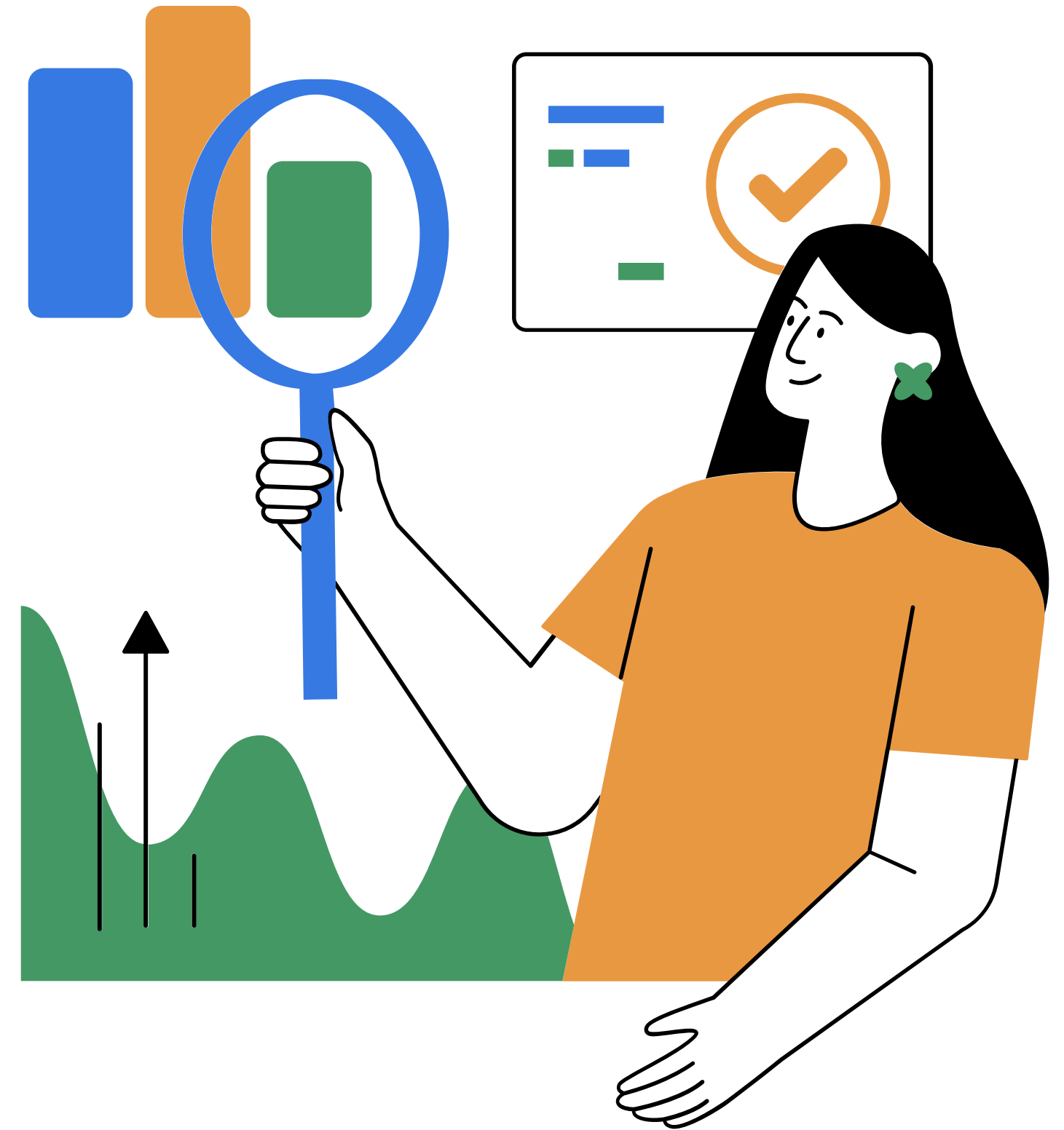
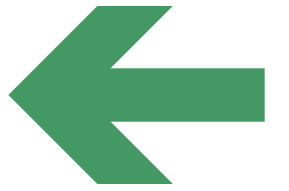


Credit Card Transaction Analysis

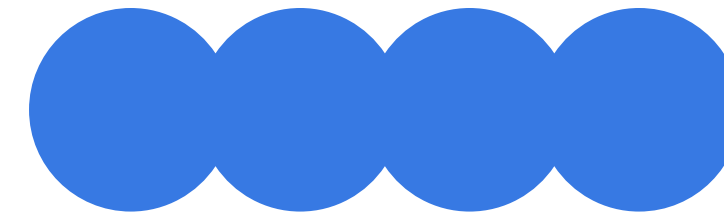
By Shifwa Luqyaanaa





List of Content

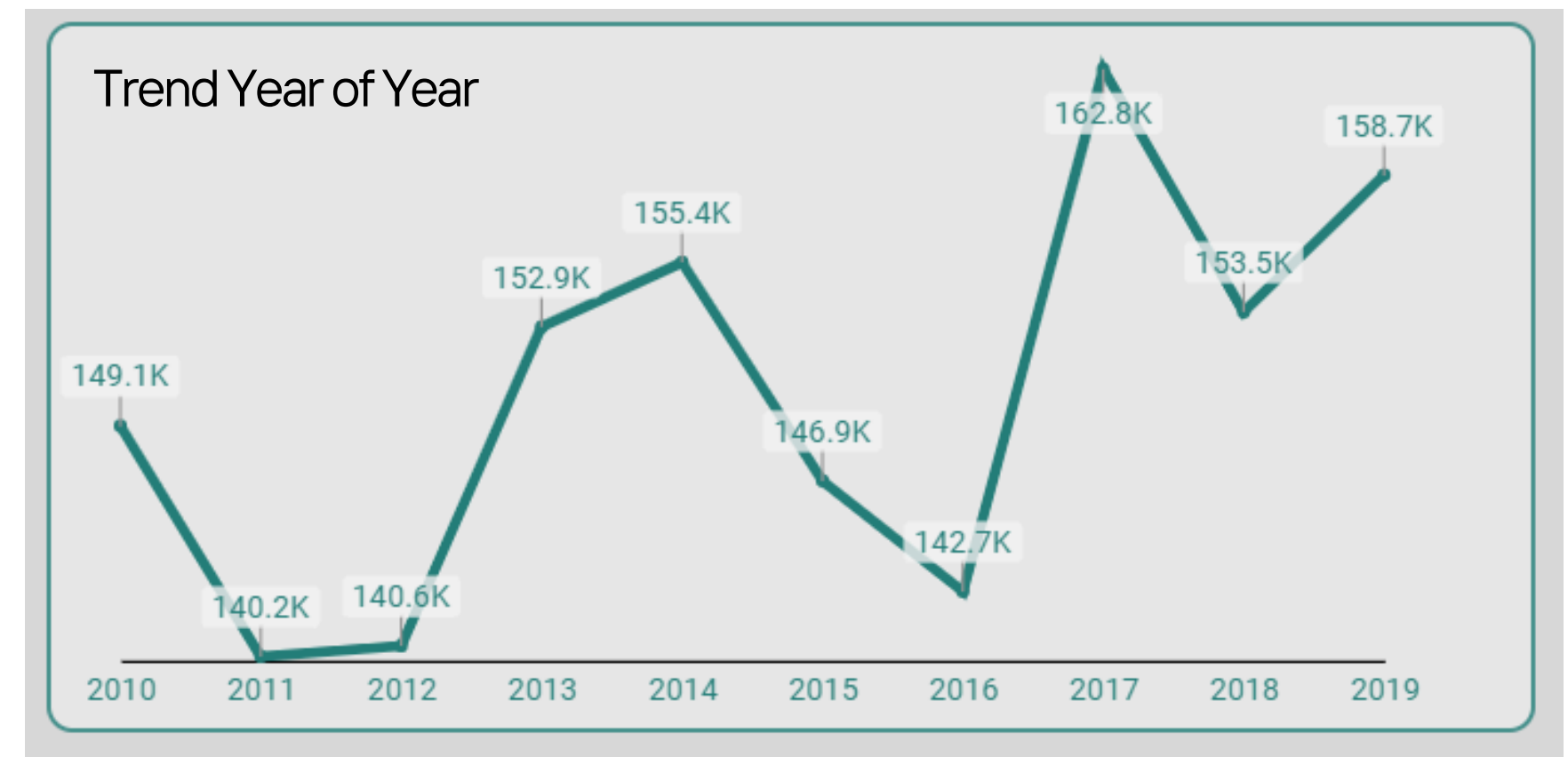
1. Data Overview
2. Scope of Problem
3. Problem Formulation
4. Methodology
5. Data Source
6. Analysis Result
7. Dashboard
8. Insight and Recommendation



Data Overview

Dataset yang digunakan merupakan data historis transaksi kartu kredit selama rentang waktu 10 tahun (2010 - 2019) dengan volume lebih dari 1 juta transaksi records. Dataset ini terdiri dari 3 tabel utama yaitu:

- Tabel Transaksi: Mencatat detail setiap transaksi (ID pelanggan, Amount, tanggal dan waktu, lokasi, merchant, dll).
- Tabel Pengguna: Berisi profil demografis dan finansial pelanggan (usia, pendapatan, skor kredit, dll).
- Tabel Kartu: Merinci atribut setiap kartu (merek, jenis, limit kredit, dll).



Scope Of Problem

Analisis ini dilakukan untuk mendapatkan insight dari dataset yang tersedia. Ruang lingkup proyek ini didefinisikan sebagai berikut:

Analisis Deskriptif:

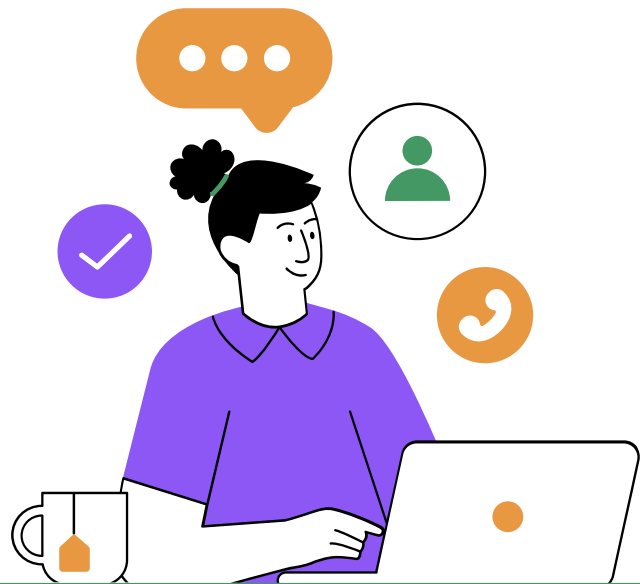
- Mengidentifikasi tren transaksi berdasarkan waktu (tahunan, per jam).
- Menganalisis distribusi transaksi berdasarkan kategori, seperti metode pembayaran, merek kartu, dan kategori merchant (MCC).

Analisis Perilaku Pelanggan:

- Memahami pola belanja dan loyalitas pelanggan berdasarkan demografi (gender dan usia) .

Segmentasi Pelanggan (Machine Learning):

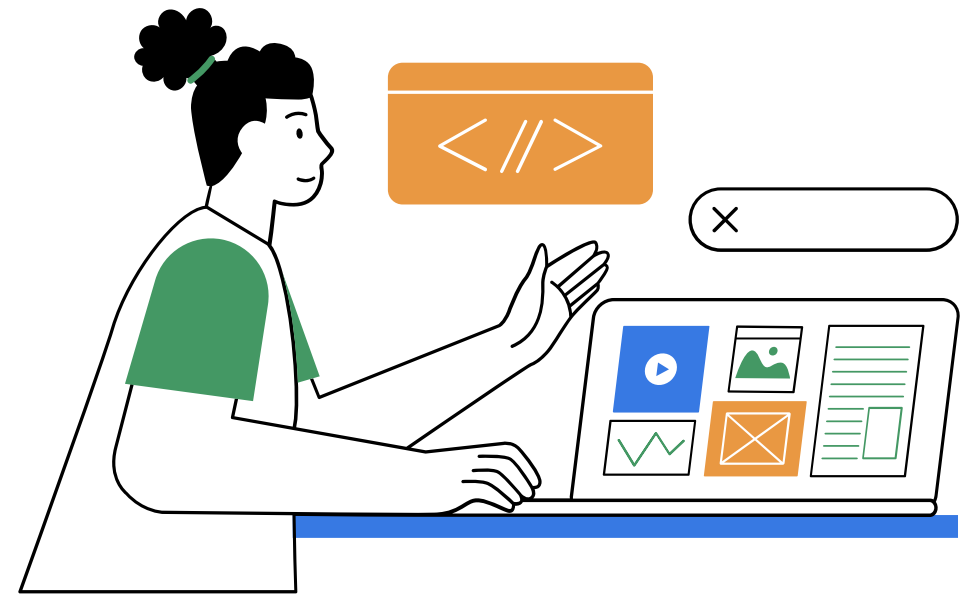
- Membangun model clustering (K-Means) untuk mengelompokkan pelanggan ke dalam beberapa kelompok berdasarkan profil finansial (pendapatan, utang, skor kredit, limit kredit).



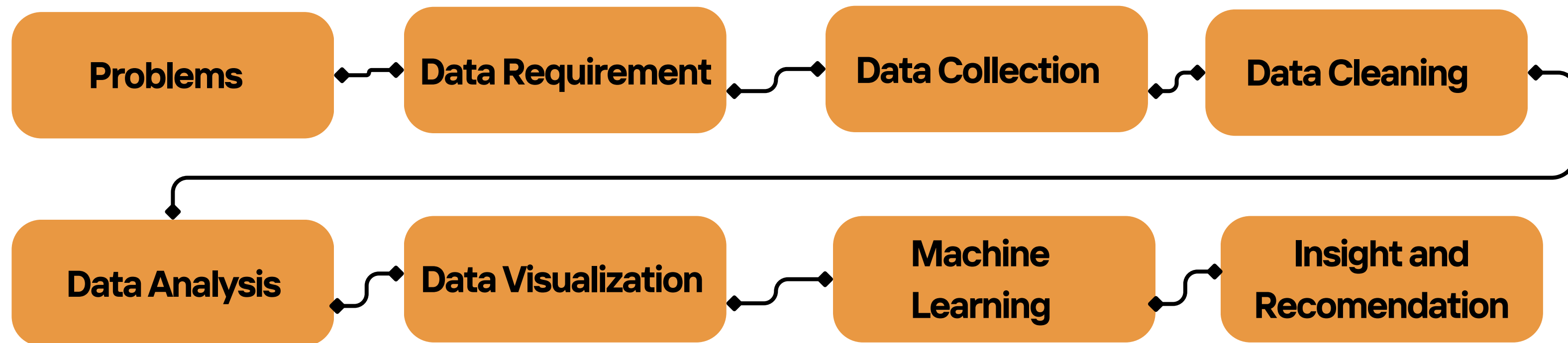
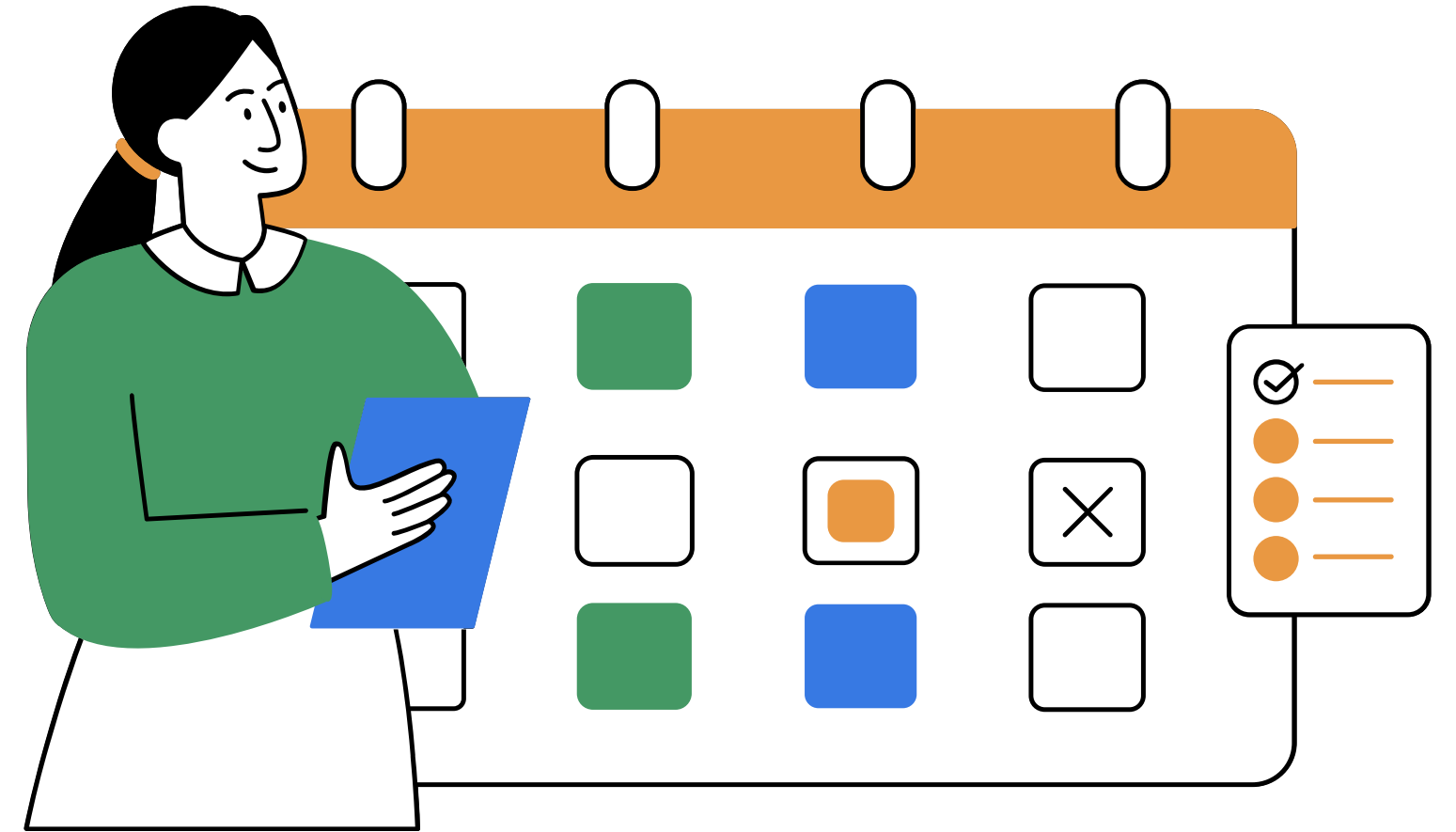
Problem Formulation



1. Bagaimana membersihkan data transaksi, pengguna, dan kartu agar siap untuk analisis?
2. Bagaimana pola transaksi nasabah berdasarkan waktu, lokasi, dan jenis kartu?
3. Metode transaksi (Chip/Gesek/Online) dan jenis kartu (Debit/Kredit) mana yang paling dominan?
4. Bagaimana membangun model untuk mengelompokkan pelanggan kedalam beberapa kelompok berdasarkan profil finansial?
5. Bagaimana menyajikan visualisasi data agar dapat membantu pemahaman mendalam mengenai data yang telah diolah ?



Methodology

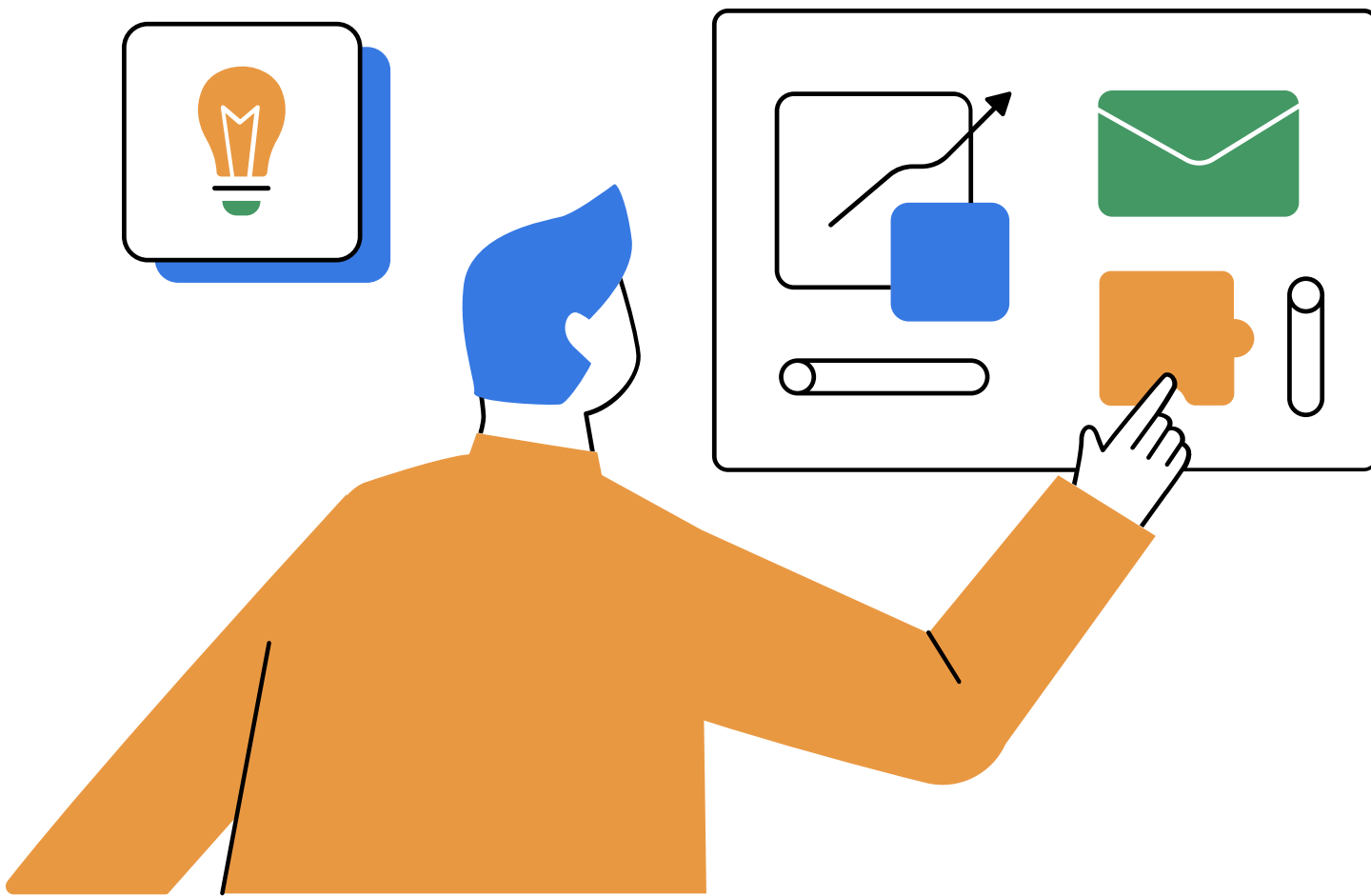


Data Source

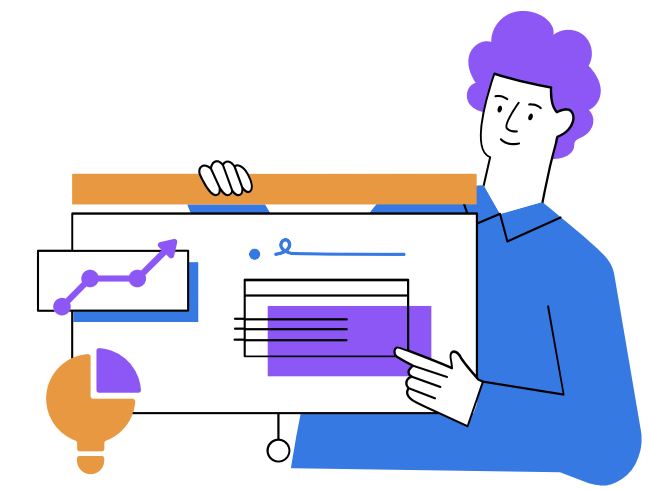
Dataset yang digunakan merupakan data sekunder yang disediakan oleh tim rekrutmen sebagai bagian dari studi kasus untuk proses seleksi Data Analyst. Dataset ini mencakup ~1 juta catatan transaksi kartu kredit historis dalam periode 10 tahun (2010-2019) dan terdiri dari tiga tabel terpisah: transaksi, users, dan cards.

Dataset lengkap dapat diakses pada link berikut:

https://drive.google.com/drive/folders/190CuIAO12_uu_9KgflQl6QiGU00qj8h8



Analysis Result



1. Cleaning Data

```
[ ] # --- 1. Baca CSV ---
df_transactions = pd.read_csv("transactions_data.csv")
df_users = pd.read_csv("users_data.csv")
df_cards = pd.read_csv("cards_data.csv")

# --- 2. Cleaning kolom numerik yang ada tanda $ ---
df_transactions['amount'] = df_transactions['amount'].replace(r'[\$,]', '', regex=True).astype(float)
df_users['per_capita_income'] = df_users['per_capita_income'].replace(r'[\$,]', '', regex=True).astype(float)
df_users['yearly_income'] = df_users['yearly_income'].replace(r'[\$,]', '', regex=True).astype(float)
df_users['total_debt'] = df_users['total_debt'].replace(r'[\$,]', '', regex=True).astype(float)
df_cards['credit_limit'] = df_cards['credit_limit'].replace(r'[\$,]', '', regex=True).astype(float)

# --- 3. Ubah kolom tanggal jadi datetime ---
df_transactions['date'] = pd.to_datetime(df_transactions['date'])
df_cards['expires'] = pd.to_datetime(df_cards['expires'], format='%m/%Y', errors='coerce')
df_cards['acct_open_date'] = pd.to_datetime(df_cards['acct_open_date'], format='%m/%Y', errors='coerce')
```

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

⇒ Tampilkan output tersembunyi

```
[ ] df.duplicated().sum()
```

⇒ np.int64(0)

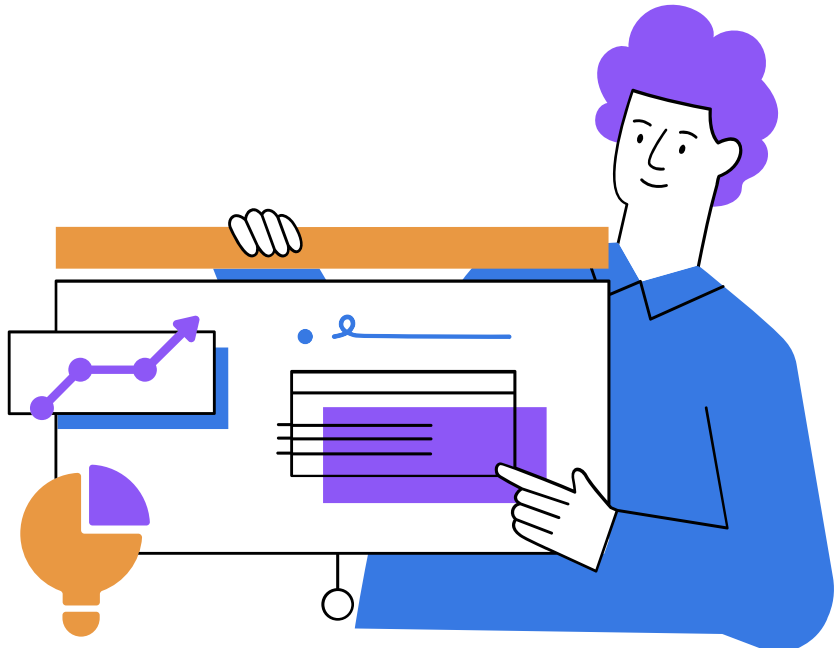
Proses pembersihan dan transformasi data secara menyeluruh menggunakan **Python** dengan menggunakan library **Pandas**. Proses ini mencakup beberapa beberapa hal yaitu:

- Konversi Tipe Data: Mengubah kolom-kolom finansial (seperti amount dan credit_limit) dari format teks yang mengandung simbol mata uang menjadi tipe data numerik.
- Kolom tanggal juga distandardisasi ke format datetime menjaga format tetap konsisten.
- Pengecekan raw yang duplicated dan menangani missing Value.

Proses ini memastikan dataset yang digunakan untuk analisis dan visualisasi merupakan data yang telah bersih agar hasil yang didapatkan lebih bagus.

Analysis Result

2. Data Insight with SQL



- Distribution Age of Users

	age_group	total_users
0	30-50	765
1	51-70	552
2	70+	203
3	<30	480

- Brand of Cards

	card_brand	total_cards
0	Mastercard	3209
1	Visa	2326
2	Amex	402
3	Discover	209

- Type of Cards

	card_type	avg_credit_limit
0	Credit	11174.380165
1	Debit	18557.888636
2	Debit (Prepaid)	64.448097

- Average Income of Users by Gender

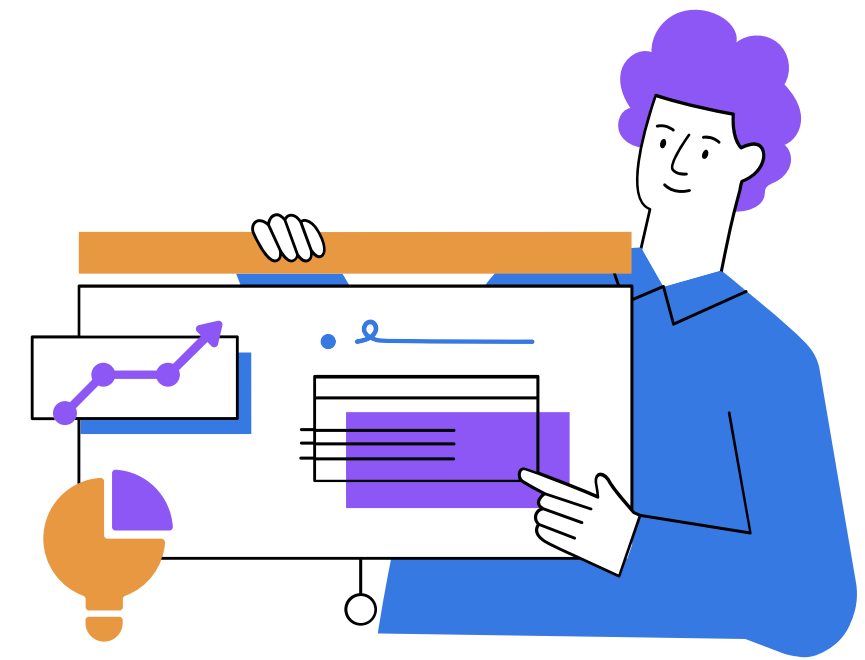
	gender	avg_income	total_users
0	Female	46048.314961	1016
1	Male	45372.638211	984

Analysis Result

2. Data Insight with SQL

- Segmentation of Users

	segmen_pendapatan	segmen_risiko	jumlah_pelanggan	rata_rata_limit_kredit	rata_rata_total_utang
0	Pendapatan Menengah	Resiko Rendah	474	19604.531646	76417.166667
1	Pendapatan Menengah	Resiko Sedang	987	20131.701114	80740.825735
2	Pendapatan Menengah	Resiko Tinggi	188	16204.367021	103166.558511
3	Pendapatan Rendah	Resiko Rendah	1277	11590.528583	42878.920908
4	Pendapatan Rendah	Resiko Sedang	2532	11523.874803	45113.201027
5	Pendapatan Rendah	Resiko Tinggi	534	10400.906367	52448.061798
6	Pendapatan Tinggi	Resiko Rendah	44	41763.659091	75399.977273
7	Pendapatan Tinggi	Resiko Sedang	90	43687.033333	152655.922222
8	Pendapatan Tinggi	Resiko Tinggi	20	33383.300000	246465.450000

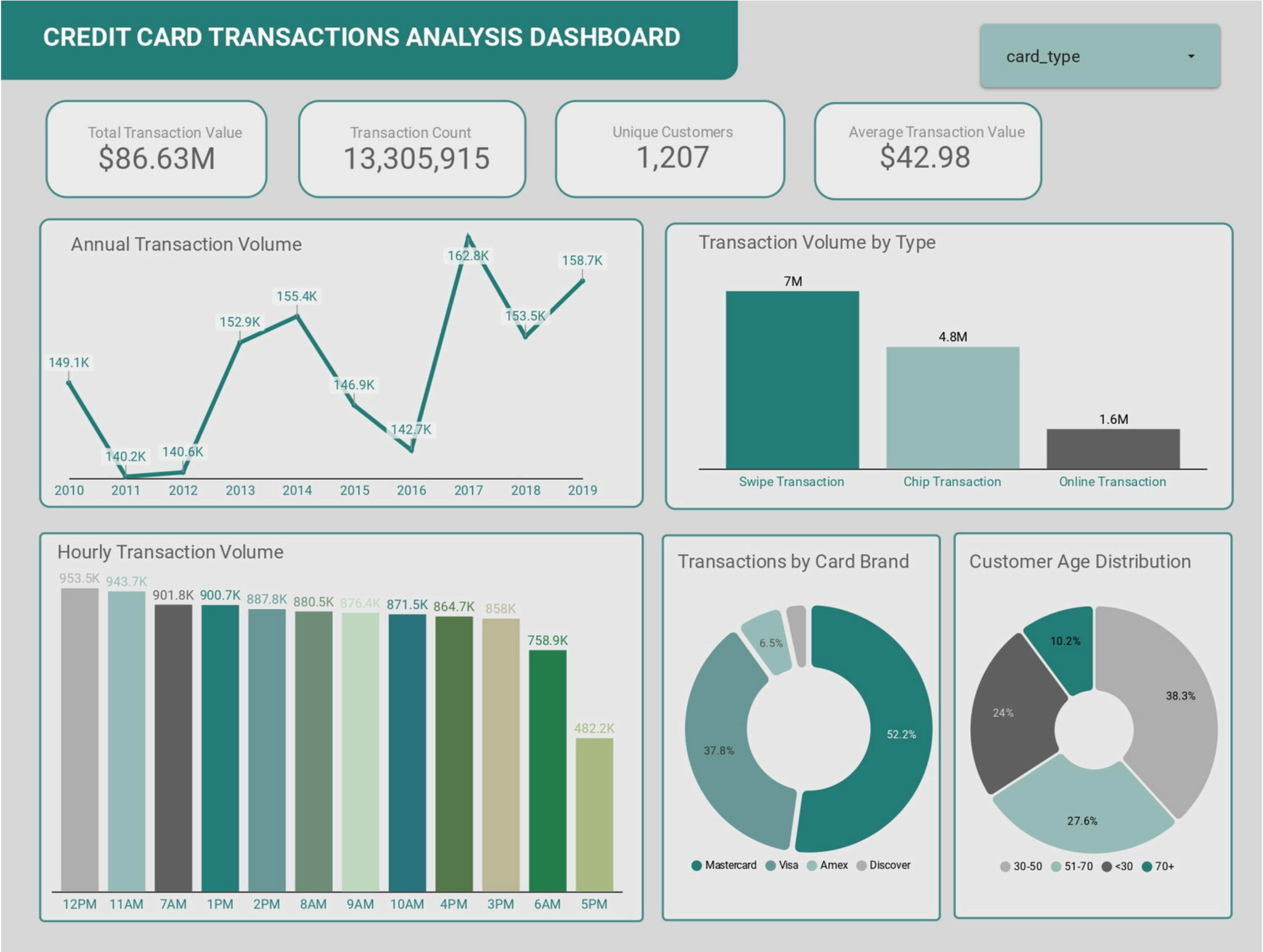


- Error of Transactions

	jenis_error	jenis_transaksi	merchant_city	jumlah_kasus
0	Insufficient Balance	Online Transaction	ONLINE	12160
1	Bad Card Number	Online Transaction	ONLINE	7766
2	Bad Expiration	Online Transaction	ONLINE	6161
3	Bad CVV	Online Transaction	ONLINE	6103
4	Technical Glitch	Online Transaction	ONLINE	3024
5	Insufficient Balance	Swipe Transaction	Riverside	1344
6	Insufficient Balance	Chip Transaction	Riverside	1275
7	Insufficient Balance	Swipe Transaction	Houston	857
8	Insufficient Balance	Swipe Transaction	Orlando	787
9	Insufficient Balance	Chip Transaction	Farmington	670

Data Visualizations

Dashboard Snapshot - Credit Card Transactions Analysis Dashboard

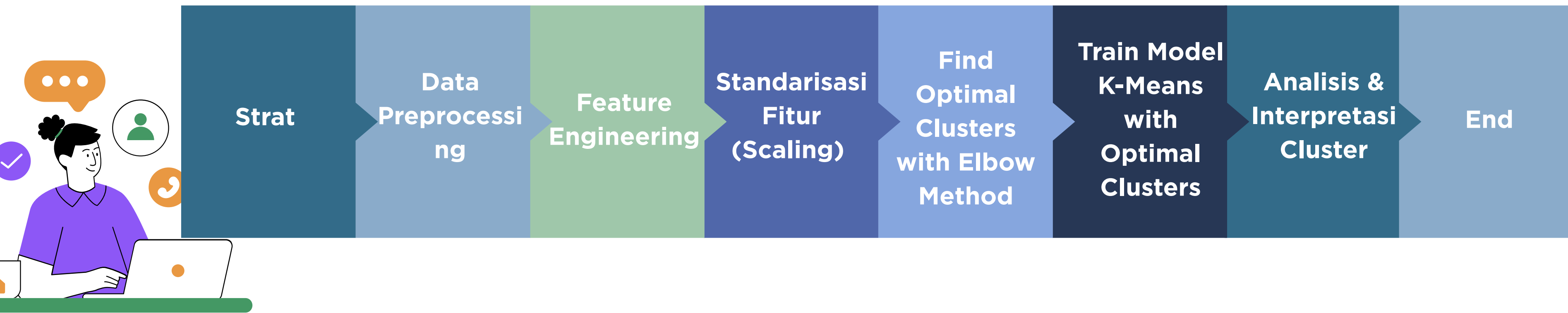


Machine Learning

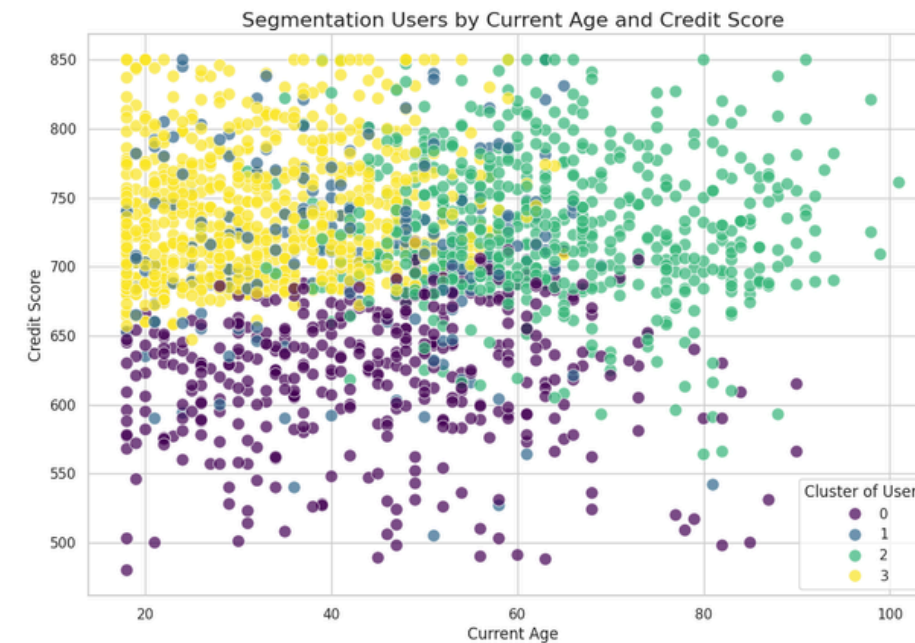
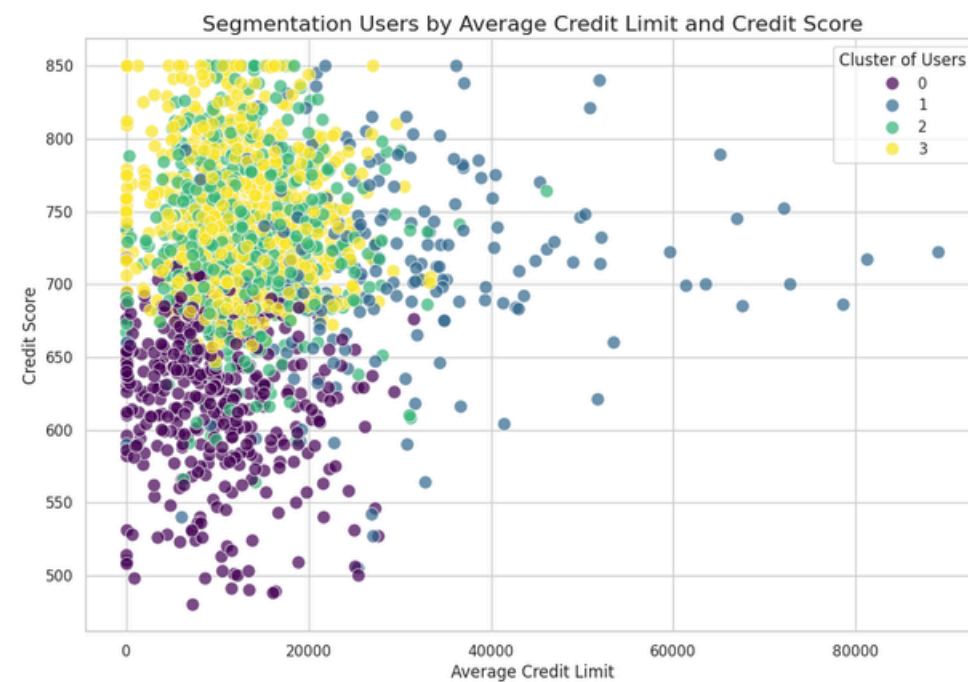
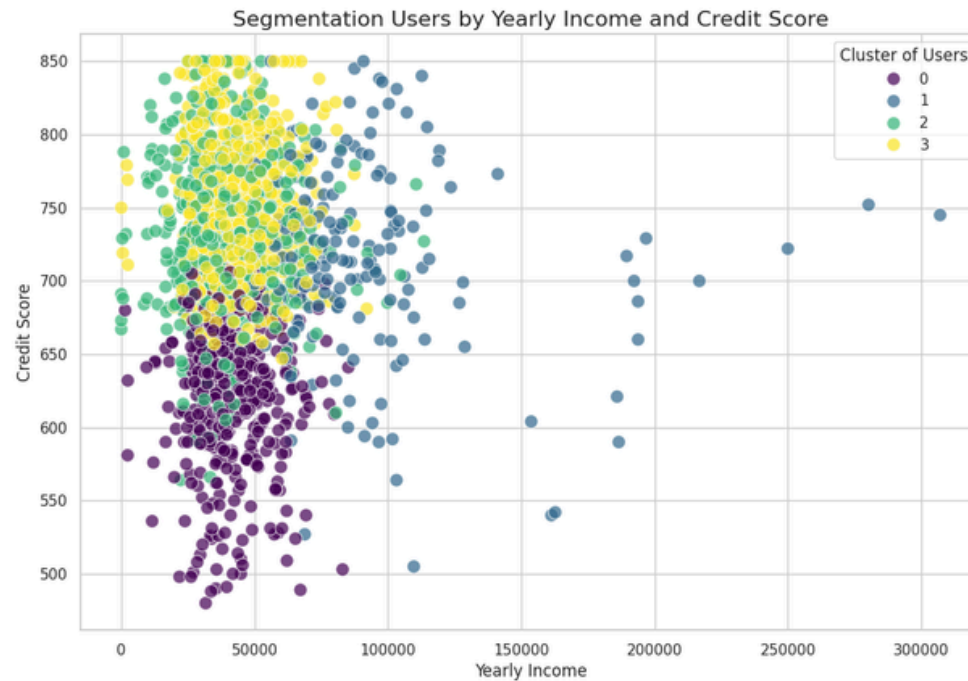
Proyek ini menggunakan pendekatan Unsupervised Learning untuk membangun model segmentasi users. Jenis pemodelan K-Means Clustering dipilih untuk secara otomatis mengelompokkan users ke dalam beberapa kelompok berbeda berdasarkan kesamaan profil finansial dan demografis mereka.

Pendekatan Unsupervised Learning dipilih karena tujuan utamanya adalah untuk menemukan pola di dalam data, bukan untuk memprediksi output yang sudah diketahui.

Tahapan Pemodelan Machine Learning



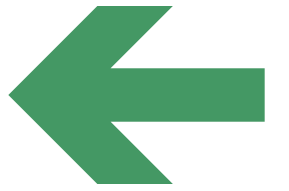
Machine Learning



Dari hasil clustering tersebut berhasil mengidentifikasi beberapa segmen users yaitu:

1. Cluster yang berwarna **kuning** merupakan user idaman, sebut saja **The Prime Users**. Dapat dilihat di semua grafik, kelompok ini secara konsisten memiliki skor kredit tertinggi, meskipun pendapatan atau limit kredit mereka tidak selalu yang paling tinggi.
2. Cluster yang berwarna **unggu** sebut sebagai **The Savers**. Grafik dengan jelas menunjukkan bahwa segmen ini berada di sisi kiri pada hampir di semua grafik. Users dengan pendapatan rendah, utang rendah, dan limit kredit terendah.
3. Cluster warna **Hijau** sebut dengan **The Established Elders**. Grafik usia di pojok kanan bawah dengan jelas menunjukkan mereka adalah kelompok users senior. Meskipun pendapatan mereka saat ini tidak tinggi (kemungkinan sudah pensiun), mereka memiliki utang yang sangat rendah dan skor kredit yang baik.
4. Cluster warna **Biru** sebut sebagai **The High Rollers**. Mereka adalah kelompok usia produktif dengan pendapatan yang baik. Namun, karakteristik utama mereka terlihat jelas di grafik utang: mereka adalah users dengan tingkat utang yang jauh paling tinggi.

Insight and Recommendation



1. Terdapat Segmen Users **Bernilai Tinggi** yang Sekaligus **Berisiko Tinggi**.

- 💡 Insight: Analisis segmentasi menunjukkan adanya kelompok users dengan **pendapatan tinggi** namun juga memiliki **profil risiko yang tinggi** faktor yang menyebabkan bisa karena total utang atau skor kredit yang kurang optimal. Kelompok users ini, meskipun dapat menguntungkan, namun diperlukan perhatian khusus.
- 🚀 Rekomendasi: Implementasikan sebuah sistem pemantauan untuk users "**Pendapatan Tinggi - Risiko Tinggi**". Tawarkan produk premium untuk memaksimalkan pendapatan, namun harus dengan evaluasi risiko berkala.

2. **Transaksi Online** Merupakan Tempat Transaksi Utama, Namun Juga Sumber **Error Terbesar**.

- 💡 Insight: Hasil analisis menunjukkan bahwa **transaksi online** merupakan salah satu metode yang paling banyak digunakan. Namun, banyak terjadi kegagalan transaksi (error) seperti **saldo tidak cukup** dan **nomor kartu salah**.
- 🚀 Rekomendasi: Perkuat sistem keamanan dan validasi untuk transaksi online (misalnya, implementasi 2-Factor Authentication).

3. Mayoritas Users Berada di **Kelompok Usia Produktif** dan **Mapan** (30-70 tahun).

- 💡 Insight: Distribusi usia menunjukkan bahwa users terbanyak adalah **kelompok usia 30-70 tahun**.
- 🚀 Rekomendasi: Fokuskan **strategi retensi** dan **cross-selling** untuk users kategori ini. Tawarkan produk yang sesuai dengan siklus hidup mereka, seperti produk investasi, atau asuransi, untuk memaksimalkan customer lifetime value.



Thank You



- Shifwa Luqyaanaa
- shifwasl27@gmail.com
- www.linkedin.com/in/shifwa-luqyaanaa