



ANALISIS CLUSTER PELANGGAN BERDASARKAN PERILAKU PELANGGAN DAN TREN PEMBELIAN PRODUK BERBASIS LOKASI

Team 23B - Badak Berkilau



Moh.Syihabuddin

Universitas Jember



Sisfy Anjani

Universitas Riau



Aprilia Tri Widiyastuti

Universitas Jenderal
Achmad Yani Yogyakarta



Hani Agustina

Universitas Komputer
Indonesia



Ikhsan Danu N.

UPN "Veteran"
Jawa Timur



Muhammad Kautsar

Universitas Negeri Jakarta



Putu Rika Sahriana

Mentor



Sarah Gracyntia Juliana

Fasilitator

Table of contents

1

Business Understanding

2

Data Understanding

3

Data Preparation

4

Modeling

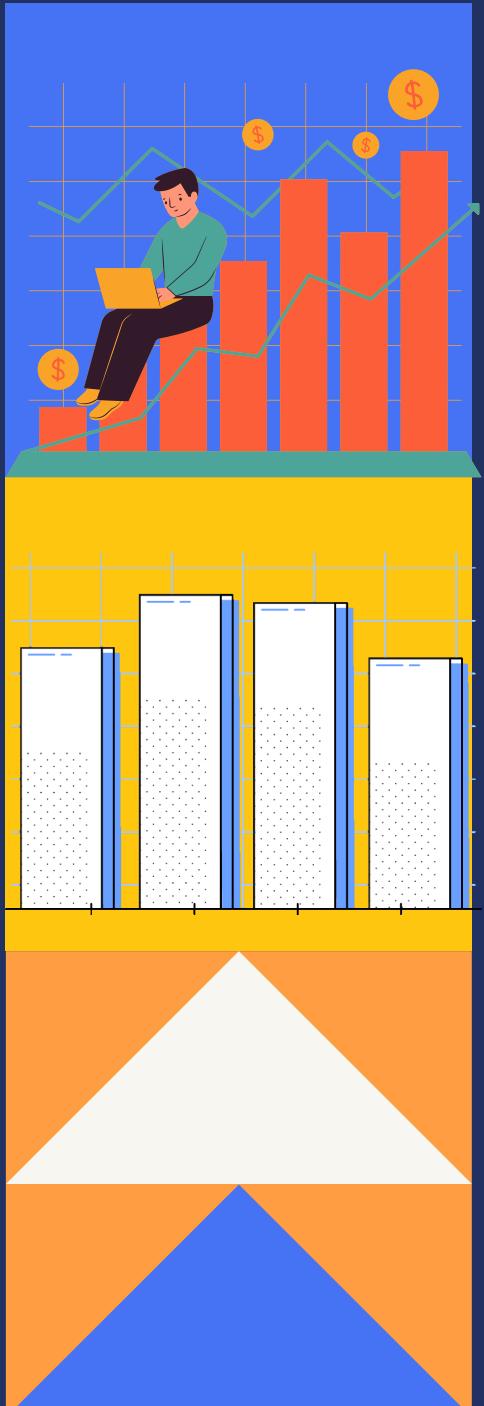
5

Evaluation

6

Insight & Recommendation

Business Understanding



Studi Kasus

SuperApp mengalami kesulitan dalam menargetkan promosi produk yang tepat kepada pelanggan karena kurangnya pemahaman tentang pola pembelian pelanggan yang berbeda di setiap wilayah.

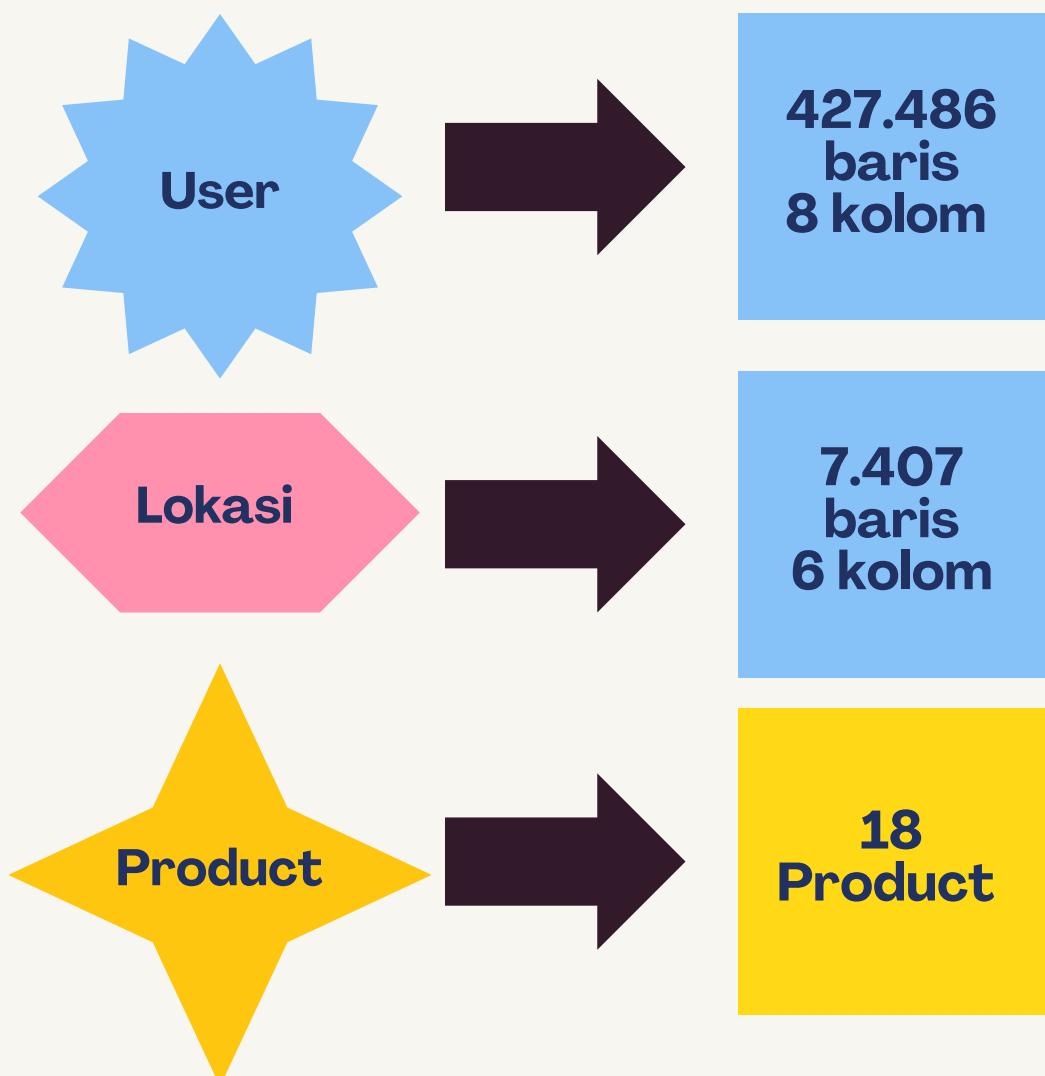


Solusi

Mengidentifikasi segmen pelanggan berdasarkan perilaku pembelian produk di 4 wilayah Provinsi untuk meningkatkan strategi pemasaran.

Data Understanding

Transaksi dari 01 Januari - 31 Desember 2024



Team Trx

Trx 35
Jawa Timur

38 Kabupaten
666 Kecamatan

Trx 18
Lampung

15 Kabupaten
229 Kecamatan

Trx 19
Kepulauan
Bangka
Belitung

7 Kabupaten
47 Kecamatan

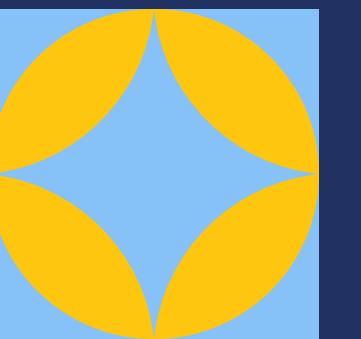
Trx 65
Kalimantan
Utara

5 Kabupaten
55 Kecamatan

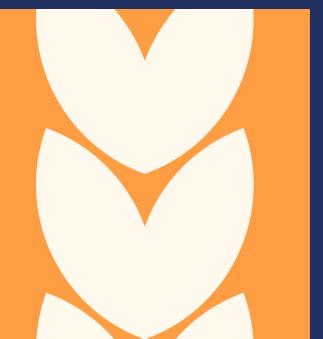
Data Preparation



Data Cleaning

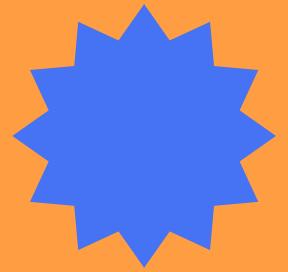


Handling Outlier

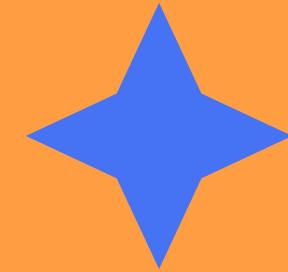


Feature
Engineering

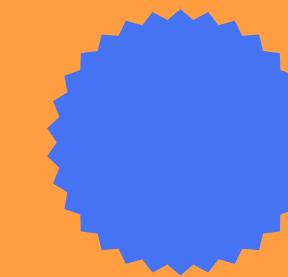
Data Cleaning



Missing Value



Duplicate Data



Negative Value

Missing Value

	Gross Amount	Discount	Baris
Trx 35	13.476	1.548.984	2.261.196
Trx 18	3.330	386.317	570.464
Trx 19	408	57.660	82.544
Trx 65	180	29.380	38.752

Solution?

gross amount di drop dan discounts diisi dengan nilai 0.

Duplicate Value

Tabel

Duplicate Value

Lokasi

2

Negative Value

Tabel

Negative Value

User

5.600

Solution?

Di lakukan drop data pada
lokasi dan user

Negative Value

Tabel	Negative Value
Trx 35	165
Trx 18	49
Trx 19	2
Trx 65	3



Startup
Campus

Kampus
Merdeka
INDONESIA JAYA

Solution?

Di ganti dengan nilai 0

Feature Engineering

Grouping Data

Mengelompokkan data transaksi berdasarkan '`user_id`'. Untuk mengclustering Pelanggan

Fitur Baru

- **Jumlah Transaksi** : Total transaksi yang dilakukan oleh seorang pelanggan
- **Produk Variety** : jenis produk berbeda yang telah dibeli oleh setiap pelanggan

Seleksi Fitur

Memilih fitur '**jumlah transaksi**' dan '**produk variety**' untuk dilakukan clustering

Feature Scaling

Menggunakan Z-score Normalization (StandarScaler)

List Produk Variety

user_id	produk variety	product list
SC1801042001002	11	[Elektronik, Food&drink, Hotel, Investasi, Keperluan rumah tangga, kesehatan, Man Fashion, otomotif, Ride Hailing, Sewa Motor/mobil, Woman Fashion]
SC1801042001006	12	[Elektronik, Food&drink, Hotel, Investasi, Keperluan Anak, Keperluan rumah tangga, kesehatan, Man Fashion, otomotif, Ride Hailing, Travel, Woman Fashion]
SC18010422002008	10	[Elektronik, Food&drink, Hotel, Investasi, Keperluan rumah tangga, kesehatan, Man Fashion, otomotif, Ride Hailing, Sewa Motor/mobil, Travel, Woman Fashion]

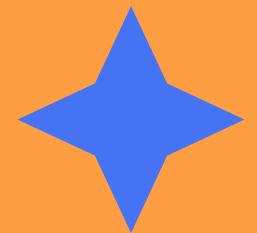
Modelling

Pilihan Metode/Algoritma



Model Yang
Digunakan

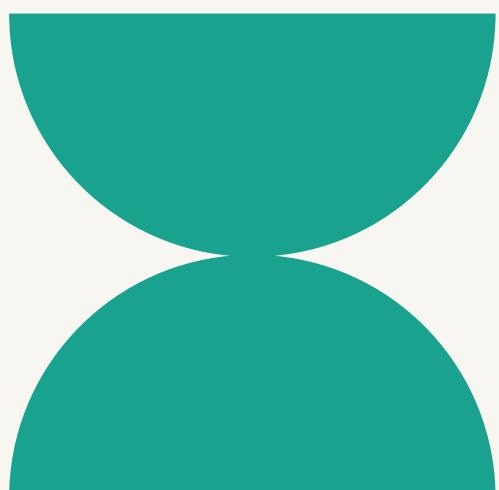
K-Means



Model
Komparasi

HDBSCAN

Hierarchical Clustering



Alasan Pemilihan Model

K-Means

sangat efisien dan cepat untuk dataset besar

HDBSCAN

hasil cluster yang dihasilkan banyak

Hierarchical Clustering

Lebih cocok digunakan untuk dataset kecil

Memanfaatkan memori dan prosesor dengan efisien

Lebih lambat dibandingkan K-means pada dataset besar.

Waktu komputasi lebih lama pada dataset besar.

Mudah divisualisasikan menggunakan scatter plot

Interpretasi hasil clustering bisa lebih rumit

Sulit menangani data dengan noise secara efektif.

Metrik Evaluasi Model

Silhouette Score sebagai Metric Evaluasi

Silhouette Score	3	4	5
K-Means	0.46	0.44	0.43
Hierarchical Clustering	0.43	0.45	0.40

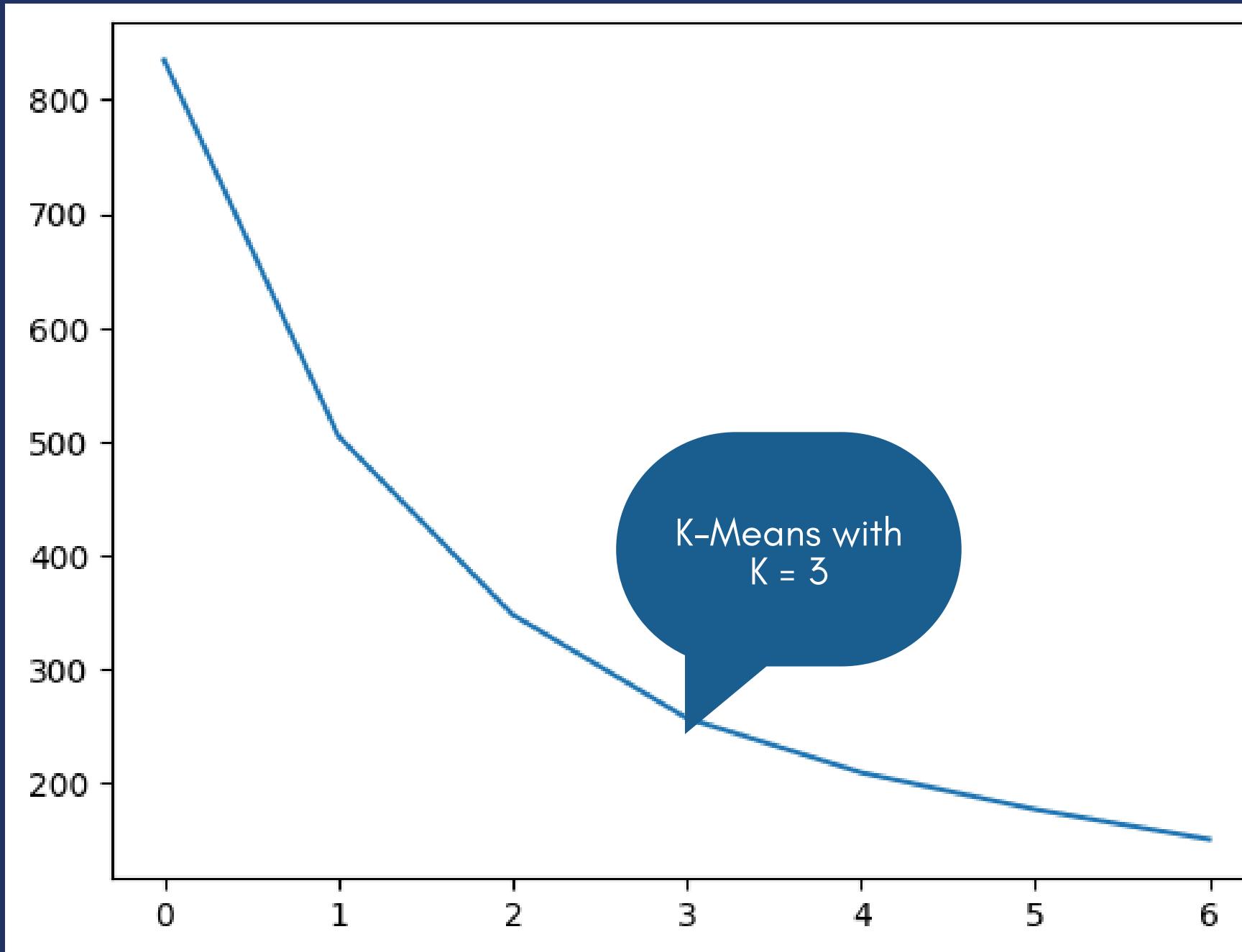
Dari hasil metrik evaluasi

K-Means memberikan kualitas klaster baik dengan 3 klaster (Silhouette Score: 0.46). HDBSCAN memiliki Silhouette Score tinggi, namun menghasilkan banyak klaster. Hierarchical Clustering menunjukkan hasil yang baik, tetapi jumlah klaster perlu dipertimbangkan. Oleh karena itu, K-Means adalah pilihan terbaik untuk analisis clustering.

HDBSCAN			
Min Cluster Size	5	50	100
Jumlah Cluster	803	381	275
Silhouette Score	0.98	0.85	0.43

K-Means Clustering

Mengelompokkan data sesuai karakteristik



Elbow Method

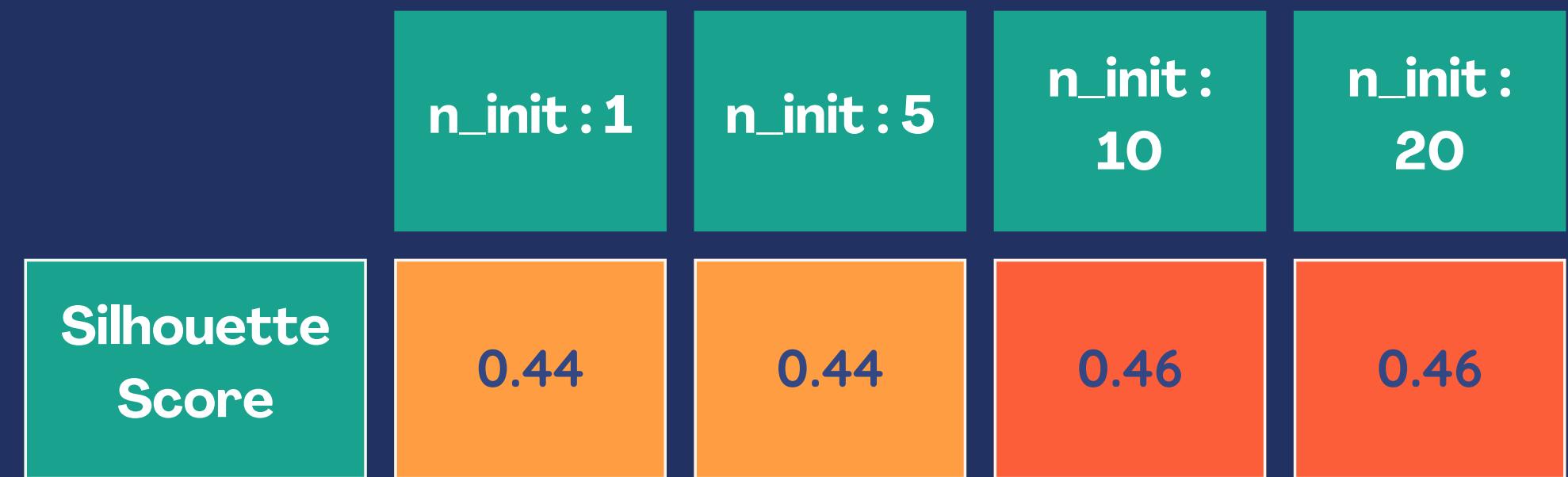
Menentukan jumlah cluster/kelompok yang optimal yaitu $K = 3$

Silhouette Score

Jumlah cluster = 3 menghasilkan nilai silhouette score **0.46**

Hyperparameter pada Model K-means Clustering

Untuk meningkatkan akurasi pengelompokan



Sudah dilakukan hyperparameter tuning, dan untuk hasil silhouette score masih tetap dan **tidak ada perubahan**

Interpretasi Model dan Hasil Evaluasi

Interpretasi :

- Nilai mendekati 1: Data point berada jauh dari cluster tetangga.
- Nilai mendekati 0: Data point berada di perbatasan antara dua cluster.
- Nilai mendekati -1: Data point mungkin ditempatkan di cluster yang salah.

Karakteristik dan Insight Clustering

Karakteristik CLuster

Cluster	Avg Jumlah Transaksi	Avg Produk Variety
0	39.80	13.03
1	22.28	7.65
2	31.60	10.62

Insight



Cluster 0: Jumlah transaksi dan produk variety yang **tinggi**. Menunjukkan bahwa mereka adalah **Pelanggan Sangat Aktif**



Cluster 1: Jumlah transaksi dan produk variety **rendah**. Menunjukkan bahwa mereka adalah **Pelanggan Kurang Aktif**



Cluster 2: Jumlah transaksi dan produk variety **sedang**. Menunjukkan bahwa mereka adalah **Pelanggan Aktif**

Rekomendasi

Strategi Promosi

Cluster 0

Lebih banyak diskon untuk pembelian dalam jumlah besar

Kustomisasi Penawaran

Menawarkan produk yang tidak pernah dibeli pada user

Cluster 1

Pemberian diskon yang lebih besar untuk pertama kali setelah lama tidak berbelanja

Memberikan penawaran pada variasi produk yang sudah dibeli untuk menarik perhatian user

Cluster 2

Memberikan promo pada produk yang sering dibeli

Menawarkan variasi produk lain yang sudah dibeli oleh user



Dashboard

ANALISIS PENGGUNA DAN TREND PEMBELIAN | Overview

By Badak Berkilau

Overview Cluster

Kampus Merdeka

Jumlah Pelanggan: 90,270

Jumlah Transaksi Pengguna: 98,285,666

Rata - Rata Diskon: 4,383

Product Name: All

Gender: All

Month of Transaction Date: All

Birth: To Null

Nama Provinsi: All

Nama Kabupaten: All

Nama Kecamatan: All

Distribusi Demografis Pengguna Di Setiap Provinsi

Menunjukkan jumlah pengguna perempuan lebih banyak yakni 2,038,630 dibandingkan laki-laki yang hanya 870,760

Distribusi Demografis Pengguna di Setiap Provinsi

Distribusi pengguna menunjukkan bahwa perempuan paling banyak di Provinsi Jawa Timur dengan jumlah 98,009, paling rendah ada di Kalimantan Utara dengan jumlah 977. Untuk laki-laki paling banyak di Jawa Timur dengan jumlah 61,271, paling rendah ada di Kalimantan Utara dengan jumlah 1,696

Tren Pembelian berdasarkan demografi pengguna di setiap provinsi

Pertumbuhan preferensi pembelian untuk anak tahun ke-2000an lebih suka belanja dibandingkan dengan generasi sebelumnya dan ketarut aman yang belum ada di Jawa

Bagaimana pengaruh discount terhadap jumlah transaksi di berbagai wilayah

Diskon tidak mempengaruhi jumlah transaksi pelanggan, jumlah transaksi pelanggan lebih banyak ketika barang sedang tidak d.

Apa tren pembelian dari waktu ke waktu untuk setiap produk atau kategori produk

Pembelian setiap produk paling laris terjadi pada bulan Januari

Produk yang paling populer

Menampilkan produk terlaris untuk meningkatkan stock barang dan promosi barang yang kurang laku, produk paling banyak yakni kategori Food and Drink

Tingkat Retensi pengguna dari waktu ke waktu di setiap provinsi

Status retensi pengguna yang baru lebih banyak dibandingkan dengan retensi lama. Data ini tidak ada status retensi tidak aktif karena masih ada aktivitas pengguna.

Transaksi Pengguna paling sering terjadi secara geografis

Treemap untuk menunjukkan mana aktivitas transaksi paling tinggi berdasarkan provinsi

Peta Provinsi

© 2024 Mapbox © OpenStreetMap

Peta interaktif yang menunjukkan lokasi geografis dari transaksi pembelian pengguna

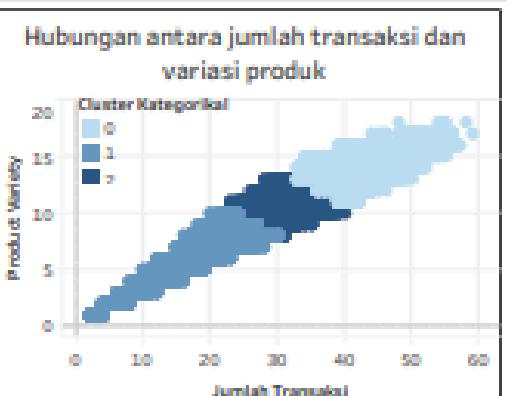
ANALISIS PENGGUNA DAN TREND PEMBELIAN | Cluster

By Badak Berkilau

Overview **Cluster**  **Kampus Mandiri**

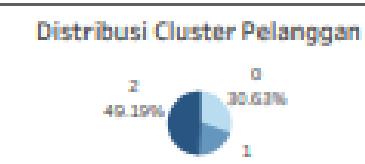
Jumlah Pelanggan	Rata-rata Jumlah Transaksi	Rata-rata Variasi Produk
90,270	33	11

Hubungan antara jumlah transaksi dan variasi produk



Terdapat Korelasi Positif antara jumlah transaksi dan variasi produk di semua cluster. Pelanggan dengan lebih banyak transaksi cenderung membeli ragam produk yang lebih besar.

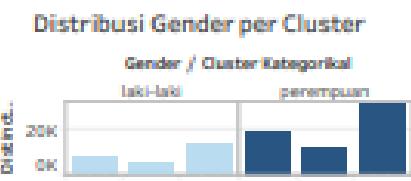
Distribusi Cluster Pelanggan



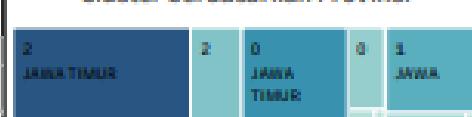
Distribusi Usia per Cluster



Distribusi Gender per Cluster

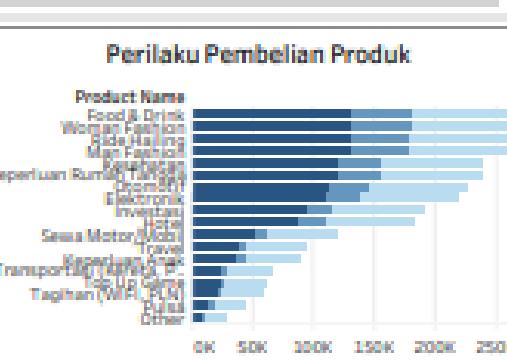


Distribusi Pelanggan untuk setiap Cluster berdasarkan Provinsi

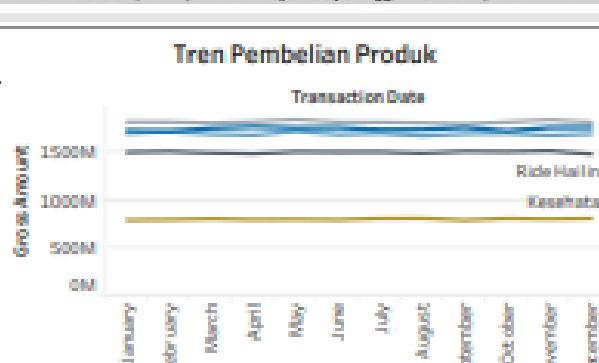


Cluster 2 memiliki jumlah pelanggan terbesar, yaitu 44.407 orang dengan proporsi 49.19% dari total pelanggan. Cluster 1 didominasi oleh perempuan dan mayoritas pelanggan berada di provinsi Jawa Timur.

Perilaku Pembelian Produk



Tren Pembelian Produk



Top 3 produk yang paling sering dibeli pelanggan adalah Food & Drink, Woman Fashion, dan Ride Hailing. Ride Hailing dan Kesehatan menunjukkan tren yang stabil dengan gross amount yang konstan sepanjang tahun...

Pengaruh diskon terhadap jumlah transaksi per cluster



Diskon secara signifikan meningkatkan jumlah transaksi di semua kelas. Pelanggan lebih cenderung melakukan pembelian saat diskon diberikan.

Profiling pelanggan

Cluster	User ID	Full Name	Avg. Jumlah Transaksi	Avg. Product Variety
0	SC1801042003006	RIVQY	40.00	14.00
0	SC1801042003012	KRESWATI ENNY	40.00	13.00
0	SC1801042004005	NASYIRUL AL	43.00	13.00
0	SC1801042004023	SUKAWI RAHMAT	36.00	12.00
0	SC1801042004025	HENDRAWAN	38.00	13.00

INSIGHT

- Cluster 0: Jumlah Transaksi dan Product Variety tinggi
- Cluster 1: Jumlah Transaksi dan Product Variety rendah
- Cluster 2: Jumlah Transaksi dan Product Variety sedang

REKOMENDASI

- Cluster 0: 1) Lebih banyak diskon untuk pembelian dalam jumlah besar 2) Memperkenalkan produk yang tidak pernah dibeli pada user
- Cluster 1: 1) Pemberian diskon yang lebih besar untuk pertama kali sejauh lama tidak berberlنجa 2) Memberikan penawaran pada varasi produk yang sudah dibeli untuk menarik perhatian user
- Cluster 2: 1) Memberikan promo pada produk yang sering dibeli..

Terima Kasih

Team 23B - Badak Berkilau

