

LAPORAN ANALISIS DATA TRANSAKSI & PREDIKSI

Fitri Fatimah¹, Nadine Valia Azzahra², Kanaya Dzikra³, Amelia Regina⁴

1. Universitas Koperasi Indonesia
Kawasan Pendidikan Tinggi Jl. Raya Jatinangor No.KM. 20
fitrinpitt0611@gmail.com
2. Universitas Koperasi Indonesia
Kawasan Pendidikan Tinggi Jl. Raya Jatinangor No.KM. 20
nadinevalia150@gmail.com
3. Universitas Koperasi Indonesia
Kawasan Pendidikan Tinggi Jl. Raya Jatinangor No.KM. 20
knydzkraa435@gmail.com
4. Universitas Koperasi Indonesia
Kawasan Pendidikan Tinggi Jl. Raya Jatinangor No.KM. 20
ameliareginapermana16@gmail.com

ABSTRAK

Dalam era digital yang didominasi oleh data, analisis dan visualisasi data menjadi aspek penting dalam pengambilan keputusan berbasis informasi. Analisis data membantu mengungkap pola dan wawasan tersembunyi yang dapat mendukung keputusan strategis di berbagai bidang, seperti bisnis, kesehatan, dan inovasi. Visualisasi data berperan dalam menyajikan informasi yang kompleks secara intuitif dan menarik, memudahkan pemahaman bagi para pengambil keputusan. Python, sebagai bahasa pemrograman yang populer, menawarkan kemudahan penggunaan dan fleksibilitas untuk berbagai kebutuhan, termasuk analisis data.

Dengan ekosistem pustaka yang kaya, seperti Pandas untuk manipulasi data dan Anaconda sebagai platform manajemen lingkungan dan paket, Python menjadi alat utama dalam analisis data modern. Pandas memungkinkan pengolahan data secara efisien melalui struktur Series dan DataFrame, sementara Anaconda menyediakan solusi pengelolaan proyek yang praktis. Kombinasi ini menjadikan Python dan alat pendukungnya sebagai fondasi yang kuat untuk memanfaatkan data secara optimal di era digital ini.

Keyword : Analisis data, Visualisasi data, Python, Pandas, Anaconda, Era digital, Pengambilan keputusan, Manipulasi data, Ekosistem pustaka, Series dan DataFrame.

PENDAHULUAN

Di era digital saat ini, analisis dan visualisasi data telah menjadi kebutuhan penting yang tidak bisa diabaikan, terutama dalam dunia yang semakin terhubung oleh teknologi. Setiap hari, data dihasilkan dalam jumlah yang sangat besar dari berbagai sumber, seperti media sosial, perangkat IoT, transaksi e-commerce, hingga aktivitas sehari-hari manusia. Namun, data mentah tersebut tidak memiliki nilai jika tidak diolah dan dianalisis dengan baik. Analisis data membantu kita menemukan pola, tren, dan wawasan yang tersembunyi, sehingga dapat digunakan untuk membuat keputusan yang lebih bijaksana dan strategis. Misalnya, dalam dunia bisnis, analisis data memungkinkan perusahaan memahami perilaku konsumen, memprediksi kebutuhan pasar, dan meningkatkan efisiensi operasional.

Sementara itu, visualisasi data menjadi alat yang sangat efektif untuk menyampaikan hasil analisis secara intuitif dan mudah dipahami. Dengan bantuan visualisasi, informasi yang kompleks dapat diubah menjadi grafik, diagram, atau peta interaktif yang menarik dan informatif. Hal ini tidak hanya membantu para ahli data untuk berbagi temuan mereka, tetapi juga memudahkan pemangku kepentingan, seperti manajer atau klien, dalam memahami informasi tanpa harus memiliki latar belakang teknis yang mendalam. Selain itu, visualisasi data memungkinkan pengambilan keputusan dilakukan dengan cepat, karena informasi dapat diserap dalam waktu singkat.

Keuntungan lainnya adalah kemampuan untuk beradaptasi dengan cepat terhadap perubahan. Dalam dunia bisnis, pemerintah, atau sektor lainnya, analisis data dapat membantu mengidentifikasi potensi risiko atau peluang secara lebih dini. Misalnya, dalam bidang kesehatan, data dapat digunakan untuk memantau penyebaran penyakit dan merancang intervensi yang tepat waktu. Sementara dalam industri kreatif, visualisasi data dapat membantu memahami tren yang sedang berkembang, sehingga memungkinkan inovasi yang relevan dengan kebutuhan pasar.

Secara keseluruhan, analisis dan visualisasi data adalah fondasi utama dalam mengambil keputusan yang berbasis fakta di era digital ini. Kemampuan untuk memahami dan memanfaatkan data bukan hanya memberikan keunggulan kompetitif, tetapi juga membuka peluang baru dalam berbagai bidang kehidupan manusia. Hal ini menjadikan analisis dan visualisasi data sebagai keterampilan yang sangat berharga di masa sekarang dan masa depan.

TINJAUAN PUSTAKA

A. PYTHON

Python adalah bahasa pemrograman tingkat tinggi yang bersifat interpretatif dan dirancang untuk kemudahan penggunaan serta keterbacaan. Diciptakan oleh Guido van Rossum dan dirilis pertama kali pada tahun 1991, Python telah berkembang menjadi salah satu bahasa pemrograman yang paling populer di dunia. Keunggulan utama Python terletak pada sintaksnya yang sederhana dan mirip dengan bahasa Inggris, yang memungkinkan programmer untuk menulis kode dengan lebih sedikit baris dibandingkan dengan banyak bahasa lainnya. Python mendukung berbagai paradigma pemrograman, termasuk pemrograman prosedural, berorientasi objek, dan fungsional, sehingga memberikan fleksibilitas kepada pengembang dalam menyusun solusi untuk berbagai masalah. Selain itu, Python memiliki ekosistem yang kaya dengan pustaka dan kerangka kerja, menjadikannya pilihan ideal untuk berbagai aplikasi, mulai dari pengembangan web hingga analisis data dan kecerdasan buatan.

B. ANACONDA

Anaconda adalah distribusi open-source dari Python yang dirancang khusus untuk komputasi ilmiah dan analisis data. Anaconda menyediakan alat manajemen paket dan lingkungan yang sangat berguna bagi para ilmuwan data dan pengembang. Dengan lebih dari 1.000 paket data sains yang sudah terinstal sebelumnya, Anaconda memudahkan pengguna untuk mengatur proyek mereka dalam lingkungan terisolasi menggunakan conda, manajer pakatnya. Ini memungkinkan pengguna untuk mengelola dependensi perangkat lunak secara efisien tanpa risiko konflik versi. Salah satu fitur unggulan Anaconda adalah Anaconda Navigator, antarmuka grafis yang memungkinkan pengguna untuk meluncurkan aplikasi pengembangan dan notebook dari lingkungan yang dikelola dengan mudah. Anaconda juga mendukung berbagai platform, termasuk Windows, macOS, dan Linux, sehingga dapat diakses oleh banyak pengguna di berbagai sistem operasi.

C. PANDAS

Pandas adalah pustaka Python yang sangat populer untuk analisis data, diciptakan oleh Wes McKinney pada tahun 2008. Pustaka ini menyediakan dua struktur data utama: *Series* dan *DataFrame*. *Series* adalah struktur data satu dimensi yang mirip dengan array atau daftar, sedangkan *DataFrame* adalah struktur dua dimensi yang menyerupai tabel dalam basis data atau spreadsheet. Pandas memungkinkan pengguna untuk melakukan manipulasi data secara efisien, termasuk pembersihan data, penggabungan dataset, pengelompokan berdasarkan kriteria tertentu, serta analisis statistik. Dengan kemampuannya untuk menangani data terstruktur dan tidak terstruktur secara intuitif, Pandas menjadi alat penting dalam analisis data modern. Pustaka ini juga dibangun di atas dua pustaka inti lainnya: NumPy untuk operasi matematis dan Matplotlib untuk visualisasi data, sehingga memberikan kemampuan analisis yang lebih mendalam kepada pengguna.

HASIL PENELITIAN DAN PEMBAHASAN

Data yang di analisis dan di visualisasikan adalah data mengenai data kartu kredit client, dimana dari data tersebut diketahui nama-nama brand kartu yang akan di analisis, id client, nomor kartu, tipe kartu, banyaknya kesalahan pada kartu, kapan para pengguna nya membuka kartu tersebut, limit kredit nya, kapan kartu akan memasuki masa kadaluarsa, serta apakah kartu tersebut diperjualbelikan secara illegal di dark web. Format dari data yang di analisis adalah *csv*, sumber dari data tersebut adalah *Kaggle*, platform yang menyediakan banyak data untuk di analisis. Dimensi dari data tersebut memiliki 6164 baris dan 13 kolom.

Insight/Informasi yang ingin didapat dari menganalisis data tersebut adalah diantaranya mengidentifikasi kartu yang akan kadaluarsa, menemukan pengguna dengan limit kredit tinggi, menganalisis limit kredit dan perilaku pelanggan untuk mendeteksi risiko seperti kemungkinan gagal bayar, memahami distribusi limit kredit di antara berbagai kelompok pengguna, mendeteksi anomaly seperti mengidentifikasi pola penggunaan atau data yang mencurigakan, mengetahui bagaimana perubahan dalam data dari tahun ke tahun, memahami hubungan korelasi antara dua atau lebih variable, statistic deskriptif dari data tersebut, serta insight lainnya yang di dapatkan pada analisis dan visualisasi data yang telah dilakukan.

Berikut merupakan hasil dari Analisis dan Visualisasi

A. MEMBACA DATA

id	client_id	card_brand	card_type	card_number	expires	cvv	has_chip	num_cards_issued	credit_limit	acct_open_date	year_joined	last_changed	card_on_dark_web
0	4534	825	Visa	Debit	4344676511950444	12/2022	623	YES	2	\$24295	09/2002	2008	No
1	2731	825	Visa	Debit	4956965974959986	12/2020	393	YES	2	\$21968	04/2014	2014	No
2	3701	825	Visa	Debit	4582115478255491	02/2024	719	YES	2	\$46414	07/2003	2004	No
3	42	825	Visa	Credit	4879494103069057	08/2024	693	NO	1	\$12400	01/2003	2012	No
4	4659	625	Mastercard	Debit (Prepaid)	5722874738736011	03/2009	75	YES	1	\$28	09/2006	2009	No
...
6141	5361	185	Amex	Credit	300609782832003	01/2024	663	YES	1	\$6900	11/2000	2013	No
6142	2711	185	Visa	Credit	4710517475996018	01/2021	492	YES	2	\$3700	04/2013	2012	No
6143	1305	1007	Mastercard	Credit	5829512204785914	08/2020	257	NO	2	\$9300	02/2013	2012	No
6144	743	1110	Mastercard	Debit	5589768928167462	01/2020	630	YES	1	\$29074	01/2020	2020	No
6145	3198	1110	Visa	Credit	48940111318543994	12/2020	120	YES	1	\$14400	05/2017	2017	No

6146 rows x 13 columns

B. MENCARI SAMPLE DATA

1. Jumlah Data

```
import pandas as pd

data = pd.read_csv('C:/TUGAS SEMESTER 3/Pemrograman Lanjut/cards_data.csv')

jumlah_data = len(data)

jumlah_data_alt = data.shape[0]

print("Jumlah data:", jumlah_data)
```

Jumlah data: 6146

2. Sampel 10 Baris Data

Sample data 10 baris:

id	client_id	card_brand	card_type	card_number	expires	cvv	has_chip	num_cards_issued	credit_limit	acct_open_date	year_pin_last_changed	card_on_dark_web
5398	1572	Mastercard	Debit	5146838421358947	09/2021	676	YES	1	\$946	11/2003	2015	No
3414	597	Amex	Credit	389433518398778	08/2023	534	YES	1	\$27500	06/2009	2009	No
845	740	Visa	Debit	4456142342115989	03/2023	189	YES	2	\$4009	01/2020	2020	No
4133	592	Mastercard	Credit	5886410881258720	02/2020	684	YES	1	\$5600	08/2004	2019	No
2257	1974	Mastercard	Debit	5522110815947841	02/2024	663	YES	1	\$14747	03/2018	2018	No
1664	239	Mastercard	Debit	5615383905529869	08/2021	586	NO	2	\$21449	02/2020	2020	No
951	1318	Visa	Credit	4763790665708312	04/2024	729	YES	1	\$10600	01/2020	2020	No
486	1820	Amex	Credit	375947824371679	05/2024	928	YES	1	\$12000	01/2020	2020	No
1151	669	Mastercard	Debit	5816338605169387	06/2020	932	YES	1	\$11164	02/2008	2009	No
5336	871	Mastercard	Debit	5125159184089047	08/2020	945	YES	2	\$20123	10/2019	2019	No

3. Sampel Data Setiap Kolom

```
Rata-rata dari setiap kolom numerik:
id                3.072500e+03
client_id         9.949396e+02
card_number       4.820426e+15
cvv               5.062208e+02
num_cards_issued  1.503091e+00
year_pin_last_changed  2.013437e+03
dtype: float64

Standar deviasi dari setiap kolom numerik:
id                1.774342e+03
client_id         5.786146e+02
card_number       1.328582e+15
cvv               2.894311e+02
num_cards_issued  5.191909e-01
year_pin_last_changed  4.270699e+00
dtype: float64
```

4. Mengurutkan Data Terkecil ke Data Terbesar (Ascending)

	id	client_id	card_brand	card_type	card_number	expires	cvv	has_chip	num_cards_issued	credit_limit	acct_open_date	year_pin_last
0	0	0	Amex	Credit	300105541992311	01/2002	0	NO	1	\$0	01/1991	
1	1	0	Amex	Credit	300609782832003	01/2006	0	NO	1	\$0	01/1994	
2	2	0	Amex	Credit	300717031458937	01/2007	0	NO	1	\$0	01/1995	
3	3	0	Amex	Credit	302014253634948	01/2008	0	NO	1	\$0	01/1995	
4	4	1	Amex	Credit	302031764623099	01/2008	1	NO	1	\$0	01/1997	
...
6141	6141	1998	Visa	Debit (Prepaid)	6987061746149197	12/2024	998	YES	3	\$9956	12/2019	
6142	6142	1998	Visa	Debit (Prepaid)	6988770500713547	12/2024	999	YES	3	\$9956	12/2019	
6143	6143	1998	Visa	Debit (Prepaid)	6994217664929322	12/2024	999	YES	3	\$9957	12/2019	
6144	6144	1999	Visa	Debit (Prepaid)	6994498161251218	12/2024	999	YES	3	\$9984	12/2019	
6145	6145	1999	Visa	Debit (Prepaid)	6997197066610978	12/2024	999	YES	3	\$9998	12/2019	

6146 rows x 13 columns

5. Mengurutkan Data Terkecil ke Data Terbesar (Descending)

	id	client_id	card_brand	card_type	card_number	expires	cvv	has_chip	num_cards_issued	credit_limit	acct_open_date	year_pin_last
0	6145	1999	Visa	Debit (Prepaid)	6897197066610978	12/2024	999	YES	3	\$9998	12/2019	
1	6144	1999	Visa	Debit (Prepaid)	6994498161251218	12/2024	999	YES	3	\$9984	12/2019	
2	6143	1998	Visa	Debit (Prepaid)	6994217664929322	12/2024	999	YES	3	\$9957	12/2019	
3	6142	1998	Visa	Debit (Prepaid)	6988770500713547	12/2024	999	YES	3	\$9956	12/2019	
4	6141	1998	Visa	Debit (Prepaid)	6987061746149197	12/2024	998	YES	3	\$9956	12/2019	
...
6141	4	1	Amex	Credit	302031764623099	01/2008	1	NO	1	\$0	01/1997	
6142	3	0	Amex	Credit	302014253634948	01/2008	0	NO	1	\$0	01/1995	
6143	2	0	Amex	Credit	300717031458937	01/2007	0	NO	1	\$0	01/1995	
6144	1	0	Amex	Credit	300609782832003	01/2006	0	NO	1	\$0	01/1994	
6145	0	0	Amex	Credit	300105541992311	01/2002	0	NO	1	\$0	01/1991	

6146 rows x 13 columns

C. STATISTIK DESKRIPTIF

1. Rata-rata dan Standar Deviasi

```

Rata-rata dari setiap kolom numerik:
id                3.072500e+03
client_id         9.949396e+02
card_number       4.820426e+15
cvv               5.062208e+02
num_cards_issued  1.503091e+00
year_pin_last_changed  2.013437e+03
dtype: float64

Standar deviasi dari setiap kolom numerik:
id                1.774342e+03
client_id         5.786146e+02
card_number       1.328582e+15
cvv               2.894311e+02
num_cards_issued  5.191909e-01
year_pin_last_changed  4.270699e+00
dtype: float64

```

2. Modus (data yang paling sering muncul) dari setiap kolom data Non-Numerik

```

Modus (data yang paling sering muncul) dari setiap kolom non-numerik:
card_brand        Mastercard
card_type         Debit
expires           02/2020
has_chip          YES
credit_limit       $0
acct_open_date    02/2020
card_on_dark_web  No
Name: 0, dtype: object

```


3. Percentile

Percentile (10%, 25%, 50%, 75%, 90%) dari setiap kolom:

	id	client_id	card_number	cvv	num_cards_issued	year_pin_last_changed
0.10	614.50	197.00	4.086688e+15	99.0	1.0	2009.0
0.25	1536.25	492.25	4.486365e+15	257.0	1.0	2010.0
0.50	3072.50	992.00	5.108957e+15	516.5	1.0	2013.0
0.75	4608.75	1495.00	5.585237e+15	756.0	2.0	2017.0
0.90	5530.50	1799.00	5.879344e+15	902.5	2.0	2020.0

4. Nilai Ekstremum dari setiap kolom data Numerik

Nilai Minimum dari setiap kolom numerik:

id	0
client_id	0
card_number	300105541992311
cvv	0
num_cards_issued	1
year_pin_last_changed	2002

dtype: int64

Nilai Maksimum dari setiap kolom numerik:

id	6145
client_id	1999
card_number	6997197066610978
cvv	999
num_cards_issued	3
year_pin_last_changed	2020

dtype: int64

5. Distribusi Frekuensi untuk data 'card_type' dan 'year_pin_last_changed'

Distribusi Frekuensi untuk kolom 'card_type':

card_type	
Debit	3511
Credit	2057
Debit (Prepaid)	578

Name: count, dtype: int64

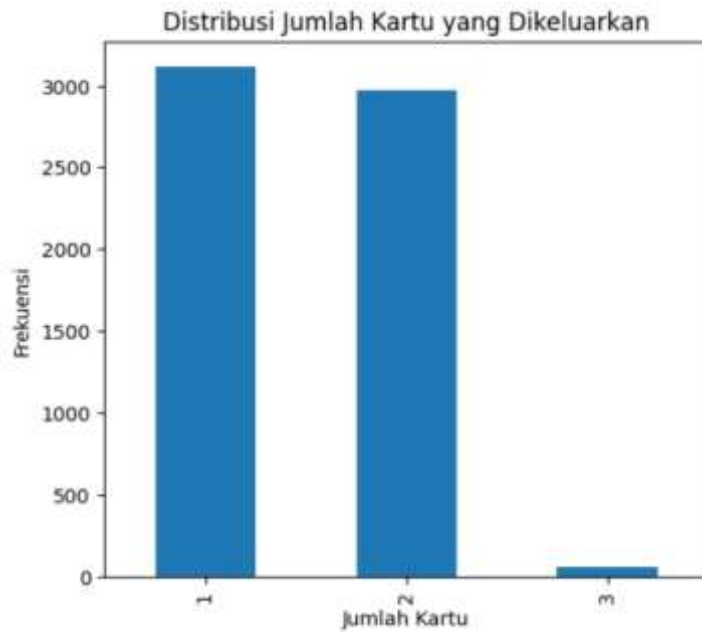
Distribusi Frekuensi untuk kolom 'year_pin_last_changed':

year_pin_last_changed	
(2001.982, 2003.8]	13
(2003.8, 2005.6]	48
(2005.6, 2007.4]	250
(2007.4, 2009.2]	759
(2009.2, 2011.0]	1521
(2011.0, 2012.8]	477
(2012.8, 2014.6]	862
(2014.6, 2016.4]	575
(2016.4, 2018.2]	318
(2018.2, 2020.0]	1323

Name: count, dtype: int64

D. VISUALISASI DATA

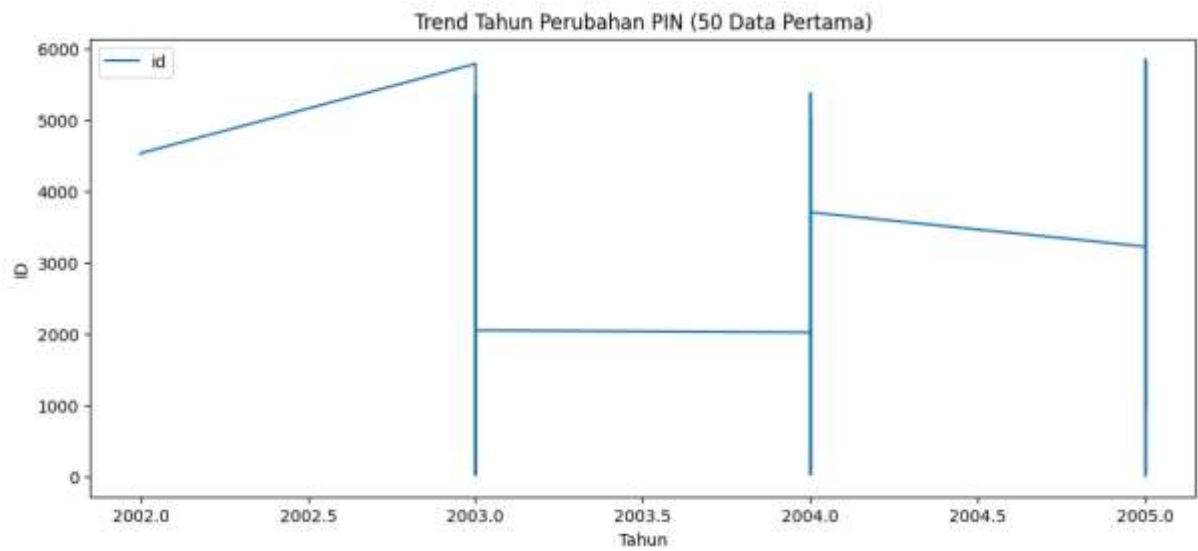
1. Bar Chart: Perbandingan Kategori



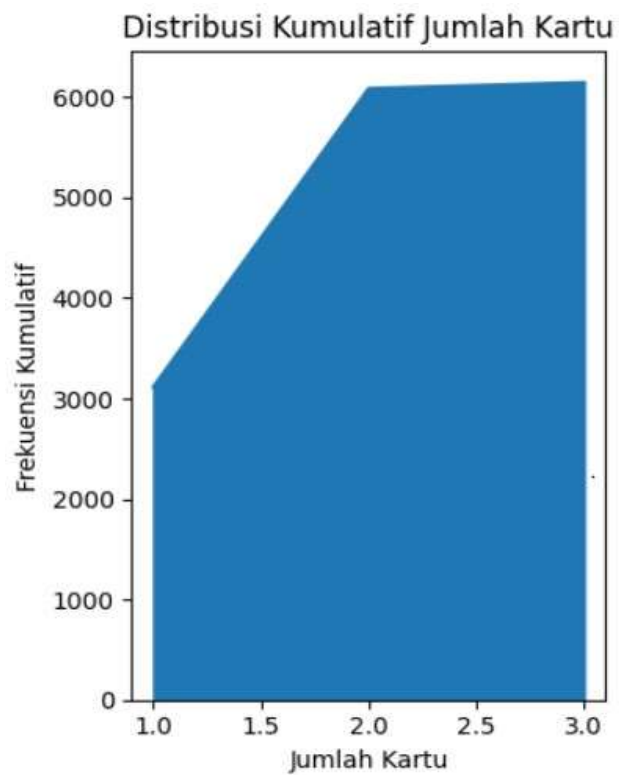
2. Pie Chart: Proporsi Tahun Perubahan Pin



3. Perubahan Terhadap Waktu



4. Area Plot: Distribusi Kumulatif Jumlah Kartu

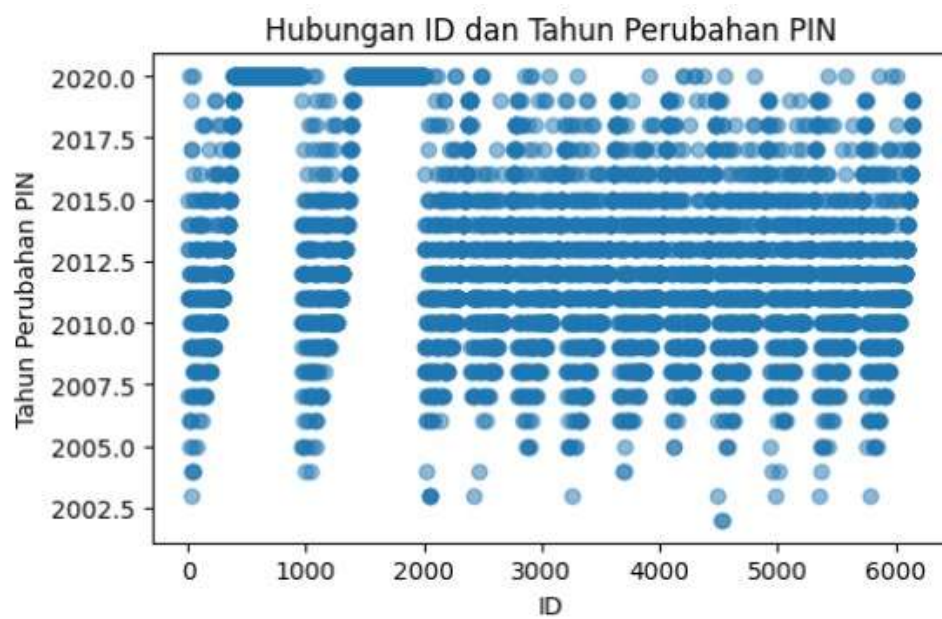


5. Grouped Bar: Hubungan Antara Jumlah Kartu dengan Tahun Perubahan PIN

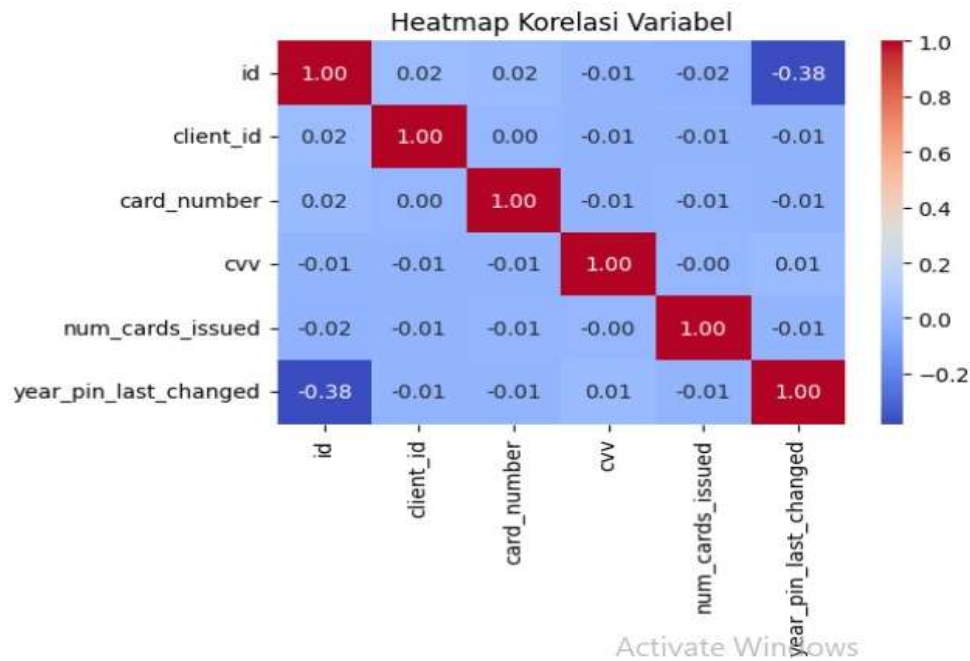


6. Plotting Relationship

Scatter Plot: Hubungan Antara ID dengan Tahun Perubahan PIN



Heatmap: Korelasi Antar Variabel Numerik



```

Statistik Deskriptif:
      id      client_id      card_number      cvv      num_cards_issued \
count  6146.000000    6146.000000    6.146000e+03    6146.000000    6146.000000
mean   3072.500000    994.939636    4.820426e+15    506.220794    1.503091
std    1774.341709    578.614626    1.328582e+15    289.431123    0.519191
min     0.000000      0.000000    3.001055e+14     0.000000    1.000000
25%    1536.250000    492.250000    4.486365e+15    257.000000    1.000000
50%    3072.500000    992.000000    5.108957e+15    516.500000    1.000000
75%    4608.750000    1495.000000    5.585237e+15    756.000000    2.000000
max    6145.000000    1999.000000    6.997197e+15    999.000000    3.000000

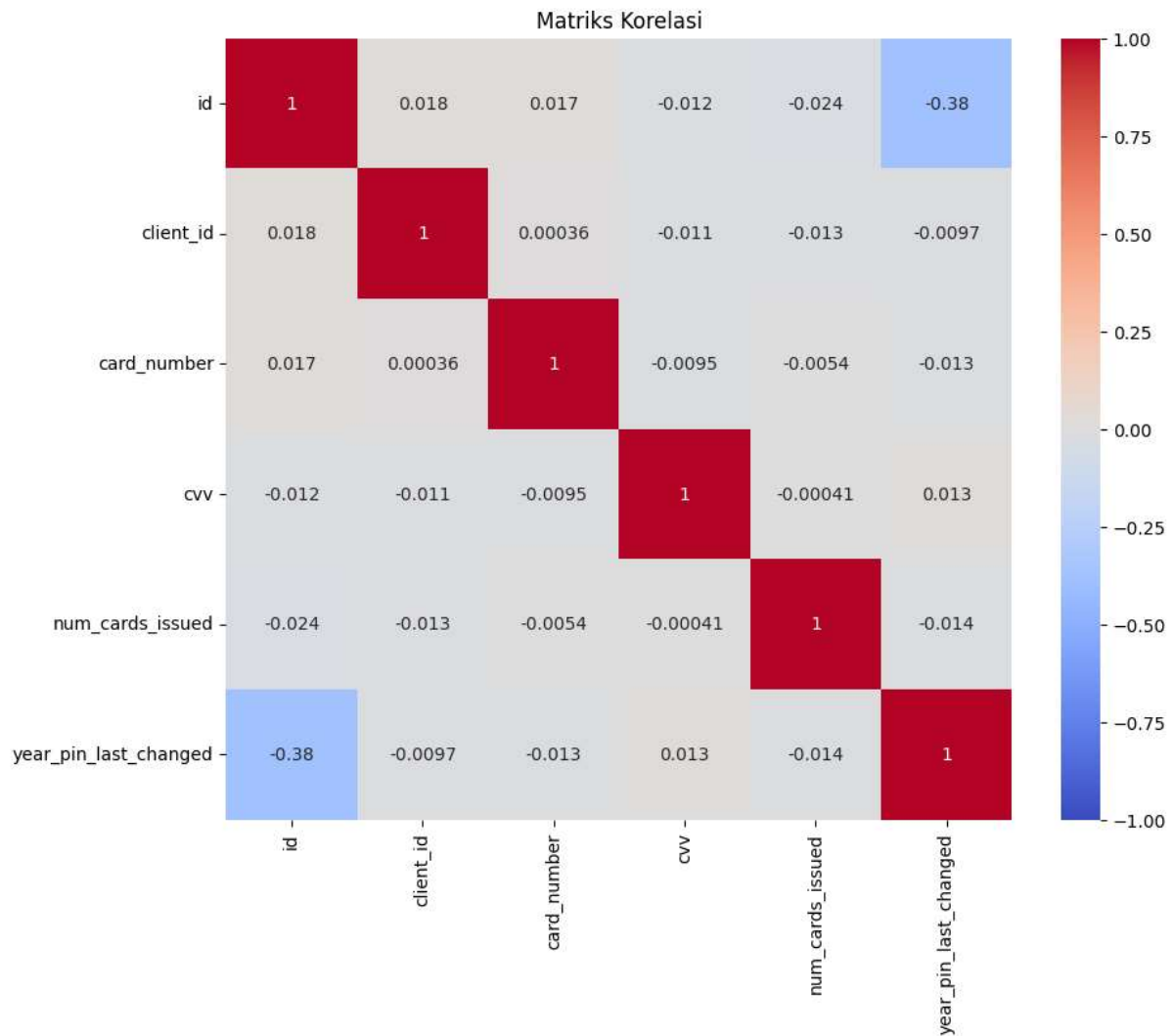
      year_pin_last_changed
count      6146.000000
mean      2013.436707
std         4.270699
min        2002.000000
25%        2010.000000
50%        2013.000000
75%        2017.000000
max        2020.000000

Informasi Missing Values:
id          0
client_id   0
...
acct_open_date  0
year_pin_last_changed  0
card_on_dark_web  0
dtype: int64
Output is truncated. View as a scrollable element or open in a text editor. Adjust cell output settings...

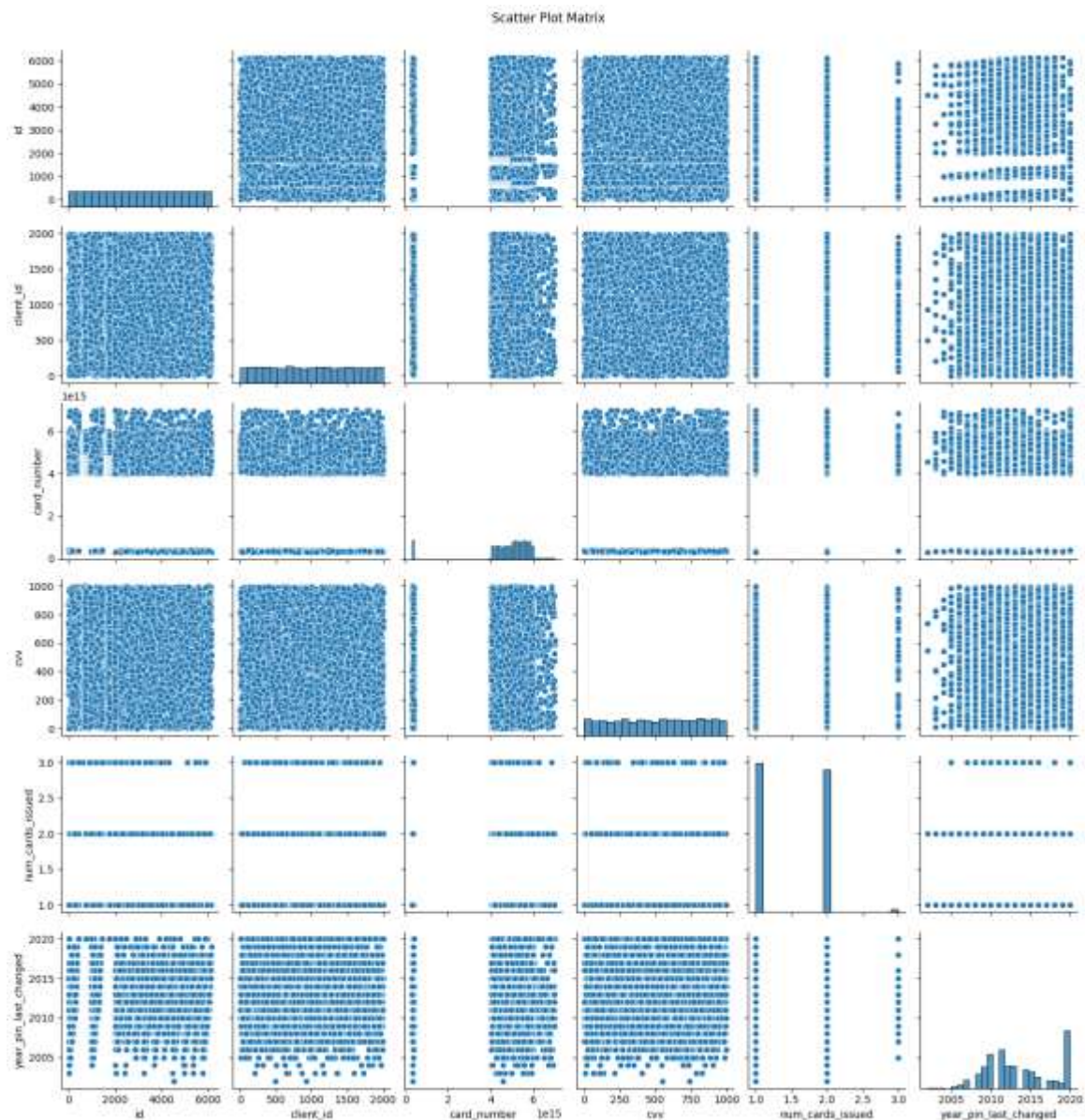
```

E. NILAI KORELASI (Antara kolom 'id' dengan kolom 'id_client')

Matriks Korelasi



Scatter Plot Matiks



Interpretasi Korelasi:

id dengan num_cards_issued: -0.024 (Korelasi sangat lemah)
 id dengan year_pin_last_changed: -0.381 (Korelasi lemah)
 client_id dengan id: 0.018 (Korelasi sangat lemah)
 client_id dengan cvv: -0.011 (Korelasi sangat lemah)
 client_id dengan num_cards_issued: -0.013 (Korelasi sangat lemah)
 client_id dengan year_pin_last_changed: -0.010 (Korelasi sangat lemah)
 card_number dengan id: 0.017 (Korelasi sangat lemah)
 card_number dengan client_id: 0.000 (Korelasi sangat lemah)
 card_number dengan cvv: -0.009 (Korelasi sangat lemah)
 card_number dengan num_cards_issued: -0.005 (Korelasi sangat lemah)
 card_number dengan year_pin_last_changed: -0.013 (Korelasi sangat lemah)
 cvv dengan id: -0.012 (Korelasi sangat lemah)
 cvv dengan num_cards_issued: -0.000 (Korelasi sangat lemah)
 cvv dengan year_pin_last_changed: 0.013 (Korelasi sangat lemah)
 num_cards_issued dengan year_pin_last_changed: -0.014 (Korelasi sangat lemah)