

PROJECT AKHIR DATA MINING DATA WAREHOUSE

CLUSTERING CREDIT CARD USERS



DOSEN PENGAMPU:

Nindy Irzavika, S.SI., M.T.

DISUSUN OLEH:

Risma Nurcahyani	2110512102
Muhammad Zaki Pradana	2110512106
Kemas Ghani Sampurna	2110512113
Angeline Mega Kristina	2110512120
Vinessa Gabby Asyfa	2110512121

S1 SISTEM INFORMASI

FAKULTAS ILMU KOMPUTER

UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA

2023

DAFTAR ISI

DAFTAR ISI.....	2
BAB I PENDAHULUAN.....	3
A. Latar Belakang.....	3
B. Rumusan Masalah.....	4
C. Tujuan.....	4
D. Manfaat.....	4
BAB II LANDASAN TEORI.....	5
A. Pengertian Data Mining.....	5
B. Fungsi Data Mining.....	5
C. Metode dalam Data Mining.....	6
D. Mengenal Clustering dalam Data Mining.....	8
BAB III PROGRAM DAN ANALISIS DATA KAGGLE.....	9
A. Penjelasan Data Set.....	9
B. Impor Data Set dari Kaggle.....	10
C. Create Model and Cluster Data.....	11
BAB IV PEMBAHASAN.....	25
A. Hasil Analisis.....	25
B. Strategi dan Rekomendasi.....	33
BAB V PENUTUP.....	35
A. Kesimpulan.....	35
B. Kritik dan Saran.....	35
DAFTAR PUSTAKA.....	36

BAB I

PENDAHULUAN

A. Latar Belakang

Seiring perkembangan zaman dan teknologi yang semakin pesat, evolusi teknologi gudang data dan pertumbuhan *big data* juga menjadi salah satu kenyataan serta masalah yang harus kita hadapi di zaman ini. Hal tersebut tentu saja mengingat data yang telah meningkat pesat selama beberapa dekade terakhir. Dalam mengolah data-data yang pertumbuhannya semakin pesat pastinya membutuhkan teknik untuk mengolah data-data tersebut baik data terstruktur maupun data tidak terstruktur. Teknik tersebut disebut juga sebagai *data mining* dimana data mentah yang berukuran sangat besar akan diolah menjadi pengetahuan yang berguna sebagai pengambilan keputusan.

Data mining digunakan sebuah organisasi untuk pengambilan keputusan (*decision maker*) melalui analisis data yang mendalam. Data yang dianalisis merupakan data yang berasal dari *database* organisasi, data publik, data teks dan dokumen, data transaksi organisasi, dan data-data lainnya. Akan tetapi semua sumber data organisasi tergantung pada konteks organisasi tersebut. Metode yang ada di dalam *data mining* tujuannya adalah untuk mendeskripsikan kumpulan data target atau memprediksi hasil melalui penggunaan algoritma.

Metode yang digunakan dalam menganalisis *data mining* memiliki tujuan untuk mengatur dan memfilter data, menampilkan informasi yang bermanfaat, mendeteksi kejahatan yang ada di sistem sehingga dapat mencegah bahkan mengatasi kejahatan yang ada di sistem. Tujuan-tujuan yang telah disebutkan nanti diharapkan dapat mengembangkan bisnis menjadi lebih menguntungkan, efisien, atau lebih kuat secara operasional.

Salah satu metode *data mining* yang kami implementasikan pada laporan akhir kami kali ini adalah metode *clustering*. Metode tersebut mengidentifikasi kesamaan antar objek, kemudian mengelompokkan *item* pada data tersebut berdasarkan kategori tertentu sehingga data akan dikelompokkan berdasarkan kesamaannya.

Tujuan kami menggunakan *clustering* sebagai metode *data mining* adalah untuk mengidentifikasi *item* yang ada pada data set sesuai dengan kategorinya sehingga memudahkan pembaca data dalam melakukan analisis data. Selain itu kami menggunakan metode ini agar mudah dalam menemukan pola sehingga mengetahui korelasi atau relasi antara variabel yang tidak jelas sebelumnya. Dan juga agar kami dapat menentukan rekomendasi strategi segmentasi risiko, segmentasi pelanggan, dan strategi lainnya.

B. Rumusan Masalah

Berdasarkan latar belakang diatas, kami menetapkan rumusan-rumusan masalah dalam laporan yang kami buat ini yaitu:

1. Bagaimana cara menggunakan Jupyter dengan bahasa *python* untuk mengolah dan menganalisis data?
2. Bagaimana proses pengolahan data set menggunakan Jupyter dengan bahasa *python*?
3. Apa dan bagaimana proses dari metode yang digunakan dalam mengolah data set?
4. Setelah melakukan analisis terhadap data, apa sajakah strategi dan rekomendasi yang dapat diberikan?
5. Apa kesimpulan dari strategi dan rekomendasi yang diberikan?

C. Tujuan

Setelah kami merumuskan masalah-masalah yang ada maka kami mengetahui tujuan dibuatnya laporan ini adalah untuk:

1. Untuk mengetahui cara menggunakan Jupyter dengan bahasa *python* untuk mengolah data dan menganalisis data.
2. Untuk mengetahui proses pengolahan data set menggunakan Jupyter dengan bahasa *python*.
3. Untuk mengetahui proses dari metode yang digunakan dalam mengolah data set.
4. Untuk mengetahui strategi dan rekomendasi yang dapat diberikan.
5. Untuk mengetahui kesimpulan dari strategi dan rekomendasi yang diberikan.

D. Manfaat

Maka dari itu kami sebagai pembuat laporan ini berharap laporan ini dapat bermanfaat untuk kami sendiri sebagai pembuat laporan dan pembaca laporan ini, yaitu:

1. Untuk mengimplementasikan materi mata kuliah *data mining* yang sudah kami pelajari selama satu semester sehingga nantinya dapat diimplementasikan kembali pada profesi masa depan.
2. Untuk menambah wawasan kepada para pembaca mengenai penggunaan *text editor* Jupyter untuk pengolahan data set sehingga hasil analisisnya dapat digunakan sebagai pengambilan keputusan organisasi.

BAB II

LANDASAN TEORI

A. Pengertian Data Mining

Menurut (Suntoro, 2019), data mining merupakan proses untuk mendapatkan informasi yang berguna dari basis data yang besar dan perlu diekstraksi agar menjadi informasi baru dan dapat membantu dalam pengambilan keputusan. Data mining atau penambangan data adalah proses yang digunakan untuk menggali informasi penting atau pola tersembunyi dari data besar secara otomatis atau setengah otomatis. Tujuannya adalah untuk menemukan wawasan atau pengetahuan yang tidak terlihat secara langsung atau tersembunyi di dalam data yang ada.

Untuk melakukan data mining, kita menggunakan berbagai teknik dan algoritma dari statistik, matematika, kecerdasan buatan, dan pembelajaran mesin. Proses ini melibatkan langkah-langkah seperti memproses data, memilih atribut yang relevan, membangun model, dan mengevaluasi hasilnya untuk mendapatkan informasi berharga.

B. Fungsi Data Mining

Data mining memiliki peran penting dalam analisis data. Fungsinya adalah untuk mengungkapkan wawasan berharga dan pengetahuan tersembunyi dari kumpulan data yang besar. Dengan menggunakan berbagai teknik dan algoritma, seperti statistik, matematika, kecerdasan buatan, dan pembelajaran mesin, data mining memungkinkan kita untuk menemukan pola, tren, dan hubungan yang mungkin tidak terlihat secara langsung dalam data tersebut.

Salah satu fungsi utama dari data mining adalah mengenali pola dan tren. Dengan menganalisis data secara menyeluruh, kita dapat menemukan pola menarik dan tren yang terjadi seiring waktu. Ini membantu kita memahami perubahan dalam data, memprediksi perilaku di masa depan, dan membuat keputusan yang lebih baik.

Selain itu, data mining juga berfungsi untuk prediksi dan ramalan. Dengan memanfaatkan pola historis dalam data, kita dapat membuat model prediktif yang berguna untuk meramalkan kejadian atau perilaku di masa depan. Ini bermanfaat dalam perencanaan strategis, pengambilan keputusan yang akurat, dan pengelolaan risiko yang lebih baik.

Data mining juga penting dalam optimasi proses bisnis. Dengan menganalisis data operasional, kita dapat mengidentifikasi peluang untuk meningkatkan efisiensi dan efektivitas proses bisnis. Dengan pemahaman yang lebih dalam tentang data, kita dapat mengoptimalkan alur kerja, mengidentifikasi masalah, dan mengimplementasikan perubahan yang diperlukan.

Selain itu, data mining membantu dalam mendeteksi anomali atau pola yang tidak biasa dalam data. Dengan menggunakan teknik deteksi anomali, kita dapat mengenali kegiatan mencurigakan, kesalahan sistem, atau perilaku yang tidak normal. Ini membantu

dalam pengelolaan risiko, meningkatkan keamanan, dan mengambil langkah-langkah pencegahan yang efektif.

Secara keseluruhan, data mining berperan penting dalam mengungkap wawasan berharga dari data. Dengan menerapkan teknik dan algoritma yang tepat, data mining membantu kita mengenali pola, meramalkan masa depan bisnis, mengoptimalkan proses bisnis, mengelola risiko, dan mendeteksi anomali. Hal ini memberikan manfaat yang signifikan dalam pengambilan keputusan, perencanaan strategis, dan peningkatan kinerja bisnis.

C. Metode dalam Data Mining

1. Clustering (Pengelompokan)

Metode clustering dalam data mining merupakan teknik yang digunakan untuk mengelompokkan objek atau data ke dalam kelompok-kelompok berdasarkan kesamaan atau kemiripan tertentu. Tujuannya adalah untuk mengungkapkan struktur yang tersembunyi dalam data tanpa memiliki label klasifikasi sebelumnya.

Dalam proses clustering, dilakukan analisis yang sistematis terhadap atribut dan karakteristik objek dalam kumpulan data. Algoritma clustering akan menghitung jarak atau kesamaan antara objek-objek tersebut, dan selanjutnya mengelompokkannya berdasarkan kesamaan tersebut. Objek yang memiliki atribut yang mirip akan dikelompokkan bersama dalam satu kelompok, sedangkan objek yang memiliki atribut yang berbeda akan ditempatkan dalam kelompok yang berbeda. Metode clustering memiliki berbagai pendekatan, seperti K-means, Hierarchical clustering, DBSCAN, dan lainnya.

2. Klasifikasi

Metode klasifikasi dalam data mining merupakan pendekatan yang digunakan untuk membangun model prediktif yang dapat mengklasifikasikan objek ke dalam kelas atau kategori yang telah ditentukan sebelumnya. Tujuannya adalah untuk melakukan pengelompokan objek berdasarkan atribut yang dimiliki dan membuat prediksi terhadap objek yang belum diketahui kelasnya.

Proses klasifikasi melibatkan analisis yang sistematis terhadap data yang telah dilabeli sebelumnya. Algoritma klasifikasi akan menggunakan contoh-contoh data yang berlabel untuk melatih model dan mempelajari pola atau aturan yang menghubungkan atribut dengan kelas yang sesuai. Setelah model dilatih, kita dapat menggunakannya untuk memprediksi kelas dari objek yang belum diketahui dengan mengamati atribut-atribut yang dimiliki oleh objek tersebut.

3. Asosiasi

Metode asosiasi dalam data mining digunakan untuk menemukan hubungan dan pola tersembunyi antara item atau atribut dalam data. Tujuannya adalah untuk

mengidentifikasi keterkaitan antara item yang sering muncul bersama atau berhubungan dalam suatu konteks.

Dalam proses asosiasi, dilakukan analisis mendalam terhadap data yang berisi transaksi atau daftar item terkait. Algoritma asosiasi akan mencari aturan atau pola yang mengungkapkan hubungan antara item tersebut. Sebagai contoh, bisa mencari aturan yang menyatakan bahwa jika seseorang membeli produk A dan B, maka kemungkinan besar mereka juga akan membeli produk C. Metode asosiasi populer yang sering digunakan adalah Algoritma Apriori.

4. Regresi

Metode regresi dalam data mining digunakan untuk memodelkan dan memprediksi hubungan antara variabel dependen (variabel target) dan satu atau lebih variabel independen (variabel prediktor). Tujuannya adalah untuk memahami dan menggambarkan hubungan fungsional antara variabel-variabel tersebut, serta melakukan prediksi nilai variabel dependen berdasarkan nilai-nilai variabel independen yang diberikan.

Dalam proses regresi, dilakukan analisis statistik yang mendalam terhadap data yang telah dikumpulkan. Algoritma regresi akan mengidentifikasi dan memodelkan pola hubungan antara variabel dependen dan variabel independen melalui pendekatan matematis. Metode regresi yang digunakan dapat berupa regresi linier, regresi logistik, regresi non-parametrik, dan sejenisnya, tergantung pada karakteristik data dan tujuan analisis yang ingin dicapai.

5. Analisis Sekuensial

Metode Analisis Sekuensial dalam data mining digunakan untuk mengungkap pola dan hubungan sekuensial antara elemen-elemen dalam data yang diurutkan atau berhubungan secara temporal. Tujuannya adalah untuk memahami dan menganalisis pola urutan yang ada dalam data, serta mengidentifikasi aturan atau tren yang timbul dari urutan tersebut.

Dalam proses Analisis Sekuensial, dilakukan analisis mendalam terhadap urutan data yang terdiri dari serangkaian kejadian atau transaksi yang saling terkait. Algoritma Analisis Sekuensial akan mencari pola atau aturan yang konsisten dalam urutan tersebut, seperti kejadian yang sering muncul bersama atau urutan peristiwa yang umum terjadi.

6. Rule Mining

Rule mining merupakan salah satu teknik penting dalam data mining yang digunakan untuk menemukan pola atau aturan tersembunyi dalam dataset. Tujuan utamanya adalah untuk mengungkapkan hubungan yang ada antara atribut dalam data dan menghasilkan aturan yang bisa digunakan untuk memprediksi atau menjelaskan

perilaku data tersebut. Proses rule mining melibatkan pencarian pola atau aturan yang penting dan berulang dalam data yang relevan dengan tujuan analisis.

Awalnya, dataset dianalisis secara menyeluruh untuk mengidentifikasi semua item atau atribut yang mungkin ada. Kemudian, digunakan teknik seperti algoritma Apriori atau FP-Growth untuk menghasilkan himpunan itemset yang sering muncul bersama dalam dataset. Setelah mendapatkan item set tersebut, langkah selanjutnya adalah menghasilkan aturan asosiasi dari itemset tersebut.

D. Mengenal Clustering dalam Data Mining

Clustering dalam data mining adalah suatu teknik yang digunakan untuk mengelompokkan objek atau data ke dalam kelompok-kelompok yang serupa berdasarkan kemiripan mereka. Tujuan utama dari clustering adalah untuk mengidentifikasi pola-pola yang ada dalam data, menemukan struktur tersembunyi, dan mengelompokkan data menjadi kelompok-kelompok yang saling terpisah.

Fungsi utama clustering adalah mengelompokkan data berdasarkan kesamaan fitur atau karakteristik tertentu. Hal ini memungkinkan untuk mengidentifikasi kelompok-kelompok yang homogen dalam data, sehingga memudahkan pemahaman dan analisis lebih lanjut terhadap data tersebut. Clustering juga dapat digunakan untuk pemrosesan data yang lebih efisien, pengurangan dimensi, dan pemilihan fitur.

Kelebihan dari clustering adalah:

1. Clustering dapat membantu mengidentifikasi pola dan hubungan dalam data yang tidak terlihat secara langsung.
2. Clustering dapat membantu dalam pemahaman dan interpretasi data dengan mengelompokkan objek-objek yang serupa bersama-sama.
3. Clustering dapat mengurangi kompleksitas data dengan mengelompokkan objek-objek yang mirip ke dalam kelompok-kelompok yang lebih kecil.

Kekurangan dari clustering adalah:

1. Hasil clustering dapat bervariasi tergantung pada metode dan parameter yang digunakan, sehingga pemilihan yang tidak tepat dapat menghasilkan kelompok yang tidak relevan.
2. Penafsiran hasil clustering dapat bersifat subyektif dan bergantung pada persepsi dan pemahaman analis.
3. Clustering dapat menjadi sulit untuk diterapkan pada dataset yang sangat besar, karena membutuhkan waktu dan sumber daya komputasi yang signifikan.
4. Dalam memilih teknik clustering yang tepat, perlu mempertimbangkan karakteristik data, tujuan analisis, dan kemampuan komputasi yang tersedia.

BAB III

PROGRAM DAN ANALISIS DATA KAGGLE

A. Penjelasan Data Set

Dataset yang digunakan pada laporan project ini adalah data credit card users, dimana dataset ini digunakan untuk melakukan analisis pengelompokan pengguna kartu kredit berdasarkan aktivitas kartu mereka. Tujuan dari analisis ini adalah untuk memahami basis pelanggan dengan lebih baik dan memberikan pelayanan yang lebih baik kepada subkelompok dalam basis tersebut. Hal ini diharapkan dapat meningkatkan pengalaman pelanggan serta metrik bisnis seperti omset dan penggunaan kartu.

Berikut adalah ringkasan kolom-kolom dan definisinya dalam dataset tersebut:

1. *CustomerID* : Nomor identifikasi unik untuk setiap pelanggan.
2. *CustomerName* : Nama pelanggan.
3. *Gender* : Jenis kelamin pelanggan (Pria atau Wanita).
4. *Age* : Usia pelanggan dalam tahun.
5. *EducationLevel* : Tingkat pendidikan pelanggan (misalnya, Pendidikan Dasar, Pendidikan Menengah, Sarjana).
6. *MaritalStatus* : Status pernikahan pelanggan (Menikah, Belum Menikah, atau Lainnya).
7. *IncomeCategory* : Kategori pendapatan pelanggan (misalnya, Rendah, Sedang, Tinggi).
8. *CardCategory* : Kategori kartu yang dimiliki pelanggan (Blue, Silver, Gold, Platinum).
9. *MonthsOnBook* : Jumlah bulan pelanggan telah menjadi anggota.
10. *TotalRelationshipCount* : Jumlah hubungan pelanggan dengan bank.
11. *MonthsInactive* : Jumlah bulan sejak pelanggan terakhir kali menggunakan kartu kredit.
12. *ContactsCountLast12Months* : Jumlah kontak yang pelanggan lakukan dalam 12 bulan terakhir.
13. *CreditLimit* : Batas kredit yang ditetapkan untuk pelanggan.
14. *TotalRevolvingBal* : Jumlah saldo yang belum dibayar pada akhir setiap bulan.
15. *AvgOpenToBuy* : Rata-rata jumlah kredit yang masih tersedia untuk pelanggan.
16. *TotalAmtChngQ4Q1* : Persentase perubahan jumlah total transaksi kartu kredit pelanggan dari kuartal keempat ke kuartal pertama.
17. *TotalTransAmt* : Total jumlah transaksi kartu kredit pelanggan dalam setahun.
18. *TotalTransCount* : Total jumlah transaksi kartu kredit pelanggan dalam setahun.
19. *TotalTransCountCardNotPresent* : Total jumlah transaksi kartu kredit pelanggan dalam setahun di mana kartu fisik tidak hadir.
20. *TotalTransCountCardPresent* : Total jumlah transaksi kartu kredit pelanggan dalam setahun di mana kartu fisik hadir.

21. *TotalTransCountBoth* : Total jumlah transaksi kartu kredit pelanggan dalam setahun dengan atau tanpa kartu fisik.
22. *AvgUtilizationRatio* : Rasio penggunaan rata-rata kartu kredit pelanggan.

Dengan menggunakan dataset ini, dapat dilakukan analisis yang lebih mendalam untuk memahami pola dan tren dalam aktivitas pengguna kartu kredit serta mengidentifikasi karakteristik unik dari setiap kelompok pengguna kartu kredit.

B. Impor Data Set dari Kaggle

1. Melakukan impor library yang dibutuhkan yaitu library pandas, numpy, matplotlib, dan seaborn

```
In [1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
sns.set_theme()
```

2. Membaca file csv kemudian menghapus 2 kolom terakhir, serta menampilkan dataset

```
In [2]: cc = pd.read_csv('./BankChurners.csv')

# Last two columns not needed
cc = cc.iloc[:, :-2]

pd.set_option('display.max_columns', None)
cc.head()
```

```
Out[2]:
```

	CLIENTNUM	Attrition_Flag	Customer_Age	Gender	Dependent_count	Education_Level	Marital_Status	Income_Category	Card_Category	Months_on_book
0	768805383	Existing Customer	45	M	3	High School	Married	60K–80K	Blue	39
1	818770008	Existing Customer	49	F	5	Graduate	Single	Less than \$40K	Blue	44
2	713982108	Existing Customer	51	M	3	Graduate	Married	80K–120K	Blue	36
3	769911858	Existing Customer	40	F	4	High School	Unknown	Less than \$40K	Blue	34
4	709106358	Existing Customer	40	M	3	Uneducated	Married	60K–80K	Blue	21

3. Menampilkan ringkasan statistik dari data numerik credit card users

```
In [3]: cc.describe()
```

```
Out[3]:
```

	CLIENTNUM	Customer_Age	Dependent_count	Months_on_book	Total_Relationship_Count	Months_Inactive_12_mon	Contacts_Count_12_mon	Credit_L
count	1.012700e+04	10127.000000	10127.000000	10127.000000	10127.000000	10127.000000	10127.000000	10127.000000
mean	7.391776e+08	46.325960	2.346203	35.928409	3.812580	2.341167	2.455317	8631.950
std	3.690378e+07	8.016814	1.298908	7.986416	1.554408	1.010622	1.106225	9088.770
min	7.080821e+08	26.000000	0.000000	13.000000	1.000000	0.000000	0.000000	1438.300
25%	7.130368e+08	41.000000	1.000000	31.000000	3.000000	2.000000	2.000000	2555.000
50%	7.179264e+08	46.000000	2.000000	36.000000	4.000000	2.000000	2.000000	4549.000
75%	7.731435e+08	52.000000	3.000000	40.000000	5.000000	3.000000	3.000000	11067.500
max	8.283431e+08	73.000000	5.000000	56.000000	6.000000	6.000000	6.000000	34516.000

4. Menampilkan jumlah baris dan kolom dari dataset

```
In [4]: cc.shape
```

```
Out[4]: (10127, 21)
```

5. Menghitung jumlah nilai null (kosong) dalam setiap kolom dataset

```
In [5]: cc.isnull().sum()
```

```
Out[5]: CLIENTNUM                0
Attrition_Flag                  0
Customer_Age                   0
Gender                         0
Dependent_count                0
Education_Level                0
Marital_Status                 0
Income_Category                0
Card_Category                  0
Months_on_book                 0
Total_Relationship_Count       0
Months_Inactive_12_mon        0
Contacts_Count_12_mon         0
Credit_Limit                   0
Total_Revolving_Bal           0
Avg_Open_To_Buy               0
Total_Amt_Chng_Q4_Q1          0
Total_Trans_Amt                0
Total_Trans_Ct                 0
Total_Ct_Chng_Q4_Q1           0
Avg_Utilization_Ratio         0
dtype: int64
```

C. Create Model and Cluster Data

1. Melakukan impor modul dari berbagai pustaka yang dibutuhkan dimana diantaranya modul Scikit-learn (sklearn) digunakan untuk melakukan klasterisasi data dengan algoritma K-Means, melakukan reduksi dimensi menggunakan PCA, mengevaluasi klaster dengan menggunakan metrik silhouette_score, dan melakukan pra pemrosesan data menggunakan teknik StandardScaler.

```
In [6]: from sklearn.cluster import KMeans
from sklearn.decomposition import PCA
from sklearn.metrics import silhouette_score
from sklearn.pipeline import Pipeline
from sklearn.preprocessing import StandardScaler

cols = cc.iloc[:, 9:].select_dtypes(['uint8', 'int64', 'float64']).columns
```

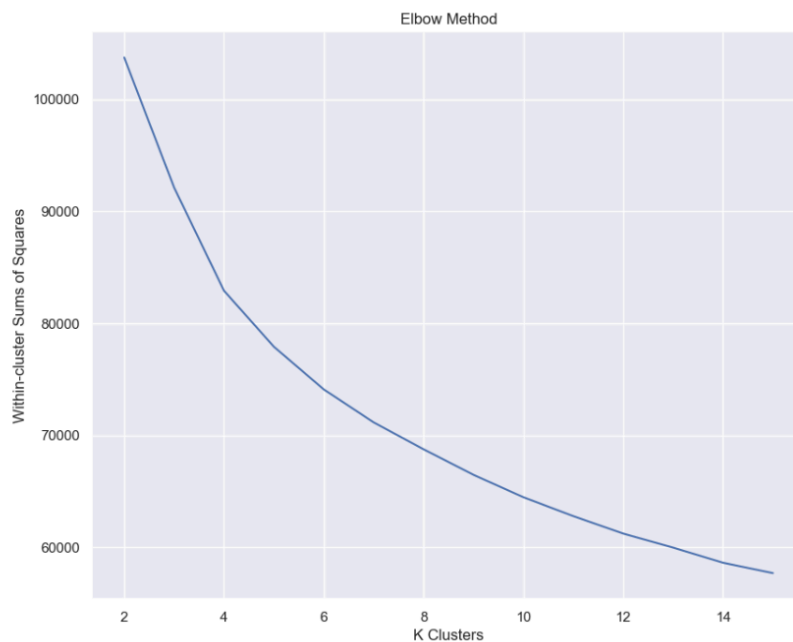
2. Melakukan pengimplementasian metode "Elbow" untuk memilih jumlah klaster optimal dalam algoritma K-Means dengan memplotkan grafik yang menunjukkan perubahan inertia (Within-cluster Sums of Squares) terhadap jumlah klaster.

```
In [7]: wcss = []

for i in range(2,16):
    km = KMeans(n_clusters = i, init = 'k-means++', max_iter = 500, n_init = 10, random_state = 1)
    data = StandardScaler().fit_transform(cc[cols])
    km.fit(data)
    wcss.append(km.inertia_)

fig = plt.figure(figsize=(10,8))
plt.plot(range(2,16), wcss)
plt.title('Elbow Method')
plt.xlabel('K Clusters')
plt.ylabel('Within-cluster Sums of Squares')
plt.show()
```

3. Plot ini menampilkan jumlah kluster pada sumbu x dan inertia pada sumbu y. Tujuannya adalah untuk mencari elbow (siku) dalam grafik, yang merupakan titik di mana penurunan inertia menjadi lebih lambat. Elbow ini dapat menjadi indikasi jumlah kluster yang optimal.



4. Membuat pipeline yang terdiri dari pra pemrosesan data dengan standard scaler dan PCA, serta algoritma klasterisasi K-Means dengan 6 kluster, dan kemudian melakukan pemrosesan dan klasterisasi pada dataset cc menggunakan pipeline yang telah dibuat.

```
In [8]: preprocessor = Pipeline(
        [
            ("scaler", StandardScaler()),
            ("pca", PCA(n_components=2, random_state=1))
        ]
    )

    clusterer = Pipeline(
        [
            (
                "kmeans",
                KMeans(
                    n_clusters=6,
                    init="k-means++",
                    n_init=50,
                    max_iter=1000,
                    random_state=1
                )
            )
        ]
    )

    pipe = Pipeline(
        [
            ("preprocessor", preprocessor),
            ("clusterer", clusterer)
        ]
    )

    pipe.fit(cc[cols])
```

```
Out[8]: Pipeline(steps=[('preprocessor',
                          Pipeline(steps=[('scaler', StandardScaler()),
                                           ('pca',
                                            PCA(n_components=2, random_state=1))])),
                        ('clusterer',
                         Pipeline(steps=[('kmeans',
                                           KMeans(max_iter=1000, n_clusters=6, n_init=50,
                                                  random_state=1))]))])
```

5. Mencetak rasio variansi yang dijelaskan oleh komponen pertama dan kedua dari analisis PCA yang dilakukan sebagai bagian dari pra pemrosesan data sebelum klusterisasi.

```
In [9]: print('PCA Component 1 Ratio of Explained Variance')
        print(round(preprocessor['pca'].explained_variance_ratio_[0], 2))
        print()
        print('PCA Component 2 Ratio of Explained Variance')
        print(round(preprocessor['pca'].explained_variance_ratio_[1], 2))
```

```
PCA Component 1 Ratio of Explained Variance
0.21
```

```
PCA Component 2 Ratio of Explained Variance
0.17
```

6. Menghasilkan skor Silhouette untuk evaluasi kualitas klusterisasi yang dilakukan dengan menggunakan pipeline yang telah dibuat, dengan menggunakan data yang telah diproses sebelumnya.

```
In [10]: preprocessed_data = pipe["preprocessor"].transform(cc[cols])
         predicted_labels = pipe["clusterer"]["kmeans"].labels_
         print('Silhouette Score: ', round(silhouette_score(preprocessed_data, predicted_labels), 3))

Silhouette Score: 0.395
```

7. Membuat dataframe baru (*pcadf*) yang berisi hasil transformasi data menggunakan komponen pertama dan kedua dari analisis PCA, menambahkan kolom "predicted_cluster" yang berisi label kluster yang diprediksi. Kemudian, kode tersebut memvisualisasikan hasil klusterisasi dengan menggunakan scatter plot, di mana sumbu x dan y adalah komponen pertama dan kedua, warna titik-titik merepresentasikan kluster yang diprediksi, dan warna yang ditampilkan menunjukkan label kluster.

```
In [11]: pcadf = pd.DataFrame(
    pipe["preprocessor"].transform(cc[cols]),
    columns=["component_1", "component_2"]
)

pcadf["predicted_cluster"] = pipe["clusterer"]["kmeans"].labels_

plt.style.use("fivethirtyeight")
plt.figure(figsize=(8, 8))

sp = sns.scatterplot(
    x="component_1",
    y="component_2",
    s=50,
    data=pcadf,
    hue="predicted_cluster",
    palette="tab10"
)

sp.set_title(
    "Clustering results from credit card use"
)
plt.legend(bbox_to_anchor=(1.05, 1), loc=2, borderaxespad=0.0)

plt.show()
```



8. Menambahkan kolom "cluster" ke dataset cc yang berisi label kluster yang diprediksi untuk setiap baris data. Kemudian, lima baris pertama dari dataset cc yang sudah diperbarui ditampilkan.

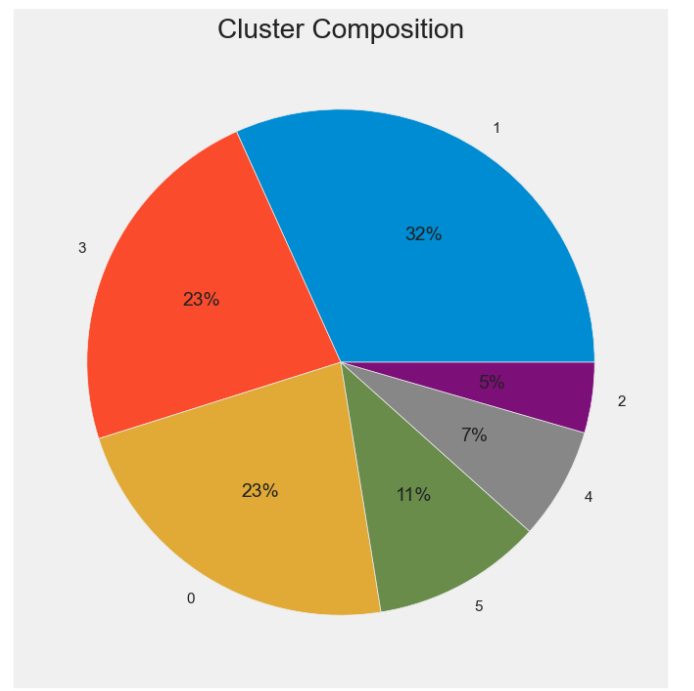
```
In [12]: cc['cluster'] = pipe['clusterer']['kmeans'].labels_
cc.head()
```

Out[12]:

	CLIENTNUM	Attrition_Flag	Customer_Age	Gender	Dependent_count	Education_Level	Marital_Status	Income_Category	Card_Category	Months_on_book	1
0	768805383	Existing Customer	45	M	3	High School	Married	60K–80K	Blue	39	
1	818770008	Existing Customer	49	F	5	Graduate	Single	Less than \$40K	Blue	44	
2	713982108	Existing Customer	51	M	3	Graduate	Married	80K–120K	Blue	36	
3	769911858	Existing Customer	40	F	4	High School	Unknown	Less than \$40K	Blue	34	
4	709106358	Existing Customer	40	M	3	Uneducated	Married	60K–80K	Blue	21	

9. Membuat pie chart untuk menampilkan komposisi kluster dalam dataset cc, dengan label kluster pada sumbu x dan persentase komposisi pada setiap kluster pada bagian pie chart.

```
In [13]: fig, axes = plt.subplots(figsize=(12,8))
ax = cc['cluster'].value_counts().plot.pie(title='Cluster Composition', autopct='%1.0f%%')
plt.title=False
ax.set_ylabel('')
plt.show()
```



D. Examine Differences Between Clusters

1. Visualisasi diagram batang horizontal untuk masing-masing kluster, dapat dilihat distribusi status atrisi (*attrition flag*) yang merupakan kondisi dimana pengguna memutuskan untuk berhenti atau meninggalkan penggunaan kartu kredit dalam kategori pengguna kartu kredit untuk setiap kluster pengguna.



- Melakukan standarisasi pada kolom-kolom yang dipilih dari dataset cc menggunakan StandardScaler(), kemudian hasil standarisasi disimpan dalam dataframe scaled_cc. Selanjutnya, kolom "cluster" yang berisi label klaster yang diprediksi ditambahkan ke dataframe scaled_cc. Menampilkan lima baris pertama dari dataframe scaled_cc yang telah diperbarui

```
In [15]: scaler = StandardScaler()
scaled_cc = pd.DataFrame(scaler.fit_transform(cc[cols]), columns=cols)
scaled_cc['cluster'] = pipe['clusterer']['kmeans'].labels_
scaled_cc.head()
```

```
Out[15]:
```

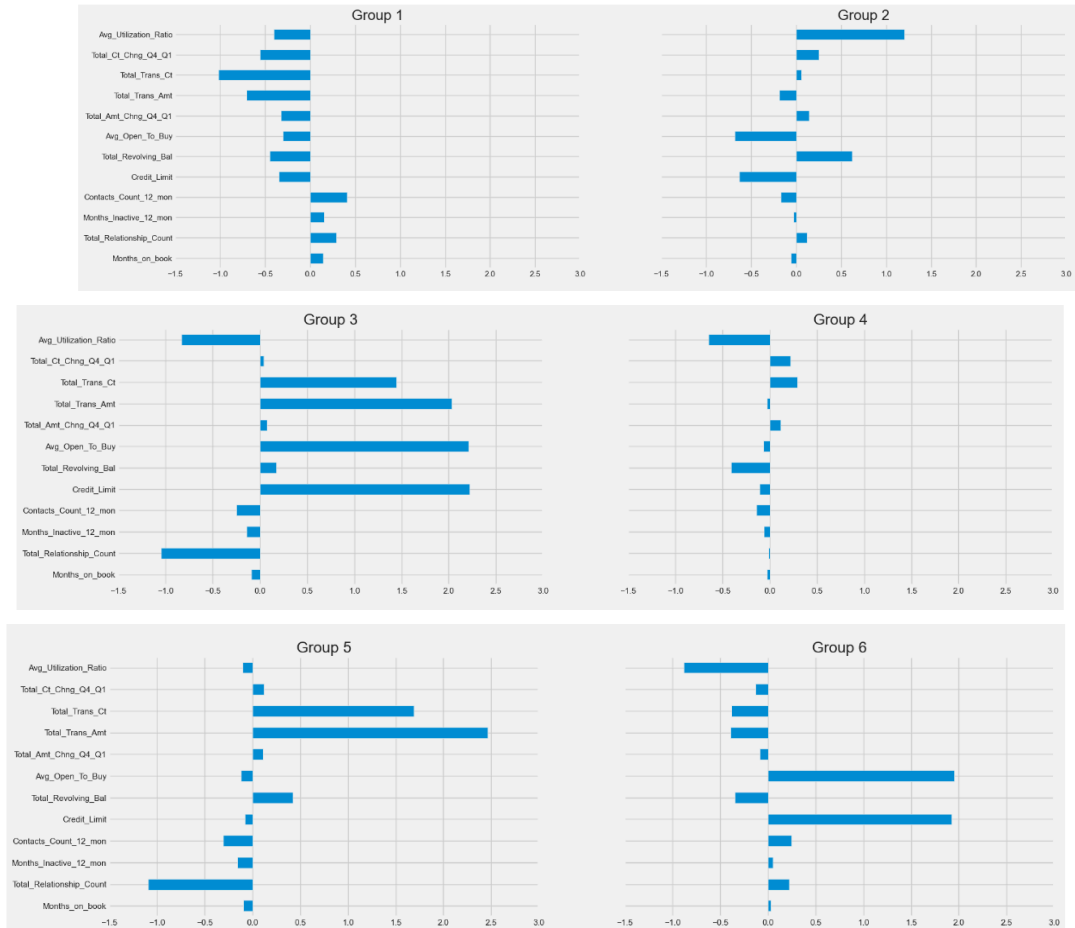
	Months_on_book	Total_Relationship_Count	Months_Inactive_12_mon	Contacts_Count_12_mon	Credit_Limit	Total_Revolving_Bal	Avg_Open_To_Buy	Total_Ar
0	0.384621	0.763943	-1.327136	0.492404	0.446622	-0.473422	0.488971	
1	1.010715	1.407306	-1.327136	-0.411616	-0.041367	-0.366667	-0.008496	
2	0.008965	0.120579	-1.327136	-2.219655	-0.573698	-1.426858	-0.445658	
3	-0.241473	-0.522785	1.641478	-1.315636	-0.585251	1.661686	-0.734100	
4	-1.869317	0.763943	-1.327136	-2.219655	-0.430877	-1.426858	-0.302868	

- Menampilkan bar plot untuk menampilkan rata-rata fitur-fitur numerik yang telah di-standarisasi untuk setiap klaster dalam dataset yang menunjukkan aktivitas akun, dengan setiap subplot mewakili satu klaster dan sumbu x menunjukkan rata-rata fitur serta sumbu y menunjukkan nama fitur-fitur tersebut.


```
In [16]: fix, axs = plt.subplots(ncols=2,rows=3, figsize=(20,32))

scaled_cc[scaled_cc['cluster'] == 0][cols].mean().plot.barh(ax=axs[0,0], xlim=(-1.5, 3), figsize=(20,20), sharey=True, title='Group 1')
scaled_cc[scaled_cc['cluster'] == 1][cols].mean().plot.barh(ax=axs[0,1], xlim=(-1.5, 3), figsize=(20,20), sharey=True, title='Group 2')
scaled_cc[scaled_cc['cluster'] == 2][cols].mean().plot.barh(ax=axs[1,0], xlim=(-1.5, 3), figsize=(20,20), sharey=True, title='Group 3')
scaled_cc[scaled_cc['cluster'] == 3][cols].mean().plot.barh(ax=axs[1,1], xlim=(-1.5, 3), figsize=(20,20), sharey=True, title='Group 4')
scaled_cc[scaled_cc['cluster'] == 4][cols].mean().plot.barh(ax=axs[2,0], xlim=(-1.5, 3), figsize=(20,20), sharey=True, title='Group 5')
scaled_cc[scaled_cc['cluster'] == 5][cols].mean().plot.barh(ax=axs[2,1], xlim=(-1.5, 3), figsize=(20,20), sharey=True, title='Group 6')
```

Out[16]: <Axes: title='center': 'Group 6'>

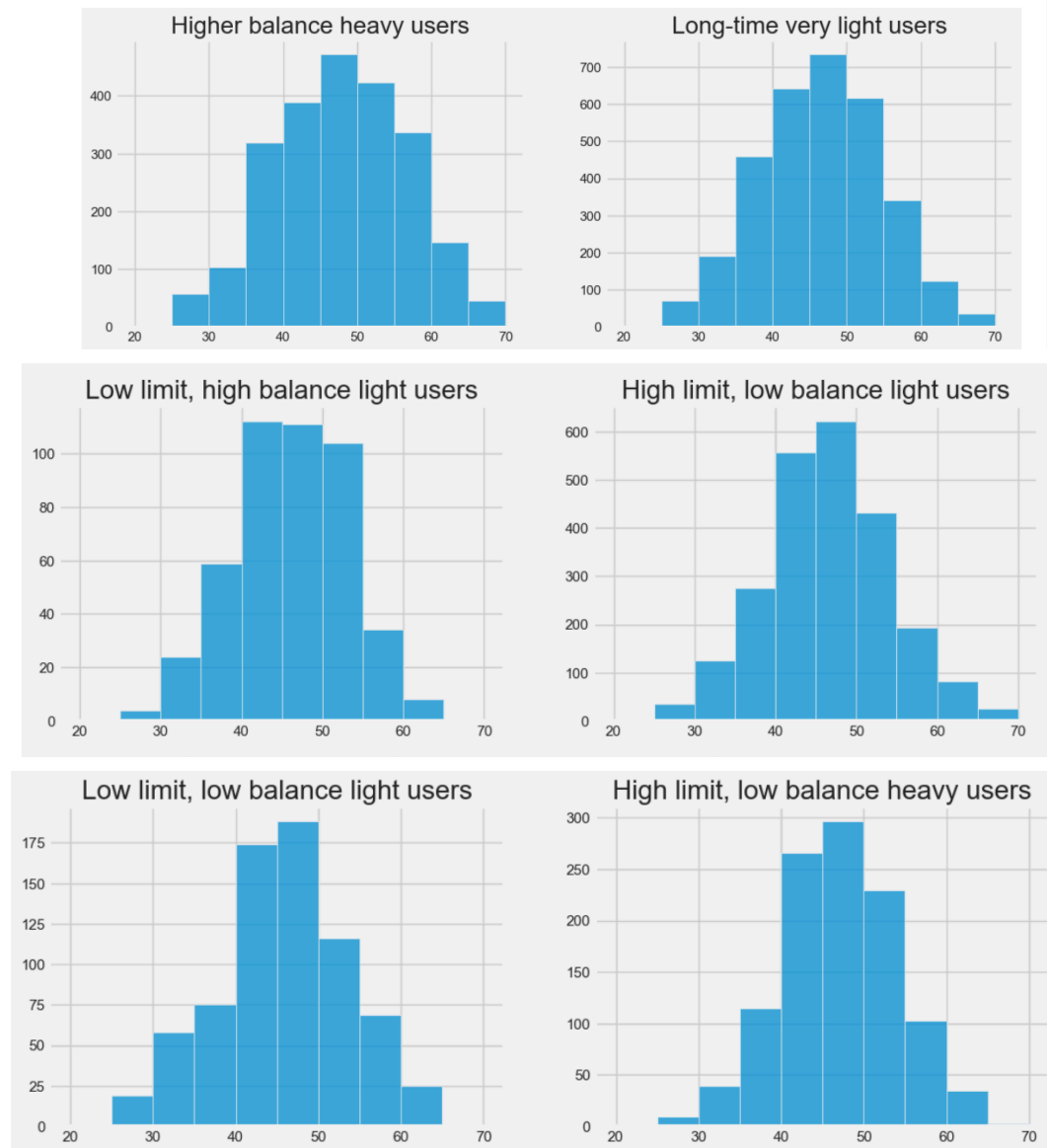


4. Membuat subplot dengan histogram yang menunjukkan distribusi usia pelanggan dalam setiap klaster.

```
In [17]: fig, axs = plt.subplots(ncols=2,nrows=3, figsize=(12,14))

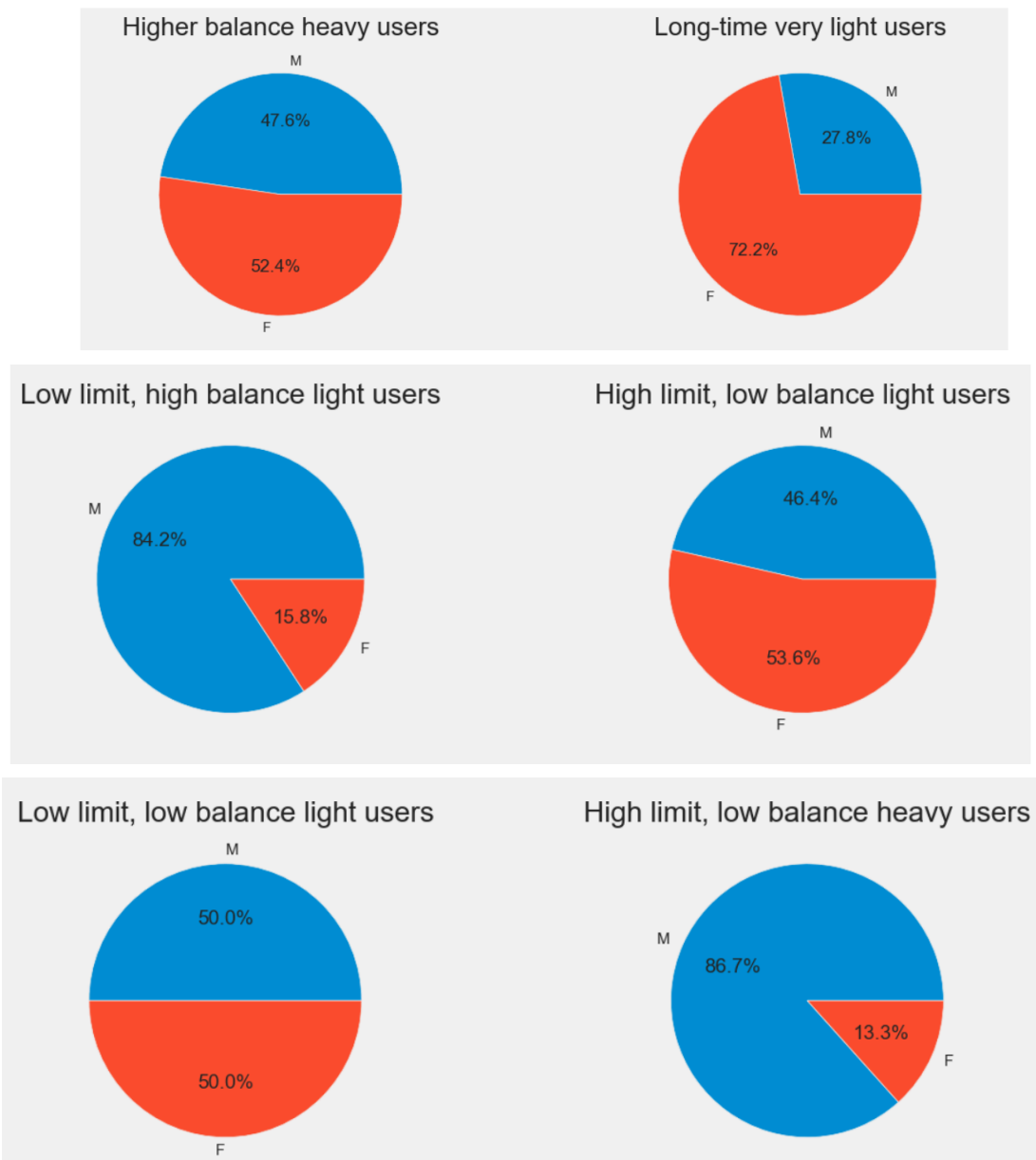
sns.histplot(cc[cc['cluster'] == 0]['Customer_Age'], ax=axs[0,0], binrange=(20,70), binwidth=5).set(title='Higher balance heavy u
sns.histplot(cc[cc['cluster'] == 1]['Customer_Age'], ax=axs[0,1], binrange=(20,70), binwidth=5).set(title='Long-time very light u
sns.histplot(cc[cc['cluster'] == 2]['Customer_Age'], ax=axs[1,0], binrange=(20,70), binwidth=5).set(title='Low limit, high balanc
sns.histplot(cc[cc['cluster'] == 3]['Customer_Age'], ax=axs[1,1], binrange=(20,70), binwidth=5).set(title='High limit, low balanc
sns.histplot(cc[cc['cluster'] == 4]['Customer_Age'], ax=axs[2,0], binrange=(20,70), binwidth=5).set(title='Low limit, low balance
sns.histplot(cc[cc['cluster'] == 5]['Customer_Age'], ax=axs[2,1], binrange=(20,70), binwidth=5).set(title='High limit, low balanc
```

```
Out[17]: [Text(0.5, 1.0, 'High limit, low balance heavy users'),
Text(0, 0.5, ''),
Text(0.5, 0, '')]
```



- Menampilkan pie chart untuk variabel "Gender" pada setiap klaster, dengan persentase masing-masing kategori "Male" dan "Female" ditampilkan di dalam pie chart. Setiap subplot diberi judul yang mencerminkan karakteristik khusus dari setiap klaster yang ditampilkan.

```
Out[18]: <Axes: title=['center': 'High limit, low balance heavy users']>
```

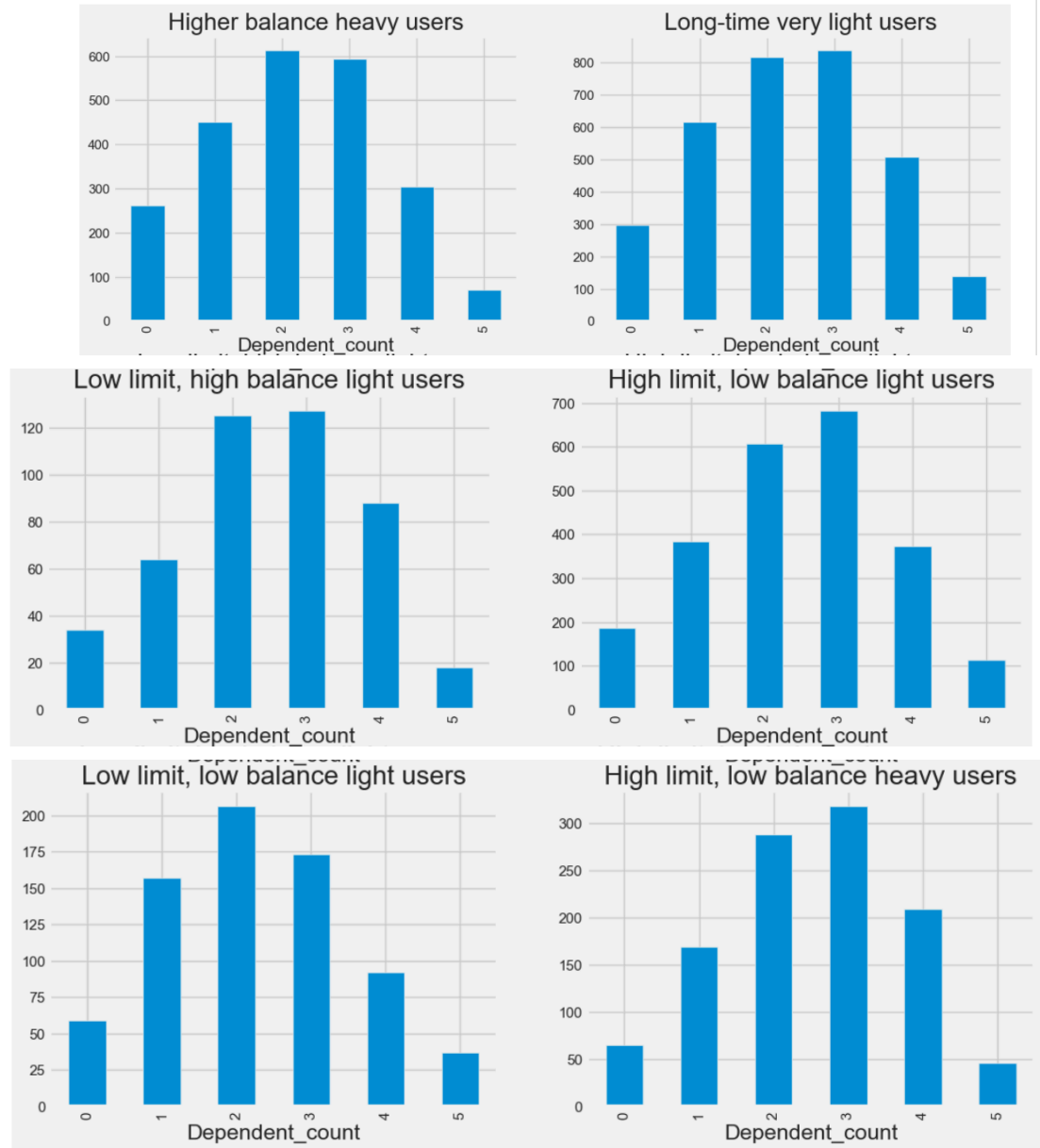


6. Visualisasi diagram batang untuk masing-masing klaster, dapat dilihat pola distribusi jumlah tanggungan (*dependent count*) dalam kategori pengguna kartu kredit untuk setiap klaster pengguna.

```
In [19]: fix, axs = plt.subplots(ncols=2,rows=3, figsize=(12,14))

cc[cc['cluster'] == 0]['Dependent_count'].value_counts().sort_index().plot.bar(ax=axs[0,0], title='Higher balance heavy users')
cc[cc['cluster'] == 1]['Dependent_count'].value_counts().sort_index().plot.bar(ax=axs[0,1], title='Long-time very light users')
cc[cc['cluster'] == 2]['Dependent_count'].value_counts().sort_index().plot.bar(ax=axs[1,0], title='Low limit, high balance light users')
cc[cc['cluster'] == 3]['Dependent_count'].value_counts().sort_index().plot.bar(ax=axs[1,1], title='High limit, low balance light users')
cc[cc['cluster'] == 4]['Dependent_count'].value_counts().sort_index().plot.bar(ax=axs[2,0], title='Low limit, low balance light users')
cc[cc['cluster'] == 5]['Dependent_count'].value_counts().sort_index().plot.bar(ax=axs[2,1], title='High limit, low balance heavy users')
```

```
Out[19]: <Axes: title={'center': 'High limit, low balance heavy users'}, xlabel='Dependent_count'>
```

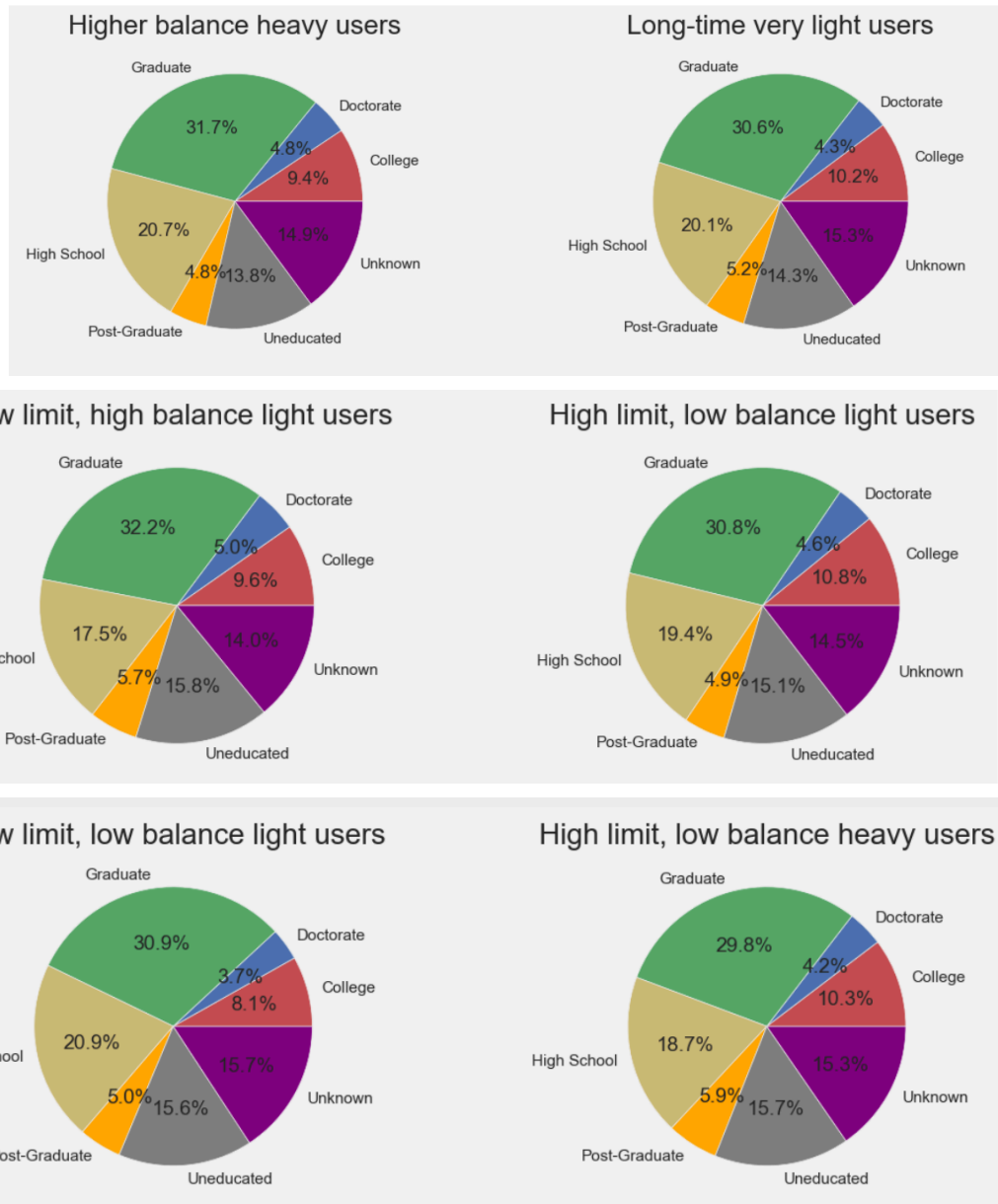


- Visualisasi diagram pie untuk masing-masing kluster, dilihat dari proporsi tingkat pendidikan (*education level*) dalam kategori pengguna kartu kredit untuk setiap kluster pengguna.

```
In [20]: fix, axs = plt.subplots(ncols=2,rows=3, figsize=(12,14))

cc[cc['cluster'] == 0]['Education_Level'].value_counts().sort_index().plot.pie(ax=axs[0,0], ylabel='', title='Higher balance heavy
cc[cc['cluster'] == 1]['Education_Level'].value_counts().sort_index().plot.pie(ax=axs[0,1], ylabel='', title='Long-time very light
cc[cc['cluster'] == 2]['Education_Level'].value_counts().sort_index().plot.pie(ax=axs[1,0], ylabel='', title='Low limit, high bal
cc[cc['cluster'] == 3]['Education_Level'].value_counts().sort_index().plot.pie(ax=axs[1,1], ylabel='', title='High limit, low bal
cc[cc['cluster'] == 4]['Education_Level'].value_counts().sort_index().plot.pie(ax=axs[2,0], ylabel='', title='Low limit, low bal
cc[cc['cluster'] == 5]['Education_Level'].value_counts().sort_index().plot.pie(ax=axs[2,1], ylabel='', title='High limit, low bal
```

```
Out[20]: <Axes: title={'center': 'High limit, low balance heavy users'}>
```

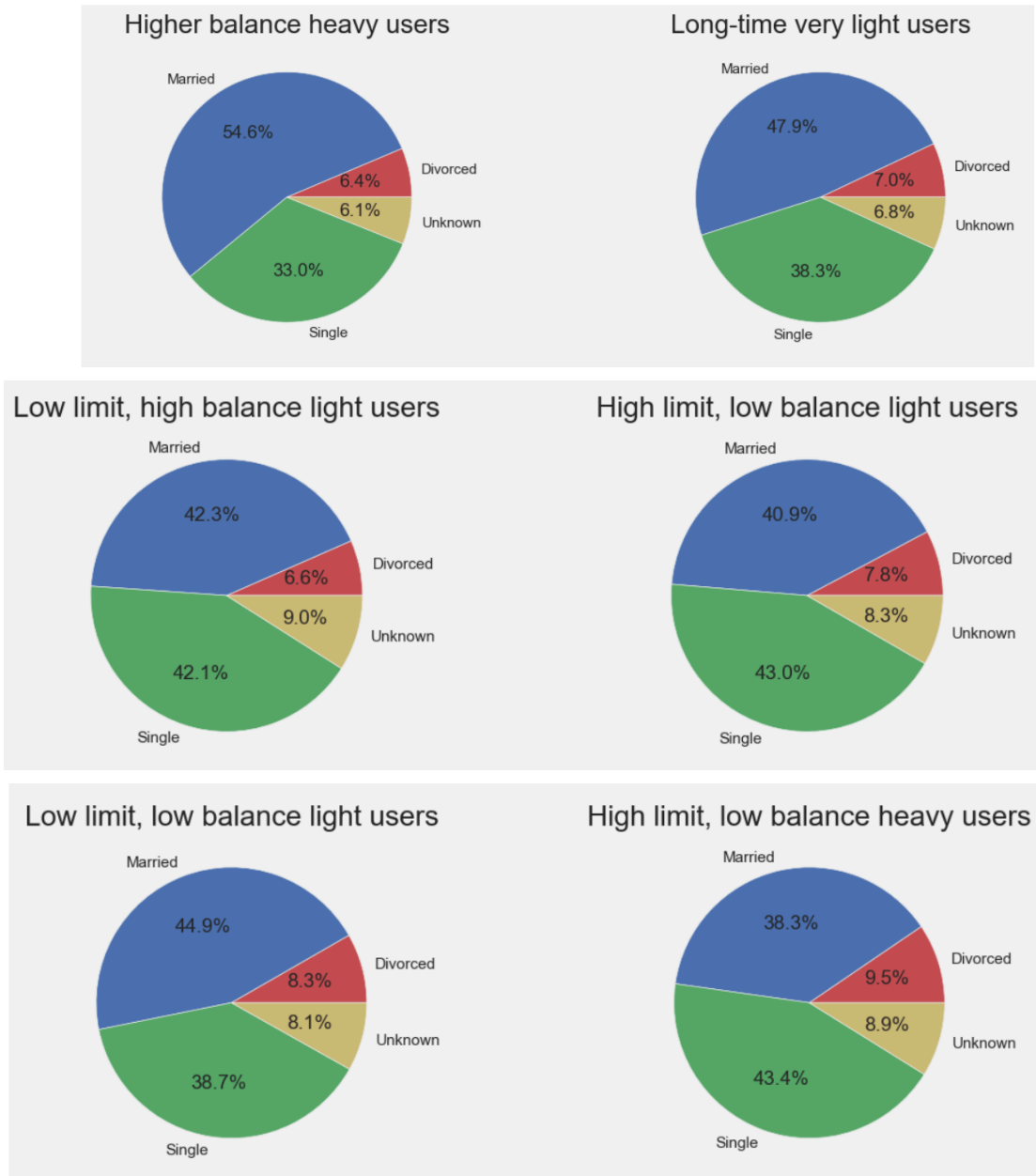


- Visualisasi diagram pie untuk masing-masing klaster, dilihat dari proporsi status pernikahan (*marital status*) dalam kategori pengguna kartu kredit untuk setiap klaster pengguna

```
In [21]: fix, axs = plt.subplots(ncols=2,rows=3, figsize=(12,14))

cc[cc['cluster'] == 0]['Marital_Status'].value_counts().sort_index().plot(kind='pie', ax=axs[0,0], ylabel='', title='Higher balance heavy users')
cc[cc['cluster'] == 1]['Marital_Status'].value_counts().sort_index().plot(kind='pie', ax=axs[0,1], ylabel='', title='Long-time very light users')
cc[cc['cluster'] == 2]['Marital_Status'].value_counts().sort_index().plot(kind='pie', ax=axs[1,0], ylabel='', title='Low limit, high balance light users')
cc[cc['cluster'] == 3]['Marital_Status'].value_counts().sort_index().plot(kind='pie', ax=axs[1,1], ylabel='', title='High limit, low balance light users')
cc[cc['cluster'] == 4]['Marital_Status'].value_counts().sort_index().plot(kind='pie', ax=axs[2,0], ylabel='', title='Low limit, low balance light users')
cc[cc['cluster'] == 5]['Marital_Status'].value_counts().sort_index().plot(kind='pie', ax=axs[2,1], ylabel='', title='High limit, low balance heavy users')
```

```
Out[21]: <Axes: title={'center': 'High limit, low balance heavy users'}>
```

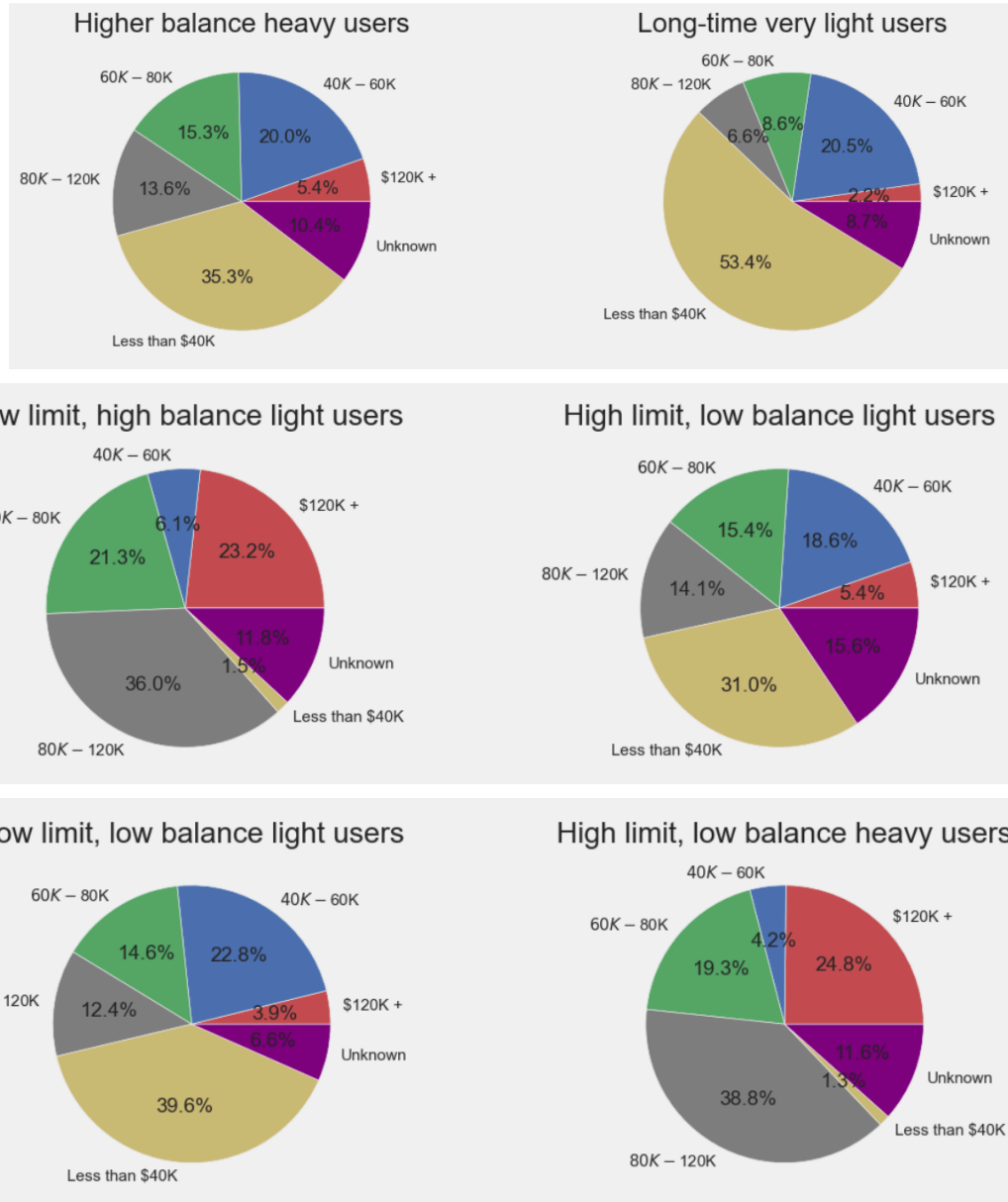


- Visualisasi diagram pie untuk masing-masing kluster, dilihat dari proporsi status penghasilan (*income category*) dalam kategori pengguna kartu kredit untuk setiap kluster pengguna

```
In [22]: fix, axs = plt.subplots(ncols=2, nrows=3, figsize=(12,14))

cc[cc['cluster'] == 0]['Income_Category'].value_counts().sort_index().plot(kind='pie', ax=axs[0,0], ylabel='', title='Higher balance heavy users')
cc[cc['cluster'] == 1]['Income_Category'].value_counts().sort_index().plot(kind='pie', ax=axs[0,1], ylabel='', title='Long-time very light users')
cc[cc['cluster'] == 2]['Income_Category'].value_counts().sort_index().plot(kind='pie', ax=axs[1,0], ylabel='', title='Low limit, high balance light users')
cc[cc['cluster'] == 3]['Income_Category'].value_counts().sort_index().plot(kind='pie', ax=axs[1,1], ylabel='', title='High limit, high balance light users')
cc[cc['cluster'] == 4]['Income_Category'].value_counts().sort_index().plot(kind='pie', ax=axs[2,0], ylabel='', title='Low limit, low balance light users')
cc[cc['cluster'] == 5]['Income_Category'].value_counts().sort_index().plot(kind='pie', ax=axs[2,1], ylabel='', title='High limit, low balance heavy users')
```

```
Out[22]: <Axes: title={'center': 'High limit, low balance heavy users'}>
```

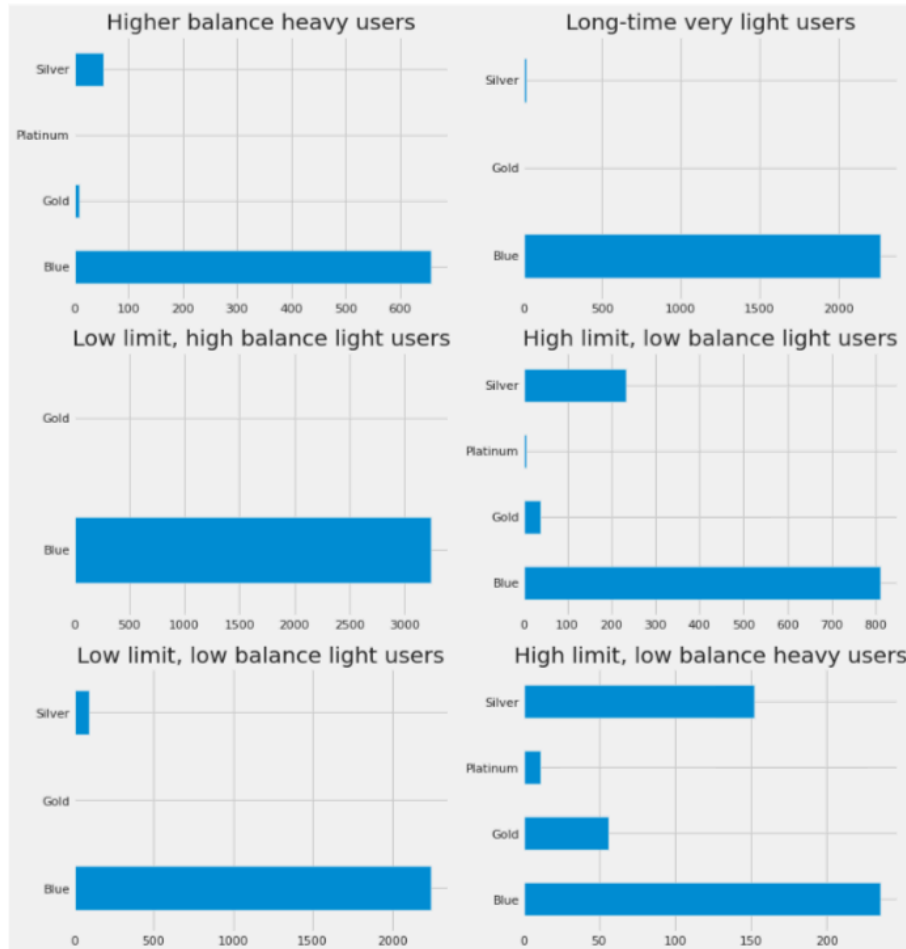


10. Visualisasi diagram batang horizontal untuk masing-masing kluster, dilihat dari distribusi kategori jenis kartu kredit (*card category*) dalam kategori pengguna kartu kredit untuk setiap kluster pengguna

```
In [23]: fix, axs = plt.subplots(ncols=2,rows=3, figsize=(12,14))

cc[cc['cluster'] == 0]['Card_Category'].value_counts().sort_index().plot.barh(ax=axs[0,0], ylabel='', title='Higher balance heavy
cc[cc['cluster'] == 1]['Card_Category'].value_counts().sort_index().plot.barh(ax=axs[0,1], ylabel='', title='Long-time very light
cc[cc['cluster'] == 2]['Card_Category'].value_counts().sort_index().plot.barh(ax=axs[1,0], ylabel='', title='Low limit, high bala
cc[cc['cluster'] == 3]['Card_Category'].value_counts().sort_index().plot.barh(ax=axs[1,1], ylabel='', title='High limit, low bala
cc[cc['cluster'] == 4]['Card_Category'].value_counts().sort_index().plot.barh(ax=axs[2,0], ylabel='', title='Low limit, low bala
cc[cc['cluster'] == 5]['Card_Category'].value_counts().sort_index().plot.barh(ax=axs[2,1], ylabel='', title='High limit, low bala
```

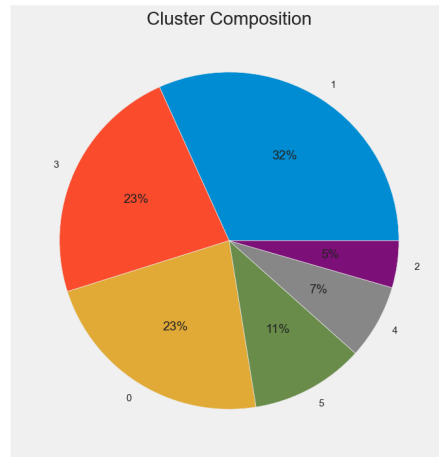
Out[23]: <Axes: title={'center': 'High limit, low balance heavy users'}>



BAB IV PEMBAHASAN

A. Hasil Analisis

1) Komposisi Klaster

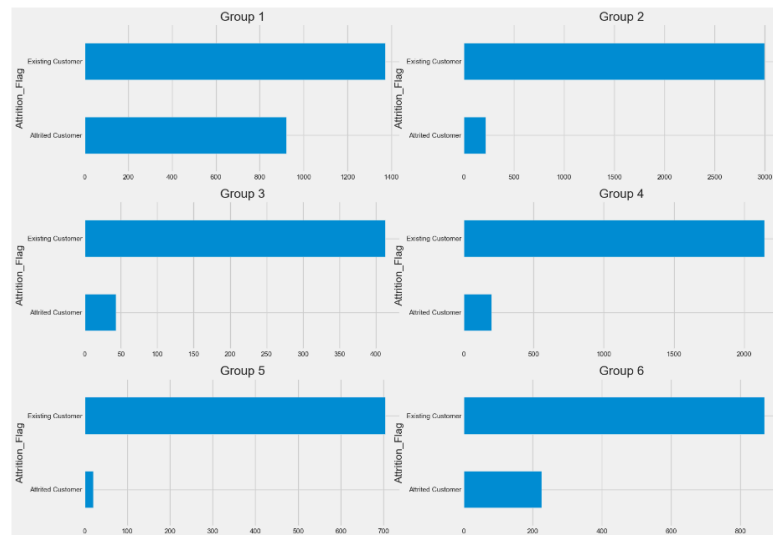


Berdasarkan piechart komposisi klaster, disimpulkan bahwa klaster 0, 1, dan 3 memiliki jumlah user yang lebih banyak dibandingkan dengan klaster 2, 4, dan 5 dengan detail sebagai berikut:

- Klaster 0 / Kelompok 1 memiliki komposisi pengguna sebanyak 23%
- Klaster 1 / Kelompok 2 memiliki komposisi pengguna sebanyak 32%
- Klaster 2 / Kelompok 3 memiliki komposisi pengguna sebanyak 5%
- Klaster 3 / Kelompok 4 memiliki komposisi pengguna sebanyak 23%
- Klaster 4 / Kelompok 5 memiliki komposisi pengguna sebanyak 7%
- Klaster 5 / Kelompok 6 memiliki komposisi pengguna sebanyak 11%

2) Turnover

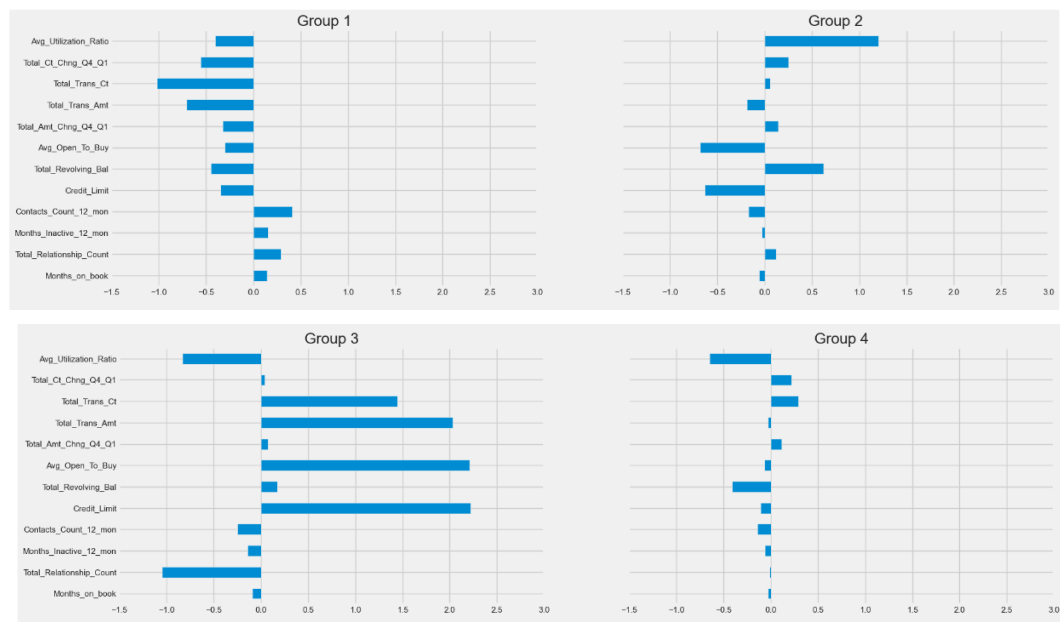
```
Out[14]: <Axes: title={'center': 'Group 6'}, ylabel='Attrition_Flag'>
```

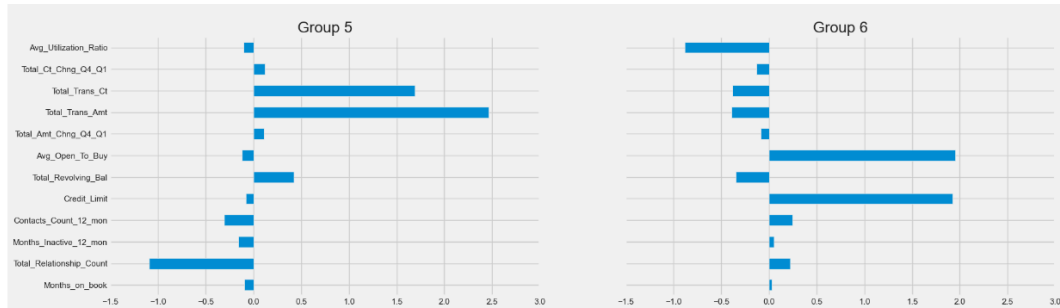


Berdasarkan hasil diagram turnover atau tingkat atrisi customer di atas, disimpulkan bahwa Kelompok 1 dan 6 yang memiliki tingkat atrisi pelanggan yang signifikan lebih tinggi, sementara Kelompok 5 memiliki tingkat atrisi yang signifikan lebih rendah. Hal ini berarti bahwa pengguna dalam Kelompok 1 dan 6 cenderung lebih mungkin untuk berhenti menjadi pelanggan dibandingkan dengan pengguna dalam Kelompok 5 sehingga menunjukkan bahwa ada faktor-faktor tertentu yang membuat pengguna dalam Group 1 dan 6 lebih rentan terhadap atrisi pelanggan.

3) Aktivitas Akun

```
<Axes: title={'center': 'Group 6'}>
```



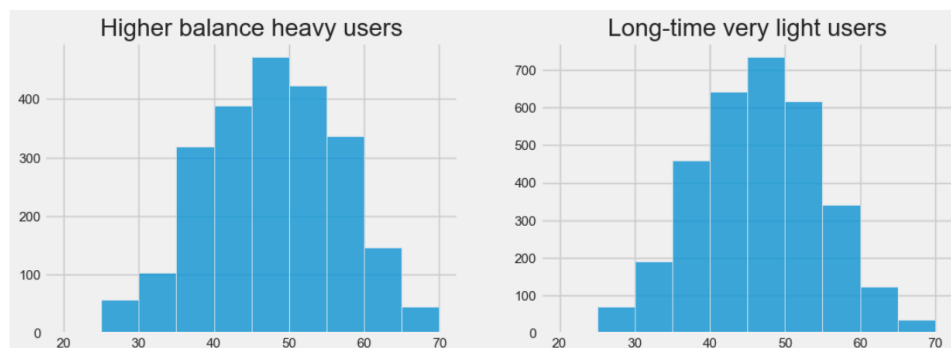


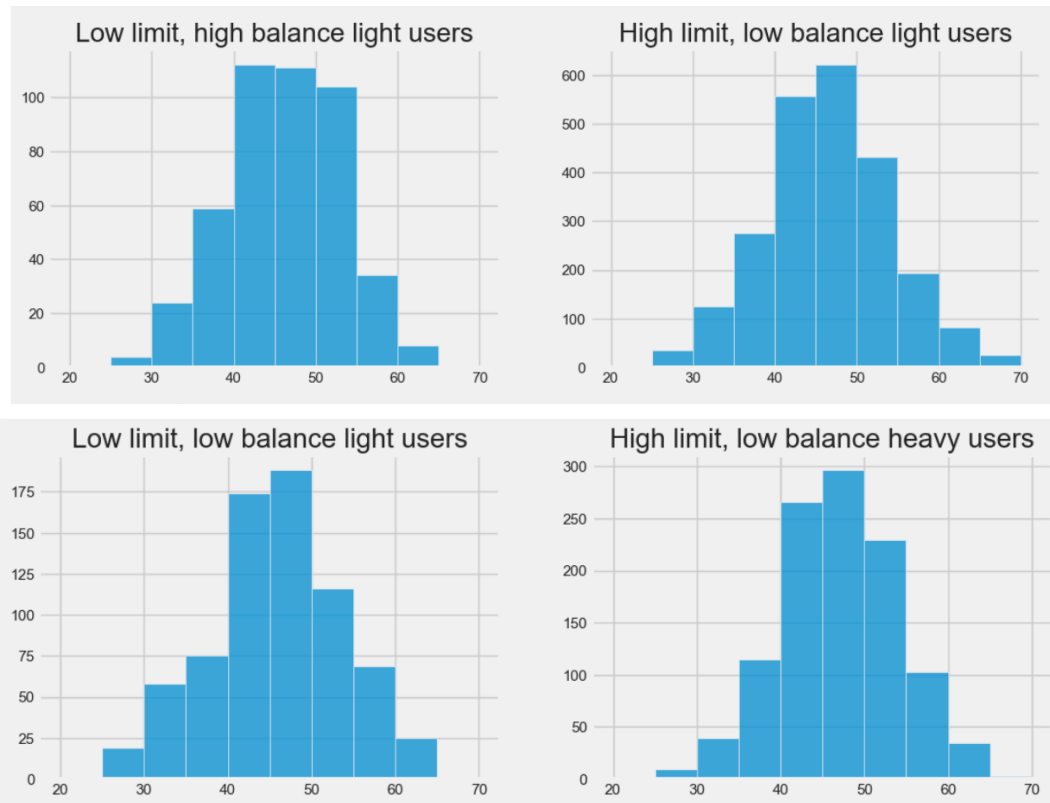
Berdasarkan diagram aktivitas akun di atas, dapat disimpulkan bahwa

- Kelompok 1 ("*Higher balance heavy users*") terdiri dari pengguna dengan saldo tinggi yang sering bertransaksi
- Kelompok 2 ("*Long-time very light users*") terdiri dari pengguna dengan tingkat penggunaan kartu kredit yang rendah namun telah menjadi pelanggan yang sudah lama
- Kelompok 3 ("*Low limit, high balance light users*") terdiri dari pengguna dengan limit kredit yang rendah namun memiliki saldo yang tinggi, dan transaksi yang sedikit
- Kelompok 4 ("*High limit, low balance light users*") terdiri dari pengguna dengan limit kredit yang tinggi namun saldo yang rendah, dan transaksi yang sedikit
- Kelompok 5 ("*Low limit, low balance light users*") terdiri dari pengguna dengan limit kredit dan saldo yang rendah, dan transaksi yang sedikit
- Kelompok 6 ("*High limit, low balance heavy users*") terdiri dari pengguna dengan limit kredit yang tinggi namun saldo yang rendah, dan transaksi yang sering

4) Umur

```
Out[17]: [Text(0.5, 1.0, 'High limit, low balance heavy users'),
Text(0, 0.5, ''),
Text(0.5, 0, '')]
```

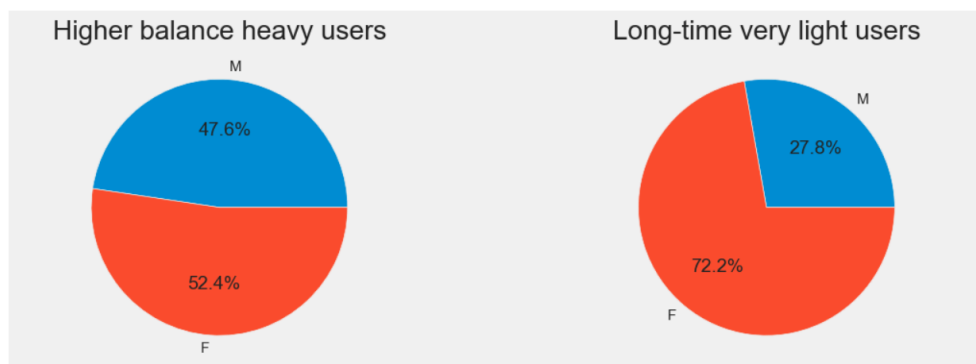


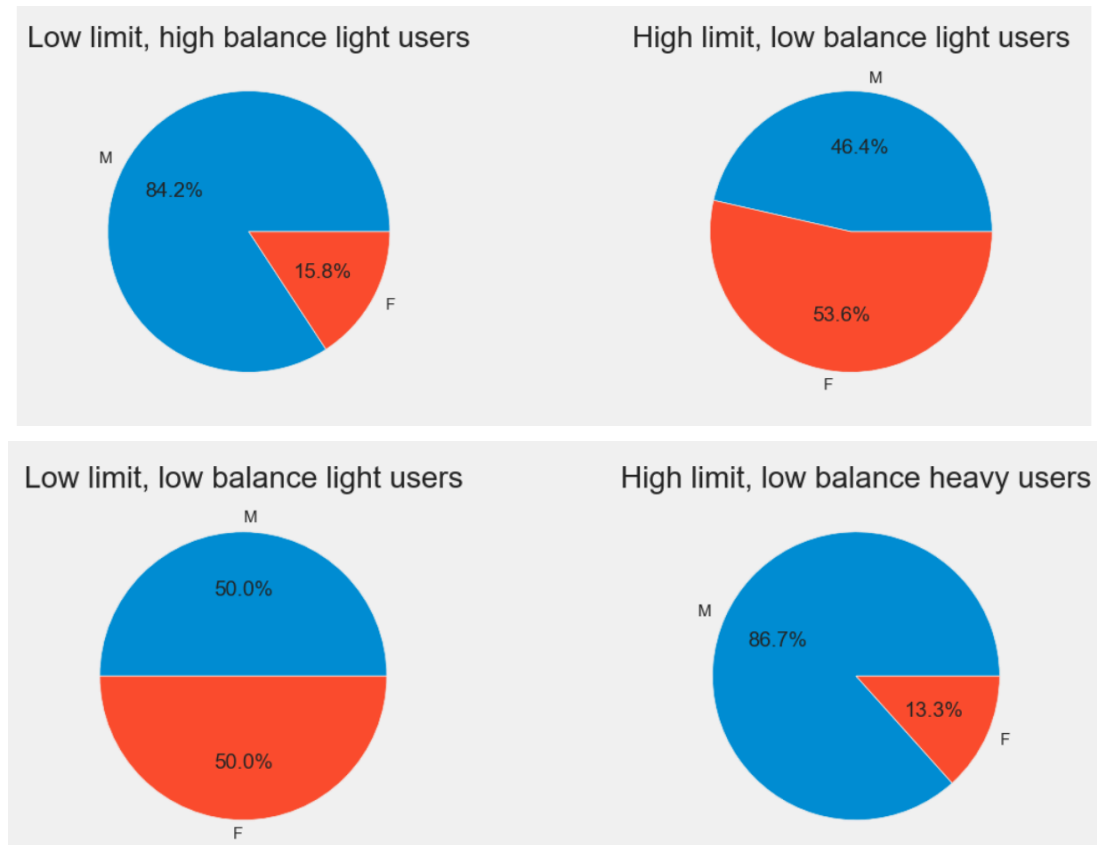


Dari diagram keenam kelompok di atas, sebagian besar pengguna berada pada rentang usia 40-an dan 50-an untuk semua klaster. Pengguna dalam klaster Higher balance heavy user memiliki jumlah pengguna yang lebih tinggi pada rentang usia 60-an dibandingkan dengan klaster lainnya. Pengguna dalam klaster Higher balance heavy user dan Long-time very light users juga memiliki jumlah pengguna yang lebih tinggi pada rentang usia 30-an.

5) Gender

Out[18]: <Axes: title={'center': 'High limit, low balance heavy users'}>

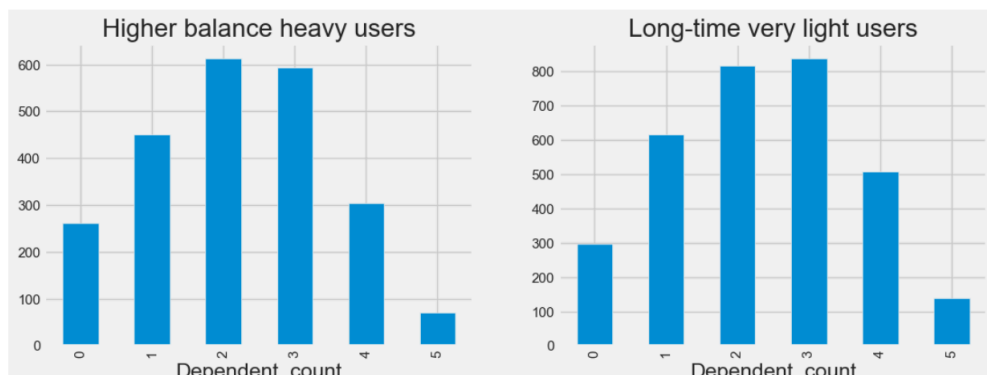


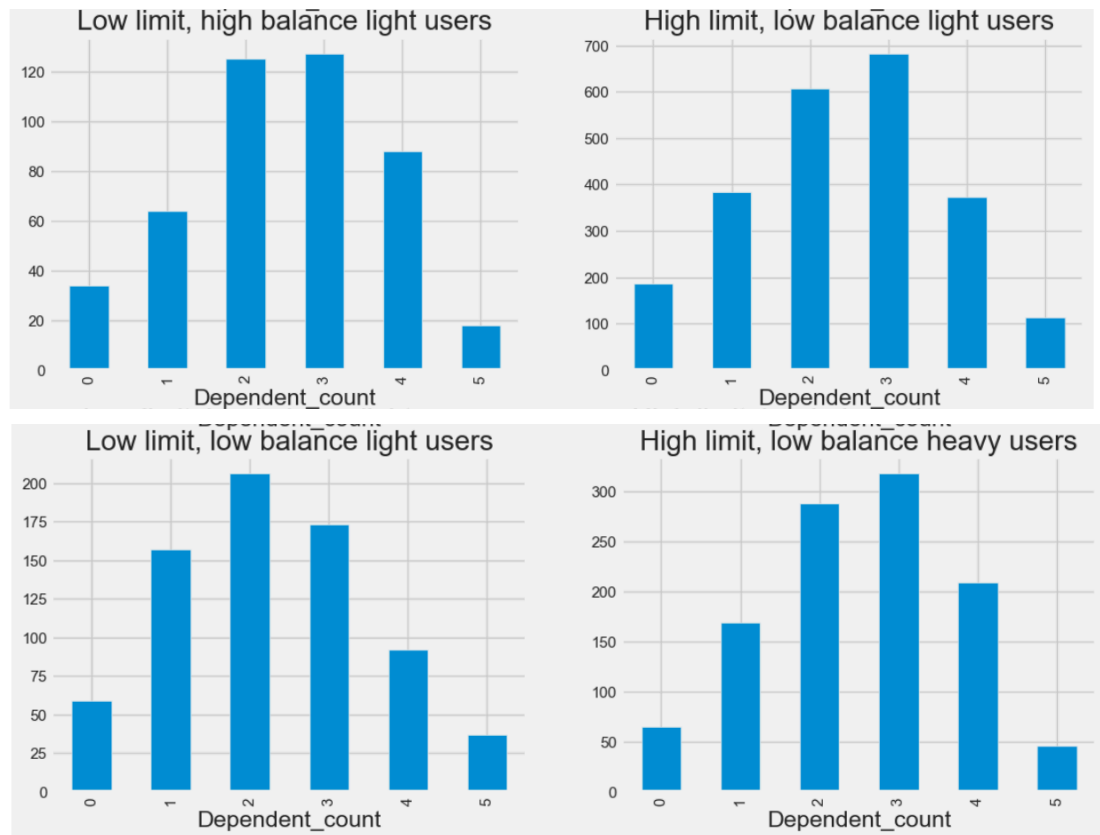


Dari hasil diagram di atas, sebagian besar pria merupakan pengguna dalam klaster Low limit high balance light users, dan High limit low balance heavy users. Sedangkan, sebagian besar wanita merupakan pengguna dalam klaster Long-time very light users dan High limit low balance light users

6) Jumlah Tanggungan

```
Out[19]: <Axes: title={'center': 'High limit, low balance heavy users'}, xlabel='Dependent_count'>
```

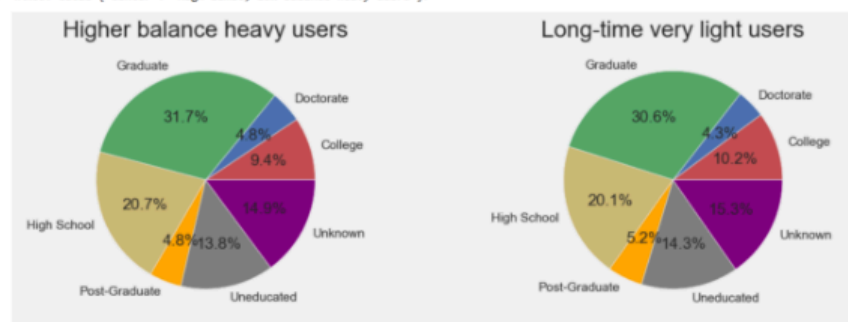


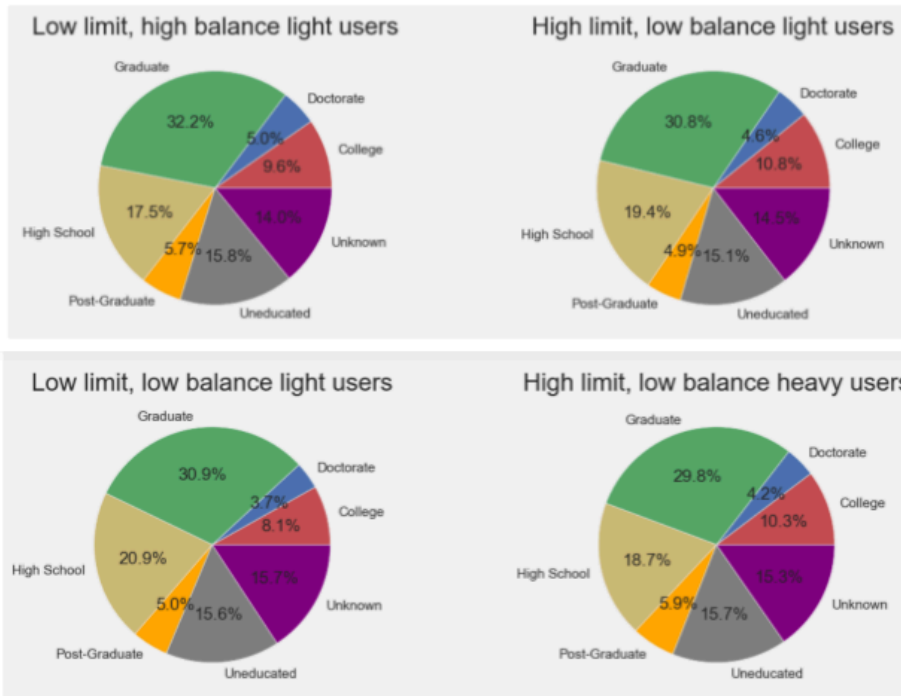


Dari diagram di atas, Tidak terlihat adanya perbedaan yang signifikan dalam jumlah tanggungan antara kelompok-kelompok tersebut.

7) Tingkat Pendidikan

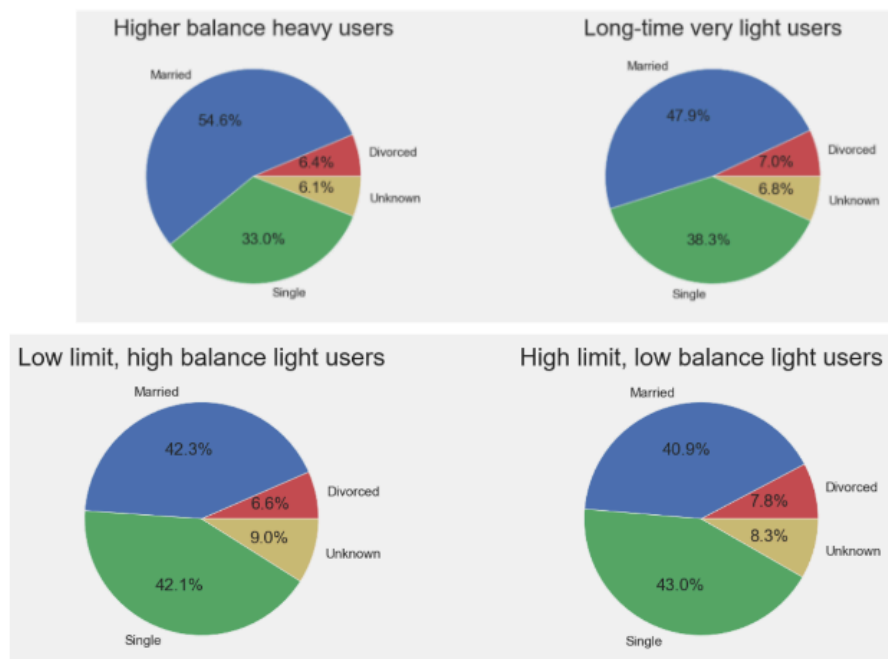
```
Out[20]: <Axes: title={'center': 'High limit, low balance heavy users'}>
```

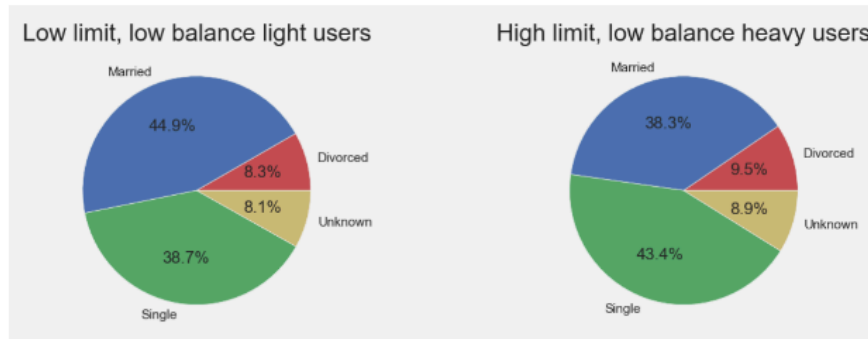




Dari diagram di atas juga tidak terlihat adanya perbedaan yang signifikan dalam tingkat pendidikan antara kelompok-kelompok tersebut.

8) Status Pernikahan

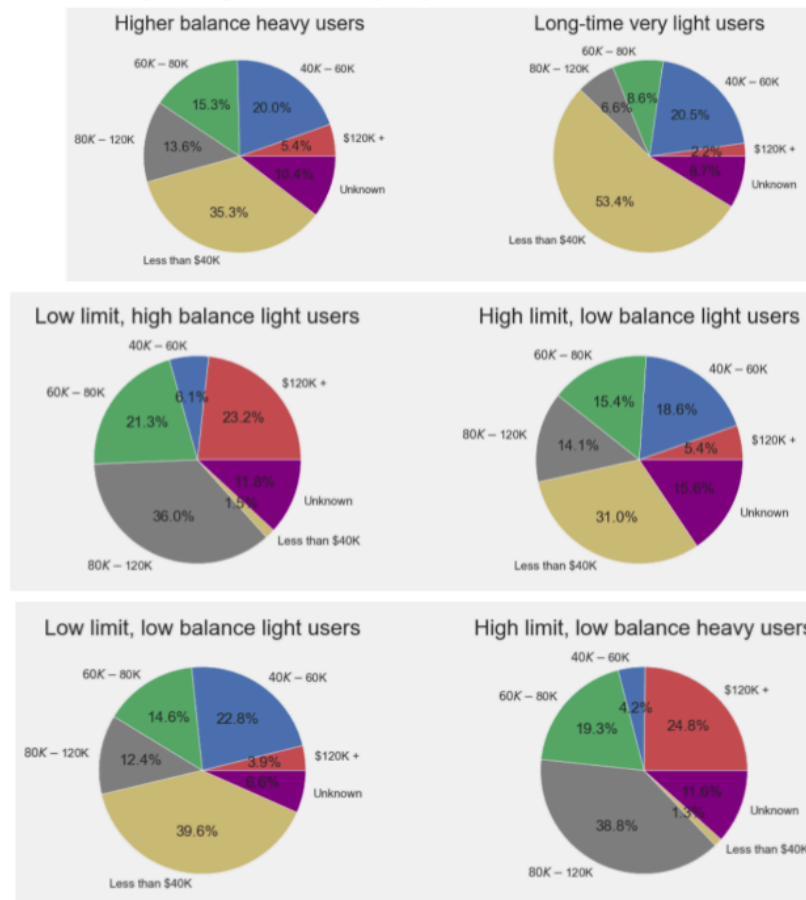




Dari diagram lingkaran di atas, pengguna dalam kelompok high balance heavy users memiliki jumlah pengguna yang menikah secara signifikan lebih tinggi dibandingkan dengan kelompok lainnya.

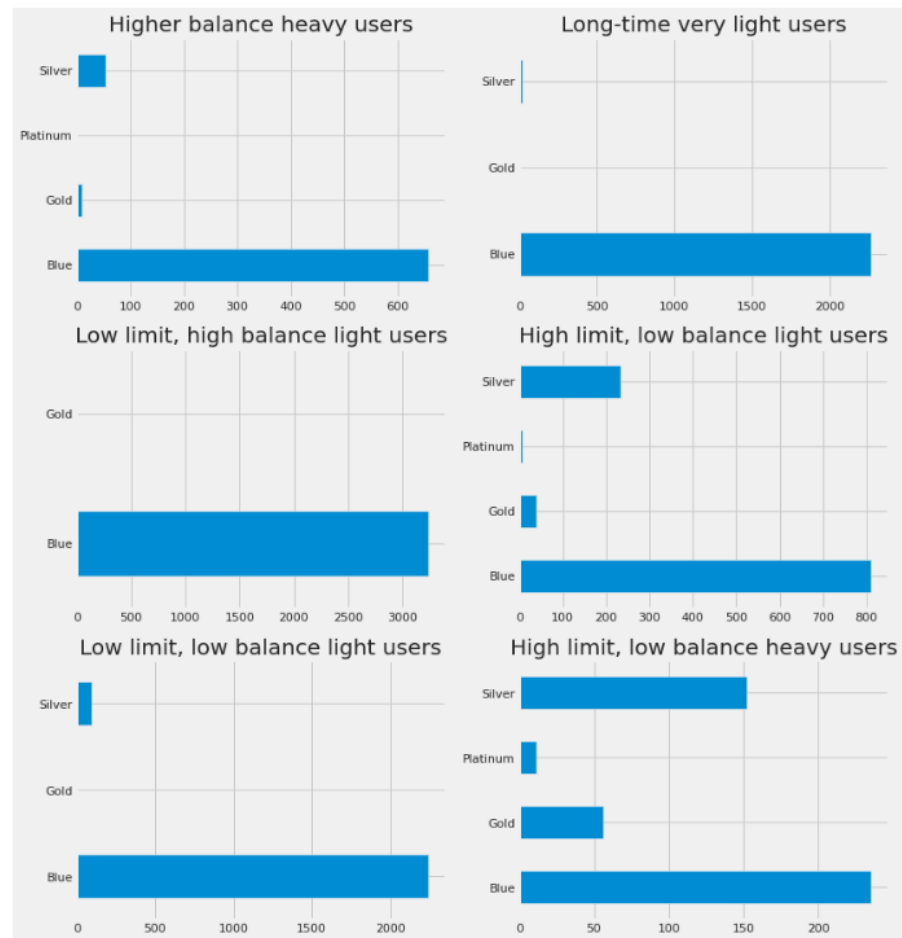
9) Tingkat Pendapatan

Out[22]: <Axes: title=[center]: 'High limit, low balance heavy users'>



Dari diagram di atas, terlihat batas kredit dan pendapatan berkorelasi secara signifikan, dengan setidaknya 30% pengguna dalam kelompok high limit memiliki pendapatan minimal 60K, dan setidaknya 50% pengguna dalam kelompok long time very light memiliki pendapatan kurang dari 60 ribu.

10) Tipe Kartu



Dari hasil diagram di atas, seperti yang diharapkan, pemegang kartu Silver, Gold, dan Platinum kebanyakan tergabung dalam kelompok batas kredit tinggi.

B. Strategi dan Rekomendasi

Dari hasil analisis di atas, terdapat strategi dan rekomendasi yang bisa dilakukan terhadap setiap kelompok pengguna kartu kredit, diantaranya adalah sebagai berikut.

1) Kelompok 1: Pengguna dengan saldo tinggi

- Tawarkan imbalan dan manfaat yang disesuaikan dengan kebiasaan dan preferensi pengeluaran mereka.
- Berikan penawaran personal yang dapat meningkatkan penggunaan kartu dan saldo mereka.
- Pertimbangkan untuk menyediakan opsi transfer saldo dengan suku bunga rendah untuk mendorong mereka menggabungkan utang-utang mereka.

2) Kelompok 2: Pengguna dengan penggunaan kartu yang sangat rendah namun telah lama menjadi pengguna

- Terapkan strategi untuk meningkatkan keterlibatan dan penggunaan kartu pada kelompok ini.
 - Tawarkan insentif seperti pengembalian uang tunai atau diskon khusus untuk mendorong penggunaan kartu mereka.
- 3) Kelompok 3: Pengguna dengan batas kredit rendah namun saldo tinggi
- Berikan penawaran atau promosi khusus untuk membantu pengguna mengelola saldo mereka dan mengurangi utang.
 - Ajukan peningkatan batas kredit bagi mereka yang memiliki rekam jejak penggunaan kartu yang baik.
- 4) Kelompok 4: Pengguna dengan batas kredit tinggi namun saldo rendah
- Tawarkan program hadiah atau manfaat eksklusif yang menarik bagi kelompok ini untuk mendorong penggunaan kartu mereka.
 - Berikan layanan pelanggan yang unggul untuk mempertahankan kepuasan mereka sebagai pengguna dengan batas kredit tinggi.
- 5) Kelompok 5: Pengguna dengan batas kredit rendah dan saldo rendah
- Tawarkan program penghargaan atau diskon yang relevan dengan kebutuhan dan preferensi pengguna.
 - Berikan pendidikan keuangan atau saran untuk membantu mereka meningkatkan manajemen keuangan mereka dan memperbaiki profil kredit.
- 6) Kelompok 6: Pengguna dengan batas kredit tinggi namun saldo rendah
- Berikan penawaran khusus seperti penawaran bunga rendah atau promosi untuk mendorong penggunaan lebih lanjut dari kartu mereka.
 - Tawarkan insentif tambahan yang relevan dengan gaya hidup dan preferensi mereka untuk mendorong penggunaan kartu yang lebih aktif.

Selain itu, penting juga untuk melakukan survei atau penelitian lebih lanjut guna memahami kebutuhan, preferensi, dan masalah yang spesifik dalam setiap kelompok pengguna. Hal ini akan membantu dalam mengembangkan strategi yang lebih efektif dan terarah.

BAB V

PENUTUP

A. Kesimpulan

Berdasarkan hasil pembahasan tersebut, maka dapat disimpulkan bahwa data mining menjadi salah satu teknik yang digunakan dalam mengolah data yang efektif untuk menghasilkan pengetahuan yang berguna dalam pengambilan keputusan. Hasil analisis model kami, berdasarkan sepuluh analisis diantaranya tentang komposisi kluster, turnover, aktivitas akun, umur, gender, jumlah tanggungan, tingkat pendidikan, status pernikahan, tingkat pendapatan, dan tipe kartu bahwa pengguna kartu kredit dapat dikelompokkan ke dalam enam kelompok dengan karakteristik yang berbeda. Pengguna dengan saldo tinggi, pengguna dengan penggunaan kartu yang sangat rendah namun telah lama menjadi pengguna, dan pengguna dengan batas kredit rendah namun saldo tinggi adalah kelompok yang perlu mendapatkan perhatian khusus.

Rekomendasi strategis termasuk memberikan insentif dan manfaat yang disesuaikan dengan kebutuhan dan kebiasaan pengguna, meningkatkan keterlibatan dan penggunaan kartu, serta memberikan program pengelolaan saldo dan pendidikan keuangan. Di sisi lain, pengguna dengan batas kredit tinggi namun saldo rendah dan pengguna dengan batas kredit rendah dan saldo rendah membutuhkan penawaran khusus, insentif tambahan, dan program penghargaan yang relevan untuk mendorong penggunaan aktif kartu kredit. Penting untuk melakukan survei lebih lanjut guna memahami kebutuhan dan masalah spesifik dalam setiap kelompok pengguna agar dapat mengembangkan strategi yang lebih efektif dan sesuai.

B. Kritik dan Saran

Meskipun laporan analisis ini memberikan pengetahuan dan wawasan tentang kelompok pengguna kartu kredit, ada beberapa kritik dan saran yang dapat diberikan walaupun kelompok pengguna telah diberi label berdasarkan karakteristik mereka, tidak dijelaskan dengan jelas mengapa pengguna termasuk ke dalam kelompok tertentu dan bagaimana pengelompokan tersebut dapat memberikan pemahaman yang lebih mendalam tentang perilaku pengguna. Disarankan untuk menyelidiki lebih lanjut mengapa pengguna dalam kelompok tertentu lebih cenderung berpindah atau mempertahankan kartu mereka, dan menyusun strategi yang lebih personal dan disesuaikan untuk setiap kelompok. Selain itu, penekanan yang lebih besar pada segmentasi demografis dapat membantu dalam mengidentifikasi kebutuhan dan preferensi pengguna yang lebih spesifik. Untuk pemahaman yang lebih baik, akan berguna jika laporan juga mencakup faktor-faktor ekonomi dan sosial yang dapat mempengaruhi perilaku pengguna kartu kredit.

DAFTAR PUSTAKA

- Suntoro, J. (2019). *Data Mining : Algoritma dan Implementasi dengan Pemrograman PHP*. Jakarta: PT Elex Media Komputindo.
- F.N.R. Fauzan, J. Aziz, B.D. Setiawan, dan I. Arwani, “Implementasi Algoritma K-Means untuk Klasterisasi Kinerja Akademik Mahasiswa”, *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 2. No. 6, 2018.