Data Analytics for Business 2024

**Project Based Learning**

Study Case

[Paper ID]

CS 04 - 5:

| 1. | Muhammad Fahmi Hutomo | KM-CS04151 |
| --- | --- | --- |
| 2. | Marsyanda Nur Zahra | KM-CS04340 |
| 3. | Syahirotul Ambar Maulidiyah | KM-CS04067 |
|  |  |  |

# **Kata Pengantar**

Puji syukur kami panjatkan kepada Tuhan Yang Maha Esa atas limpahan rahmat dan karunia-Nya, sehingga kami dapat menyelesaikan laporan Project-Based Learning Bitlabs Academy yang menggunakan studi kasus dari Paper ID ini dengan baik. Laporan ini merupakan salah satu bentuk implementasi pembelajaran berbasis proyek yang bertujuan untuk mengasah kemampuan analitis dan pemecahan masalah secara terstruktur.

Kami menyampaikan terima kasih sebesar-besarnya kepada pihak Bitlabs Academy yang telah memberikan bimbingan dan fasilitas dalam proses pembelajaran ini. Ucapan terima kasih juga kami sampaikan kepada para mentor dan instruktur yang telah memberikan arahan serta dukungan selama pelaksanaan proyek ini. Tak lupa, kami haturkan rasa terima kasih kepada rekan-rekan peserta yang telah bekerja sama dan berbagi pengalaman, sehingga proses pembelajaran menjadi lebih bermakna.

Melalui proyek ini, kami berharap dapat mengembangkan pemahaman yang lebih mendalam terhadap analisis studi kasus, mengaplikasikan teori-teori yang relevan, serta memperkuat kemampuan kami dalam menghadapi tantangan dunia profesional.

Kami menyadari bahwa laporan ini masih jauh dari sempurna. Oleh karena itu, kami sangat terbuka terhadap saran dan kritik yang membangun untuk pengembangan ke depan. Semoga laporan ini dapat bermanfaat bagi para pembaca dan menjadi referensi yang berguna bagi kegiatan serupa di masa yang akan datang.

Surabaya, 08 Desember 2024

Kelompok 05 - CS 04

# **Daftar Isi**

[**Kata Pengantar 2**](#_bn85tq6pthr4)

[**Daftar Isi 3**](#_gjdgxs)

[**Profil Perusahaan 5**](#_1fob9te)

[**Latar Belakang Masalah 6**](#_hzoap0bimocm)

[**Data dan Sumber Data 6**](#_ufx4hg8261yt)

[**ALUR PENGERJAAN STUDYCASE 7**](#_82nlt1l4gt57)

[A. Tools yang digunakan dalam proses menganalisis data 7](#_wq63ej1ettbn)

[B. Proses Analisis Data 7](#_lwgho7y6xgrj)

[**BAB I 8**](#_30j0zll)

[**[Business] Business Analysis 8**](#_xylr7nksr6ge)

[**BAB II 11**](#_tyjcwt)

[**Data Analytics using Python Programming 11**](#_oowbtw9hmmxf)

[2.1 Data Cleaning 11](#_itd0aqaheccs)

[2.2 Feature Engineering 12](#_631d88v8btj6)

[2.3 Scaling and Normalization 13](#_duak042rdcsc)

[**BAB III 14**](#_mtocb2risnnd)

[**Exploratory Data Analysis (EDA) 14**](#_e31321xeakbh)

[**3.1 Exploratory Data Analysis 14**](#_ggj7xioo6qxo)

[3.1.1 Analisis Distribusi Transaction\_amount 14](#_5h5gn68sdmy6)

[3.1.2 Analisis Frekuensi Transaksi Pasangan Buyer-Seller 15](#_n7hy6u5voo3r)

[3.1.3 Analisis Penggunaan Promo 16](#_lqpiel49mbx)

[3.1.4 Analisis Pola Transaksi Berulang, Transaksi dengan Nilai Tinggi, dan Self-transaction 17](#_tlke4vwx4e33)

[**3.2 Visualization 18**](#_owp6c3ndeh3x)

[3.2.1 Visualisasi Total Transaksi Harian dari Waktu ke Waktu 18](#_kngf0qqxybx9)

[3.2.2 Graf Jaringan Buyer-Seller 18](#_m4z83tvpgkvg)

[3.2.3 Tren Transaksi Harian yang Menggunakan Promo 20](#_s3jzg59q2vnj)

[3.2.4 Visualisasi Exploitasi Penggunaan Promo 21](#_jqoq0iqytn2v)

[**BAB IV 22**](#_r8l65head3j0)

[**[SQL] Advanced SQL Queries and Stored Procedures 22**](#_48aywn485hfa)

[4.1 Advanced SQL Queries for Fraud Detection 22](#_iqcrrakeddhx)

[4.1.1 Advanced SQL Queries 22](#_e6mxsuby32y8)

[4.1.1.1 Transaksi Jauh di Luar Rentang Normal 22](#_c0i3h25hrlrg)

[4.1.1.2 Analisis Hubungan Pembeli-Penjual 24](#_w36e64ii2a5n)

[4.1.1.3 Deteksi Penyalahgunaan Promosi 26](#_6gho65hmwjdg)

[4.1.1.4 Waktu yang Mencurigakan 27](#_3hd1vspdh51)

[4.1.1.5 Koneksi Pengguna yang Ditandai 28](#_iwikob21sfjg)

[4.2 SQL Joins for User-Company Fraud Insights 30](#_cjqx0gje3jkd)

[**BAB V 32**](#_dfqz4f7rnmaq)

[**SQL Views and Stored Procedures 32**](#_upftnmbm6izu)

[5.1 SQL View 32](#_tuaxhz9x4l1c)

[5.1.1 View Pasangan buyer-seller paling mencurigakan 32](#_qu506u817kmi)

[5.1.2 View - Pengguna yang Ditandai dan Transaksi Mereka 33](#_y3s6dosbbw7c)

[5.2 Stored Procedures 34](#_1rr2n93t1jyz)

[5.2.1 Laporan Penipuan Bulanan 34](#_pgxkxmxfw1wt)

[5. 2.2 Deteksi Penyalahgunaan Otomatis 37](#_fneelz3npswk)

[**BAB VI 39**](#_2s8eyo1)

[**[Python] Advanced Fraud Analysis and Network Insights 39**](#_op1dywxadvfz)

[6.1 Social Network Analysis 39](#_eqp628y3kcel)

[6.1.1 Analisis Hubungan Pembeli-Penjual untuk Mengidentifikasi Pola Interaksi dan Terlibat Fraud (using SQL) 39](#_syos2n8dzle2)

[6.1.2 Analisis Hubungan Pembeli-Penjual untuk Mengidentifikasi Pola Interaksi dan Terlibat Fraud (Visualization using Phyton) 40](#_5s6okc8bs9uu)

[6.2 Cohort Analysis 41](#_o9xmg6211uy9)

[6.2.1 Kelompokkan pembeli berdasarkan tanggal transaksi pertama mereka dan mengukur aktivitas berkelanjutan mereka dari waktu ke waktu. 41](#_r21ifm3w8dm9)

[6.2.2 Identifikasi apakah pembeli tertentu terlibat dalam perilaku penipuan setelah periode tidak aktif atau berulang kali berinteraksi dengan penjual yang sama. 43](#_pupc0snrapf3)

[6.3 Insight Generation 44](#_new816yvjr1f)

[**BAB VII 47**](#_on9cbr8yk304)

[**[Visualization] Tableau for Fraud Monitoring and Dashboard Creation 47**](#_t3lzltxxkooh)

[7.1 Interactive Fraud Detection Dashboards 47](#_fzoik0l2stb5)

[7.1.1 Tableau Dashboards 47](#_su9t91z9l3r1)

[7.1.1.1Tren Transaksi Penipuan 47](#_cxma3yd9dqzh)

[7.1.1. 2 Visualisasi Hubungan Pembeli-Penjual 48](#_9gceew1hdrbs)

[7.1.1.3 Penyalahgunaan Promosi 49](#_rrhokv27u508)

[7.2 Dynamic Filtering and Drill-Downs 49](#_k3hmgvmhcstl)

[**BAB VIII 50**](#_stfui1rkj80s)

[**BAB IX 51**](#_26in1rg)

[**Kesimpulan dan Saran 51**](#_789gb4xyb0k2)

[A. Kesimpulan 51](#_hoxbhix58fe1)

[B. Saran 51](#_vd4dsetv3zmb)

[**LAMPIRAN 52**](#_ub1ibdfpi3da)

[**A. Online Diagram 52**](#_e7hbbhdzt9uk)

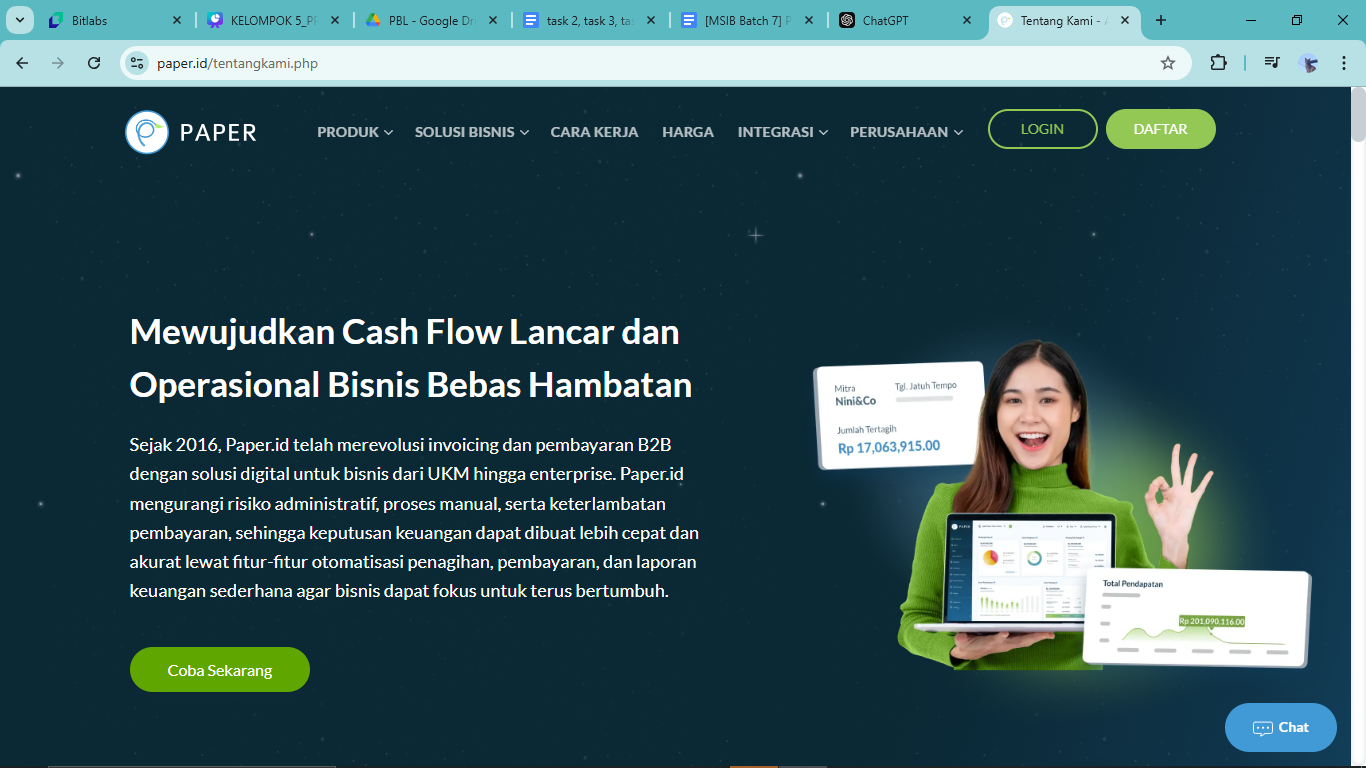
[**B. Python Code 52**](#_fc7sj1f5r52w)

[**C. Recording 52**](#_fjhp24af9s01)

# **Profil Perusahaan**

Paper.id adalah platform penyedia jasa penerbitan faktur dan pembayaran digital yang dirancang khusus untuk bisnis di Indonesia. Platform ini memungkinkan pengguna untuk membuat faktur tanpa batas dengan mudah, mengotomatiskan pengingat pembayaran, dan mengintegrasikan berbagai metode pembayaran seperti kartu kredit dan kode QR. Dengan fokus pada penyederhanaan transaksi keuangan, Paper.id memberdayakan bisnis untuk mengelola proses faktur mereka dengan lebih efisien dan akurat.

Sejak diluncurkan, Paper.id telah mencapai pertumbuhan yang luar biasa, dengan lebih dari 600.000 pengguna yang memanfaatkan layanannya. Platform ini telah memfasilitasi pemrosesan lebih dari 8 juta faktur, dengan total nilai pembayaran digital melebihi Rp 10 triliun. Dampak signifikan ini menunjukkan bahwa Paper.id berperan dalam mentransformasi pengelolaan keuangan bagi usaha kecil dan menengah di Indonesia.

Dengan antarmuka yang ramah pengguna dan fitur yang tangguh, Paper.id menjadi alat penting bagi bisnis yang ingin mengoptimalkan operasional keuangannya. Perusahaan ini terus berinovasi dan beradaptasi dengan kebutuhan pengguna yang terus berkembang, memastikan posisinya sebagai sumber daya berharga dalam ekonomi digital Indonesia. Untuk informasi lebih lanjut, kunjungi situs web mereka di [www.paper.id](http://www.paper.id/).

**Gambar 1.** *Website Paper.id*

# **Latar Belakang Masalah**

Baru-baru ini, perusahaan mengalami peningkatan transaksi fraud yang berdampak negatif pada pendapatan dan kepercayaan pelanggan. Tugas Anda adalah melakukan analisis lanjutan terhadap fraud untuk mengidentifikasi pola perilaku yang mencurigakan serta mendeteksi hubungan yang mencurigakan antara pembeli dan penjual. Anda juga akan menggunakan query SQL tingkat lanjut dan analisis mendalam untuk memberikan wawasan yang dapat ditindaklanjuti serta rekomendasi guna mengurangi risiko fraud.

# **Data dan Sumber Data**

Adapun data pada perusahaan sebagai berikut:

1. **Digital Payment Transaction Data:**

Columns: dpt\_id, buyer\_id, seller\_id, transaction\_amount, payment\_method\_name, payment\_provider\_name, transaction\_created\_datetime, transaction\_updated\_datetime, dpt\_promotion\_id

1. **Digital Payment Request Data:**

Columns: dpt\_id, total\_fee\_amount, document\_type\_name

1. **Promotion Data:**Columns: dpt\_promotion\_id, promotion\_code, promotion\_name, transaction\_promo\_cashback\_amount
2. **Company Data:**Columns: company\_id, company\_kyc\_status\_name, company\_kyb\_status\_name, company\_type\_group, company\_phone\_verified\_flag, company\_email\_verified\_flag, user\_fraud\_flag, testing\_account\_flag, blacklist\_account\_flag, company\_registered\_datetime

# **Alur Pengerjaan Study Case**

## **Tools yang digunakan dalam proses menganalisis data**

Dalam proses analisis data, berbagai alat digunakan untuk menangani tugas spesifik yang berbeda. BPMN dengan Draw.io digunakan untuk memodelkan proses bisnis secara visual, memudahkan pemangku kepentingan dalam memahami alur dan interaksi antar proses. DBeaver yang terhubung dengan PostgreSQL dipilih karena kemampuannya dalam mengelola dan mengeksekusi query SQL yang optimal untuk database ini, serta menyediakan antarmuka pengguna yang memudahkan manipulasi dan visualisasi data. Python digunakan karena kekuatannya dalam mengolah dan menganalisis data besar melalui pustaka seperti Pandas dan NumPy, serta kemampuannya untuk terintegrasi dengan berbagai alat dan database. Sementara itu, Tableau digunakan untuk membuat visualisasi data yang interaktif dan mudah dipahami, memungkinkan pemangku kepentingan mengeksplorasi data secara real-time. Kombinasi alat-alat ini memungkinkan alur kerja yang efisien dalam memproses, menganalisis, dan menyajikan data secara menyeluruh.

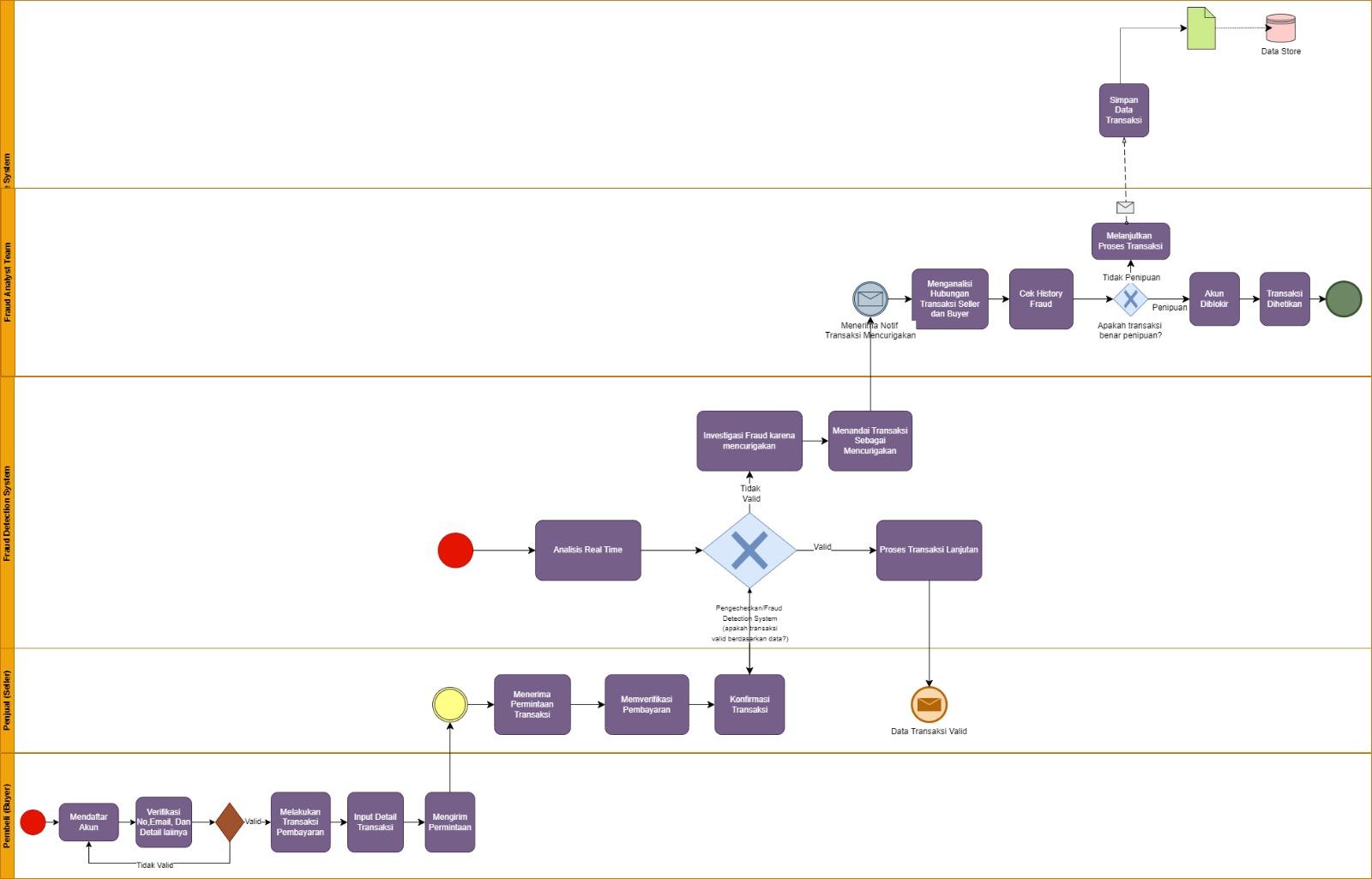
## **Proses Analisis Data**

Proses analisis dimulai dengan data cleaning, di mana data yang hilang, outlier, atau inkonsistensi seperti nilai yang tidak sesuai pada kolom terkait fraud (misalnya, user\_fraud\_flag) diidentifikasi dan diperbaiki. Setelah itu, dilakukan feature engineering untuk membuat fitur turunan seperti Buyer-Seller Relationship Score dan Transaction Frequency Metrics untuk mendeteksi anomali dalam interaksi pembeli-penjual atau pola transaksi yang tidak biasa. Selanjutnya, data yang telah dibersihkan dan ditransformasi melalui scaling dan normalization (misalnya, normalisasi jumlah transaksi) dianalisis lebih lanjut melalui Exploratory Data Analysis (EDA), yang melibatkan analisis distribusi transaksi, frekuensi pasangan pembeli-penjual, dan penggunaan promosi. Visualisasi data seperti grafik transaksi dari waktu ke waktu dan jaringan pembeli-penjual digunakan untuk mengidentifikasi pola aktivitas mencurigakan. Pada tahap berikutnya, analisis lebih lanjut dilakukan menggunakan SQL queries untuk mendeteksi anomali transaksi, kolusi pembeli-penjual, serta penyalahgunaan promosi, dengan menggunakan query SQL yang kompleks dan stored procedures untuk otomatisasi laporan fraud. Social Network Analysis kemudian digunakan untuk mengidentifikasi hubungan antara pembeli dan penjual yang terlibat dalam fraud melalui visualisasi jaringan dan cohort analysis untuk melacak perilaku transaksi yang berulang. Terakhir, hasil analisis tersebut disajikan dalam bentuk Tableau dashboards untuk memantau tren fraud secara real-time, dengan kemampuan untuk menelusuri dan memfilter data untuk mendalami hubungan yang mencurigakan dan memberikan rekomendasi tindakan pencegahan.

# **BAB I**

# **[Business] Business Analysis**

Proses deteksi penipuan melalui pendekatan Business Process Model Notation (*BPMN*) dirancang untuk memetakan alur kerja secara menyeluruh, mulai dari inisiasi transaksi hingga penyelesaian kasus penipuan. Model ini mencakup langkah-langkah kunci seperti inisiasi transaksi, deteksi penipuan secara real-time, penandaan pengguna mencurigakan, hingga resolusi kasus, dengan tujuan untuk meningkatkan keandalan sistem pembayaran. Dengan visualisasi BPMN, perusahaan dapat lebih mudah mengidentifikasi area yang perlu dioptimalkan, sehingga mendukung upaya menjaga kepercayaan pelanggan dan keberlanjutan operasional di ekosistem digital. Berikut adalah penjelasan rinci dari model BPMN yang mengilustrasikan proses deteksi penipuan (fraud flow analysis) di Paper.id.

**Gambar 1.1** *BPMN fraud flow analysis*

Berdasarkan **Gambar 1.1** *BPMN fraud flow analysis* diatas menunjukan bisnis proses sebagai berikut *:*

1. **Proses Pendaftaran**: Proses ini merupakan langkah awal yang harus dilalui oleh pengguna sebelum dapat melakukan transaksi di platform. Pada tahap ini, pengguna (baik pembeli maupun penjual) diwajibkan untuk:

* Melakukan Pendaftaran Akun: Pengguna mengisi data pribadi seperti nama, alamat email, dan nomor telepon.
* Verifikasi Email dan Nomor Telepon: Sistem mengirimkan kode OTP (One-Time Password) atau tautan verifikasi ke email dan nomor telepon pengguna. Langkah ini dilakukan untuk memastikan bahwa data yang diberikan valid dan dapat dihubungi.

1. **Proses Inisiasi Transaksi** : Setelah pendaftaran dan verifikasi selesai, pengguna dapat memulai transaksi. Proses ini melibatkan dua pihak utama, yaitu pembeli (Buyer) dan penjual (Seller):

* Pembeli (Buyer):

Melakukan Transaksi Pembayaran: Pembeli memulai dengan menginput detail pembayaran seperti jumlah, metode pembayaran, dan informasi terkait lainnya.

Mengirim Permintaan Transaksi: Setelah detail transaksi lengkap, permintaan dikirimkan kepada penjual untuk diproses lebih lanjut.

* Penjual (Seller):

Menerima Permintaan Transaksi: Penjual menerima notifikasi bahwa pembeli telah mengirimkan detail transaksi.

Memverifikasi Pembayaran: Penjual memastikan bahwa pembayaran sudah dilakukan dan valid.

Mengonfirmasi Transaksi: Setelah pembayaran terverifikasi, transaksi dikonfirmasi untuk diproses lebih lanjut.

1. **Proses Analisis Real-Time oleh Sistem Deteksi Fraud** : Tahap ini menjadi inti dari sistem keamanan untuk mendeteksi potensi kecurangan dalam transaksi. Prosesnya meliputi:

* Penerimaan Data Transaksi: Sistem menerima data transaksi dari penjual, termasuk informasi seperti jumlah transaksi, identitas pembeli, dan riwayat transaksi.
* Analisis Real-Time: Sistem menggunakan algoritma dan model analitik untuk memeriksa transaksi dalam waktu nyata. Indikator seperti pola transaksi mencurigakan, hubungan antara penjual dan pembeli. Dimana jika transaksi valid, ditandai sebagai aman dan jika transaksi mencurigakan, sistem memberikan notifikasi kepada tim analisis fraud untuk tindakan lebih lanjut.

1. **Investigasi dan Tindak Lanjut Fraud** : Ketika sebuah transaksi ditandai sebagai mencurigakan, tim analisis fraud akan mengambil alih untuk melakukan investigasi lebih mendalam. Proses ini melibatkan langkah-langkah seperti:

* Investigasi Manual oleh Tim Analisis Fraud: Tim menganalisis data transaksi yang mencurigakan menggunakan metode analitik lanjutan dan hubungan antara pembeli dan penjual diperiksa untuk mengidentifikasi adanya pola fraud, seperti transaksi fiktif atau kolusi.
* Penandaan Fraud: Jika ditemukan bukti kecurangan, transaksi akan ditandai sebagai fraud.
* Keputusan Akhir: Jika transaksi dinyatakan valid, proses dilanjutkan seperti biasa. Jika terbukti fraud, transaksi dihentikan, akun diblokir, dan pelaku dicatat dalam sistem blacklist.

Dari analisis fraud flow yang diterapkan dalam bisnis proses Paper.id, terlihat bahwa sistem telah dirancang untuk mendeteksi dan menangani potensi kecurangan. Proses pendaftaran dengan verifikasi email untuk memastikan hanya pengguna valid yang dapat mengakses platform. Sistem deteksi fraud real-time yang mengandalkan analitik data untuk mengidentifikasi pola mencurigakan. Proses ini diperkuat oleh investigasi manual oleh tim analisis fraud untuk memastikan keakuratan hasil deteksi. Penandaan akun mencurigakan dan pemblokiran otomatis menciptakan lapisan perlindungan tambahan, sementara pencatatan pelaku dalam sistem blacklist memperkuat keamanan ekosistem. Pendekatan ini tidak hanya meminimalkan risiko penipuan tetapi juga menjaga kepercayaan pengguna terhadap layanan Paper.id.

# **BAB II**

# **Data Analytics using Python Programming**

## **2.1 Data Cleaning**

Data cleaning dilakukan menggunakan Python dan library Pandas, dengan seluruh proses dilakukan di Google Colab. Dataset yang digunakan terdiri dari empat tabel, yang masing-masing dimuat dari Google Drive ke dalam lingkungan Google Colab. Proses pembersihan data bertujuan untuk meningkatkan kualitas data untuk menangani nilai duplikat, nilai hilang dan outliers. Sehingga siap digunakan dalam analisis lebih lanjut. Beberapa langkah utama yang dilakukan selama proses ini sebagai berikut:

1. Import data
2. Baca Info data
3. Hapus data duplikat.
4. Handling missing values:

* dpt\_promotion\_id kosong di dataframe transaction dianggap transaksi tidak menggunakan promo, diisi dengan ’ no promotion’.
* baris dengan dpt\_promotion\_id kosong di dataframe promotion dihapus karena hanya satu baris dengan promotion\_code dan promotion\_name yang juga kosong disertai cashback 0.
* promotion\_code dan promotion\_name yang kosong akan diisi dengan 'no promotion' jika promotion['dpt\_promotion\_id] == 'no promotion'.
* promotion\_code dan promotion\_name yang kosong akan diisi dengan 'unknown' jika promotion['dpt\_promotion\_id] != 'no promotion'.
* company\_type\_group kosong pada dataframe user diisi dengan ‘unknown’.

1. Outliers seperti transaction\_amount yang terlampau besar nilainya tetap dipertahankan untuk dianalisis lebih lanjut

Selain langkah-langkah utama dalam proses pembersihan data, tahap ini juga mencakup identifikasi terhadap data yang hilang atau tidak konsisten, khususnya pada kolom-kolom yang berkaitan dengan transaksi fraud. Langkah ini bertujuan untuk mengungkap pola atau anomali yang dapat memberikan gambaran lebih jelas terkait aktivitas mencurigakan di dalam transaksi. Berdasarkan analisis yang dilakukan, beberapa temuan penting berhasil diidentifikasi, yaitu:

1. User Tidak Terdaftar Sebagai Seller

Terdapat 318 user yang tidak terdaftar namun tercatat melakukan transaksi sebagai seller.

1. User Fraud Tidak Di-blacklist

Sebanyak 660 user yang ditandai fraud tidak dimasukkan ke dalam daftar blacklist.

1. User Fraud dengan Status KYC Tidak Dibekukan

Terdapat 11 user yang terdeteksi fraud tetapi status KYC (*Know Your Customer*)nya tidak dibekukan.

1. Transaksi dengan Identitas Tidak Tervalidasi

Lebih dari 50% dari total transaksi melibatkan seller yang memiliki identitas belum tervalidasi, termasuk nomor telepon dan email yang tidak diverifikasi, serta status KYC (*Know Your Customer*) dan KYB (*Know Your Business*) yang belum tervalidasi, dan akun tersebut bukan termasuk akun uji coba (testing account)

1. Self-Transaction (Transaksi Diri Sendiri)

Sebanyak 46% dari total transaksi yang terjadi adalah self-transaction, di mana buyer dan seller merupakan entitas (company) yang sama.

* Dari jumlah tersebut, 24 buyer melakukan self-transaction lebih dari 100 kali, yang memperkuat indikasi pola perilaku yang tidak wajar.
* Namun, self-transaction yang menggunakan promo tercatat hanya sebesar 0,19% dari total transaksi, sehingga dampak penyalahgunaan promo relatif kecil.

## **2.2 Feature Engineering**

Pada tahap ini dibuat beberapa fitur turunan atau kolom baru yang dapat digunakan untuk melakukan analisis lebih lanjut.

1. Pertama, buyer-seller relationship\_score. Relationship\_score 1(sangat lemah), 2(lemah), 3(sedang), 4(kuat), dan 5(sangat kuat).
2. Kedua, transaction frequency metrics. Ada tiga metrik yang dibuat, yaitu burst\_activity, unusual\_gap, dan burst\_amount. Ketiganya hanya bernilai 1 (ada) atau 0 (tidak ada).
3. Ketiga, promotion exploitation indicator. Metrik is\_promotion\_exploitation bernilai 1 (ada) atau 0 (tidak ada).

## **2.3 Scaling and Normalization**

Pada transaction\_amount dilakukan tranformasi logaritmik sebelum normalisasi karena distribusi datanya memiliki nilai skewness 139. Sedangkan transaction\_created\_datetime setelah diubah ke detik distribusi datanya mendekati normal dengan skewness hanya -0,24 sehingga tidak dilakukan transformasi. Tahap terakhir dari proses ini adalah melakukan normalisasi transaction\_amount yang sudah ditransformasi dan transaction\_created\_datetime dalam detik menggunakan MinMaxScaler().

# **BAB III**

# **Exploratory Data Analysis (EDA)**

## **3.1 Exploratory Data Analysis**

Pada tahap ini dilakukan analisis eksplorasi data (Exploratory Data Analysis/EDA) secara mendalam untuk mendeteksi pola yang berkaitan dengan perilaku fraud. Analisis mencakup beberapa aspek utama, seperti distribusi nilai transaksi, frekuensi pasangan buyer-seller, serta penggunaan promo. Selain itu, investigasi lebih lanjut dilakukan untuk mengidentifikasi pola pada transaksi yang berulang atau nilai transaksi yang secara abnormal tinggi. Hasil dari analisis ini memberikan beberapa informasi penting sebagai berikut:

### **3.1.1 Analisis Distribusi Transaction\_amount**

Analisis ini bertujuan untuk memahami distribusi nilai transaksi, termasuk rentang transaksi yang mencakup nilai minimum hingga maksimum. Distribusi yang tidak merata dengan sejumlah kecil transaksi bernilai tinggi menjadi salah satu perhatian utama. Sehingga tahap ini diperoleh beberapa informasi di antaranya:

* Nilai transaksi minimum sebesar 0 dan maksimum sekitar 20 Milyar.
* Nilai transaksi 0 terjadi sebanyak 26 transaksi.
* Nilai transaksi lebih dari 0 hingga 1K terjadi sebanyak 16 transaksi atau 0.032% dari keseluruhan transaksi.
* Nilai transaksi lebih dari 1K hingga 10K terjadi sebanyak 538 transaksi atau 1.076% dari keseluruhan transaksi.
* Nilai transaksi lebih dari 10K hingga 100K terjadi sebanyak 12454 transaksi atau 24.908% dari keseluruhan transaksi.
* Nilai transaksi lebih dari 100K hingga 1M terjadi sebanyak 8726 transaksi atau 17.452% dari keseluruhan transaksi.
* Nilai transaksi lebih dari 1M hingga 10M terjadi sebanyak 13380 transaksi atau 26.760% dari keseluruhan transaksi.
* Lebih dari 10M hingga 100M: 13228 transaksi 26.456% dari keseluruhan transaksi
* Lebih dari 100M hingga 1B: 1628 transaksi 3.256% dari keseluruhan transaksi
* Lebih dari 1B hingga 10B: 3 transaksi 0.006% dari keseluruhan transaksi
* Lebih dari 10B hingga 30B: 1 transaksi 0.002% dari keseluruhan transaksi

### **3.1.2 Analisis Frekuensi Transaksi Pasangan Buyer-Seller**

Analisis ini fokus pada frekuensi transaksi berulang antara pasangan buyer-seller. Frekuensi transaksi yang sangat tinggi dari beberapa pasangan buyer-seller mengindikasikan adanya potensi pola transaksi yang mencurigakan.Dimana dapat diperoleh beberapa informasi diantaranya:

* Terdapat 10354 pasangan buyer-seller pada data transaction.
* Transaksi berulang lebih dari 0 hingga 5 kali dilakukan oleh 8398 pasangan (81.11% dari total pasangan buyer-seller).
* Transaksi berulang lebih dari 5 hingga 10 kali dilakukan oleh 1017 pasangan (9.82% dari total pasangan buyer-seller).
* Transaksi berulang lebih dari 10 hingga 20 kali dilakukan oleh 587 pasangan (5.67% dari total pasangan buyer-seller).
* Transaksi berulang lebih dari 20 hingga 30 kali dilakukan oleh 158 pasangan (1.53% dari total pasangan buyer-seller)
* Transaksi berulang lebih dari 30 hingga 40 kali dilakukan oleh 60 pasangan (0.58% dari total pasangan buyer-seller)
* Transaksi berulang lebih dari 40 hingga 50 kali dilakukan oleh 39 pasangan (0.38% dari total pasangan buyer-seller)
* Transaksi berulang lebih dari 50 hingga 100 kali dilakukan oleh 63 pasangan (0.61% dari total pasangan buyer-seller)
* Transaksi berulang lebih dari 100 hingga 150 kali dilakukan oleh 15 pasangan (0.14% dari total pasangan buyer-seller)
* Transaksi berulang lebih dari 150 hingga 200 kali dilakukan oleh 7 pasangan (0.07% dari total pasangan buyer-seller)
* Transaksi berulang lebih dari 200 hingga 250 kali dilakukan oleh 5 pasangan (0.05% dari total pasangan buyer-seller)
* Transaksi berulang lebih dari 250 hingga 300 kali dilakukan oleh 3 pasangan (0.03% dari total pasangan buyer-seller)
* Transaksi berulang lebih dari 300 hingga 350 kali dilakukan oleh 1 pasangan (0.01% dari total pasangan buyer-seller)
* Transaksi berulang lebih dari 350 hingga 1250 kali dilakukan oleh 0 pasangan (0.00% dari total pasangan buyer-seller)
* Transaksi berulang lebih dari 1250 hingga 1300 kali dilakukan oleh 1 pasangan (0.01% dari total pasangan buyer-seller)

### **3.1.3 Analisis Penggunaan Promo**

Penggunaan promo dalam transaksi dievaluasi untuk mengidentifikasi pola dan potensi penyalahgunaan. Top 3 kode promo paling sering digunakan juga diidentifikasi dalam analisis ini. Dimana dapat diperoleh beberapa informasi diantaranya:

* Transaksi yang menggunakan promo terjadi sebanyak 937 kali atau hanya 1,87% dari total transaksi.
* Top 3 promo paling banyak digunakan yaitu, promotion-219036467 sebanyak 243 kali, promotion-214984720 sebanyak 88 kali, dan promotion-188676794 sebanyak 69 kali.
* Buyer menggunakan promo sebanyak 8 kali: 1 (0.14%)
* Buyer menggunakan promo sebanyak 6 kali: 2 (0.27%)
* Buyer menggunakan promo sebanyak 5 kali: 2 (0.27%)
* Buyer menggunakan promo sebanyak 4 kali: 7 (0.96%)
* Buyer menggunakan promo sebanyak 3 kali: 26 (3.56%)
* Buyer menggunakan promo sebanyak 2 kali: 109 (14.93%)
* Buyer menggunakan promo sebanyak 1 kali: 583 (79.86%)
* Buyer menggunakan satu kode promo sebanyak 5 kali: 2 (0.23%)
* Buyer menggunakan satu kode promo sebanyak 4 kali: 1 (0.11%)
* Buyer menggunakan satu kode promo sebanyak 3 kali: 4 (0.46%)
* Buyer menggunakan satu kode promo sebanyak 2 kali: 44 (5.03%)
* Buyer menggunakan satu kode promo sebanyak 1 kali: 823 (94.16%)

### **3.1.4 Analisis Pola Transaksi Berulang, Transaksi dengan Nilai Tinggi, dan Self-transaction**

Analisis ini mendalami pola transaksi berulang, termasuk transaksi dengan nilai tinggi dan self-transaction, di mana buyer dan seller merupakan entitas yang sama. Beberapa rentang transaksi menunjukkan proporsi yang signifikan dari self-transaction, mengindikasikan potensi penyalahgunaan sistem. Dimana tahap ini dapat diperoleh beberapa informasi diantaranya:

* Terdapat 32 pasangan buyer-seller yang melakukan transaksi berulang lebih dari 100 kali dan 24 di antaranya (75%) nya merupakan self-transaction.
* Nilai transaksi 0 terjadi sebanyak 26 transaksi, 1 transaksi merupakan self-transaction, (3.85%) dari keseluruhan transaksi di rentang ini.
* Nilai transaksi lebih dari 0 hingga 1K terjadi sebanyak 16 transaksi, 8 transaksi merupakan self-transaction, (50.00%) dari keseluruhan transaksi di rentang ini.
* Nilai transaksi lebih dari 1K hingga 10K terjadi sebanyak 538 transaksi, 500 transaksi merupakan self-transaction, (92.94%) dari keseluruhan transaksi di rentang ini.
* Nilai transaksi lebih dari 10K hingga 100K terjadi sebanyak 12454 transaksi, 6643 transaksi merupakan self-transaction, (53.34%) dari keseluruhan transaksi di rentang ini.
* Nilai transaksi lebih dari 100K hingga 1M terjadi sebanyak 8726 transaksi, 5749 transaksi merupakan self-transaction, (65.88%) dari keseluruhan transaksi di rentang ini.
* Nilai transaksi lebih dari 1M hingga 10M terjadi sebanyak 13380 transaksi, 6874 transaksi merupakan self-transaction, (51.38%) dari keseluruhan transaksi di rentang ini.
* Nilai transaksi lebih dari 10M hingga 100M terjadi sebanyak 13228 transaksi, 3364 transaksi merupakan self-transaction, (25.43%) dari keseluruhan transaksi di rentang ini.
* Nilai transaksi lebih dari 100M hingga 1B terjadi sebanyak 1628 transaksi, 85 transaksi merupakan self-transaction, (5.22%) dari keseluruhan transaksi di rentang ini.
* Nilai transaksi lebih dari 1B hingga 10B terjadi sebanyak 3 transaksi, 3 transaksi merupakan self-transaction, (100.00%) dari keseluruhan transaksi di rentang ini.
* Nilai transaksi lebih dari 10B hingga 30B terjadi sebanyak 1 transaksi, 1 transaksi merupakan self-transaction, (100.00%) dari keseluruhan transaksi di rentang ini.

## **3.2 Visualization**

Visualisasi data dilakukan untuk memberikan gambaran yang lebih jelas terkait pola transaksi, hubungan antar pengguna, serta penggunaan promo. Dengan pendekatan ini, potensi anomali atau indikasi fraud dapat lebih mudah diidentifikasi secara intuitif. Beberapa fokus utama dalam visualisasi data dirangkum sebagai berikut

### **3.2.1 Visualisasi Total Transaksi Harian dari Waktu ke Waktu**

****

**Gambar 3.2.1** *Visualisasi Total Transaksi Harian dari Waktu ke Waktu*

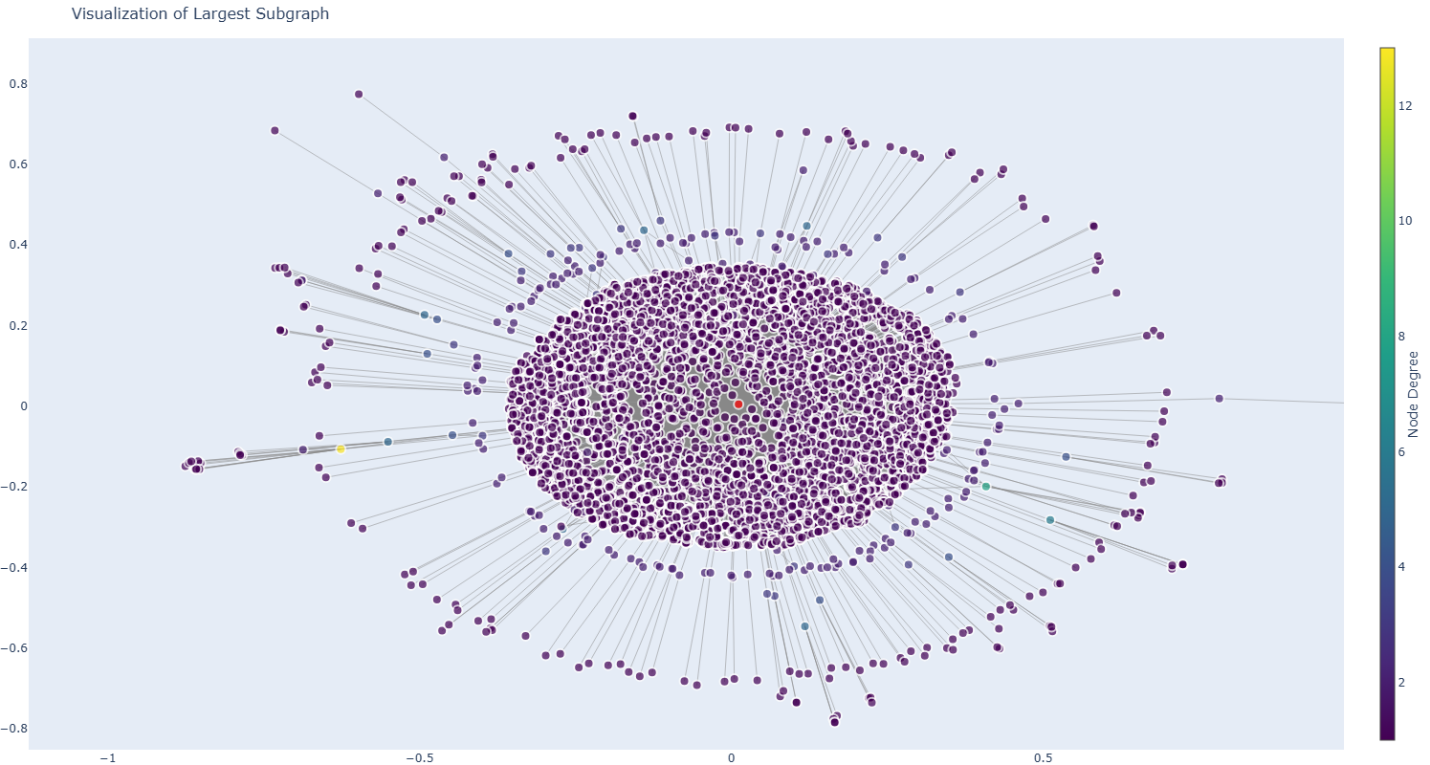
Pada kuarter kedua tahun 2023 terjadi dua kali lonjakan total nilai transaksi harian.

* Lonjakan pertama pada tanggal 14 April terjadi karena ada satu transaksi dengan nilai sangat tinggi, sekitar 20 Milyar. Sebenarnya separuh transaksi pada tanggal tersebut hanya berada di kisaran 5 hingga 6 digit dan total transaksi yang terjadi hanya 110 transaksi.
* Lonjakan kedua pada tanggal 18 Juni tidak sebesar lonjakan pertama dan lebih disebabkan karena separuh transaksi pada hari tersebut bernilai sedikitnya 8 digit dan transaksi yang terjadi pada tanggal tersebut sebanyak 266 kali.

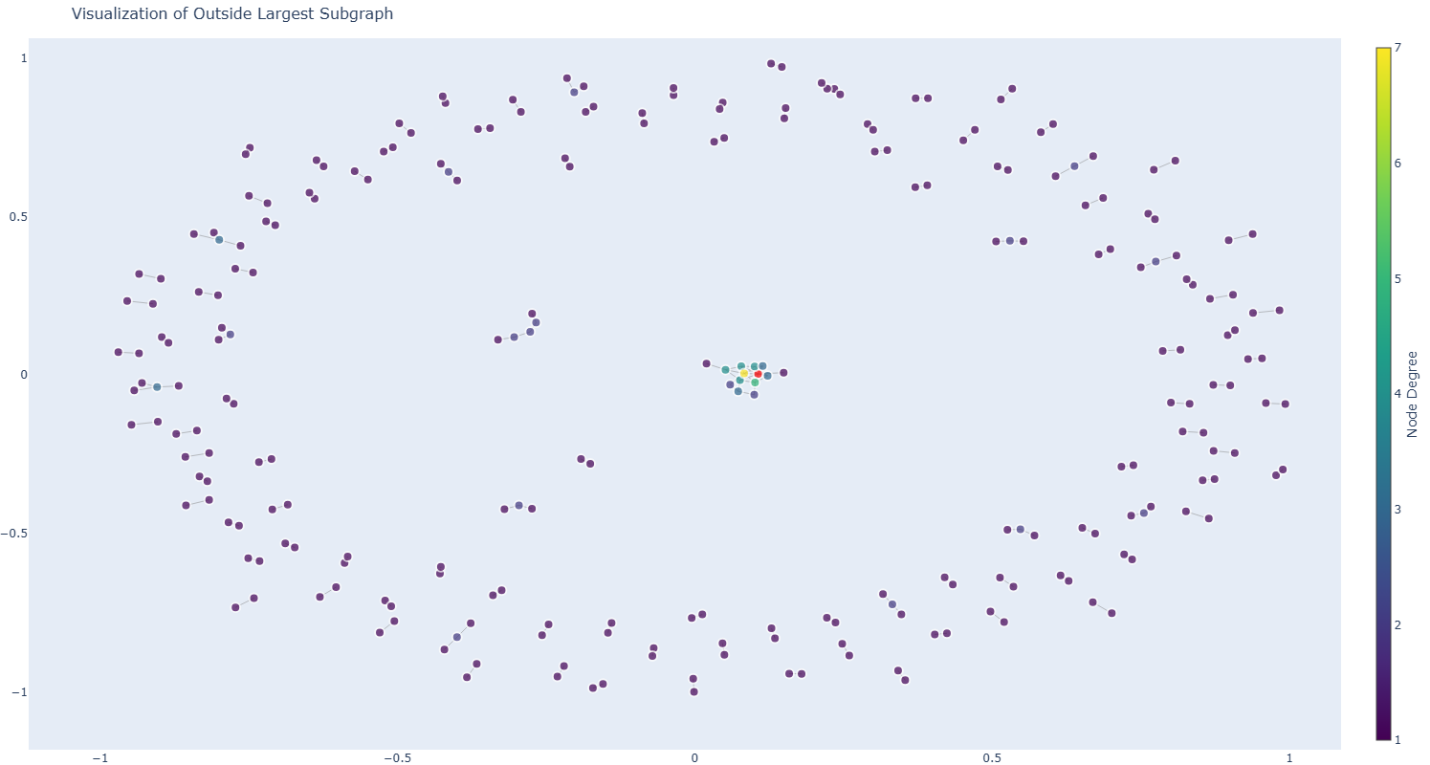
Secara umum, frekuensi transaksi tidak memengaruhi total nilai transaksi harian.

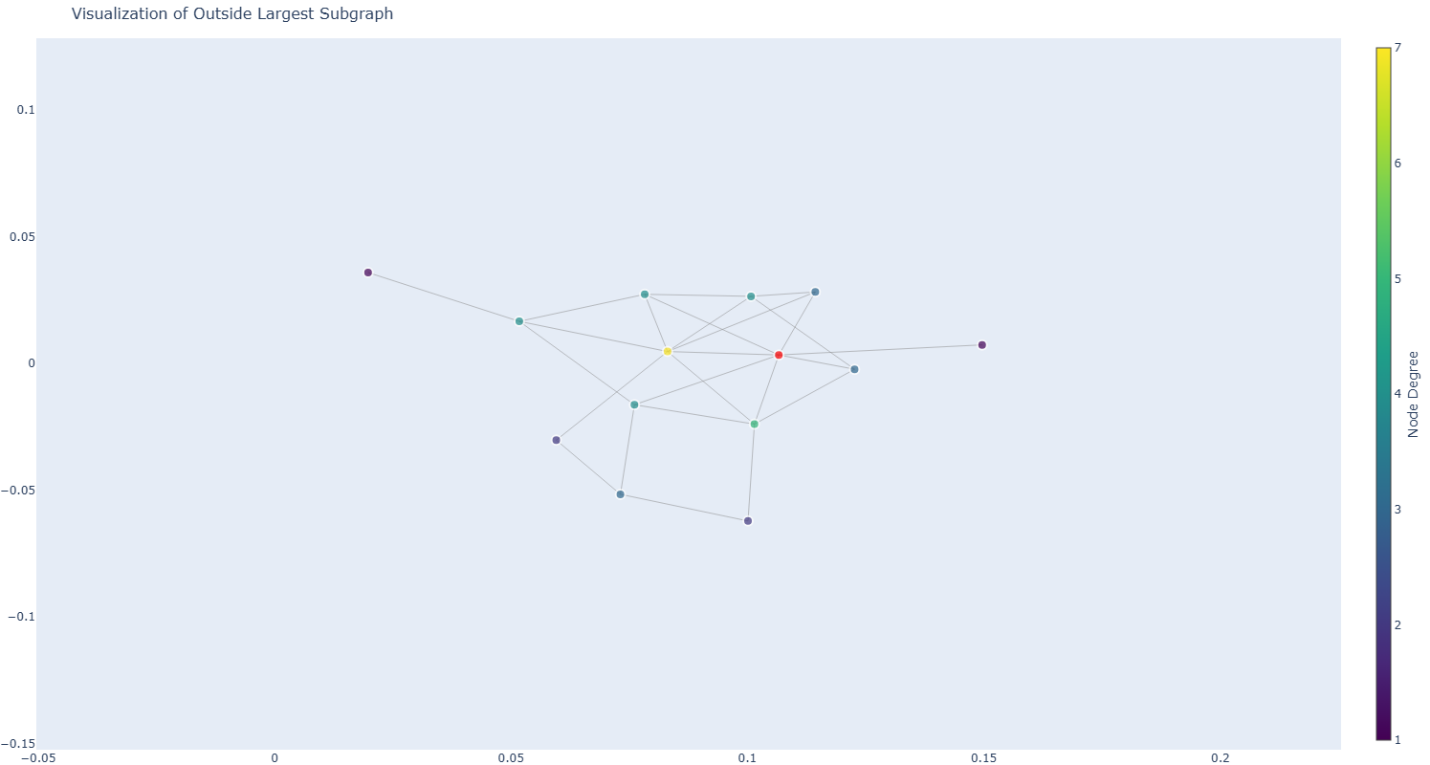
### **3.2.2 Graf Jaringan Buyer-Seller**

Transaksi buyer-seller pada data transaction membentuk banyak graf kecil yang hanya terdiri dari beberapa node,dengan total node dari semua grafik kecil ini hanya 252 node. Di samping itu, ada satu graf yang sangat besar dengan jumlah node 6591 node, dengan node pusat terhubung ke 6356 node lain. Visualisasinya seperti yang gambar di bawah ini.



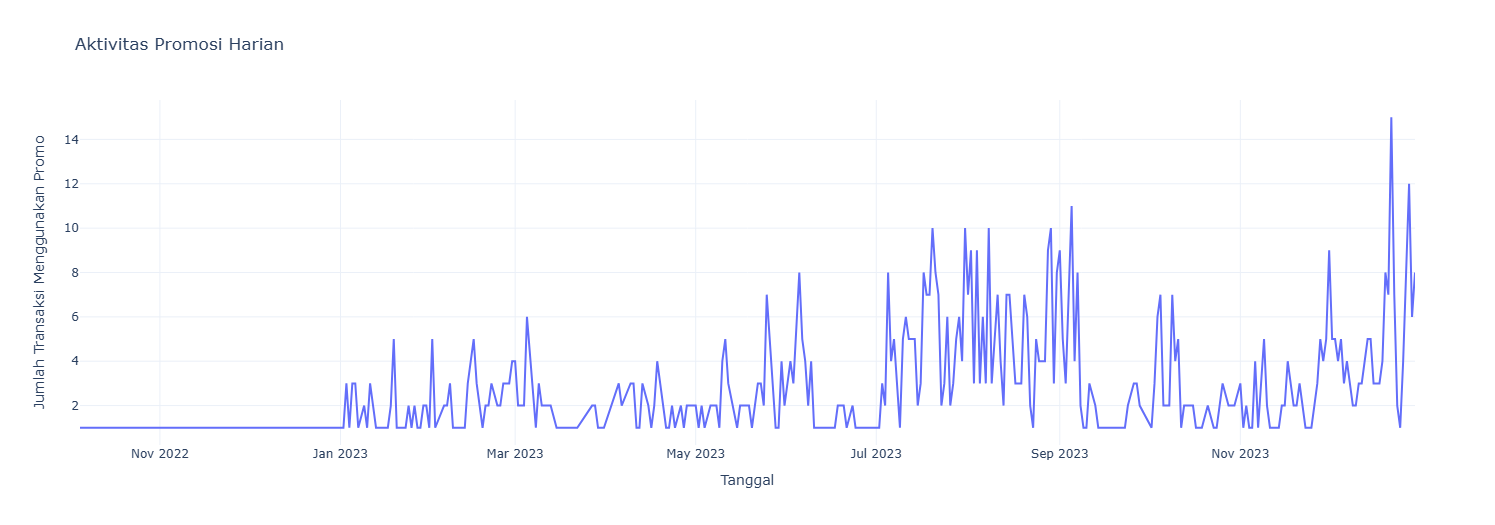
Node pusat berwarna merah berada ditengah, terhubung dengan node lainnya yang sebagian besar hanya memiliki satu atau dua edge, tetapi ada juga beberapa node yang memiliki edge lebih dari 5 hingga 13. Pada bagian selanjutnya akan dilakukan analisis jaringan sosial untuk mengetahui perilaku transaksi fraud berdasarkan graf ini.



Kebanyakan graf-graf kecil ini hanya terdiri dari 2 sampai 3 node, tetapi ada satu graf yang cukup besar jika dibanding yang lain. Untuk lebih jelasnya seperti gambar di bawah ini  


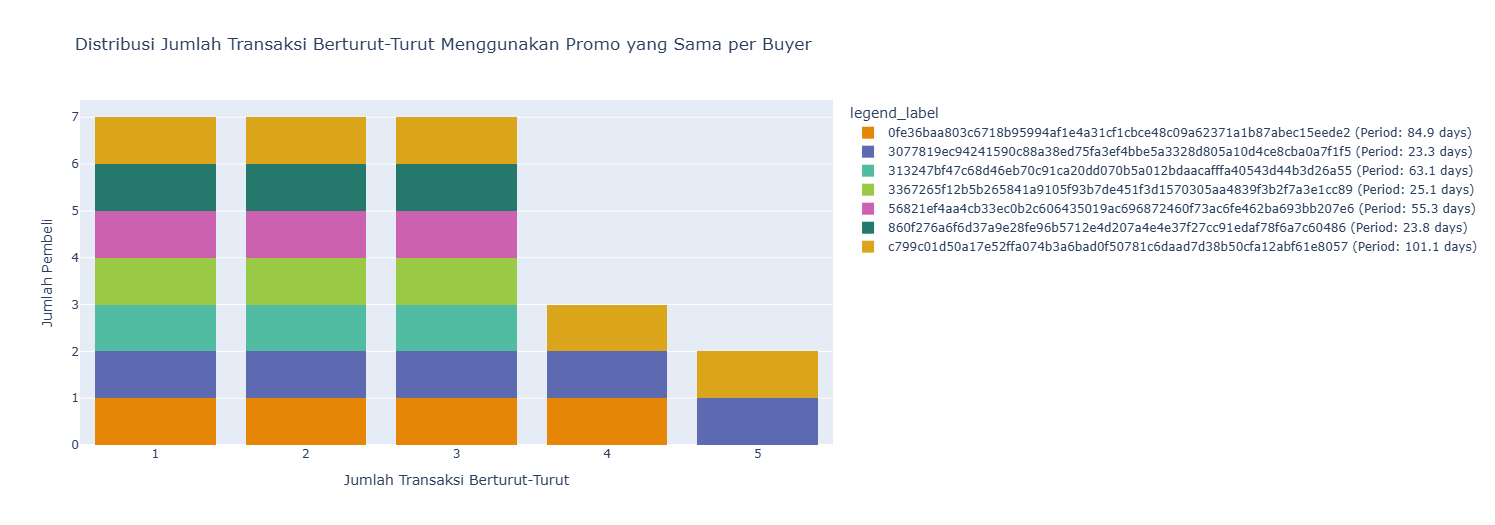
Setelah dilakukan analisis tidak ditemukan hal-hal mencurigakan dalam transaksi user-user dari jaringan ini. Status user pada data user juga tidak masalah, tidak ada yang tergolong fraud atau di blacklist.

### **3.2.3 Tren Transaksi Harian yang Menggunakan Promo**



Secara umum sepanjang tahun 2023 penggunaan promo tampak fluktuatif dari hari ke hari, tetapi pada tiga kuartal pertama, penggunaan promo paling banyak terjadi pada tanggal 5 bulan terakhir di kuartal tersebut. Sedangkan pada kuartal ke-empat, penggunaan promo terbesar masih terjadi di bulan terakhir kuartal, tetapi tidak terjadi pada tanggal 5 seperti biasanya, kali ini terjadi pada tanggal 22 Desember dan menjadi aktivitas penggunaan promo tertinggi sepanjang tahun 2023.

### **3.2.4 Visualisasi Exploitasi Penggunaan Promo**



Dari visualisasi di atas (buyer berwarna biru tua), tampak adanya transaksi berturut-turut yang menggunakan sebanyak 5 kali dalam rentang waktu 23 hari. Hal ini menunjukkan adanya eksploitasi promosi dalam waktu yang relatif singkat.

# **BAB IV**

# **[SQL] Advanced SQL Queries and Stored Procedures**

## **4.1 Advanced SQL Queries for Fraud Detection**

### **4.1.1 Advanced SQL Queries**

Identifikasi transaksi yang mencurigakan atau fraud menjadi salah satu aspek penting dalam menjaga integritas sistem dan keamanan data, terutama dalam sistem transaksi digital yang semakin kompleks. Dengan adanya potensi risiko penyalahgunaan dan kegiatan fraud, sangat penting bagi organisasi untuk memantau setiap transaksi yang terjadi dan mendeteksi pola-pola yang tidak biasa. Salah satu cara yang efektif untuk mendeteksi transaksi yang mencurigakan adalah dengan menggunakan analisis berbasis data, di mana SQL (Structured Query Language. Adapun cara untuk mendeteksinya sebagai berikut:

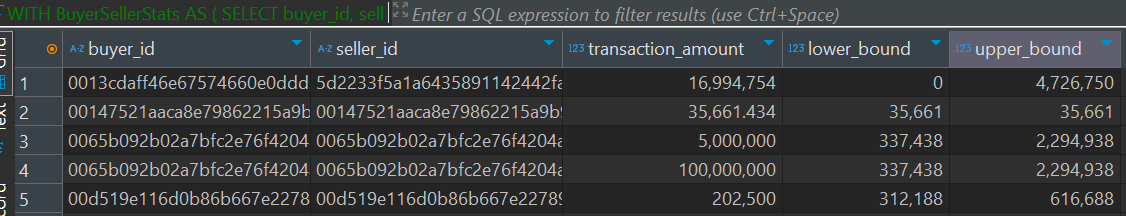
#### **4.1.1.1 Transaksi Jauh di Luar Rentang Normal**

Untuk mendeteksi transaksi yang tidak biasa atau anomali adalah dengan menganalisis perilaku transaksi berdasarkan pasangan pembeli (buyer) dan penjual (seller). Salah satu metode yang dapat digunakan untuk mendeteksi anomali ini adalah dengan menggunakan query SQL sebagai berikut :

**Kode Program 4.1.1.** *Exploratory Data Analysis* pada Python.

| 1. -- 4.1.a Transaction Anomalies: Transactions with values far outside the normal range for each buyer/seller pair. 2. -- Transaksi Jauh di Luar Rentang Normal 3. WITH BuyerSellerStats AS ( 4. SELECT 5. buyer\_id, 6. seller\_id, 7. PERCENTILE\_CONT(0.25) WITHIN GROUP (ORDER BY transaction\_amount) AS Q1, 8. PERCENTILE\_CONT(0.75) WITHIN GROUP (ORDER BY transaction\_amount) AS Q3 9. FROM transaction 10. GROUP BY buyer\_id, seller\_id 11. ), 12. IQRCalculation AS ( 13. SELECT 14. bss.buyer\_id, 15. bss.seller\_id, 16. bss.Q1, 17. bss.Q3, 18. (bss.Q3 - bss.Q1) AS IQR, 19. GREATEST(0, ROUND((bss.Q1 - 1.5 \* (bss.Q3 - bss.Q1)))) AS lower\_bound, 20. ROUND((bss.Q3 + 1.5 \* (bss.Q3 - bss.Q1))) AS upper\_bound 21. FROM BuyerSellerStats bss 22. ), 23. Anomalies AS ( 24. SELECT 25. t.buyer\_id, 26. t.seller\_id, 27. t.transaction\_amount, 28. ic.lower\_bound, 29. ic.upper\_bound 30. FROM transaction t 31. JOIN IQRCalculation ic 32. ON t.buyer\_id = ic.buyer\_id AND t.seller\_id = ic.seller\_id 33. WHERE t.transaction\_amount < ic.lower\_bound 34. OR t.transaction\_amount > ic.upper\_bound 35. ) 36. SELECT 37. buyer\_id, 38. seller\_id, 39. transaction\_amount, 40. lower\_bound, 41. upper\_bound 42. FROM Anomalies 43. ORDER BY buyer\_id, seller\_id, transaction\_amount; 44. select count(\*) from Anomalies; |
| --- |

Melalui analisis menggunakan Interquartile Range (IQR), kita dapat menentukan batas-batas bawah (lower\_bound) dan atas (upper\_bound) dari transaksi yang wajar untuk setiap pasangan pembeli-penjual. Dengan membandingkan nilai transaksi aktual terhadap batas-batas ini, kita dapat mengidentifikasi transaksi yang memiliki nilai jauh di luar rentang normal dan dianggap mencurigakan

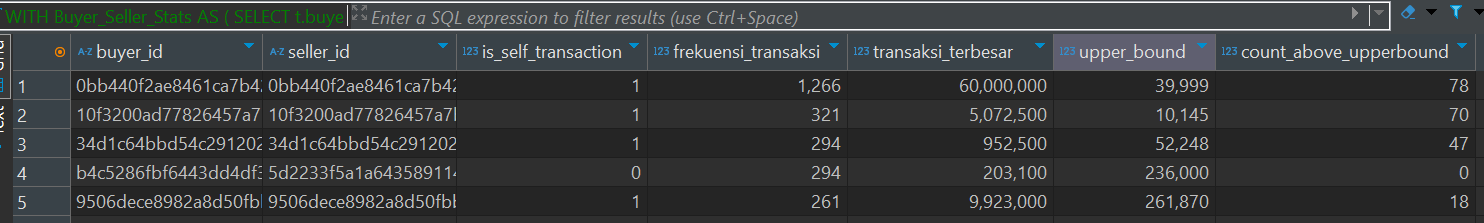


Dalam konteks ini, hasil query yang dihasilkan akan menampilkan pasangan buyer-seller yang memiliki transaksi dengan nilai "transaction\_amount" yang tergolong anomali, di luar batas IQR yang telah ditentukan. Selain itu, hasil query juga akan menampilkan nilai lower\_bound dan upper\_bound untuk memberikan gambaran lebih jelas mengenai rentang transaksi yang wajar bagi masing-masing pasangan.

#### **4.1.1.2 Analisis Hubungan Pembeli-Penjual**

**Kode Program 4.1.2.** *Exploratory Data Analysis* pada Python.

| 1. -- 4.1.b Buyer-Seller Relationship Analysis: buyer-seller pairs with unusually high transaction frequencies or amounts 2. -- Analisis Hubungan Pembeli-Penjual 3. WITH Buyer\_Seller\_Stats AS ( 4. SELECT 5. t.buyer\_id, 6. t.seller\_id, 7. COUNT(t.dpt\_id) AS frekuensi\_transaksi, 8. MAX(t.transaction\_amount) AS transaksi\_terbesar, 9. PERCENTILE\_CONT(0.25) WITHIN GROUP (ORDER BY t.transaction\_amount) AS Q1, 10. PERCENTILE\_CONT(0.75) WITHIN GROUP (ORDER BY t.transaction\_amount) AS Q3 11. FROM 12. transaction t 13. GROUP BY 14. t.buyer\_id, t.seller\_id 15. ), 16. IQR\_Calculations AS ( 17. SELECT 18. bss.buyer\_id, 19. bss.seller\_id, 20. bss.frekuensi\_transaksi, 21. bss.transaksi\_terbesar, 22. bss.Q1, 23. bss.Q3, 24. (bss.Q3 - bss.Q1) AS IQR, 25. (bss.Q3 + 1.5 \* (bss.Q3 - bss.Q1)) AS upper\_bound 26. FROM 27. Buyer\_Seller\_Stats bss 28. ), 29. Anomalies AS ( 30. SELECT 31. ic.buyer\_id, 32. ic.seller\_id, 33. ic.frekuensi\_transaksi, 34. ic.transaksi\_terbesar, 35. ic.upper\_bound, 36. COUNT(CASE WHEN t.transaction\_amount > ic.upper\_bound THEN 1 END) AS count\_above\_upperbound 37. FROM 38. transaction t 39. JOIN 40. IQR\_Calculations ic 41. ON t.buyer\_id = ic.buyer\_id AND t.seller\_id = ic.seller\_id 42. GROUP BY 43. ic.buyer\_id, ic.seller\_id, ic.frekuensi\_transaksi, ic.transaksi\_terbesar, ic.upper\_bound 44. ), 45. Final\_Output AS ( 46. SELECT 47. a.\*, 48. CASE WHEN a.buyer\_id = a.seller\_id THEN 1 ELSE 0 END AS is\_self\_transaction, 49. bu.user\_fraud\_flag AS buyer\_fraud\_flag, 50. bu.blacklist\_account\_flag AS buyer\_blacklist\_flag, 51. su.user\_fraud\_flag AS seller\_fraud\_flag, 52. su.blacklist\_account\_flag AS seller\_blacklist\_flag 53. FROM 54. Anomalies a 55. LEFT JOIN 56. users bu ON a.buyer\_id = bu.company\_id 57. LEFT JOIN 58. users su ON a.seller\_id = su.company\_id 59. ) 60. SELECT 61. buyer\_id, 62. seller\_id, 63. is\_self\_transaction, 64. frekuensi\_transaksi, 65. ROUND(transaksi\_terbesar) AS transaksi\_terbesar, 66. ROUND(upper\_bound) AS upper\_bound, 67. count\_above\_upperbound 68. -- buyer\_fraud\_flag, 69. -- buyer\_blacklist\_flag, 70. -- seller\_fraud\_flag, 71. -- seller\_blacklist\_flag 72. FROM 73. Final\_Output 74. ORDER BY 75. frekuensi\_transaksi DESC, transaksi\_terbesar DESC; |
| --- |

Hasil queri ini menampilkan pasangan buyer-seller dengan frekuensi transaksi tinggi, transaksi terbesar mereka, batas atas transaksi yang dianggap anomali berdasarkan analisis IQR ”upper\_bound”, dan total transaksi anomalinya”count\_above\_upperbound”. Is\_self\_transaction disertakan juga untuk memperkuat sentimen kecurigaan pada setiap pasangan jika diketahui melakukan self-transaction.  


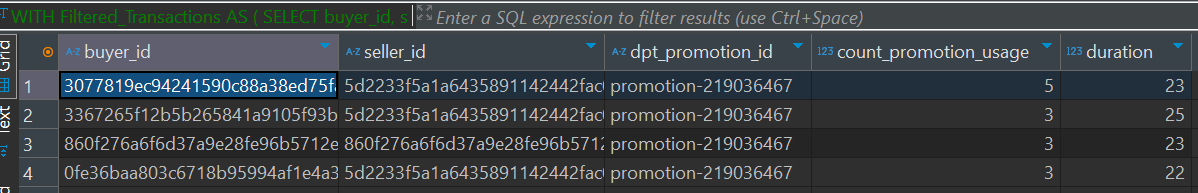
Pasangan buyer-seller nomor satu dianggap mencurigakan dan berpotensi melakukan kolusi karena banyak sekali melakukan self-transaction dan banyak juga nilai transaksi yang terhitung anomali jika dibandingkan dengan nilai-nilai transaksi yang biasanya

#### **4.1.1.3 Deteksi Penyalahgunaan Promosi**

**Kode Program 4.1.2.** *Exploratory Data Analysis* pada Python.

| 1. -- Promotion Misuse Detection: Users excessively used promotions within a short period of time 2. -- Deteksi Penyalahgunaan Promosi 3. WITH Filtered\_Transactions AS ( 4. SELECT 5. buyer\_id, 6. seller\_id, 7. dpt\_promotion\_id, 8. transaction\_created\_datetime::timestamp AS created\_at 9. FROM 10. transaction 11. WHERE 12. dpt\_promotion\_id <> 'no promotion' 13. ), 14. Ranked\_Transactions AS ( 15. SELECT 16. buyer\_id, 17. seller\_id, 18. dpt\_promotion\_id, 19. created\_at, 20. ROW\_NUMBER() OVER ( 21. PARTITION BY buyer\_id, seller\_id, dpt\_promotion\_id 22. ORDER BY created\_at 23. ) AS rank, 24. COUNT(\*) OVER (PARTITION BY buyer\_id, seller\_id, dpt\_promotion\_id) AS count\_promotion\_usage 25. FROM 26. Filtered\_Transactions 27. ), 28. Duration\_Calculation AS ( 29. SELECT 30. buyer\_id, 31. seller\_id, 32. dpt\_promotion\_id, 33. COUNT(\*) AS count\_promotion\_usage, 34. MIN(created\_at) AS first\_usage, 35. MAX(created\_at) AS last\_usage, 36. EXTRACT(DAY FROM (MAX(created\_at) - MIN(created\_at))) AS duration 37. FROM 38. Ranked\_Transactions 39. GROUP BY 40. buyer\_id, seller\_id, dpt\_promotion\_id 41. ) 42. SELECT 43. buyer\_id, 44. seller\_id, 45. dpt\_promotion\_id, 46. count\_promotion\_usage, 47. duration 48. FROM 49. Duration\_Calculation 50. where     1. (count\_promotion\_usage >= 3) and (duration <= 30) 51. ORDER BY 52. count\_promotion\_usage DESC, duration DESC; |
| --- |

Hasil queri ini menampilkan pasangan buyer-seller, promo id unik yang digunakan berkali-kali, banyaknya promo digunakan, dan durasi penggunaan promo berulang.



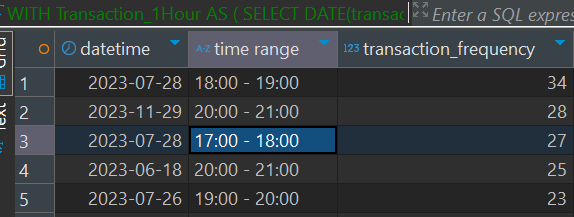
Pasangan buyer-seller nomor 1 dianggap menyalahgunakan atau mengeksploitasi promo karena menggunakan satu jenis promo yang sama sebanyak lima kali dalam waktu relatif singkat, yaitu 23 hari.

#### **4.1.1.4 Waktu yang Mencurigakan**

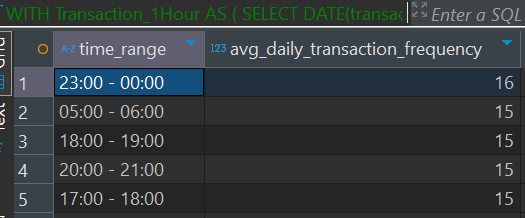
**Kode Program 4.1.2.** *Exploratory Data Analysis* pada Python.

| 1. --Suspicious Timing: Transactions at irregular hours or intervals (e.g., many transactions in a short time span). 2. -- waktu yang mencurigakan 3. WITH Transaction\_1Hour AS ( 4. SELECT 5. DATE(transaction\_created\_datetime::timestamp) AS transaction\_date, 6. DATE\_TRUNC('hour', transaction\_created\_datetime::timestamp) 7. + INTERVAL '1 hour' \* FLOOR(EXTRACT(MINUTE FROM transaction\_created\_datetime::timestamp) / 60) AS transaction\_1hour\_start, 8. COUNT(\*) AS transaction\_frequency 9. FROM 10. transaction 11. WHERE 12. EXTRACT(HOUR FROM transaction\_created\_datetime::timestamp) NOT BETWEEN 9 AND 16 13. GROUP BY 14. DATE(transaction\_created\_datetime::timestamp), 15. DATE\_TRUNC('hour', transaction\_created\_datetime::timestamp) 16. + INTERVAL '1 hour' \* FLOOR(EXTRACT(MINUTE FROM transaction\_created\_datetime::timestamp) / 60) 17. ), 18. Formatted\_1Hour AS ( 19. SELECT 20. transaction\_date, 21. TO\_CHAR(transaction\_1hour\_start, 'HH24:MI') || ' - ' || 22. TO\_CHAR(transaction\_1hour\_start + INTERVAL '1 hour', 'HH24:MI') AS time\_range, 23. transaction\_frequency 24. FROM 25. Transaction\_1Hour 26. where transaction\_frequency >= 12 27. ) 28. --SELECT 29. -- transaction\_date AS datetime, 30. -- time\_range AS "time range", 31. -- transaction\_frequency 32. --FROM 33. -- Formatted\_1Hour 34. --ORDER BY 35. -- /\*datetime, "time range",\*/ transaction\_frequency DESC; 37. select     1. time\_range,     2. round(avg(transaction\_frequency)) avg\_daily\_transaction\_frequency 38. from Formatted\_1Hour 39. group by time\_range 40. order by avg\_daily\_transaction\_frequency desc; |
| --- |

Hasil queri ini menampilkan tanggal dan rentang waktu per jam pada waktu ireguler, yang mana pada rentang waktu dan tanggal tersebut terjadi transaksi sedikitnya 12 kali (di luar 09.00-17.00).



Ternyata tidak jarang transaksi dilakukan pada waktu ireguler.



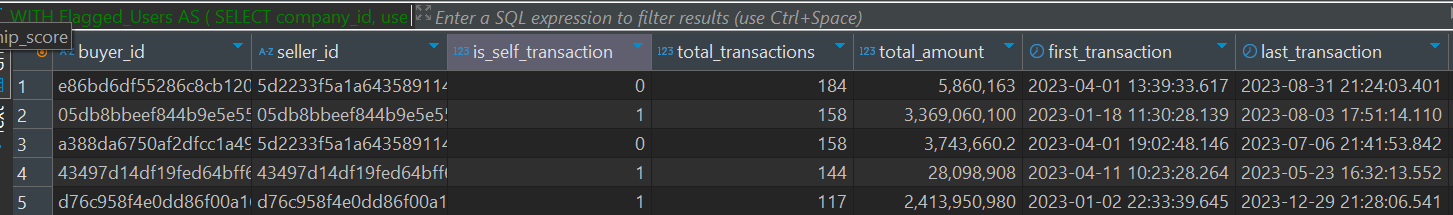
Rentang waktu 23.00-00.00 dianggap paling mencurigakan karena rata-rata transaksi hariannya terbesar.

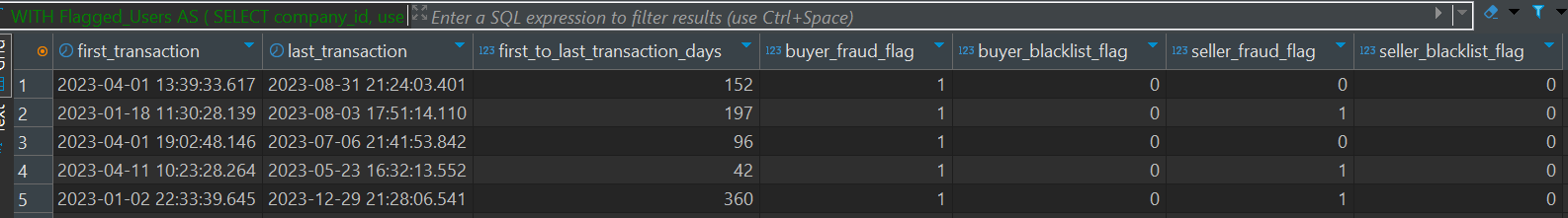
#### **4.1.1.5 Koneksi Pengguna yang Ditandai**

**Kode Program 4.1.2.** *Exploratory Data Analysis* pada Python.

| 1. ------------------------------------ Flagged User Connections 1--------------------------------------------- 2. WITH Flagged\_Users AS ( 3. SELECT 4. company\_id, 5. user\_fraud\_flag, 6. blacklist\_account\_flag 7. FROM 8. users 9. WHERE 10. user\_fraud\_flag = 1 OR blacklist\_account\_flag = 1 11. ), 12. User\_Connections AS ( 13. SELECT 14. ft.dpt\_id, 15. ft.buyer\_id, 16. ft.seller\_id, 17. fu\_buyer.user\_fraud\_flag AS buyer\_fraud\_flag, 18. fu\_buyer.blacklist\_account\_flag AS buyer\_blacklist\_flag, 19. fu\_seller.user\_fraud\_flag AS seller\_fraud\_flag, 20. fu\_seller.blacklist\_account\_flag AS seller\_blacklist\_flag, 21. ft.transaction\_amount, 22. ft.transaction\_created\_datetime::timestamp AS transaction\_time 23. FROM 24. transaction ft 25. LEFT JOIN 26. users fu\_buyer ON ft.buyer\_id = fu\_buyer.company\_id 27. LEFT JOIN 28. users fu\_seller ON ft.seller\_id = fu\_seller.company\_id 29. WHERE 30. fu\_buyer.user\_fraud\_flag = 1 OR 31. fu\_buyer.blacklist\_account\_flag = 1 OR 32. fu\_seller.user\_fraud\_flag = 1 OR 33. fu\_seller.blacklist\_account\_flag = 1 34. ), 35. Flagged\_Transactions AS ( 36. SELECT 37. buyer\_id, 38. seller\_id, 39. CASE WHEN buyer\_id = seller\_id THEN 1 ELSE 0 END AS is\_self\_transaction, 40. COUNT(\*) AS total\_transactions, 41. SUM(transaction\_amount) AS total\_amount, 42. MIN(transaction\_time) AS first\_transaction, 43. MAX(transaction\_time) AS last\_transaction, 44. DATE\_PART('day', MAX(transaction\_time) - MIN(transaction\_time)) AS first\_to\_last\_transaction\_days, 45. MAX(buyer\_fraud\_flag) AS buyer\_fraud\_flag, 46. MAX(buyer\_blacklist\_flag) AS buyer\_blacklist\_flag, 47. MAX(seller\_fraud\_flag) AS seller\_fraud\_flag, 48. MAX(seller\_blacklist\_flag) AS seller\_blacklist\_flag 49. FROM 50. User\_Connections 51. GROUP BY 52. buyer\_id, seller\_id 53. ) 54. SELECT 55. buyer\_id, 56. seller\_id, 57. is\_self\_transaction, 58. total\_transactions, 59. total\_amount, 60. first\_transaction, 61. last\_transaction, 62. first\_to\_last\_transaction\_days, 63. buyer\_fraud\_flag, 64. buyer\_blacklist\_flag, 65. seller\_fraud\_flag, 66. seller\_blacklist\_flag 67. FROM 68. Flagged\_Transactions 69. ORDER BY 70. total\_transactions DESC, total\_amount DESC; |
| --- |

Hasil queri ini menampilkan rekam jejak transaksi user yang fraud atau di-blacklist





Pasangan buyer-seller nomor 4 merupakan self-transaction yang mana usernya termasuk user fraud. User ini melakukan 144 transaksi hanya dalam 42 hari tetapi belum diblacklist.

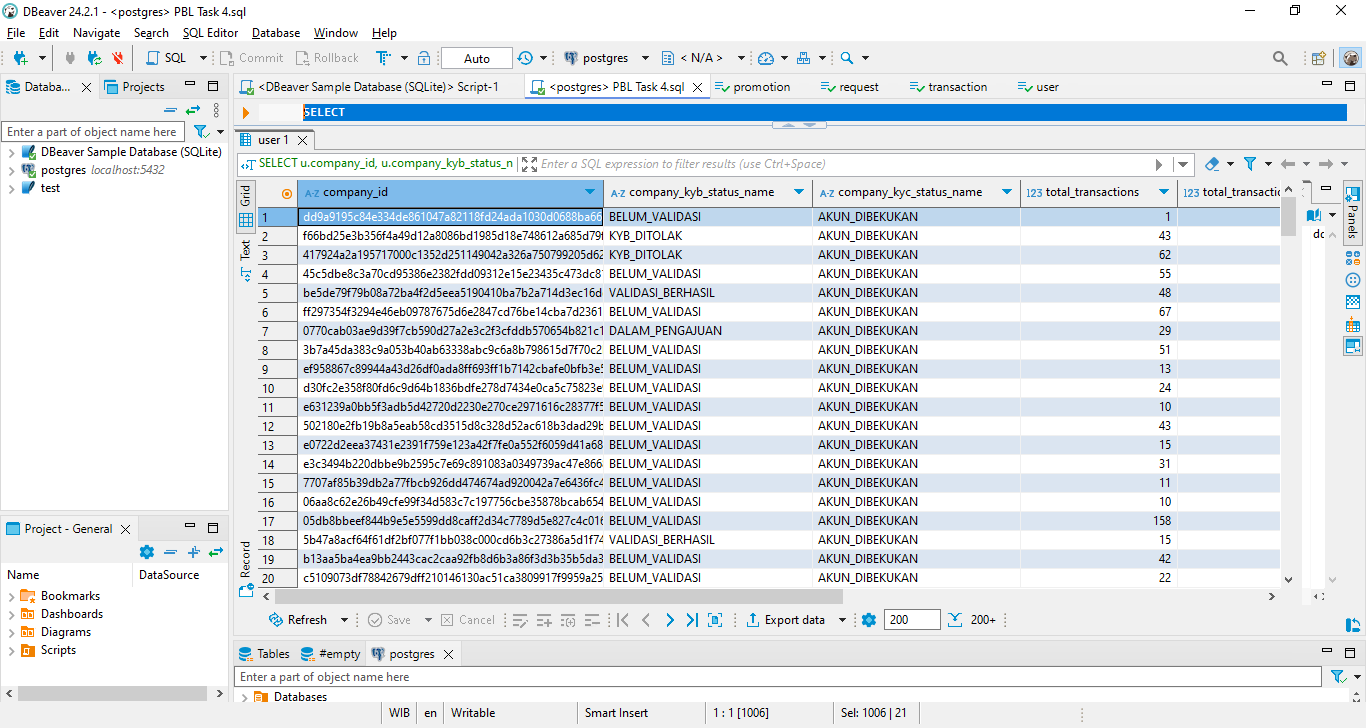
## **4.2 SQL Joins for User-Company Fraud Insights**

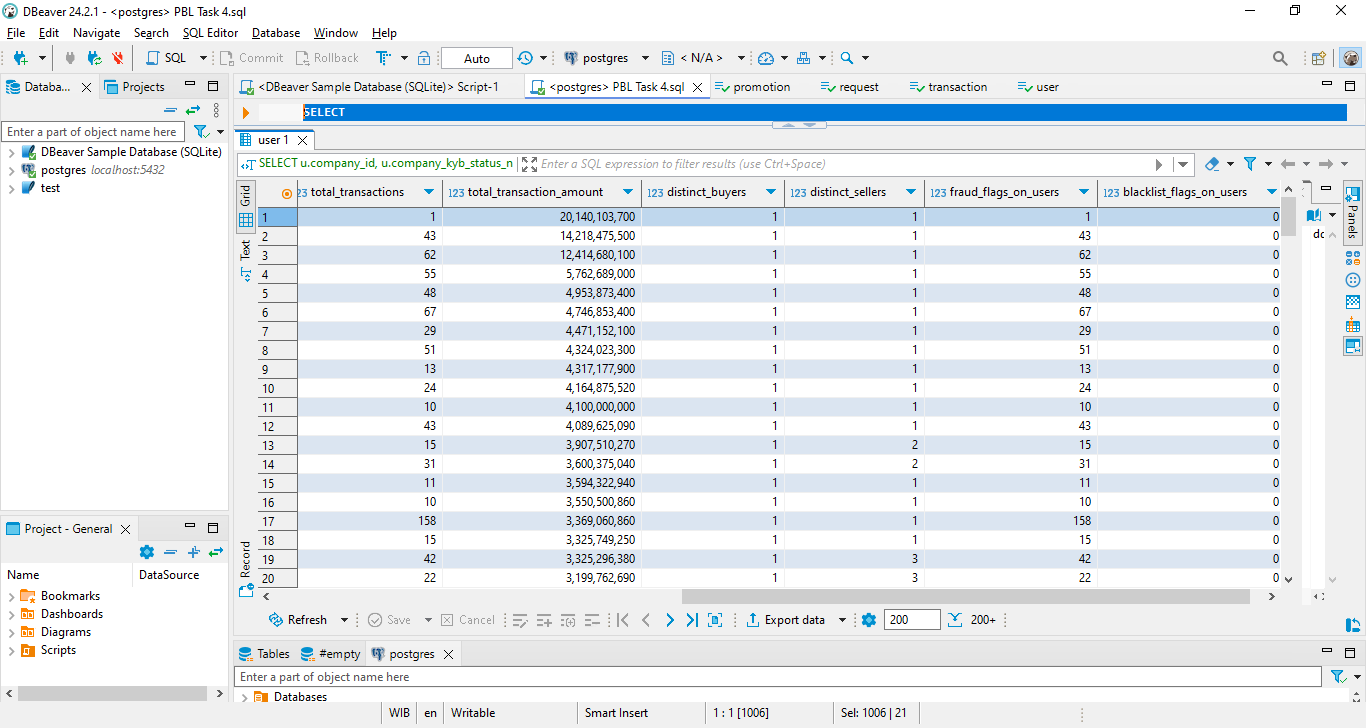
Untuk mengidentifikasi dan mencegah tindakan fraud dalam sistem transaksi, sangat penting untuk menganalisis tidak hanya perilaku transaksi pengguna, tetapi juga hubungan antara pengguna dengan perusahaan tempat mereka bertransaksi. Dengan menggabungkan data transaksi dengan informasi terkait status KYC (Know Your Customer), KYB (Know Your Business), dan riwayat flag fraud yang dimiliki oleh perusahaan, kita dapat memperoleh wawasan lebih dalam mengenai pola-pola yang dapat mengindikasikan potensi fraud. Adapun querynya sebagai berikut:

**Kode Program 4.2**. *User-Company Fraud Insights*

| 1. **SELECT** 2. *u*.company\_id, 3. *u*.company\_kyb\_status\_name, 4. *u*.company\_kyc\_status\_name, 5. **COUNT**(**DISTINCT** *t*.dpt\_id) **AS** *total\_transactions*, -- Menghitung transaksi unik 6. **SUM**(*t*.transaction\_amount) **AS** *total\_transaction\_amount*, 7. **COUNT**(**DISTINCT** *t*.buyer\_id) **AS** *distinct\_buyers*, 8. **COUNT**(**DISTINCT** *t*.seller\_id) **AS** *distinct\_sellers*, 9. **SUM**(**CASE** **WHEN** *u*.user\_fraud\_flag = 1 **THEN** 1 **ELSE** 0 **END**) **AS** *fraud\_flags\_on\_users*, 10. **SUM**(**CASE** **WHEN** *u*.blacklist\_account\_flag = 1 **THEN** 1 **ELSE** 0 **END**) **AS** *blacklist\_flags\_on\_users* 11. **FROM** transaction *t* 12. **LEFT** **JOIN** 13. "user" *u* **ON** *t*.buyer\_id = *u*.company\_id 14. **LEFT** **JOIN** 15. "user" *u\_seller* **ON** *t*.seller\_id = *u\_seller*.company\_id 16. **WHERE** *t*.transaction\_created\_datetime **BETWEEN** '2022-01-01' **AND** '2023-12-31' 17. **GROUP** **BY** *u*.company\_id, *u*.company\_kyb\_status\_name, *u*.company\_kyc\_status\_name 18. **HAVING** 19. **SUM**(**CASE** **WHEN** *u*.user\_fraud\_flag = 1 **THEN** 1 **ELSE** 0 **END**) > 0 20. **OR** **SUM**(**CASE** **WHEN** *u*.blacklist\_account\_flag = 1 **THEN** 1 **ELSE** 0 **END**) > 0 21. **ORDER** **BY** *total\_transaction\_amount* **DESC**; |
| --- |

Dari hasil query tersebut didapat hasil sebagai berikut:





# **BAB V**

# **SQL Views and Stored Procedures**

## **5.1 SQL View**

Melalui pembuatan SQL Views, kita dapat merancang laporan fraud secara reguler dengan dua jenis tampilan yang sangat berguna

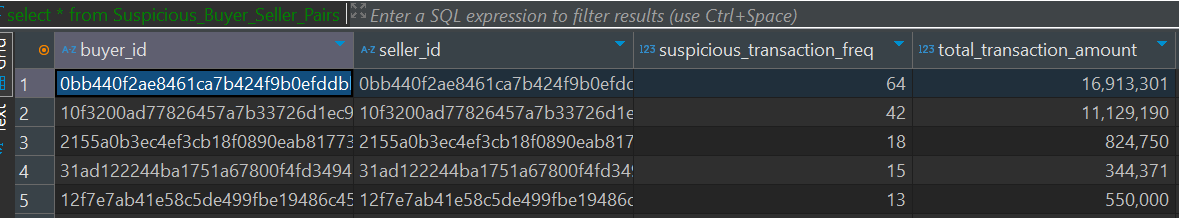
### **5.1.1 View Pasangan buyer-seller paling mencurigakan**

**V**iew ini memberikan ringkasan tentang hubungan pembeli-penjual yang paling mencurigakan, berdasarkan frekuensi transaksi dan jumlah transaksi yang dilakukan.. Adapun querynya sebagai berikut :

**Kode Program 4.3.1.1**.

| 1. use pbl\_paper\_id; 2. select count(\*) from transaction\_frequency\_metrics; 3. /\*------------------------------------ SQL View --------------------------------------------\*/ 4. -- 1.1 Membuat view untuk pasangan pembeli-penjual yang paling mencurigakan berdasarkan fitur baru 5. -- burst\_activity, unusual\_gap, burst\_amount dibuat pada tahap feature engineering di task 2 6. -- Pasangan buyer-seller paling mencurigakan 7. CREATE VIEW Suspicious\_Buyer\_Seller\_Pairs AS 8. SELECT 9. buyer\_id, 10. seller\_id, 11. COUNT(dpt\_id) AS suspicious\_transaction\_freq, 12. ROUND(SUM(transaction\_amount)) AS total\_transaction\_amount 13. FROM 14. transaction\_frequency\_metrics 15. WHERE 16. (burst\_activity = 1 AND burst\_amount = 1) 17. OR (unusual\_gap = 1 AND burst\_amount = 1) 18. GROUP BY 19. buyer\_id, 20. seller\_id 21. ORDER BY 22. suspicious\_transaction\_freq DESC; 24. select \* from Suspicious\_Buyer\_Seller\_Pairs; |
| --- |

Dimana didapat hasil sebagai berikut:



View ini menampilkan pasangan buyer-seller yang dianggap mencurigakan berdasarkan indikator burs\_activity, unusual\_gap, dan burst\_amount yang dihasilkan dari bagian feature engineering, juga frekuensi transaksi mencurigakan yang mereka lakukan serta total transaksi mencurigakannya.

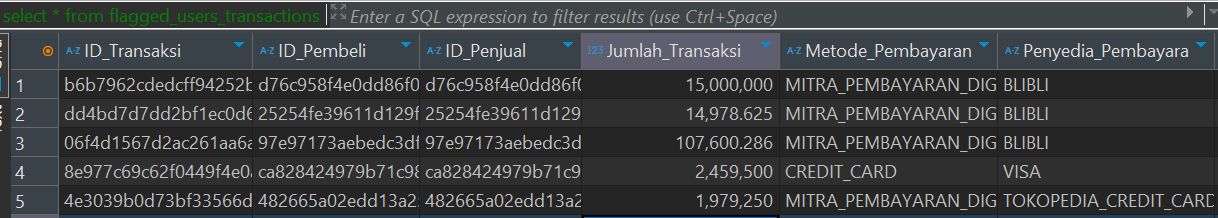
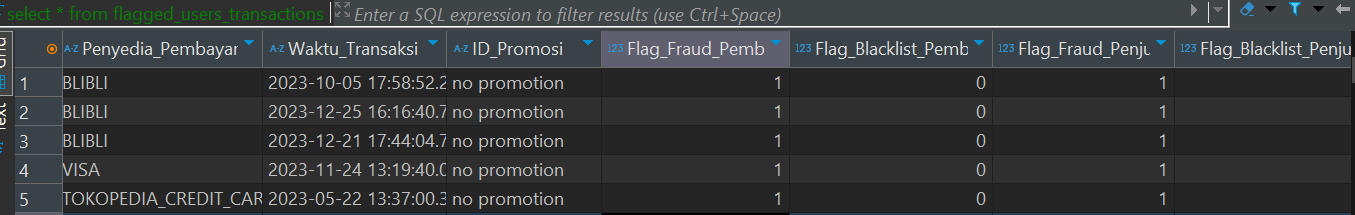
### **5.1.2 View - Pengguna yang Ditandai dan Transaksi Mereka**

View ini berfokus pada pengguna yang telah terflag atau dimasukkan dalam daftar hitam, bersama dengan transaksi yang melibatkan mereka. Adapun querynya sebagi berikut :

**Kode Program 4.3.1.2**.

| 1. -- 1.2 Membuat View untuk pengguna yang terindikasi fraud atau di-blacklist dan transaksi mereka 2. -- Pengguna yang ditandai dan transaksi mereka 3. CREATE VIEW Flagged\_Users\_Transactions AS 4. SELECT 5. t.dpt\_id AS ID\_Transaksi, 6. t.buyer\_id AS ID\_Pembeli, 7. t.seller\_id AS ID\_Penjual, 8. t.transaction\_amount AS Jumlah\_Transaksi, 9. t.payment\_method\_name AS Metode\_Pembayaran, 10. t.payment\_provider\_name AS Penyedia\_Pembayaran, 11. t.transaction\_created\_datetime AS Waktu\_Transaksi, 12. t.dpt\_promotion\_id AS ID\_Promosi, 13. ub.user\_fraud\_flag AS Flag\_Fraud\_Pembeli, 14. ub.blacklist\_account\_flag AS Flag\_Blacklist\_Pembeli, 15. us.user\_fraud\_flag AS Flag\_Fraud\_Penjual, 16. us.blacklist\_account\_flag AS Flag\_Blacklist\_Penjual, 17. CASE 18. WHEN t.buyer\_id = t.seller\_id THEN 1 19. ELSE 0 20. END AS Flag\_Self\_Transaction 21. FROM 22. transaction t 23. LEFT JOIN 24. user ub 25. ON 26. t.buyer\_id = ub.company\_id 27. LEFT JOIN 28. user us 29. ON 30. t.seller\_id = us.company\_id 31. WHERE 32. (ub.user\_fraud\_flag = 1 OR ub.blacklist\_account\_flag = 1) 33. OR 34. (us.user\_fraud\_flag = 1 OR us.blacklist\_account\_flag = 1); 36. select count(\*) from flagged\_users\_transactions; 37. select \* from flagged\_users\_transactions; |
| --- |

Dimana didapat hasil sebagai berikut:



View ini menampilkan detail transaksi beserta status buyer dan seller sehingga dapat diketahui apakah transaksi tertentu dilakukan oleh user yang fraud atau tidak.

## **5.2 Stored Procedures**

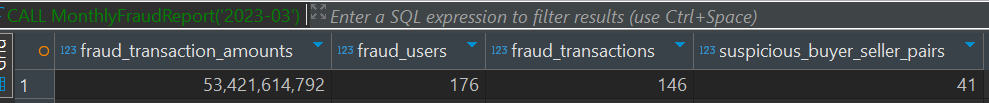
Untuk meningkatkan efisiensi dan akurasi dalam deteksi fraud, stored procedures dapat digunakan untuk mengotomatisasi proses analisis dan identifikasi transaksi mencurigakan. Dengan menggunakan stored procedures, organisasi dapat menjalankan serangkaian query atau logika deteksi fraud secara otomatis, tanpa perlu melibatkan intervensi manual setiap kali analisis dilakukan. Adapaun Store procedures yang dilakukanialah:

### **5.2.1 Laporan Penipuan Bulanan**

Prosedur ini menghasilkan report transaksi bulanan yang tergolong fraud atau mencurigakan. Dari prosedur ini kita dapat melihat total nilai transaksi fraud, berapa fraud user yang terlibat dalam transaksi fraud, berapa berapa pasang buyer-seller ditandai fraud yang bertransaksi, dan berapa pasang buyer-seller yang terindikasi melakukan transaksi mencurigakan (berdasarkan indikator burs\_activity, unusual\_gap, dan burst\_amount) tetapi belum ditandai sebagai fraud user. Harapannya prosedur ini dapat digunakan untuk mendeteksi user-user yang belum ditandai walau sebenarnya mereka melakukan transaksi fraud. Adapun querynya sebagai berikut :

**Kode Program 4.1.2.**

| 1. use pbl\_paper\_id; 2. /\*------------------------------------------SQL Stored Procedures-------------------------------------\*/ 3. -- 2.1 Monthly Fraud Report Procedure 4. -- laporan penipuan bulanan 5. DELIMITER $$ 6. CREATE PROCEDURE MonthlyFraudReport(IN report\_month VARCHAR(7)) 7. BEGIN 8. DECLARE total\_fraud\_amount DOUBLE; 9. DECLARE total\_fraud\_user\_count INT; 10. DECLARE total\_suspicious\_count INT; 11. DECLARE total\_fraud\_transaction\_count INT; 12. -- Hitung Total Fraud Amount 13. SELECT 14. SUM(tfm.transaction\_amount) 15. INTO total\_fraud\_amount 16. FROM transaction\_frequency\_metrics tfm 17. LEFT JOIN user u\_buyer ON tfm.buyer\_id = u\_buyer.company\_id 18. LEFT JOIN user u\_seller ON tfm.seller\_id = u\_seller.company\_id 19. WHERE 20. (u\_buyer.user\_fraud\_flag = 1 OR u\_seller.user\_fraud\_flag = 1) 21. AND DATE\_FORMAT(tfm.transaction\_created\_datetime, '%Y-%m') = report\_month; 22. -- Hitung Total Fraud Users 23. SELECT 24. COUNT(DISTINCT CASE WHEN u\_buyer.user\_fraud\_flag = 1 THEN tfm.buyer\_id END) 25. + COUNT(DISTINCT CASE WHEN u\_seller.user\_fraud\_flag = 1 THEN tfm.seller\_id END) 26. INTO total\_fraud\_user\_count 27. FROM transaction\_frequency\_metrics tfm 28. LEFT JOIN user u\_buyer ON tfm.buyer\_id = u\_buyer.company\_id 29. LEFT JOIN user u\_seller ON tfm.seller\_id = u\_seller.company\_id 30. WHERE 31. (u\_buyer.user\_fraud\_flag = 1 OR u\_seller.user\_fraud\_flag = 1) 32. AND DATE\_FORMAT(tfm.transaction\_created\_datetime, '%Y-%m') = report\_month; 33. -- Hitung Total Fraud Transactions (Buyer-Seller Pairs) 34. SELECT 35. COUNT(DISTINCT CONCAT(tfm.buyer\_id, '-', tfm.seller\_id)) 36. INTO total\_fraud\_transaction\_count 37. FROM transaction\_frequency\_metrics tfm 38. LEFT JOIN user u\_buyer ON tfm.buyer\_id = u\_buyer.company\_id 39. LEFT JOIN user u\_seller ON tfm.seller\_id = u\_seller.company\_id 40. WHERE 41. (u\_buyer.user\_fraud\_flag = 1 OR u\_seller.user\_fraud\_flag = 1) 42. AND DATE\_FORMAT(tfm.transaction\_created\_datetime, '%Y-%m') = report\_month; 43. -- Hitung Total Suspicious Buyer-Seller Pairs 44. SELECT 45. COUNT(DISTINCT CONCAT(tfm.buyer\_id, '-', tfm.seller\_id)) 46. INTO total\_suspicious\_count 47. FROM transaction\_frequency\_metrics tfm 48. WHERE 49. ((tfm.burst\_activity = 1 AND tfm.burst\_amount = 1) 50. OR (tfm.unusual\_gap = 1 AND tfm.burst\_amount = 1)) 51. AND DATE\_FORMAT(tfm.transaction\_created\_datetime, '%Y-%m') = report\_month 52. AND CONCAT(tfm.buyer\_id, '-', tfm.seller\_id) NOT IN ( 53. SELECT DISTINCT CONCAT(f\_tfm.buyer\_id, '-', f\_tfm.seller\_id) 54. FROM transaction\_frequency\_metrics f\_tfm 55. LEFT JOIN user f\_u\_buyer ON f\_tfm.buyer\_id = f\_u\_buyer.company\_id 56. LEFT JOIN user f\_u\_seller ON f\_tfm.seller\_id = f\_u\_seller.company\_id 57. WHERE 58. (f\_u\_buyer.user\_fraud\_flag = 1 OR f\_u\_seller.user\_fraud\_flag = 1) 59. AND DATE\_FORMAT(f\_tfm.transaction\_created\_datetime, '%Y-%m') = report\_month 60. ); 61. -- Hasil Akhir 62. SELECT 63. CAST(total\_fraud\_amount AS UNSIGNED) AS fraud\_transaction\_amounts, 64. total\_fraud\_user\_count AS fraud\_users, 65. total\_fraud\_transaction\_count AS fraud\_transactions, 66. total\_suspicious\_count AS suspicious\_buyer\_seller\_pairs; 67. END $$ 68. DELIMITER ; 69. CALL MonthlyFraudReport('2023-03'); |
| --- |



Pada bulan Maret 2023 total transaksi fraud (transaksi di mana buyer atau seller ditandai fraud) mencapai sekitar 53 Miliar. Ada 146 pasangan buyer-seller yang terlibat dalam transaksi ini. Beberapa transaksi melibatkan buyer dan seller di mana keduanya merupakan user yang ditandai fraud. Disamping itu, pada periode bulan ini, ditemukan 41 pasangan buyer-seller yang melakukan transaksi mencurigakan, tetapi keduanya belum ada yang ditandai fraud.

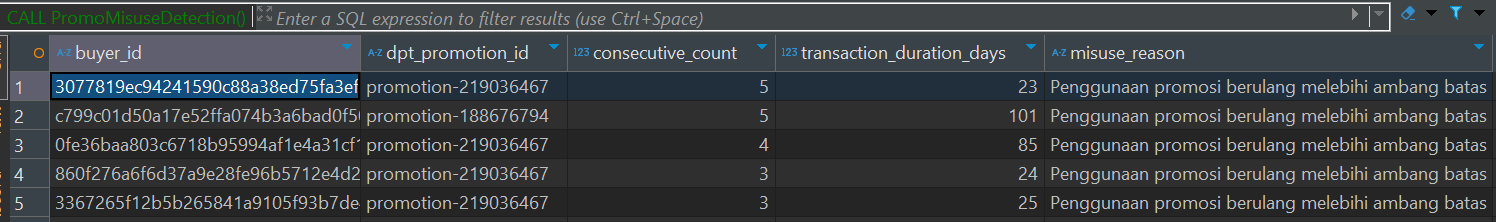
### **5. 2.2 Deteksi Penyalahgunaan Otomatis**

Prosedur ini menampilkan report buyer\_id yang menggunakan jenis promo yang sama berulang kali sedikitnya 3 kali. Sebagian buyer\_id menggunakannya dalam waktu relatif singkat sehingga dapat dianggap melakukan penyalahgunaan.. Adaun querynya sebagai berikut:

**Kode Program 4.3.2.2**. *User-Company Fraud Insights*

| 1. -- 2.2 Automated Promotion Misuse Detection 2. -- Deteksi penyalahgunaan otomatis 3. DELIMITER // 4. CREATE PROCEDURE PromoMisuseDetection() 5. BEGIN 6. -- Hapus tabel sementara jika sudah ada 7. DROP TEMPORARY TABLE IF EXISTS temp\_promo\_transactions; 8. DROP TEMPORARY TABLE IF EXISTS promo\_counts; 9. -- Tabel sementara untuk menyimpan data transaksi dengan lag calculation 10. CREATE TEMPORARY TABLE temp\_promo\_transactions AS 11. SELECT 12. buyer\_id, 13. dpt\_promotion\_id, 14. transaction\_created\_datetime, 15. -- Penanda transaksi sebelumnya 16. LAG(transaction\_created\_datetime) OVER (PARTITION BY buyer\_id, dpt\_promotion\_id ORDER BY transaction\_created\_datetime) AS prev\_transaction\_datetime 17. FROM transaction 18. WHERE 19. dpt\_promotion\_id IS NOT NULL 20. AND dpt\_promotion\_id <> 'no promotion'; -- Mengabaikan transaksi tanpa promosi 21. -- Variabel untuk menghitung consecutive promo count 22. SET @row\_num := 0; 23. SET @prev\_buyer := NULL; 24. SET @prev\_promo := NULL; 25. -- Tabel sementara kedua untuk menghitung consecutive promo count 26. CREATE TEMPORARY TABLE promo\_counts AS 27. SELECT 28. buyer\_id, 29. dpt\_promotion\_id, 30. transaction\_created\_datetime, 31. -- Menghitung transaksi berturut-turut 32. @row\_num := IF(@prev\_buyer = buyer\_id AND @prev\_promo = dpt\_promotion\_id, @row\_num + 1, 1) AS consecutive\_promo\_count, 33. @prev\_buyer := buyer\_id, 34. @prev\_promo := dpt\_promotion\_id 35. FROM temp\_promo\_transactions 36. ORDER BY buyer\_id, dpt\_promotion\_id, transaction\_created\_datetime; 37. -- Tampilkan data misuse ke output jika melebihi threshold 38. SELECT 39. buyer\_id, 40. dpt\_promotion\_id, 41. MAX(consecutive\_promo\_count) AS consecutive\_count, 42. DATEDIFF(MAX(transaction\_created\_datetime), MIN(transaction\_created\_datetime)) AS transaction\_duration\_days, 43. 'Penggunaan promosi berulang melebihi ambang batas' AS misuse\_reason 44. FROM promo\_counts 45. GROUP BY buyer\_id, dpt\_promotion\_id 46. HAVING MAX(consecutive\_promo\_count) >= 3 -- Ambang batas deteksi 47. ORDER BY consecutive\_count DESC, transaction\_duration\_days ASC; 49. END // 50. DELIMITER ; 51. -- Menjalankan Stored Procedure 52. CALL PromoMisuseDetection(); 53. select\* from log\_promosi\_misuse; |
| --- |

Dari hasil query tersebut didapat hasil sebagai berikut:



Buyer\_id pada nomor satu menggunakan promo yang sama berturut-turut selama lima kali hanya dalam 23 hari. Dapat dikatakan bahwa bahwa buyer\_id ini telah melakukan penyalahgunaan promosi.

# **BAB VI**

# **[Python] Advanced Fraud Analysis and Network Insights**

## **6.1 Social Network Analysis**

Menggunakan query SQL untuk membuat jaringan sosial yang menggambarkan hubungan antara pembeli dan penjual. Data transaksi inidigunakan untuk mengidentifikasi pola interaksi antara keduanya. Analisis ini membantu dalam memahami dinamika hubungan dan potensi risiko atau peluang bisnis, Adapaun analisis nya sebagai berikut:

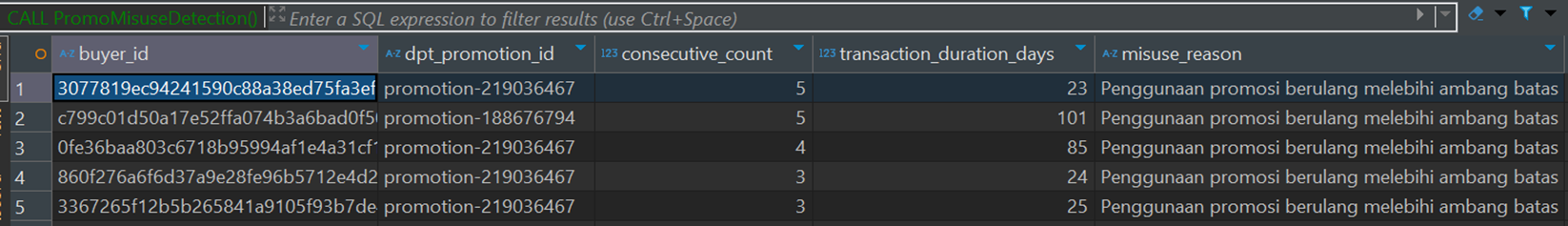
### **6.1.1 Analisis Hubungan Pembeli-Penjual untuk Mengidentifikasi Pola Interaksi dan Terlibat Fraud (using SQL)**

Query SQL telah dirancang untuk menghasilkan data pasangan pembeli-penjual yang terlibat dalam transaksi fraud. Suatu transaksi diklasifikasikan sebagai fraud apabila terdapat pembeli atau penjual yang ditandai dengan nilai 1 pada kolom fraud\_flag atau blacklist\_flag. Analisis ini bertujuan untuk memahami pola interaksi dalam jaringan sosial antara pembeli dan penjual, sekaligus mendeteksi risiko terkait aktivitas penipuan

**Kode Program 3.1.** *Exploratory Data Analysis* pada Python.

| 1. use pbl\_paper\_id; 2. -- Membuat edge\_list untuk social network analisis user yang terlibat transaksi fraud 3. WITH transaction\_with\_buyer\_flags AS ( 4. SELECT 5. t.\*, 6. u.user\_fraud\_flag AS user\_fraud\_flag\_buyer, 7. u.blacklist\_account\_flag AS blacklist\_account\_flag\_buyer 8. FROM transaction t 9. LEFT JOIN user u 10. ON t.buyer\_id = u.company\_id 11. ), 12. transaction\_with\_seller\_flags AS ( 13. SELECT 14. tb.\*, 15. COALESCE(us.user\_fraud\_flag, 1) AS user\_fraud\_flag\_seller, 16. COALESCE(us.blacklist\_account\_flag, 1) AS blacklist\_account\_flag\_seller 17. FROM transaction\_with\_buyer\_flags tb 18. LEFT JOIN user us 19. ON tb.seller\_id = us.company\_id 20. ), 21. filtered\_transactions AS ( 22. SELECT \* 23. FROM transaction\_with\_seller\_flags 24. WHERE 25. user\_fraud\_flag\_buyer = 1 OR 26. blacklist\_account\_flag\_buyer = 1 OR 27. user\_fraud\_flag\_seller = 1 OR 28. blacklist\_account\_flag\_seller = 1 29. ), 30. final\_filtered\_transactions AS ( 31. SELECT \* 32. FROM filtered\_transactions 33. WHERE buyer\_id != seller\_id 34. ) 35. SELECT buyer\_id, seller\_id 36. FROM final\_filtered\_transactions; |
| --- |

Hasil dari query ini menunjukkan terdapat 5.107 baris data yang mencakup pasangan pembeli-penjual beserta transaksi yang berpotensi fraud.

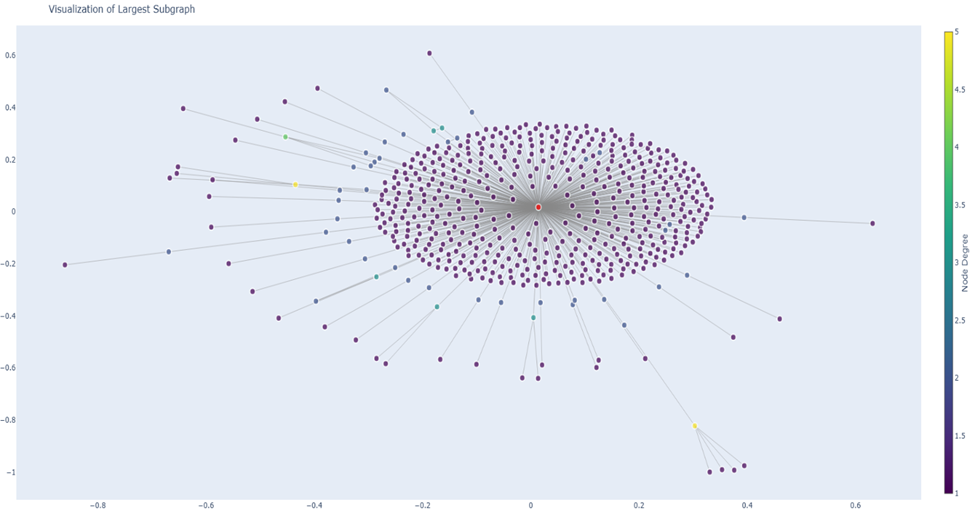


Data ini mencerminkan frekuensi dan pola hubungan antar entitas, termasuk pengelompokan entitas yang sering berinteraksi. Pengelompokan ini dapat membantu mengidentifikasi kluster pembeli dan penjual yang memiliki kemungkinan lebih tinggi untuk terlibat dalam aktivitas mencurigakan.Selanjutnya, data hasil query ini dieksport dalam format CSV untuk analisis lebih lanjut menggunakan Python.

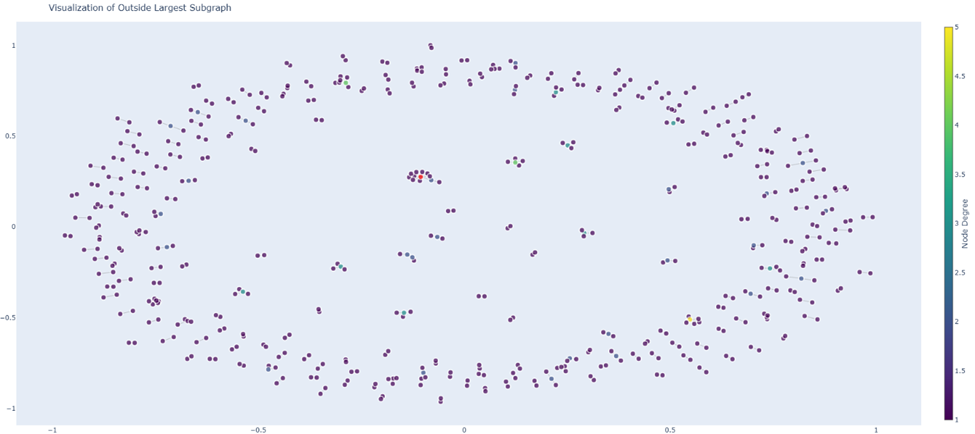
### **6.1.2 Analisis Hubungan Pembeli-Penjual untuk Mengidentifikasi Pola Interaksi dan Terlibat Fraud (Visualization using Phyton)**

Pada tahap ini digunakan beberapa library sebagian diantaranya yaitu networkx untuk membuat graf dan plotly untuk menampilkan graf yang responsif.

Data hasil queri sebelumnya diubah menjadi graf G. Graf ini mempunyai 859 edges dan 1089 nodes. Dari 1089 nodes yang terlibat dalam transaksi fraud ternyata 318 nodes tidak terdaftar di data user. Data transaksi fraud membentuk satu graf besar dengan 552 nodes dan banyak graf-graf kecil yang kebanyakan hanya terdiri dari 2 hingga 3 nodes. Oleh karena itu, visualisasinya dibagi menjadi dua, largest subgraph dan outside largest subgraph seperti di bawah ini:



Node berwarna merah merupakan central node yang terkoneksi ke 511 nodes lainnya pada graf utama. Selain central node kebanyakan node hanya terkoneksi ke satu atau dua node lainnya, tetapi ada juga beberapa node yang terkoneksi ke 3 hingga 5 nodes lain. Jumlah nodes tidak terdaftar pada graf utama G\_largest hanya 25 nodes atau sekitar 8% dari total nodes tidak terdaftar. Setelah dilakukan pengecekan pada data user ternyata central node tidak ditandai fraud atau pun tidak diblacklist.



Total nodes pada G\_ outside\_largest adalah 537 nodes dan 293 diantaranya adalah akun tidak terdaftar (92% dari total akun tidak terdaftar). Pada kumpulan graf-graf kecil ini terdapat satu graph yang lebih banyak nodes-nya dibanding yang lain, totalnya ada 12 nodes. Central node-nya terkoneksi ke 10 nodes yang lainnya. Setelah di lakukan pengecekan ternyata central node-nya adalah akun testing. Selain itu ada satu lagi akun testing pada graf ini, sementara yang 10 lainnya tidak terdaftar di data user.

## **6.2 Cohort Analysis**

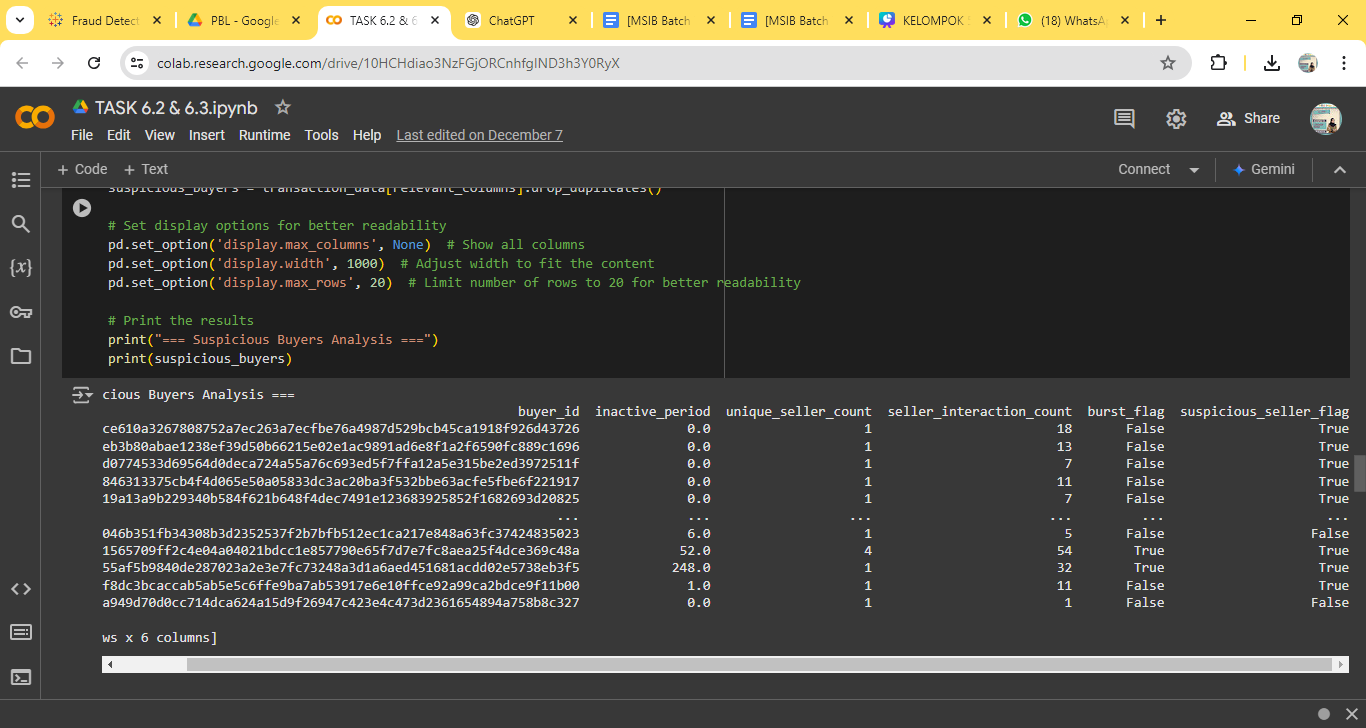
Melakukan analisis kohort untuk melacak **transaksi berulang** antara pembeli dan penjual:

### **6.2.1 Kelompokkan pembeli berdasarkan tanggal transaksi pertama mereka dan mengukur aktivitas berkelanjutan mereka dari waktu ke waktu.**

Analisis kohort menunjukkan bahwa jumlah pembeli aktif menurun seiring waktu dalam setiap kohort, mengindikasikan adanya *churn* atau penurunan retensi pelanggan. Kohort awal, seperti Februari 2022, memiliki retensi yang lebih rendah dibandingkan dengan kohort yang lebih baru. Namun, terlihat tren positif pada pertumbuhan jumlah pembeli baru di bulan-bulan akhir tahun 2023, seperti Desember 2023, yang mencatat 1.099 pembeli aktif pada bulan pertama. Hal ini menunjukkan keberhasilan dalam menarik pelanggan baru, meskipun keberlanjutan aktivitas mereka perlu dipantau lebih lanjut.

Selain itu, beberapa kohort di pertengahan 2023 menunjukkan retensi yang lebih baik, seperti kohort April dan Mei 2023, yang masih memiliki banyak pembeli aktif di bulan-bulan berikutnya. Ini bisa jadi hasil dari strategi pemasaran atau promosi yang efektif pada periode tersebut. Namun, penurunan aktivitas di bulan-bulan selanjutnya tetap menjadi perhatian, sehingga penting untuk meningkatkan retensi melalui pendekatan seperti program loyalitas, personalisasi penawaran, atau peningkatan pengalaman pelanggan.

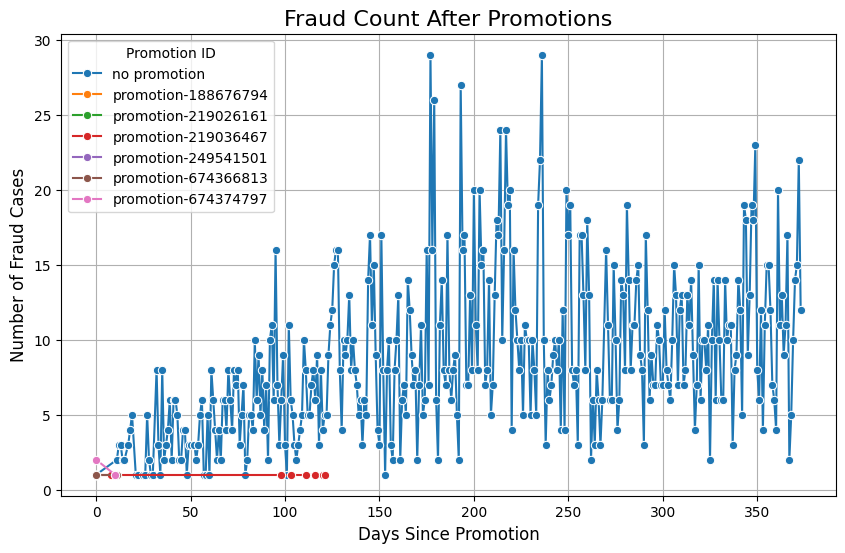
### **6.2.2 Identifikasi apakah pembeli tertentu terlibat dalam perilaku penipuan setelah periode tidak aktif atau berulang kali berinteraksi dengan penjual yang sama.**

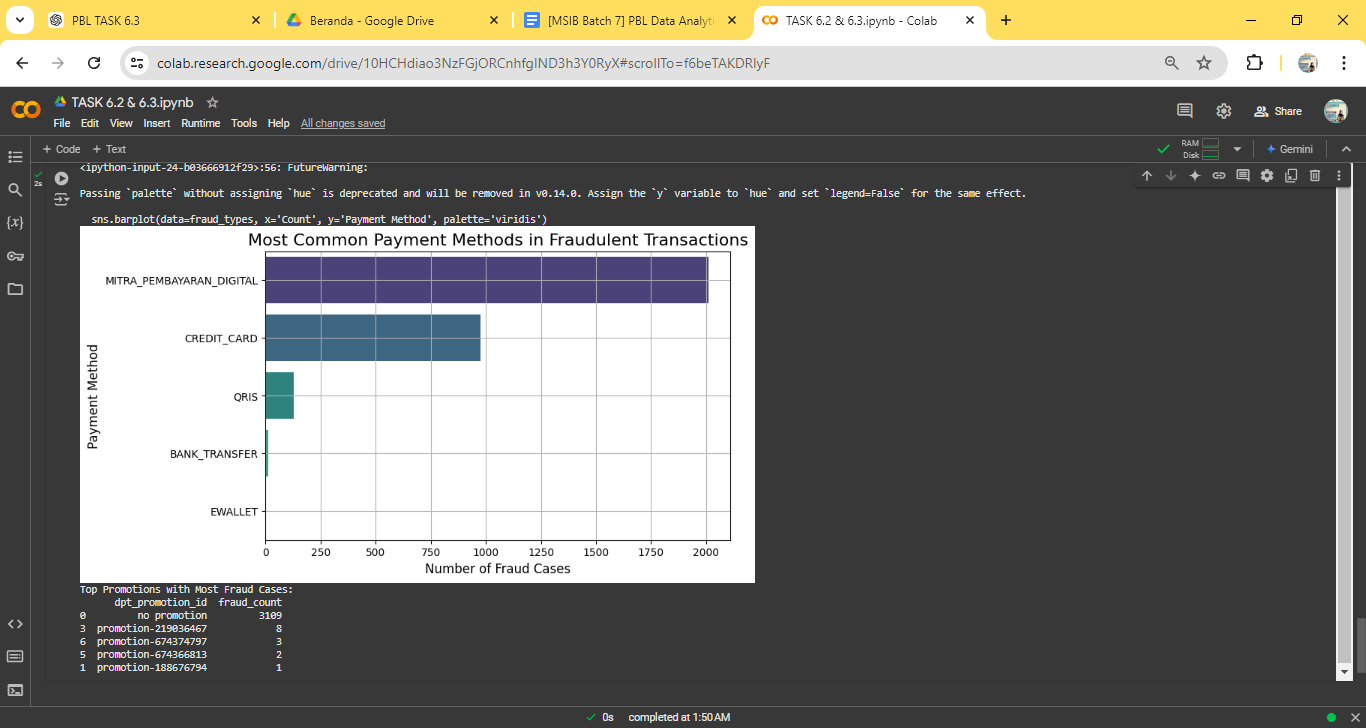


Analisis menunjukkan bahwa pembeli dengan periode tidak aktif yang panjang atau yang sering berinteraksi dengan penjual yang sama cenderung memiliki perilaku mencurigakan. Misalnya, pembeli dengan **inactive\_period** tinggi yang tiba-tiba kembali dengan jumlah interaksi yang signifikan atau memiliki **unique\_seller\_count** rendah dapat mengindikasikan potensi kolusi atau penipuan. Kehadiran **burst\_flag**, yang menandakan lonjakan aktivitas mendadak, juga dapat menjadi indikator perilaku tidak wajar, terutama jika diikuti oleh transaksi dengan penjual yang ditandai sebagai mencurigakan (**suspicious\_seller\_flag**).

Langkah penting berikutnya adalah mengevaluasi apakah pola ini meningkat setelah kampanye promosi tertentu, seperti diskon besar atau event spesial. Hal ini karena pelaku penipuan sering memanfaatkan momen dengan aktivitas tinggi untuk menyembunyikan aktivitas mereka. Dengan memahami korelasi ini, dapat diterapkan langkah pencegahan, seperti pemantauan lebih ketat selama periode promosi atau penerapan sistem peringatan untuk aktivitas yang tidak biasa.

## **6.3 Insight Generation**

Menganalisis perilaku kelompok untuk mendeteksi pola penipuan yang berulang. Misalnya, apakah ada lonjakan penipuan setelah kampanye promosi?



Dari grafik terlihat bahwa promosi tertentu cenderung meningkatkan jumlah kasus penipuan, meskipun kasus terbanyak justru terjadi tanpa promosi. Lonjakan penipuan signifikan terlihat pada promosi seperti *promotion-219036467* dan *promotion-674347497*, mengindikasikan adanya pola penyalahgunaan promosi oleh pelaku. Di sisi lain, metode pembayaran seperti *MITRA\_PEMBAYARAN\_DIGITAL* dan *CREDIT\_CARD* lebih sering digunakan dalam transaksi penipuan dibandingkan metode lain seperti QRIS atau bank transfer.

Insight ini menunjukkan bahwa promosi dapat menjadi celah bagi pelaku penipuan, sementara tingginya kasus tanpa promosi mengindikasikan perlunya peningkatan keamanan secara sistemik. Rekomendasi utamanya adalah memperkuat pengamanan pada promosi yang rentan dan metode pembayaran populer, sekaligus menerapkan langkah pencegahan lebih ketat untuk transaksi reguler.

# **BAB VII**

# **[Visualization] Tableau for Fraud Monitoring and Dashboard Creation**

## **7.1 Interactive Fraud Detection Dashboards**

Dashboard Deteksi Fraud Interaktif memungkinkan tim pemantauan untuk memvisualisasikan dan menganalisis transaksi mencurigakan secara real-time. Dengan tampilan yang interaktif, tim dapat menyesuaikan filter dan parameter untuk menggali data lebih dalam, seperti periode waktu, jenis transaksi, atau status fraud. Dashboard ini memberikan kemudahan dalam memantau pola transaksi dan mengidentifikasi potensi fraud secara cepat. Berikut dashboard yang dapat ditampilkan berdasarkan hasil analisis:

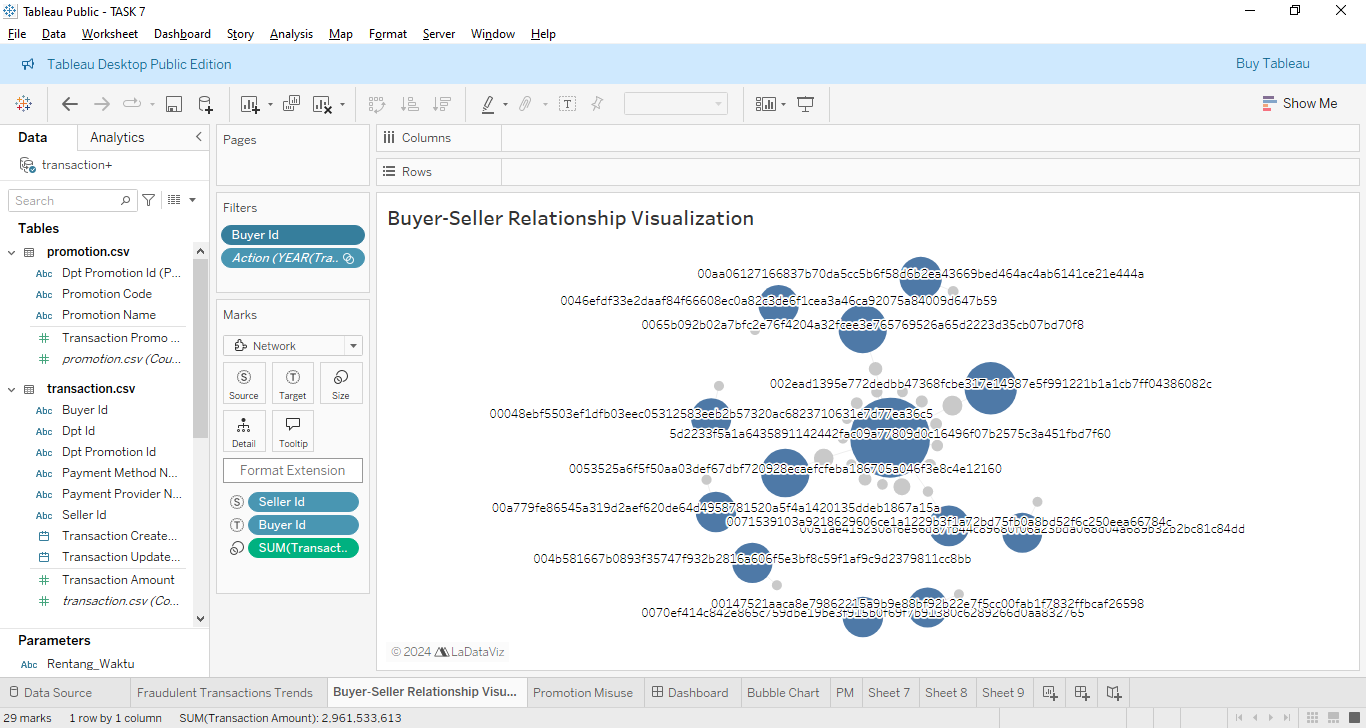
### **7.1.1 Tableau Dashboards**

#### **7.1.1.1Tren Transaksi Penipuan**

Dari visualisasi tren transaksi penipuan, dapat terlihat bahwa kejadian penipuan cenderung meningkat pada periode tertentu, khususnya pada bulan Juli 2023 yang menunjukkan jumlah transaksi penipuan tertinggi. Selain itu, fenomena penipuan juga lebih sering terjadi pada waktu-waktu tertentu, seperti pada saat akhir bulan. Grafik yang disajikan menunjukkan adanya lonjakan jumlah transaksi yang mencurigakan, terutama pada waktu-waktu tersebut.

Pada visualisasi ini, warna biru pada grafik menunjukkan jumlah transaksi yang tinggi, sementara warna merah menunjukkan nilai transaksi yang lebih rendah. Perbedaan warna ini memudahkan identifikasi waktu dan jumlah transaksi yang mencurigakan, sehingga tim pemantau penipuan dapat dengan cepat mengetahui tren penipuan yang terjadi. Tanda-tanda transaksi mencurigakan terdeteksi di transaksi dengan status penipuan, memberikan indikasi bahwa ada pola yang perlu diwaspadai.

#### **7.1.1. 2 Visualisasi Hubungan Pembeli-Penjual**

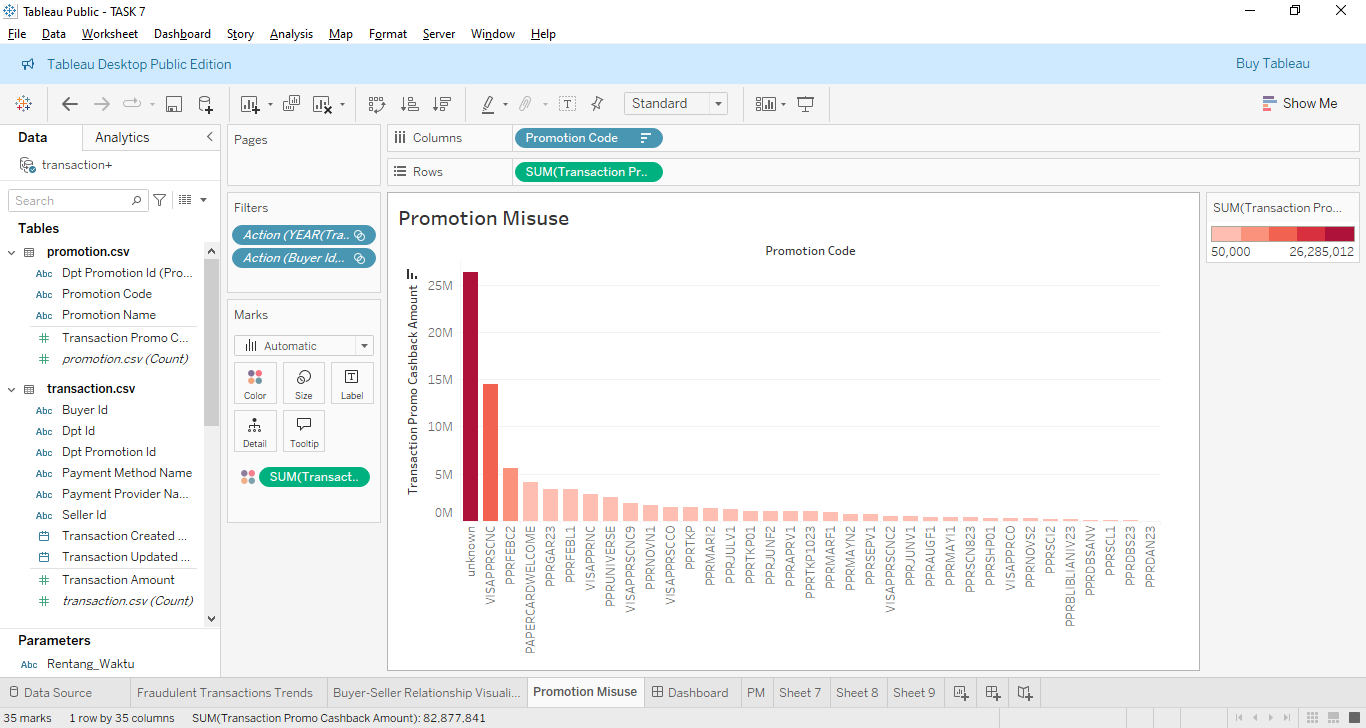


Dalam visualisasi Hubungan Pembeli-Penjual, terdapat dua pola yang perlu dicermati. Pertama, terlihat adanya hubungan yang sangat intens antara satu penjual dan satu pembeli, yang bisa menjadi indikasi adanya potensi penipuan. Kedua, ditemukan pula pola hubungan antara satu penjual dan banyak pembeli, yang menunjukkan adanya kemungkinan transaksi yang lebih luas, namun tetap patut dicurigai.

Dengan demikian, transaksi yang terjadi antara pembeli dan penjual dengan pola hubungan yang sangat sering, terutama jika dihubungkan dengan status penipuan yang terdeteksi, menunjukkan adanya potensi fraud yang perlu ditindaklanjuti.

#### **7.1.1.3 Penyalahgunaan Promosi**

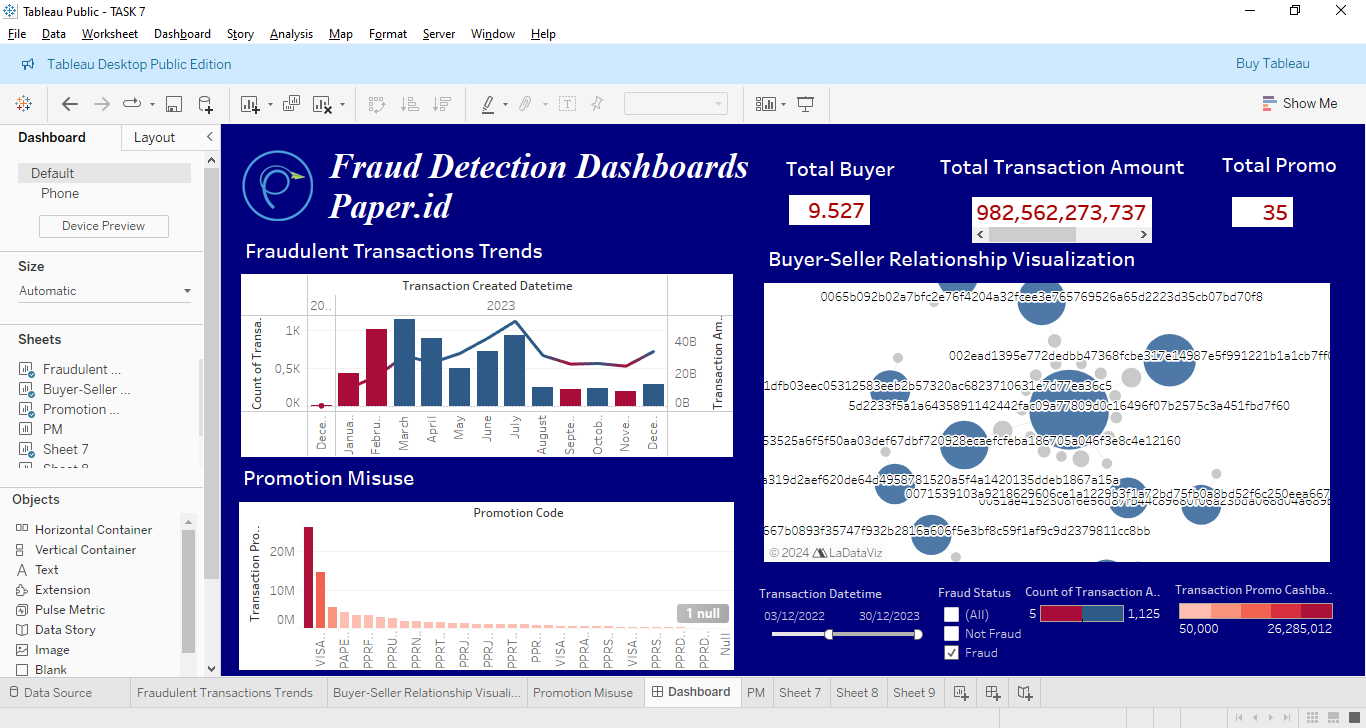
Visualisasikan bagaimana promosi dieksploitasi oleh pengguna yang melakukan penipuan.



Visualisasi yang telah dibuat menunjukkan hubungan antara kode promosi dan jumlah transaksi serta total cashback yang diterima. Dalam konteks eksploitasi promosi, terlihat bahwa kode promosi tertentu memiliki transaksi dengan cashback yang sangat besar, yang dapat menjadi indikasi adanya penyalahgunaan. Kode promosi dengan jumlah cashback tinggi, seperti yang tercatat pada kategori tertentu, mungkin menunjukkan bahwa promosi tersebut dieksploitasi secara berlebihan oleh pengguna untuk memaksimalkan keuntungan mereka.

Selain itu, status "Unknown" pada beberapa transaksi juga dapat menjadi sinyal adanya eksploitasi, karena transaksi dengan status tersebut mungkin tidak tercatat atau teridentifikasi dengan jelas. Hal ini dapat menunjukkan bahwa promosi digunakan secara tidak transparan atau bahkan oleh akun yang mencurigakan. Dengan demikian, visualisasi ini memberikan gambaran awal tentang potensi penyalahgunaan promosi yang perlu ditindaklanjuti lebih lanjut.

## **7.2 Dynamic Filtering and Drill-Downs**

Dengan mengaktifkan filter dinamis yang memungkinkan tim untuk menyaring data berdasarkan berbagai parameter, seperti periode waktu, hubungan pengguna, dan status flag fraud. Adapun hasilnya sebagai berikut:

Dasbor Deteksi Penipuan yang dibuat di Tableau memberikan gambaran mendalam mengenai tren transaksi penipuan, hubungan antara pembeli dan penjual yang mencurigakan, serta penyalahgunaan promosi dalam transaksi digital. Visualisasi pertama menampilkan tren transaksi penipuan berdasarkan waktu, dengan analisis terhadap jumlah transaksi dan total nilai transaksi yang menunjukkan periode dengan volume penipuan yang lebih tinggi. Ini memberikan wawasan tentang titik rawan penipuan yang terjadi pada bulan-bulan tertentu, yang dapat digunakan untuk memperketat pengawasan pada periode tersebut.

Selanjutnya, visualisasi hubungan pembeli-penjual menunjukkan jaringan transaksi yang mencurigakan melalui analisis hubungan antar pengguna. Pembeli dan penjual yang terlibat dalam transaksi penipuan dapat dikenali dengan lebih mudah, memberikan tim pemantauan penipuan alat yang efektif untuk mengidentifikasi pola dan anomali. Terakhir, visualisasi penyalahgunaan promosi menunjukkan bagaimana kode promosi digunakan berlebihan oleh pengguna yang berpotensi melakukan penipuan, dengan total cashback yang tidak wajar. Dasbor ini juga dilengkapi dengan pemfilteran dinamis, memungkinkan tim untuk menelusuri dan melakukan drill-down pada data sesuai dengan periode waktu, status penipuan, dan hubungan pengguna, memberikan analisis yang lebih tajam dan responsif terhadap ancaman penipuan yang muncul.

# **BAB VIII**

# **Insights and Recommendations**

## 8.1 Key Fraud Insights

Gunakan dasbor untuk mengidentifikasi:

* + - Pendorong penipuan teratas dan pengguna penipuan yang paling terhubung.

Dasbor ini memberikan wawasan penting mengenai pendorong penipuan, termasuk kode promosi yang dieksploitasi secara berulang dan transaksi yang dilakukan selama periode tertentu dengan angka penipuan yang tinggi. Kode promosi yang teridentifikasi sebagai sering digunakan oleh pengguna yang melakukan penipuan menyoroti area yang perlu diperhatikan oleh tim pemantauan. Selain itu, dengan menggunakan visualisasi hubungan pembeli-penjual, kita dapat mengidentifikasi pengguna yang memiliki keterhubungan tinggi dengan banyak pihak, baik sebagai pembeli atau penjual. Mereka ini lebih cenderung terlibat dalam jaringan penipuan yang lebih luas. Pengguna-pengguna ini, dengan berbagai keterhubungan, menjadi titik fokus utama dalam upaya pencegahan lebih lanjut.

* + - Hubungan pembeli-penjual utama yang terlibat dalam penipuan.

Hubungan pembeli-penjual yang terlibat dalam penipuan dapat dengan jelas diidentifikasi melalui visualisasi jaringan yang ada. Dalam visualisasi ini, pembeli dan penjual yang terhubung satu sama lain dengan banyak transaksi mencurigakan menonjol sebagai titik-titik utama dalam jaringan penipuan. Pembeli yang terhubung dengan banyak penjual atau sebaliknya, menunjukkan adanya pola transaksi yang mencurigakan yang melibatkan lebih dari satu pihak. Analisis ini memberikan gambaran yang lebih jelas tentang siapa yang berperan aktif dalam transaksi yang tidak sah. Mengidentifikasi hubungan ini memungkinkan untuk melakukan penyaringan yang lebih mendalam terhadap pembeli dan penjual yang terlibat, serta memfokuskan upaya pencegahan penipuan terhadap transaksi yang dilakukan oleh aktor-aktor ini. Seiring berjalannya waktu, memonitor hubungan ini dapat membantu mengurangi dampak penipuan yang lebih luas dan mencegah pelaku penipuan yang mungkin berulang.

## **8.2 Action Plans**:

Memberikan rekomendasi untuk pencegahan penipuan berdasarkan temuan:

* + Terapkan proses verifikasi pengguna yang lebih ketat.

Sebagai langkah pencegahan penipuan yang lebih efektif, sangat penting untuk memperketat proses verifikasi pengguna, terutama pada tahap pendaftaran dan transaksi pertama. Implementasi prosedur verifikasi identitas melalui multi-faktor otentikasi atau verifikasi dokumen dapat membantu memastikan bahwa pengguna yang terlibat dalam transaksi benar-benar sah. Proses ini bisa melibatkan verifikasi melalui email, SMS, atau aplikasi otentikasi untuk meminimalkan risiko pendaftaran akun palsu yang dapat digunakan untuk penipuan. Dengan melakukan verifikasi yang lebih ketat, platform dapat meminimalisir pelaku penipuan yang mencoba masuk ke dalam sistem dengan identitas palsu atau menggunakan informasi yang tidak valid.

Selain itu, proses verifikasi yang lebih ketat dapat diimplementasikan pada pengguna yang terlibat dalam transaksi dengan volume tinggi atau transaksi yang mencurigakan. Pengguna yang melakukan transaksi besar atau yang melibatkan kode promosi yang sering dieksploitasi dapat diminta untuk memberikan informasi tambahan sebagai langkah pencegahan. Ini akan membantu mengurangi potensi kerugian yang timbul dari pengguna yang berniat buruk. Dengan pengawasan yang lebih ketat pada transaksi yang mencurigakan dan pengguna baru, platform dapat lebih proaktif dalam mendeteksi dan mencegah penipuan sebelum terjadi, meningkatkan keamanan dan kepercayaan pengguna secara keseluruhan.

* + Pantau dan tandai aktivitas pembeli-penjual yang mencurigakan secara real-time.

Untuk menangani penipuan dengan lebih cepat dan efektif, sistem pemantauan real-time harus diterapkan untuk memonitor aktivitas pembeli dan penjual. Dengan menggunakan teknologi analisis data yang canggih, seperti machine learning dan algoritma prediktif, platform dapat mendeteksi perilaku mencurigakan yang mungkin terlewatkan oleh pengawasan manual. Aktivitas mencurigakan, seperti pembelian berulang dengan kode promo yang sama atau transaksi yang tidak wajar dalam periode tertentu, dapat segera ditandai dan diteruskan untuk pemeriksaan lebih lanjut. Hal ini tidak hanya membantu dalam mendeteksi penipuan lebih awal tetapi juga dapat mencegah kerugian finansial yang lebih besar dengan menghentikan transaksi yang mencurigakan sebelum diproses lebih lanjut.

Pemantauan secara real-time juga memungkinkan untuk mendeteksi jaringan hubungan pembeli dan penjual yang terlibat dalam penipuan. Pembeli atau penjual yang memiliki banyak transaksi dengan pengguna lain yang tercatat sebagai penipu dapat segera diidentifikasi dan diberi peringatan atau bahkan diblokir. Dengan pendekatan ini, sistem akan lebih siap dalam menangani aktivitas mencurigakan, memungkinkan tim pemantauan untuk bertindak dengan lebih cepat dan meminimalkan dampak penipuan. Pemantauan real-time yang tepat dapat mengurangi keterlambatan dalam deteksi dan mempercepat langkah-langkah pencegahan yang diperlukan.

* + Tetapkan kebijakan promosi yang lebih ketat untuk mencegah penyalahgunaan.

Kebijakan promosi yang lebih ketat sangat penting untuk mencegah penyalahgunaan promosi yang sering kali menjadi pendorong utama penipuan. Salah satu langkah yang dapat diambil adalah dengan membatasi penggunaan kode promosi oleh pengguna yang terdeteksi sering menggunakannya dalam jumlah yang tidak wajar. Misalnya, mengatur pembatasan seperti jumlah maksimum penggunaan kode per pengguna atau periode waktu tertentu dapat mengurangi potensi penyalahgunaan. Selain itu, platform dapat mengenakan kebijakan untuk mengkonfirmasi apakah transaksi dengan kode promosi melibatkan transaksi yang sah atau hanya digunakan untuk memanfaatkan sistem secara tidak adil. Kebijakan promosi yang lebih selektif dan terkontrol dapat membantu meminimalkan celah yang dapat dimanfaatkan oleh penipu.

Selain membatasi jumlah penggunaan, kebijakan promosi yang lebih ketat juga dapat melibatkan analisis perilaku pengguna yang menerima promosi. Promosi yang diberikan kepada pengguna dengan riwayat transaksi mencurigakan atau yang terhubung dengan pembeli atau penjual yang terlibat dalam penipuan dapat dibatasi atau dibatalkan untuk menghindari potensi kerugian lebih lanjut. Langkah-langkah preventif ini akan lebih memperkecil ruang bagi pengguna yang berniat menyalahgunakan promosi untuk melakukan penipuan, sekaligus menjaga kredibilitas dan integritas platform. Dengan kebijakan yang lebih ketat, perusahaan dapat menciptakan lingkungan transaksi yang lebih adil bagi pengguna yang sah.

# **BAB IX**

# **Kesimpulan dan Saran**

## **A. Kesimpulan**

Perusahaan fintech yang memproses jutaan transaksi digital menghadapi tantangan besar akibat meningkatnya kasus penipuan yang merugikan baik secara finansial maupun kepercayaan pelanggan. Berdasarkan analisis data, pola penipuan menunjukkan bahwa pelaku sering memanfaatkan kode promosi secara berlebihan dan memiliki jaringan hubungan antara pembeli dan penjual yang saling terkait. Penggunaan SQL tingkat lanjut memungkinkan penggalian data mendalam untuk mengidentifikasi perilaku yang tidak normal, seperti lonjakan transaksi dalam waktu singkat dan pola penggunaan promosi yang tidak wajar. Analisis mendalam ini menegaskan bahwa kurangnya pengawasan real-time dan kebijakan promosi yang lemah berkontribusi pada peningkatan kasus penipuan. Temuan ini memberikan gambaran penting bagi perusahaan untuk mengidentifikasi area kelemahan dalam sistem dan meningkatkan mekanisme keamanan mereka.

## **B. Saran**

Untuk mengurangi risiko penipuan, perusahaan perlu menerapkan langkah-langkah strategis berbasis data. Pertama, adopsi algoritma machine learning untuk mendeteksi anomali transaksi secara real-time dapat membantu dalam mengidentifikasi pola mencurigakan sejak dini. Kedua, memperketat proses verifikasi pengguna dengan menggunakan autentikasi multi-faktor akan mengurangi peluang pendaftaran akun palsu. Selanjutnya, kebijakan promosi perlu diperbaiki dengan membatasi penggunaan kode promo pada perilaku tertentu, misalnya dengan menerapkan deteksi pengguna yang sering menggandakan akun untuk memanfaatkan promosi. Dengan implementasi langkah-langkah ini, perusahaan dapat mengurangi kerugian akibat penipuan dan memulihkan kepercayaan pelanggan terhadap platform.

# **LAMPIRAN**

## **Online Diagram BPMN:**

Business Process BPMN :  
<https://drive.google.com/file/d/1ystxfq1UZwjYS9Dha4MCD6jh-zHReCMZ/view>

## **Python Code**

2.1 Data Cleaning : <https://colab.research.google.com/drive/15d9THxmybdyF7BAVqykP9Uf-1J92l71c?usp=drive_link>

2.2 Featuring Engineering : <https://colab.research.google.com/drive/1DILx9sPzDnBP_t3ojtTNEioDySU1izP5?usp=drive_link>

2.3 Scaling and Normalization : <https://colab.research.google.com/drive/1MYCN3P0NdeOlhVvNYWbTaywE7wsjja-k?usp=drive_link>

3.1 Exploratory Data Analysis : <https://colab.research.google.com/drive/1tsI1EAvnbMZs3dmw7wuptRoQAXFfmf17?usp=drive_link>   
3.2 Visualization : <https://colab.research.google.com/drive/1Tc77-V_2DSuYzSJ8PKAcZcFmH2t0FUp7?usp=drive_link>

6.1.2 Analisis Hubungan Pembeli-Penjual untuk Mengidentifikasi Pola Interaksi dan Terlibat Fraud (Visualization using Phyton) :

<https://colab.research.google.com/drive/1Uzx7Z1xqsoTPQT7VycDPe9Fa6ygH8C-0?usp=drive_link>

6.2 Cohort Analysis :

<https://colab.research.google.com/drive/10HCHdiao3NzFGjORCnhfgIND3h3Y0RyX#scrollTo=f6beTAKDRlyF>

## **Notulen Canva (PPT) dan Tableau**

Tableau: <https://public.tableau.com/app/profile/syahirotul.maulidiyah/viz/FraudDetectionDashboardsPaperID_17336703277340/Dashboard?publish=yes>

Notulen Canva PPT : <https://www.canva.com/design/DAGYt5mxznk/b-MFeZreZmkBF3XzOPcI_g/edit?utm_content=DAGYt5mxznk&utm_campaign=designshare&utm_medium=link2&utm_source=sharebutton>

## **Recording**

Link Youtube Recording :

<https://youtu.be/ggbildpcUFc?si=50XWObYJFlkdS6oH>