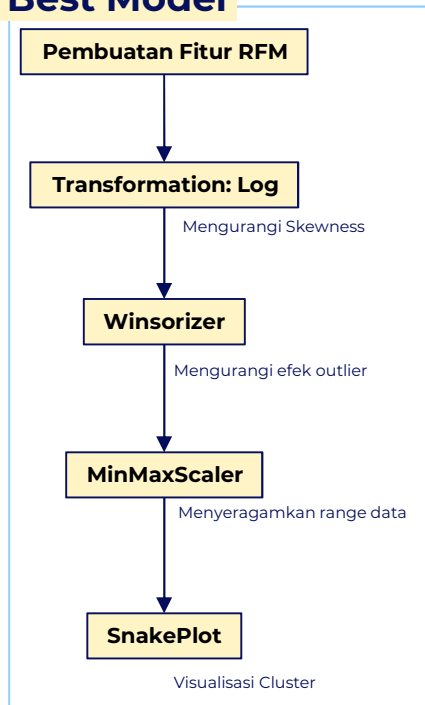
SOLUTION

K-MEANS CLUSTERING

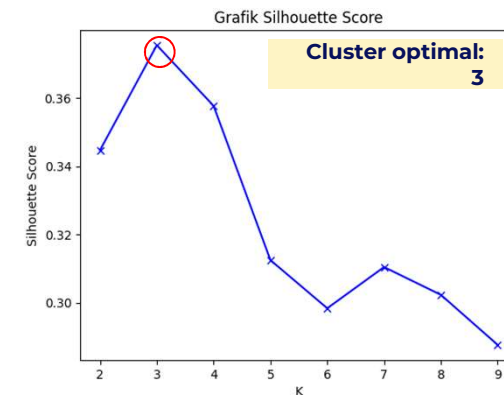
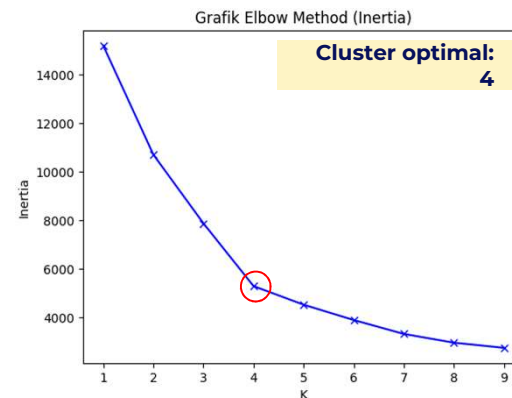
15/25

membagi titik data yang tidak berlabel ke dalam 'k' kluster berbeda

Best Model



- Data **tidak ada label** → unsupervised learning
 - Data **sangat banyak** (Ratusan ribu) → K-Means mampu mengelompokkan data besar dengan sangat cepat.
- Sumber: [media.neliti](https://media.neliti.com/media/publications/102421/102421.pdf)



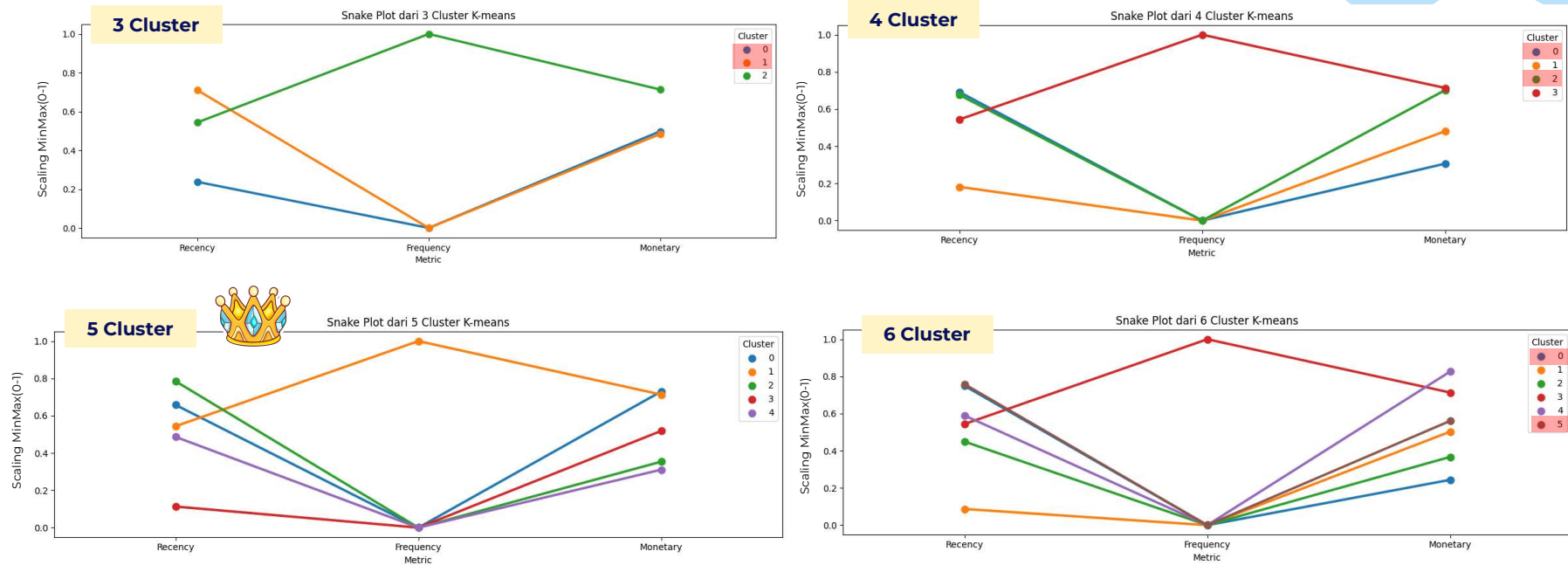
Jumlah Cluster optimal dari Metrics Elbow & Silhouette dirasa masih **belum memvisualisasikan** cluster yang terbentuk, sehingga perlu metode lain menggunakan Snake Plot.

SOLUTION

K-MEANS CLUSTERING

16/25

Snake Plot Visualisasi Nilai Rata-Rata value RFM setiap clusternya.



K-Means Clustering sudah dapat memisahkan customer dengan frekuensi order **1x** dan **> 1x**. Dipilih **k=5** karena Cluster yang terbentuk lebih heterogen antar cluster daripada k=3, k=4 & k=6.

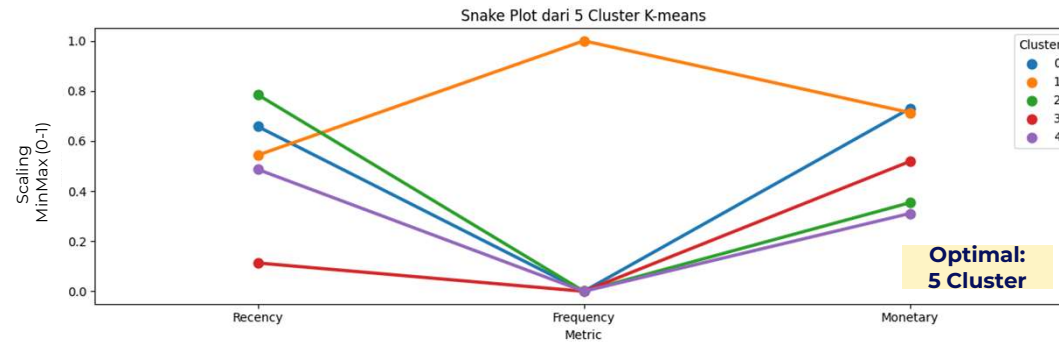
SOLUTION

K-MEANS CLUSTERING

Cluster yang terpilih k=5

17/25

Snake Plot



Average RFM

	Recency	Frequency	Monetary	
	mean	mean	mean	count
segment				
0	277.25	1.00	310.15	28633
1	226.08	2.11	310.49	2913
2	384.88	1.00	76.45	28111
3	44.51	1.00	165.51	15378
4	153.71	1.00	66.92	20384

Proporsi Cluster



SOLUTION

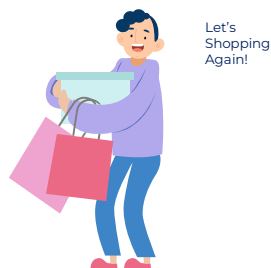
KARAKTERISTIK CLUSTER K-MEANS

ROYAL



Recency ★ ★ ★ ★ ★
 Frequency ★ ★ ★ ★ ★
 Monetary ★ ★ ★ ★ ★

SUSTAIN



Recency ★ ★ ★ ★ ★
 Frequency ★ ★ ★ ★ ★
 Monetary ★ ★ ★ ★ ★

JEOPARDY



Recency ★ ★ ★ ★ ★
 Frequency ★ ★ ★ ★ ★
 Monetary ★ ★ ★ ★ ★

FRESH



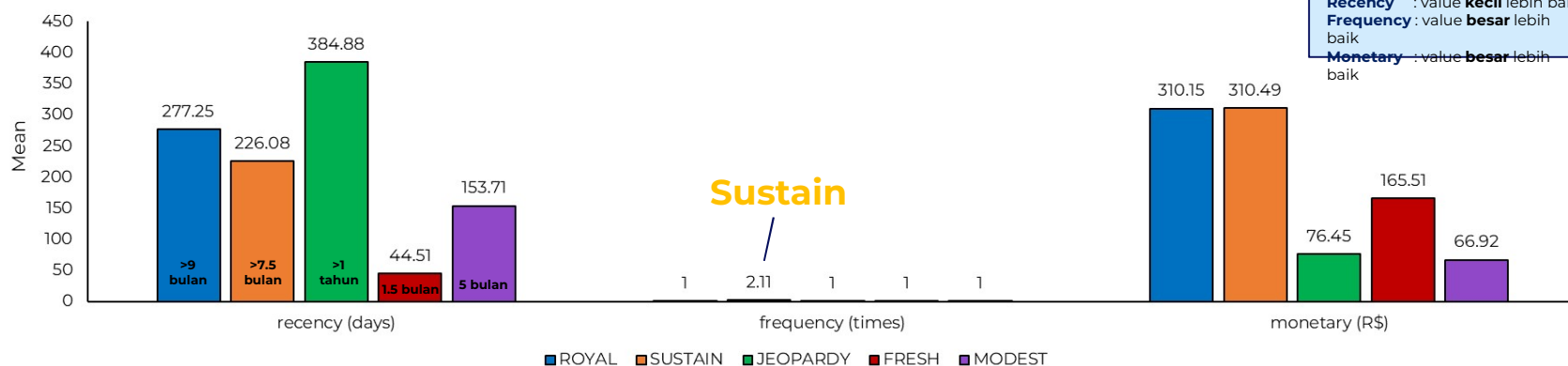
Recency ★ ★ ★ ★ ★
 Frequency ★ ★ ★ ★ ★
 Monetary ★ ★ ★ ★ ★

MODEST



Recency ★ ★ ★ ★ ★
 Frequency ★ ★ ★ ★ ★
 Monetary ★ ★ ★ ★ ★

Characteristics K-Means Cluster



SOLUTION

Perbandingan Model

Fitur RFM yang dibuat dapat diimplementasikan pada **Scoring RFM** maupun **K-Means Clustering**. Hasil pada Scoring RFM dan K-Means akan kita bandingkan untuk mendapatkan hasil yang terbaik.

1. Scoring RFM

- 9 **Cluster**
- **belum bisa** memisahkan frequency 1x dan >1x order
- Beberapa cluster memiliki **karakteristik yang sama**
- Beberapa Cluster jumlah customernya **sangat sedikit** (<3)

	Recency	Frequency	Monetary	
	mean	mean	mean	count
segment				
About to Sleep	274.57556	1.01624	89.81615	7511
Cant Lose Them	415.77942	1.04956	271.90131	18565
Champions	14.00000	16.00000	902.04000	1
Customer Needing Attention	217.33333	6.00000	811.20000	3
Hibernating	322.55713	1.00579	56.14261	7772
Lost	480.06467	1.00640	55.50745	7964
Loyal Customers	188.00000	9.00000	1172.67000	1
Promising	138.58321	1.05702	237.15379	34881
Recent Customers	122.21489	1.00433	51.59348	18721

2. K-Means Clustering

- 5 **Cluster**
- Mampu **memisahkan** frequency 1x dan >1x order dengan visualisasi menggunakan Snake Plot.
- Cluster memiliki **karakteristik yang berbeda**
- Tidak ada cluster yang jumlah customernya **sangat sedikit**

	Recency	Frequency	Monetary	
	mean	mean	mean	count
segment				
0	277.25	1.00	310.15	28633
1	226.08	2.11	310.49	2913
2	384.88	1.00	76.45	28111
3	44.51	1.00	165.51	15378
4	153.71	1.00	66.92	20384

Project ini menggunakan **K-Means Clustering** dibanding scoring RFM karena lebih mampu meng-cluster customer berdasarkan karakteristiknya

SOLUTION

RECOMMENDATION SYSTEM

Rekomendasi produk yang personalized dapat meningkatkan keuntungan bisnis online. Ketika rekomendasi produk disesuaikan dengan kebutuhan dan minat pelanggan Anda, rekomendasi produk berfungsi sebagai alat yang ampuh. **(exposebox)**

Sistem yang kita buat bisa me-rekomendasikan:

- Overall **Hot Selling** Item
- Hot Selling Item Based On **Customer State**
- Hot Selling Item Based On **Customer Cluster/ Segmentation**
- Or **Mix** Of Customer State and Customer Cluster/ Segmentation

Untuk memudahkan tim marketing dalam memberikan product recommendation, dibuatkan recommendation sytem pada WebApp berikut:

<https://olist-recommendation-system.streamlit.app/>

Olist Recommendation System

Cluster Name
Modest

Customer State
SP

Number of Product to Recommend
10

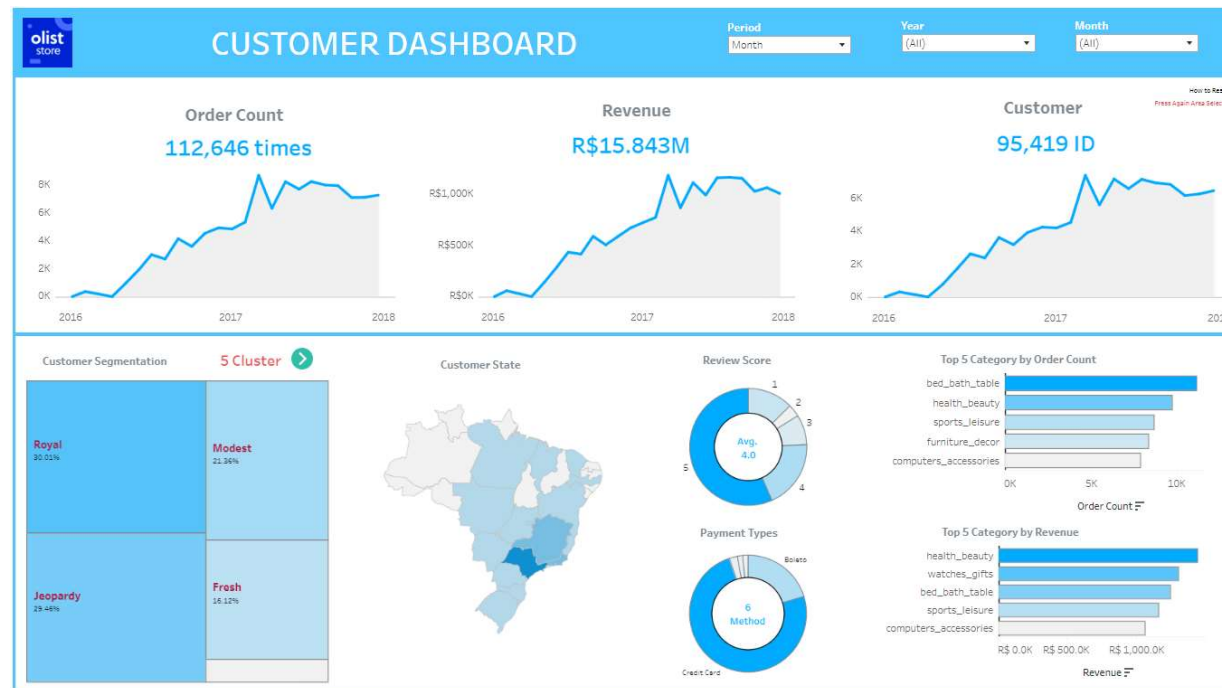
Recommend

	product_category	product_id
0	health_beauty	437c05a395e9e47f9762e677a7068ce7
1	furniture_decor	aca2eb7d00ea1a7b8ebd4e68314663af
2	cool_stuff	54d9ac713e253fa1fae9c8003b011c2a
3	health_beauty	e0cf79767c5b016251fe139915c59a26
4	watches_gifts	53b36df67ebb7c41585e8d54d6772e08
5	food	89321f94e35fc6d7903d36f74e351d40
6	garden_tools	422879e10f46682990de24d770e7f83d
7	garden_tools	368c6c730842d78016ad823897a372db
8	bed_bath_table	06edb72f1e0c64b14c5b79353f7abea3
9	bed_bath_table	99a4788cb24856965c36a24e339b6058

SOLUTION

TABLEAU DASHBOARD

Untuk **Memudahkan** Olist Store dalam **memonitor** performa, berikut produk yang dapat digunakan





06

CONCLUSION & RECOMMENDATION

CONCLUSION: Modelling & Recommendation System

1. Dengan **skor RFM** terbentuk **9** cluster: Cluster **Promising** dengan >33%, **Recent Customers** 19,62% & **Cant Lose Them** 19,46%. Total ketiga cluster ini adalah **3/4** dari total customer Olist Store.
2. Dengan **K-Means** menggunakan Snake Plot terbentuk **5 Cluster**, diurutkan dari yang terbanyak Royal, Jeopardy, Modest, Fresh, Sustain.
3. K-means clustering **lebih baik** dari skor RFM karena bisa **memisahkan** customer yang membeli produk **1x** dan **lebih dari 1x**. Selaras dengan tujuan kita yaitu ingin mengetahui **retensi pelanggan**.
4. Recommendation system yang dibuat dapat menemukan **produk yang cocok** untuk customer pada **setiap clusternya**.



CONCLUSION: Bisnis Olist Store

1. **97%** Pelanggan Olist Store hanya berbelanja **1x**.
2. Potensi Olist Store untuk terus berkembang cukup tinggi, dimana tren **revenue** cenderung **naik**. Pada 2017 terjadi kenaikan **12.431%** (Karena 2016 baru opening) & pada 2018 terjadi kenaikan **21,01%**.
3. Customer Olist Store mayoritas dari state **Sao Paulo (41,90%)**. Payment type yang digunakan **94,82%** adalah Credit Card & Boleto, dan product category paling laris adalah **bed_bath_table** dengan jumlah order 11.115 (9,88%).
4. Bisa menggunakan recommendation system untuk menemukan **produk yang cocok** untuk customer pada suatu customer_state dan/atau suatu cluster.
5. Recommendation yang sudah dibuat diatas bisa menjadi insight tambahan bagi Team Marketing Olist Store dalam **menyusun strategi** untuk meningkatkan retention rate customer terhadap Olist Store.



RECOMMENDATION: Modelling & Recommendation System

1. **Mencoba algoritma** Machine Learning lain untuk unsupervised learning seperti DBScan, K-Prototype, agglomerative clustering, dll. Namun algoritma tersebut membutuhkan spesifikasi komputer yang tinggi.
2. Mencoba membuat **algoritma yang lebih simpel** agar notebook bisa di run dengan waktu yang singkat.
3. Olist Store perlu memperbaiki Environment System supaya data yang diperoleh lebih clean.
4. Data **review score** pelanggan terhadap suatu produk akan sangat berguna untuk membangun sistem rekomendasi yang lebih baik.
5. Olist Store perlu menggunakan **Tableau Dashboard** yang sudah dibuat supaya memudahkan monitoring performa.

RECOMMENDATION: Bisnis Olist Store

Memberikan **Strategi Marketing** sesuai dengan **Karakteristik** Clusternya, berikut contoh Action Plan yang bisa dilakukan

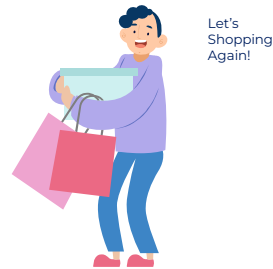
ROYAL



Recency ★★★★★
Frequency ★★★★★
Monetary ★★★★★

Promo **gratis ongkir & diskon produk**.
Approaching menggunakan **e-mail** marketing. Dari simulasi, dapat meningkatkan keuntungan sebesar **R\$ 3.920.446** atau **54,88%** dari revenue tahunan (2017)

SUSTAIN



Recency ★★★★★
Frequency ★★★★★
Monetary ★★★★★

Mengutamakan produk **bed_bath_table**, **furniture_decor**, dan **sports_leisure**.
Approaching menggunakan **e-mail** marketing.

JEOPARDY



Recency ★★★★★
Frequency ★★★★★
Monetary ★★★★★

Dikesampingkan untuk cost reduction. Dari simulasi, dapat menghemat biaya sebesar **R\$ 665.985** atau **9,32%** dari revenue tahunan (2017)

FRESH



Recency ★★★★★
Frequency ★★★★★
Monetary ★★★★★

Memberikan **diskon**.
Approaching menggunakan **Display & Mobile advertising**

MODEST



Recency ★★★★★
Frequency ★★★★★
Monetary ★★★★★

Memberikan **diskon**.
Approaching menggunakan **Display & Mobile advertising**

THANKS

by : AlphaGroup_JC_DS_OL_09

CREDITS: This presentation template was created by [Slidesgo](#), including icons by [Flaticon](#), and infographics & images by [Freepik](#).

