

# POROFOLIO APLIKASI KLUSTERING CUSTOMER

---

Proyek: Aplikasi Klustering Customer

Bahasa Pemrograman: Python

Framework: Streamlit



## Analisa Klusterisasi Customer

## **1. Deskripsi Proyek**

Aplikasi Klustering Customer ini dikembangkan menggunakan Python dan Streamlit. Tujuan aplikasi adalah melakukan segmentasi pelanggan (customer segmentation) berdasarkan data karakteristik tertentu seperti usia, pendapatan, dan perilaku belanja. Dengan teknik unsupervised learning seperti K-Means, aplikasi membantu memahami kelompok pelanggan yang memiliki kesamaan karakteristik sehingga dapat digunakan untuk strategi bisnis dan pemasaran.

## **2. Struktur Folder dan File**

Berikut struktur folder dari proyek:

- extracted\_customer\_project/
  - App\_Klustering\_Customer/
    - customer\_shopping\_data.csv
    - customer\_shopping\_data.xlsx
    - Klustering\_Customer.py
    - requirements.txt
    - utils.py
    - assets/
      - logo.png
    - pages/
      - 1\_Upload Dataset.py
      - 2\_Analisis Algoritma.py
      - 3\_Informasi Dataset.py
    - \_pycache\_/
      - utils.cpython-38.pyc

## **3. Kebutuhan Sistem (Requirements)**

Untuk menjalankan aplikasi ini, diperlukan:

- Python 3.11
- Streamlit
- Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn, Plotly
- File requirements.txt berisi daftar dependensi lengkap

## **4. Langkah Instalasi & Menjalankan Aplikasi**

1. Ekstrak file ZIP proyek ke dalam direktori pilihan Anda.
2. Buka terminal atau command prompt, lalu masuk ke folder proyek:

```
cd App_Klustering_Customer
```

3. Instal semua dependensi menggunakan perintah:

```
pip install -r requirements.txt
```

4. Jalankan aplikasi Streamlit dengan perintah:

```
streamlit run main.py (atau file utama aplikasi Anda)
```

5. Aplikasi akan terbuka di browser default, biasanya di alamat <http://localhost:8501>

## 5. Penjelasan Modul dan File Utama

- main.py: File utama Streamlit yang menjalankan aplikasi dan menampilkan halaman utama.
- utils.py: Berisi fungsi bantu seperti preprocessing data, perhitungan jarak antar data, dan penerapan algoritma klustering.
- pages/: Folder berisi halaman tambahan Streamlit seperti upload dataset, hasil klustering, dan visualisasi data.
- assets/: Berisi file aset seperti logo atau gambar tambahan.
- requirements.txt: Berisi daftar pustaka Python yang dibutuhkan.
- dataset.csv: Dataset utama yang digunakan untuk analisis dan klustering pelanggan.

## 6. Screenshot Tampilan Aplikasi

Bagian ini disediakan untuk menambahkan screenshot masing-masing fitur aplikasi.

- Halaman Utama Aplikasi (Home Page)



- Data Preparation

User dapat melihat hasil load dataset customer. Selain itu juga terdapat data missing value atau data kotor. Jika ada maka akan dibersihkan datanya menggunakan konsep median. Dan apabila data duplikat maka akan dihapus data tersebut.

The screenshot shows a user interface for data preparation. On the left, there's a sidebar with options: 'Klustering Customer', 'Data Preparation' (which is selected and highlighted in grey), 'Analisis Algoritma', and 'Informasi Dataset'. Below the sidebar is a section titled 'Analisa Klusterisasi Customer' featuring a circular icon with four colored segments (blue, yellow, green, orange) and arrows indicating data flow or clustering. The main area is titled 'Data Preparation' with a wrench icon. It displays a message: 'Jumlah missing values sebelum dibersihkan: 0' and a green success message: 'Pembersihan selesai! Missing values tersisa: 0'. Below this is a section titled 'Preview Dataset (Setelah Pembersihan)' with a magnifying glass icon. A table shows 9 rows of cleaned dataset data:

	invoice_no	customer_id	gender	age	category	quantity	price	payment_method	invoice_date	shopping_i
0	I138884	C241288	Female	28	Clothing	5	1500.4	Credit Card	5/8/2022	Kanyon
1	I317333	C111565	Male	21	Shoes	3	1800.51	Debit Card	12/12/2021	Forum Ista
2	I127801	C266599	Male	20	Clothing	1	300.08	Cash	9/11/2021	Metrocity
3	I173702	C988172	Female	66	Shoes	5	3000.85	Credit Card	16/05/2021	Metropol A
4	I337046	C189076	Female	53	Books	4	60.6	Cash	24/10/2021	Kanyon
5	I227836	C657758	Female	28	Clothing	5	1500.4	Credit Card	24/05/2022	Forum Ista
6	I121056	C151197	Female	49	Cosmetics	1	40.66	Cash	13/03/2022	Istinye Parl
7	I293112	C176086	Female	32	Clothing	2	600.16	Credit Card	13/01/2021	Mall of Ista
8	I293455	C159642	Male	69	Clothing	3	900.24	Credit Card	4/11/2021	Metrocity

- Halaman Analisis Algoritma

Pada halaman ini user dipermudah dengan ada fitur tanggal, serta pengaturan data juga terdapat aturan pilihan ataupun input manual data. Kemudian user dipermudah dalam memiliki nilai K. Pada kasus ini saya mendefinisikan nilai K hanya sampai 5:

K=1 → High Value Customers (Belanja sering, jumlah besar, nilai transaksi tinggi)

K=2 → Medium Value Customers (Belanja cukup sering, jumlah sedang)

K=3 → Low Value Customers (Belanja jarang, jumlah kecil, nilai transaksi rendah)

K=4 → Bargain Shoppers (Belanja quantity tinggi tapi dengan harga murah)

K=5 → Occasional Buyers (Belanja sesekali, jumlahnya random)

Silhouette Score digunakan untuk menilai kualitas klaster (cluster quality) pada algoritma unsupervised learning seperti K-Means, DBSCAN, atau Agglomerative Clustering.

Silhouette Score selalu berada dalam rentang -1 hingga 1:

- +1 Data sangat dekat dengan klasternya sendiri dan jauh dari klaster lain (Klaster sangat jelas)
- =0 Data berada di antara dua klaster (tidak jelas masuk ke mana) (Klaster tumpang tindih)
- 1 Data lebih dekat ke klaster lain dibanding klaster sendiri (Klaster salah terbentuk)

Nilai Silhouette Score sebesar 0.368 menunjukkan bahwa hasil pengelompokan model cukup baik, meskipun masih terdapat tumpang tindih antar klaster. Nilai ini menandakan bahwa sebagian besar data sudah berada di klaster yang sesuai, namun beberapa titik data masih berada di area perbatasan antar klaster. Model masih dapat ditingkatkan dengan menyesuaikan jumlah klaster atau fitur yang digunakan.

The screenshot shows a dark-themed user interface for customer clustering analysis. On the left, a sidebar menu includes 'Klustering Customer', 'Data Preparation', 'Analisis Algoritma' (which is highlighted), and 'Informasi Dataset'. Below the menu is a graphic of three stylized human figures in blue, green, and orange, with arrows indicating interaction. The main panel has a title 'Analisis Klusterisasi Customer' with a bar chart icon. It features a 'Filter Tanggal Invoice' section with a date range from '2021/01/01 – 2023/03/08'. Below this is a 'Pengaturan Clustering' section with two dropdown menus: 'Pilih cara menentukan jumlah data' (set to 'Pilih dari daftar') and 'Pilih jumlah data untuk clustering' (set to '100'). A slider for 'Pilih jumlah cluster (k)' is set to '5'. At the bottom, a green bar displays the 'Silhouette Score: 0.368'.

Preview hasil clustering dengan K=5

The screenshot shows a table titled 'Preview Hasil Clustering' with a magnifying glass icon. The table has 12 columns: gender, age, category, quantity, price, payment\_method, invoice\_date, shopping\_mall, cluster, and cluster\_label. The data consists of five rows:

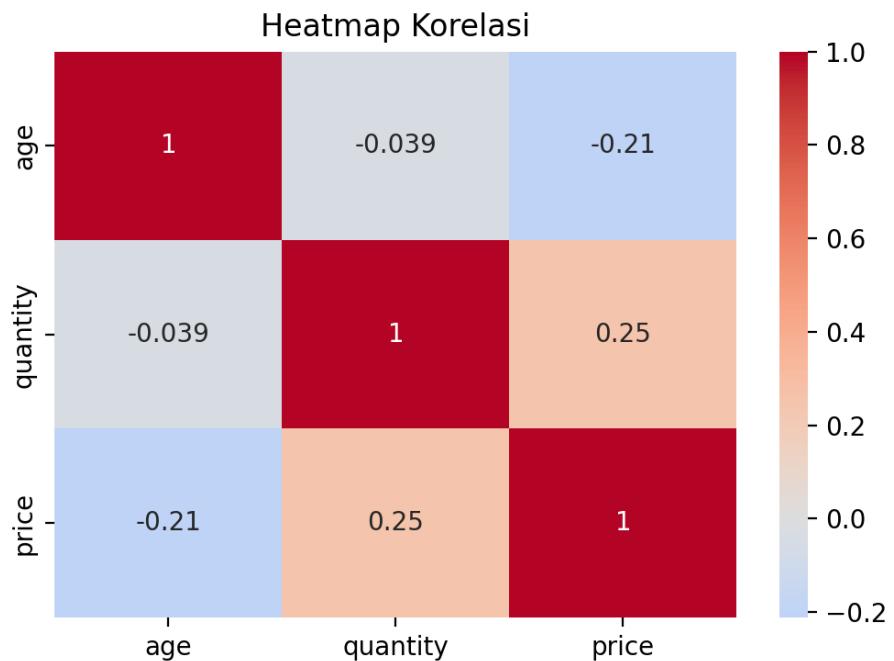
	gender	age	category	quantity	price	payment_method	invoice_date	shopping_mall	cluster	cluster_label
65475	Female	56	Food & Beverage	4	20.92	Cash	2021-05-14 00:00:00	Kanyon	4	Occasional Buyers
96548	Male	22	Food & Beverage	5	26.15	Cash	2022-09-09 00:00:00	Metrocity	0	High Value Customers
33949	Female	53	Food & Beverage	4	20.92	Debit Card	2021-04-01 00:00:00	Kanyon	4	Occasional Buyers
58198	Female	65	Food & Beverage	5	26.15	Credit Card	2022-08-03 00:00:00	Cevahir AVM	4	Occasional Buyers
71781	Female	51	Shoes	4	2400.68	Credit Card	2021-04-15 00:00:00	Metropol AVM	3	Bargain Shoppers

Visualisasi Clustering.

The screenshot shows a visualization interface for clustering. It features two dropdown menus: 'Pilih sumbu X' (set to 'age') and 'Pilih sumbu Y' (set to 'price').



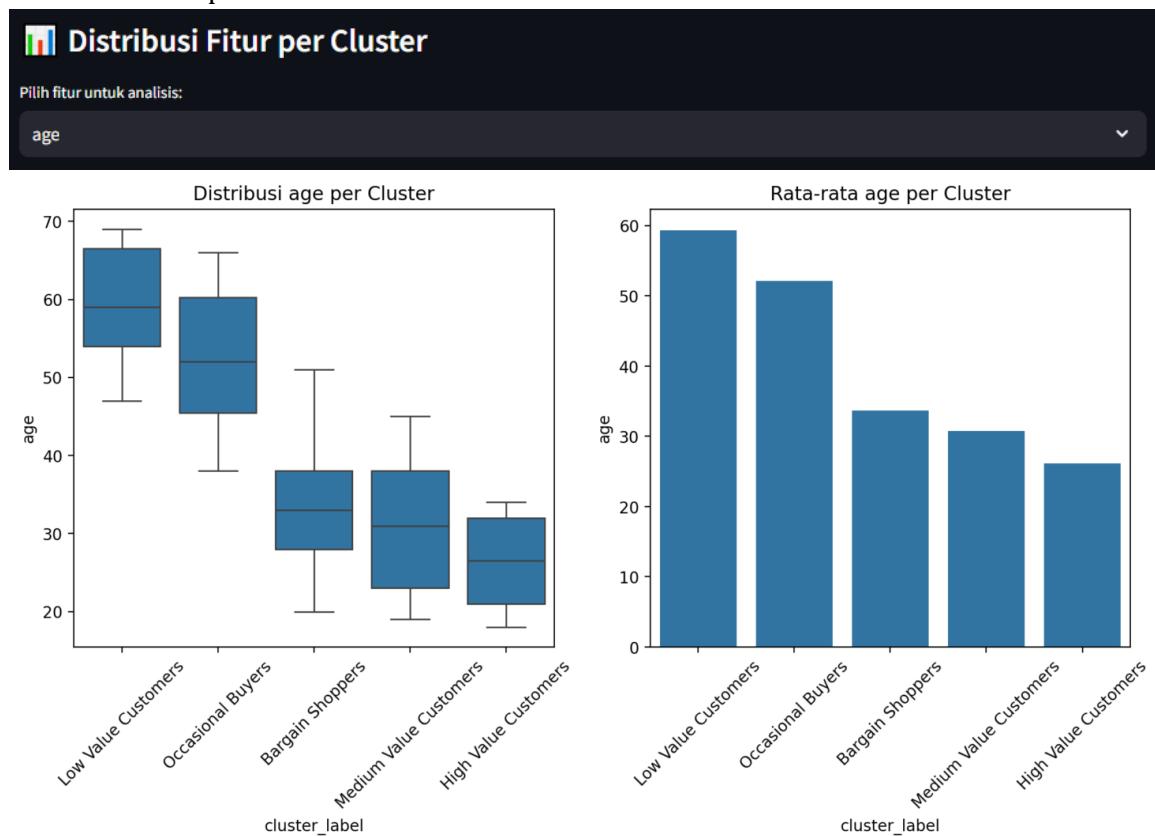
Korelasi antar variabel (Heatmap)



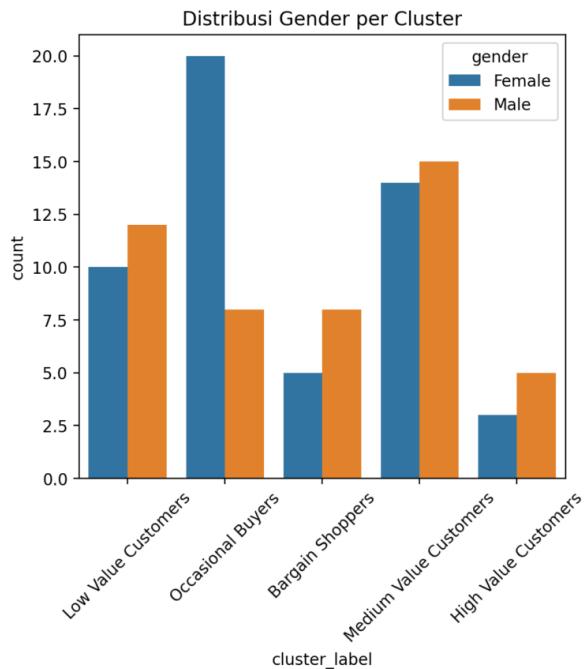
## Ringkasan Cluster

cluster_label	age		quantity		price	
	mean	count	mean	count	mean	count
Bargain Shoppers	33.6923	13	3.4615	13	2631.1485	13
High Value Customers	26.125	8	4.75	8	430.8538	8
Low Value Customers	59.3636	22	1.5455	22	203.5114	22
Medium Value Customers	30.7586	29	2	29	369.3724	29
Occasional Buyers	52.1429	28	4.0714	28	410.0579	28

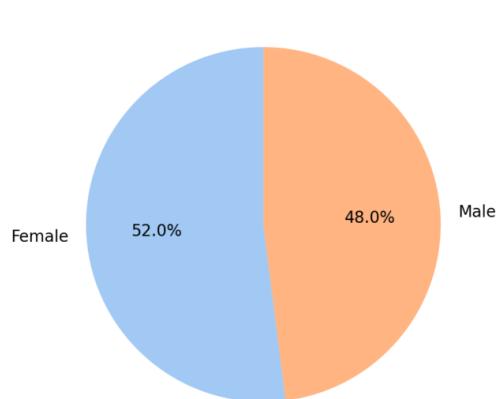
## Distribusi Fitur per Cluster



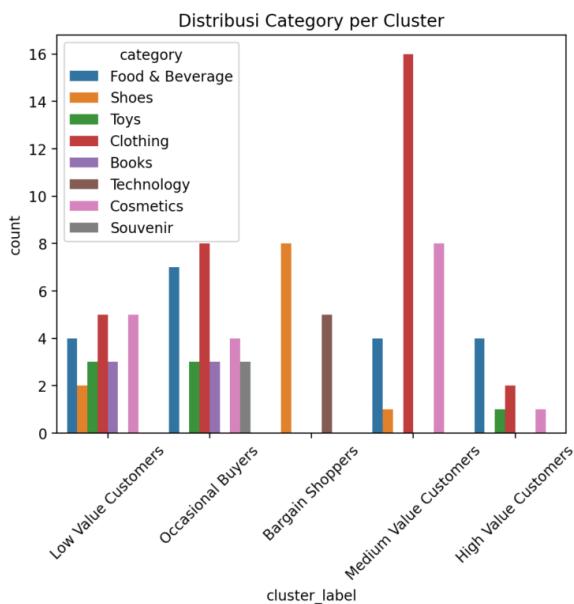
### Distribusi Gender per Cluster



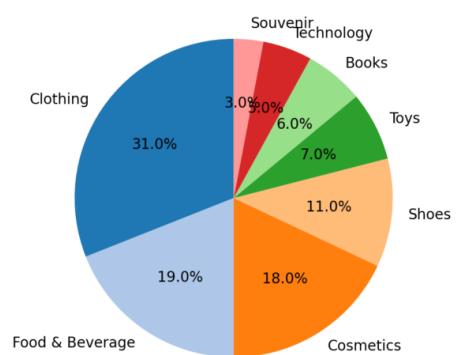
### Percentase Gender Keseluruhan



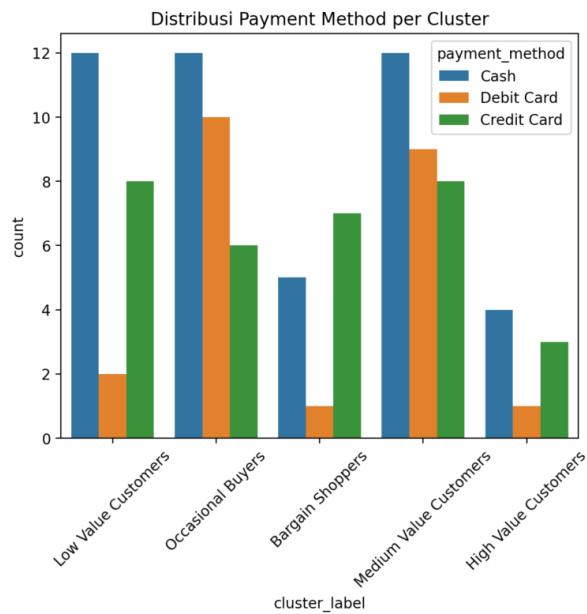
### Distribusi Category per Cluster



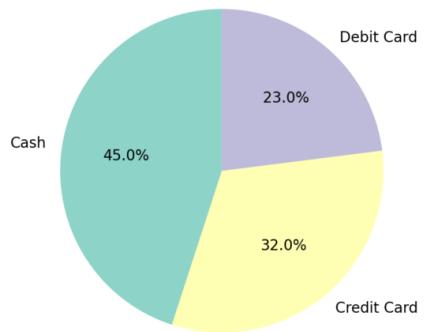
### Percentase Category Keseluruhan



### Distribusi Payment Method per Cluster

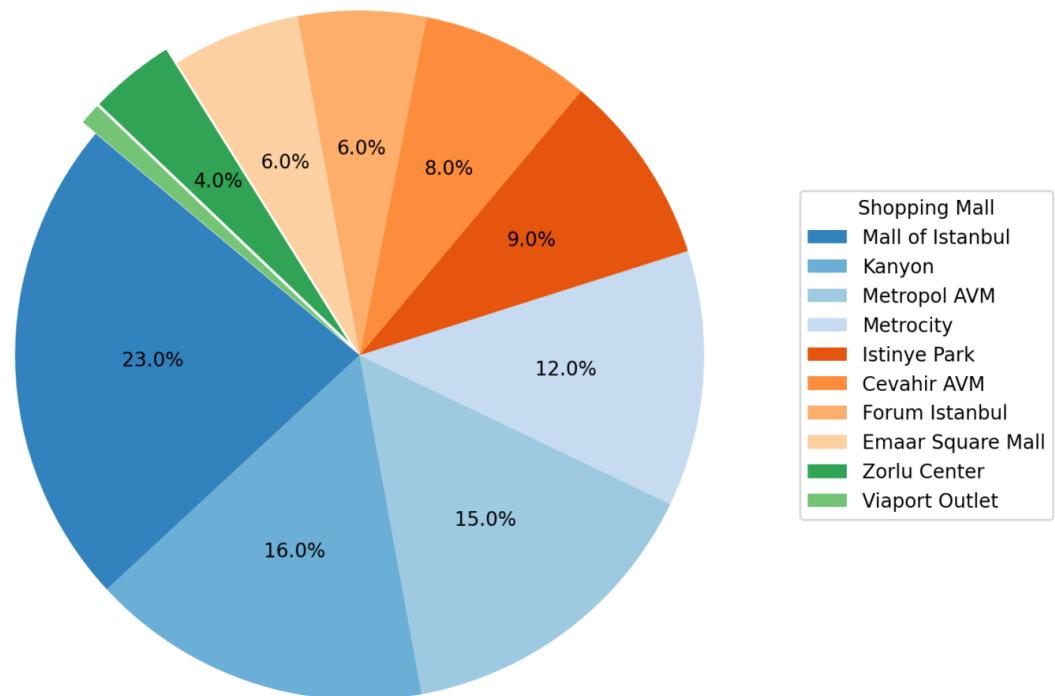


### Percentase Payment Method Keseluruhan



### Percentase Shopping Mall

#### Percentase Shopping Mall Keseluruhan



Simpan hasil clustering dalam bentuk CSV dan Excel

The screenshot shows a web interface for saving clustering results. It features a dark header with the title "Simpan Hasil Clustering". Below the header are two buttons: "Download Hasil Clustering (CSV)" and "Download Hasil Clustering (Excel + Summary)". The main area displays a Microsoft Excel spreadsheet titled "hasil\_clustering - Microsoft Excel". The spreadsheet contains a single sheet named "ClusteringResult" which lists 21 data rows. Each row includes columns for invoice, customer\_id, gender, age, category, quantity, price, payment method, and cluster\_label. The "cluster\_label" column provides a categorical grouping for each record. The Excel ribbon at the top shows standard tabs like FILE, HOME, INSERT, etc., and various toolbars for font, alignment, and data manipulation.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	invoice	customer_id	gender	age	category	quantity	price	payment method	debit_card	cluster	cluster_label									
2	1704758	C971218	Female	56	Food & Be	4	20,92	Cash	#####	Kanyon	4	Occasional Buyers								
3	1218404	C949846	Male	22	Food & Be	5	26,15	Cash	#####	Metrocity	0	High Value Customers								
4	1195516	C440398	Female	53	Food & Be	4	20,92	Debit Card	#####	Kanyon	4	Occasional Buyers								
5	1833735	C113971	Female	65	Food & Be	5	26,15	Credit Card	#####	Cevahir A'	4	Occasional Buyers								
6	1144592	C529879	Female	51	Shoes	4	2400,68	Credit Card	#####	Metropol	3	Bargain Shoppers								
7	1104653	C209092	Male	62	Toys	5	179,2	Cash	#####	Cevahir A'	4	Occasional Buyers								
8	1866357	C254260	Female	20	Clothing	1	300,08	Debit Card	#####	Mall of Ist	3	Medium Value Customers								
9	1101429	C148693	Female	50	Books	4	60,6	Credit Card	#####	Mall of Ist	4	Occasional Buyers								
10	1202480	C255862	Female	54	Clothing	2	600,16	Cash	#####	Metropol	2	Low Value Customers								
11	1179825	C787552	Female	35	Technolog	3	3150	Credit Card	#####	Mall of Ist	3	Bargain Shoppers								
12	1788067	C122409	Female	41	Cosmetics	1	40,66	Cash	#####	Metropol	1	Medium Value Customers								
13	1259820	C367305	Female	69	Books	2	30,3	Credit Card	#####	Mall of Ist	2	Low Value Customers								
14	1546029	C232750	Female	50	Books	5	75,75	Credit Card	#####	Istinye Pa	4	Occasional Buyers								
15	1285144	C266019	Female	25	Clothing	3	900,24	Debit Card	#####	Cevahir A'	1	Medium Value Customers								
16	1447841	C117147	Male	60	Cosmetics	4	162,64	Debit Card	#####	Mall of Ist	4	Occasional Buyers								
17	1847315	C207762	Male	20	Shoes	4	2400,68	Debit Card	#####	Metropol	3	Bargain Shoppers								
18	1380336	C285904	Female	67	Books	2	30,3	Cash	#####	Kanyon	2	Low Value Customers								
19	1225089	C060826	Male	35	Technolog	3	3150	Cash	#####	Metropol	3	Bargain Shoppers								
20	1559235	C179544	Male	62	Food & Be	2	10,46	Credit Card	#####	Metropol	2	Low Value Customers								
21	1124462	C385569	Male	38	Cosmetics	3	121,98	Debit Card	#####	Mall of Ist	1	Medium Value Customers								

## 7. Catatan Pengembangan

Aplikasi ini dapat dikembangkan lebih lanjut dengan fitur tambahan seperti:

- Menambahkan algoritma lain seperti DBSCAN atau Hierarchical Clustering
- Menyimpan hasil segmentasi ke database
- Menampilkan dashboard interaktif menggunakan Plotly atau Power BI
- Fitur ekspor hasil analisis ke file Excel atau PDF

## 8. Profil dan Kontak

Nama: Arjun Yuda Firwanda

Role: Team Program Liraa Job

Email: (arjunyudafirwanda@gmail.com)

LinkedIn: (<https://www.linkedin.com/in/arjun-yuda-firwanda-8095261a4/>)

GitHub: (<https://github.com/arjunyudafirwanda/>)