# BAB III
# PELAKSANAAN KERJA PROFESI

## 3.1 Bidang Kerja

Pada pelaksanaan kegiatan Kerja Profesi, Praktikan bekerja sebagai *Data Analyst* dalam divisi produk. Praktikan dipercayakan untuk melaksanakan tugas yang berkaitan dengan analisis data. Untuk mendukung kelancaran pengerjaan tugas, tools yang digunakan seperti *Google Colab* dengan bahasa pemrograman *Python, Metabase, BigQuery, Google Sheets,* dan *PowerPoint* untuk memproses, menganalisis, dan memvisualisasikan data. Selain itu, praktikan juga bertanggung jawab untuk menyajikan data dalam format yang mudah dipahami dan digunakan oleh tim *marketing*, serta memberikan wawasan yang mendukung keputusan strategis perusahaan dalam meningkatkan penjualan. Dalam melakukan analisis dan dokumentasi, praktikan melakukan beberapa pekerjaan sebagai berikut :

1. *User Behaviour Analysis* Menggunakan BigQuery
   Praktikan melakukan analisis perilaku pengguna melalui BigQuery, yang memungkinkan untuk mengumpulkan, memproses, dan menganalisis data pengguna secara efisien. Data yang dianalisis mencakup pola perilaku pengguna pada produk Reksa dana untuk mengggali penyebab utama dari lonjakan *redeem* reksa dana di bulan April.

2. *Predictive Modelling* dengan Python untuk Prediksi Pengguna yang Berpotensi Membeli Produk
   Praktikan menggunakan bahasa pemrograman Python dalam proyek *machine learning* untuk memprediksi pengguna yang berpotensi untuk membeli salah satu produk Bareksa yakni emas, berdasarkan pola perilaku mereka yang dianalisis sebelumnya. Hal ini bertujuan

untuk memberikan informasi yang lebih mendalam tentang peluang konversi dan memungkinkan pengembangan strategi pemasaran yang lebih efektif.

3. *Data Reporting* melalui Google Sheets dan PowerPoint Praktikan bertanggung jawab untuk menyajikan hasil analisis dalam format yang mudah dipahami oleh tim produk dan manajemen perusahaan. Hasil analisis disajikan melalui *Google Sheets* dan *PowerPoint* untuk memberikan gambaran yang jelas dan terstruktur tentang data dan prediksi yang telah dibuat.
4. Visualisasi Data untuk Memberikan Wawasan yang Mudah Dipahami Praktikan menggunakan *tools* seperti Metabase untuk memvisualisasikan data dan membuat grafik atau diagram yang memudahkan pemahaman dan pemantauan mengenai perilaku pengguna dan produk.

## 3.2 Pelaksanaan Kerja

Kegiatan berbentuk praktik dalam melaksanakan Kerja Profesi dimulai pada 10 Februari 2025 hingga 10 Mei 2025. Kerja Profesi ini diposisikan pada Divisi Produk sebagai Data Analisis untuk menganalisis dan memvisualisasikan data yang berkaitan dengan produk yang diperjual belikan di Bareksa. Selama masa kerja, praktikan sudah melakukan beberapa kegiatan dimulai dari pengenalan lingkungan kerja, *product knowledge*, berdiskusi dengan tim, melakukan pelatihan penggunaan tools pada perusahaan yaitu *BigQuery* dan *Metabase*, sampai pada tahap visualisasi data dan pelaporan melalui *google sheets*.

### 3.2.1 *Data and Research Sync*

*Data and Research Sync* merupakan pertemuan rutin setiap minggu yang dihadiri oleh seluruh anggota tim data dari divisi produk yang bersifat *online*, dipimpin oleh Gerry Fernando selaku *Senior*

*Data and Research Manager* yang bertujuan untuk menyelaraskan serta memperbarui perkembangan individu anggota tim dalam interval waktu satu minggu. Terdapat dua divisi dalam tim, yaitu *data science* dan *data analyst*, dan setiap individu memberikan presentasi lisan yang mencakup hasil analisis, kesimpulan, dan informasi lain yang telah diselesaikan dan dikerjakan. Setiap laporan yang dipresentasikan bertujuan untuk merangkum pekerjaan yang telah dilakukan secara spesifik dan juga untuk menjawab pertanyaan apakah hasil yang diperoleh akurat dan relevan dengan pekerjaan yang telah dilakukan. Hal ini memungkinkan anggota tim untuk berdiskusi, saling mengajarkan, dan memperbaiki masalah yang mungkin muncul selama pengerjaan. Dengan demikian, pertemuan ini bertujuan untuk memastikan apakah semua anggota memiliki pemahaman yang sama tentang tujuan utama dari tim data divisi produk sehingga informasi yang diperoleh dapat digunakan untuk meningkatkan kualitas dan efektivitas pengembangan produk.

Pada bagian kedua dari pertemuan, pekerjaan dibagi kepada setiap anggota tim secara individu. Pembagian tugas ini dilakukan sesuai dengan pekerjaan yang harus diselesaikan pada minggu mendatang dan dikategorikan sebagai hal yang sangat penting, memerlukan perhatian segera, serta berdasarkan ketersediaan dan keterampilan masing-masing anggota tim. Dengan cara ini, setiap anggota diberikan tanggung jawab yang sesuai dengan kemampuannya dan dijamin dapat diselesaikan tepat waktu. Tanggung jawab ini juga mencakup penetapan tenggat waktu yang dapat dicapai untuk memastikan bahwa target yang ditetapkan dapat dipatuhi dalam kerangka waktu yang telah ditentukan. Selain itu, pertemuan ini juga menjadi wadah untuk menyelesaikan masalah yang perlu ditangani dalam kegiatan yang sedang berlangsung dan memungkinkan kesepakatan bersama mengenai prinsip dan metode prosedur yang terkait dengan pengolahan data. Hal ini meningkatkan kesatuan di antara anggota tim dan lebih jauh lagi memastikan

bahwa hasil yang dihasilkan akurat dan memenuhi persyaratan pengembangan produk.

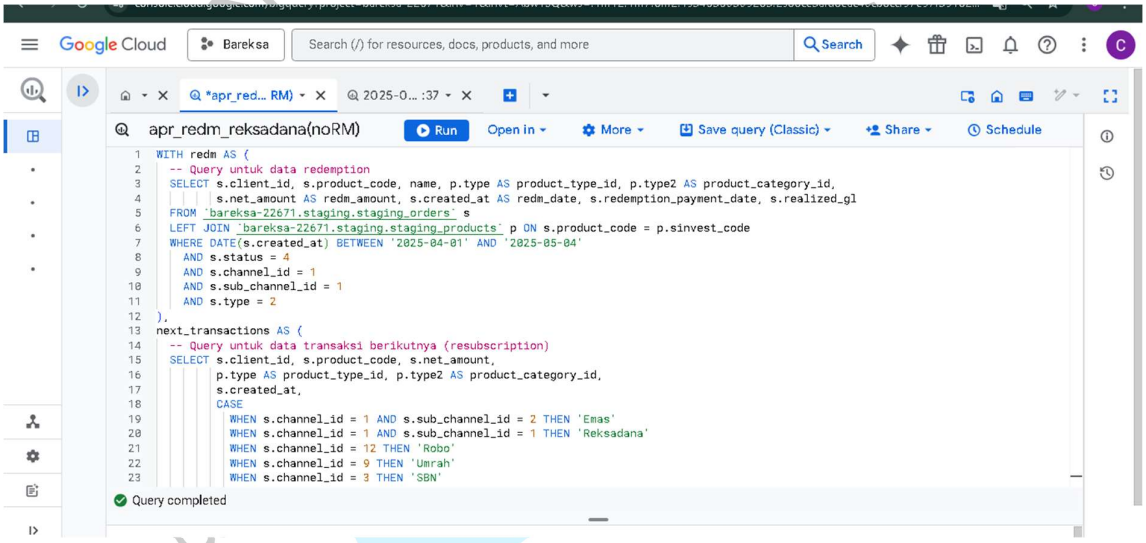
### 3.2.2 *User Behaviour Analysis*

Dalam menjalani kegiatan kerja profesi, Praktikan diberikan tugas untuk melakukan studi komprehensif tentang data pengguna yang terkait dengan salah satu produk Bareksa, khususnya produk reksa dana. Tujuan dari tugas ini adalah untuk memahami fenomena yang terjadi pada bulan April di mana terdapat lonjakan penjualan (*redemption*) pada bulan tersebut. *Redemption* adalah ketika seorang pengguna mencairkan dana dari produk investasi mereka, yang biasanya menunjukkan pergeseran strategi atau reaksi terhadap kondisi pasar. Praktikan juga diharuskan untuk menindaklanjuti pengguna yang kembali membeli produk (*resubscription*) setelah *redemption* untuk mencoba menentukan apakah ada pola atau tren tertentu yang mendorong keputusan pengguna untuk investasi kembali setelah menarik dana mereka. Karakteristik pengguna yang ingin dicari oleh Praktikan melalui analisis ini adalah pengguna yang kemungkinan besar melakukan *redemption* dan *subscription*, yang bersama dengan faktor-faktor lain akan dianalisis berkaitan dengan keputusan mereka. Selain itu, praktikan juga menghubungkan temuan ini untuk memberikan wawasan yang lebih lengkap, yang nantinya akan disajikan dalam laporan yang dapat digunakan oleh tim terkait, terutama tim marketing, untuk merancang strategi pemasaran dan peningkatan layanan produk yang lebih efektif.

#### *1. Data Extraction and Processing*

Untuk analisis ini, data dimulai dengan pengambilan dari *database* Bareksa menggunakan alat *query* seperti Google *BigQuery*. *Query* pertama bertujuan mendapatkan data *redemption* dengan menangkap bidang-bidang penting seperti ID pengguna,

kode produk, jumlah penjualan (*redemption*), nama produk, dan tanggal transaksi untuk pengguna yang menebus reksa dana mereka antara 01 April dan 4 Mei 2025. Seperti dalam kebanyakan skenario bisnis, hanya transaksi *redemption* yang valid yang dimasukkan sehingga *filter* tertentu pada status dan saluran diterapkan. Setelah itu, *query* lain dijalankan untuk mencakup data pembelian ulang(*resubs*) dan fokus beralih ke pengguna yang melakukan pembelian ulang ke produk setelah *redeem*. Berikut contoh query ekstraksi data pada **Gambar 3.1**, **Gambar 3.2** dan **Gambar 3.3**.

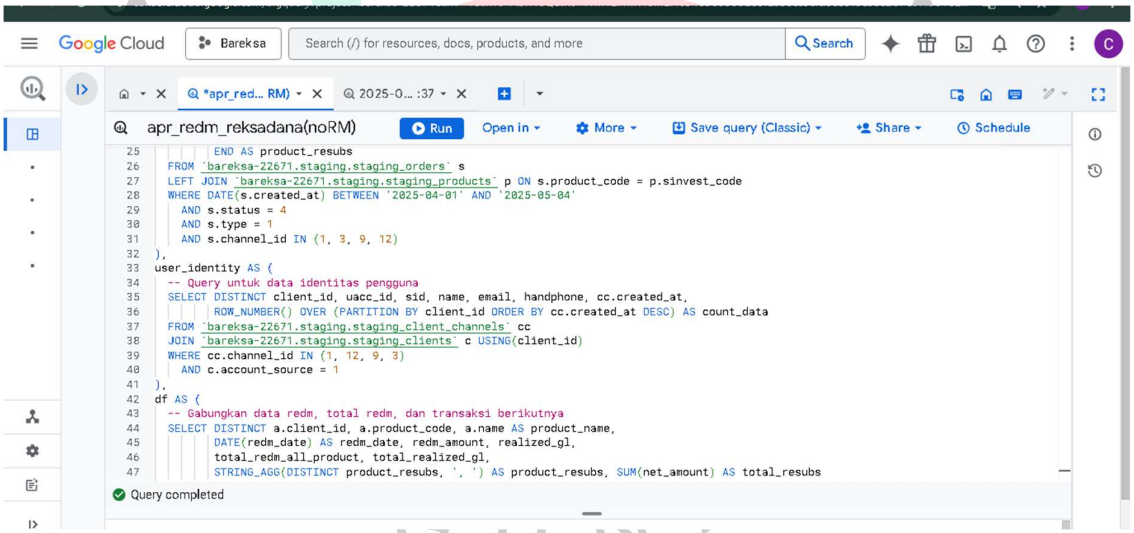



**Gambar 3.1 Proses Ekstraksi Data di *Bigquery* (a)**
Sumber : Dokumentasi yang Dihasilkan oleh Praktikan

Bagian pertama *query* ini menggunakan *Common Table Expression* (CTE) yang disebut redm, data yang ditransformasi tanggal 4 April hingga 4 Mei diambil dari tahap *filtering* dari staging_orders yang digabungkan dengan *staging_products*. Proses penggabungan ini dilakukan dengan *cumulate status* = 4 yang berarti hanya transaksi yang berhasil yang akan dipilih. Selain itu, *filter channel_id* = 1 dan *sub_channel_id* = 1 digunakan untuk mengklasifikasikan pengguna pada produk reksa dana.

Penggabungan kolom *product_code* dari *staging_orders* dan sinvest_code dari *staging_products* berfungsi untuk mendeklarasikan kode produk yang akan didefinisikan oleh nama produk yang sesuai.

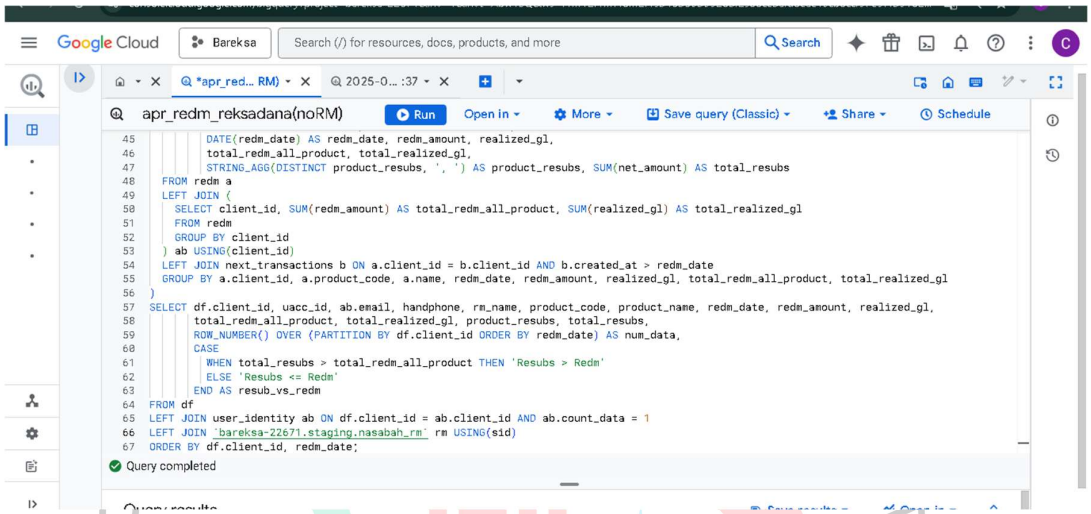
Bagian kedua dari *query* ini adalah CTE *next_transaction*, yang menangkap transaksi berikutnya setelah pengguna melakukan *redeem*, yang disebut sebagai pembelian ulang (*resubs*). Data yang diambil mencakup id pengguna, kode produk, jumlah transaksi, tipe produk, kategori produk, tanggal transaksi, dan produk yang dibeli kembali. Selain itu, klausa *CASE* dibuat untuk mendefinisikan jenis produk berdasarkan *channel_id* dan *sub_channel_id*. CTE ini hanya mencakup transaksi dengan status 4 dan tipe 1 (pembelian ulang) selama periode yang sama seperti sebelumnya.




**Gambar 3.2 Proses Ekstraksi Data di *Bigquery* (b)**
Sumber : Dokumentasi yang Dihasilkan oleh Praktikan

CTE ketiga, user_identity, berfungsi untuk mengambil informasi pengguna yang relevan dengan transaksi. Data yang diambil mencakup *client_id*, uacc_id, sid, nama pengguna, email, nomor telepon, dan tanggal pembuatan akun (terkait dengan saluran

spesifik). *ROW_NUMBER*() digunakan untuk memberikan nomor urut pada setiap client_id berdasarkan urutan tanggal pembuatan akun secara menurun, yang menjamin bahwa hanya satu entri per pengguna yang akan dipilih.

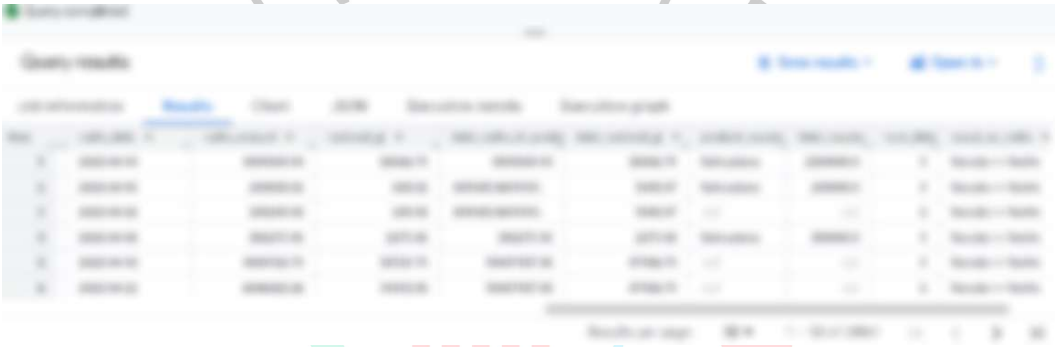



**Gambar 3.3 Output Ekstraksi Data di Bigquery (a)**
Sumber : Dokumentasi yang Dihasilkan oleh Praktikan

CTE df adalah langkah integrasi yang menghubungkan informasi dari transaksi penukaran (*redm*), total penukaran di setiap produk (total_redm_all_product), bersama dengan data transaksi pembelian ulang berikutnya. Di sini, data penukaran digabungkan dengan total penukaran dan informasi transaksi untuk transaksi selanjutnya. Data produk yang terlibat dalam pembelian ulang digabungkan menjadi satu kolom melalui penggunaan *STRING_AGG* untuk menyederhanakan analisis. Selain itu, jumlah total pembelian ulang (*total_resubs*) per pengguna dihitung.

Setelah menggabungkan data di CTE df, bagian akhir dari *query* ini mengambil informasi lebih rinci seperti id pengguna, identifikasi pengguna, nama RM, produk, tanggal penukaran, jumlah penukaran, dan total pembelian ulang. Hasil keluaran dari *query* ini

juga menambahkan kolom baru *resub_vs_redm* yang menghitung perbandingan antara pembelian ulang yang terintegrasi dengan jumlah penukaran yang terintegrasi per pengguna, mencatat apakah transaksi pembelian ulang melebihi penukaran atau tidak. Untuk setiap pengguna, *ROW_NUMBER*() diterapkan untuk mengurutkan secara berurutan berdasarkan tanggal pembelanjaan terbaik. Hasilnya diurutkan dengan *client_id* dan tanggal penukaran. Berikut contoh hasil dari *query* yang dilakukan terdapat pada **Gambar 3.4**.



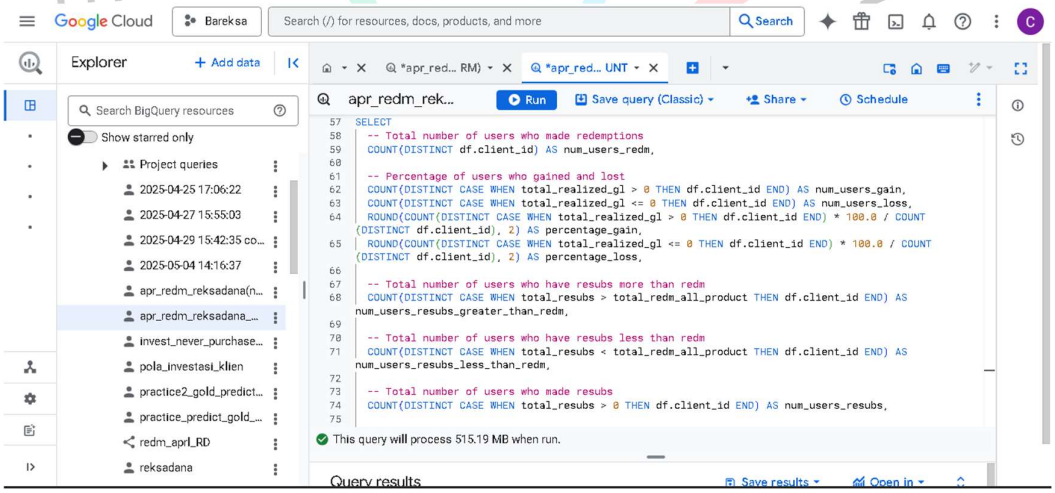
**Gambar 3.4 *Output* Ekstraksi Data di *Bigquery* (b)**
Sumber : Dokumentasi yang DIhasilkan oleh Praktikan

Output yang dihasilkan oleh *query* ini mencakup detail relevan mengenai pengguna yang melakukan pembelian investasi baru pada produk-produk seperti reksa dana, emas, dan produk keuangan lainnya setelah melakukan *redeem*. Selain itu, *query* juga dijalankan untuk mengidentifikasi pengguna dengan tujuan menangkap informasi yang diperlukan, memastikan terjadinya keterkaitan yang tepat antara *redeem events* dan *resubscription events*, serta melengkapi pengguna dengan informasi seperti nama, email, no. telepon, dan status akunnya. Dataset yang diekstrak kemudian digabung menjadi satupadu berdasarkan atribut yang relevan untuk menggambarkan secara detail aktivitas *redeem* and *resubscription* dari pengguna. Hal ini memungkinkan analisis tren

yang lebih mendalam dan memberikan wawasan lebih jauh mengenai perilaku transaksi dan pola investasi pengguna

### 2. *Trend Analysis*

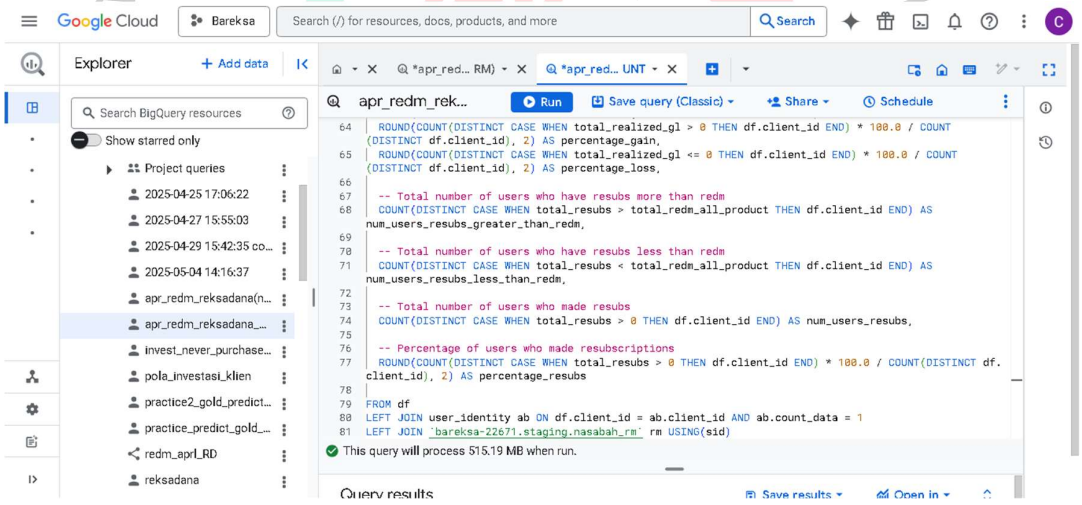
Setelah mengumpulkan data relevan dari berbagai CTE seperti redm untuk transaksi penjualan (*redemption)*, next_transactions untuk transaksi pembelian ulang, dan user_identity untuk identitas pengguna, data tersebut diintegrasikan ke dalam df CTE untuk pemeriksaan lebih lanjut terkait interaksi antara *redemption* dan *resubscription* secara terperinci. Pada tahap *query* ini, beberapa metrik tentang perilaku pengguna dibuat. Salah satu metrik tersebut berkaitan dengan penghitungan pengguna yang melakukan *redemption* dan menghitung persentase pengguna yang memperoleh atau kehilangan berdasarkan nilai realize_gl. Berikut contoh *query trend analysis* yang terdapat pada **Gambar 3.5** dan **Gambar 3.6**.




**Gambar 3.5 *Trend Analysis* pada *BigQuery* (a)**
Sumber : Dokumentasi yang Dihasilkan oleh Praktikan

Metrik-metrik penting dihimpun dan dihitung untuk memperoleh wawasan terkait perilaku pengguna, khususnya dalam

kaitannya dengan transaksi *redemption* dan *resubscription*. Pertama-tama, *query* ini menghitung jumlah pengguna yang melakukan *redemption*. Untuk keperluan ini, fungsi *COUNT*(*DISTINCT* df_*user_level.client_id*) digunakan untuk mendapatkan jumlah unik pengguna yang melakukan transaksi *redemption* sehingga tidak ada duplikasi pengguna pada kalkulasi. Selanjutnya, dilakukan analisis terhadap pengguna yang mengalami keuntungan atau kerugian berdasarkan nilai realized_gl. Pengguna dengan nilai realized_gl positif dianggap memperoleh keuntungan, sedangkan mereka yang memiliki nilai realized_gl negatif atau nol dianggap mengalami kerugian. Untuk memberikan gambaran yang lebih menyeluruh, proporsi pengguna yang mengalami keuntungan atau kerugian juga dihitung dengan membagi jumlah pengguna dalam masing-masing kategori terhadap total pengguna *redemption*, kemudian dikalikan dengan 100.




**Gambar 3.6 *Trend Analysis* pada *BigQuery* (b)**
Sumber : Dokumentasi yang Dihasilkan oleh Praktikan

*Query* ini juga menganalisis hubungan antara transaksi *resubscription* dan *redemption* di antara para pengguna. Analisis ini dilakukan untuk mengetahui sejauh mana transaksi *resubscription* pengguna melebihi, kurang dari, atau sama dengan jumlah

*redemption* yang mereka lakukan. Ini dicapai melalui pernyataan *CASE* di mana *total_resubs* dibandingkan dengan *total_redm_all_product*, dan pengguna yang memenuhi kondisi tersebut dihitung. Sebagai contoh, *num_users_resubs_greater_than_redm* menghitung pengguna yang melakukan *resubscription* lebih besar dari *redemption* mereka, *num_users_resubs_less_than_redm* menghitung pengguna yang melakukan *resubscription* lebih kecil dari jumlah *redemption*, dan *num_users_resubs_equal_redm* menghitung pengguna yang jumlah *resubscription*-nya sama dengan *redemption.* Dengan cara ini, *query* tidak hanya menyediakan angka terkait transaksi, tetapi juga menawarkan pemahaman yang lebih mendalam tentang hubungan pengguna dengan produk, yang dapat dimanfaatkan dalam strategi pemasaran dan manajemen produk di masa mendatang.

Rasio antara pengguna yang melakukan *resubscription* dan total pengguna yang melakukan *redemption* juga dihitung untuk memperkirakan proporsi pengguna *redemption* yang melanjutkan ke transaksi *resubscription*, terlepas dari apakah jumlahnya lebih besar, lebih kecil, atau sama. Tidak seperti kelompok sebelumnya, pengguna ini hanya dihitung jika mereka memiliki transaksi *resubscription* dalam bentuk apa pun. Nilai num_users_resubs mencerminkan metrik tersebut, dengan perhitungan persentase dilakukan dengan pendekatan yang sama. Berikut hasil *query trend analysis* pada **Gambar 3.7**

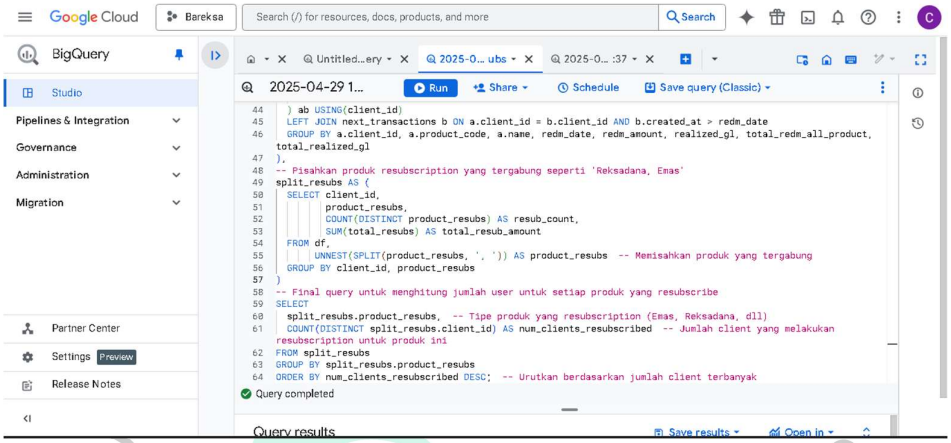

**Gambar 3.7 *Output Trend Analysis* pada *BigQuery* (a)**
Sumber : Dokumentasi yang Dihasilkan oleh Praktikan

Dengan menggabungkan semua metrik yang dijelaskan, kita memperoleh pemahaman yang lebih mendalam tentang interaksi pengguna terhadap aktivitas *redemption* dan *resubscription*, serta hubungan antara kedua aktivitas tersebut. Ini memberikan wawasan tentang perilaku pengguna dalam bertransaksi, khususnya apakah mereka cenderung melakukan *resubscription* setelah *redemption*, serta bagaimana nilai yang mereka peroleh dari *redemption* memengaruhi keputusan transaksi berikutnya.

Untuk mengidentifikasi produk yang paling banyak di*resubscription*, Praktikan mengakumulasi dari data pengguna yang melakukan *resubscription*. Isu utama yang menjadi fokus dalam *query* ini adalah pemisahan produk *resubscription* yang sebelumnya tergabung, seperti “Reksadana, Emas”, agar dapat diperlakukan sebagai entri terpisah dalam menghitung partisipasi pengguna.

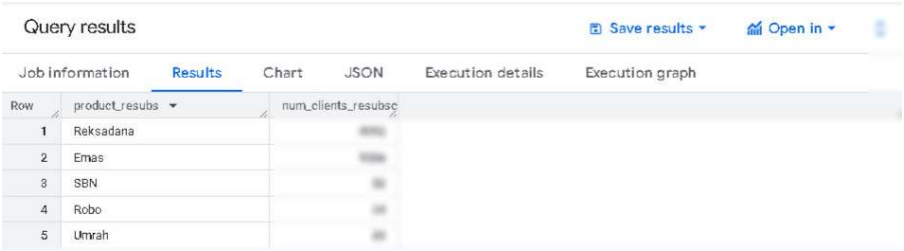
Pada bagian pertama, sebuah *Common Table Expression* (CTE) dengan alias split_resubs memproses data produk terkait dengan *resubscription*. Data produk, atau data *resubscription*, disimpan dalam tabel bernama df yang menyimpan catatan transaksi *resubscription* pengguna. Produk yang sebelumnya ditulis dalam satu kolom product_resubs sebagai “Reksadana, Emas” dan produk lainnya diubah menjadi entri terpisah menggunakan fungsi *SPLIT*(). Hasil dari pemisahan tersebut kemudian diproses lebih lanjut menggunakan fungsi *UNNEST*() yang menghasilkan setiap produk yang tergabung sebelumnya menjadi baris yang terpisah, sehingga setiap produk yang telah digabung kini ditampilkan sebagai entri yang terpisah berdasarkan *client_id*. Agregasi juga dilakukan untuk menentukan nilai untuk setiap produk *resubscription* yang terpisah. Jumlah produk yang terhitung disebut sebagai *resub_count* dan total transaksi *resubscription* untuk produk yang dihitung tersebut disebut sebagai *total_resub_amount*, keduanya dikelompokkan berdasarkan *client_id* dan produk *resubscription*, serta dihitung secara terpisah

untuk setiap produk dan menghitung jumlahnya. Berikut *query* pada **Gambar 3.8**.




**Gambar 3.8 *Query Trend Analysis* pada *BigQuery***
Sumber : Dokumentasi yang Dihasilkan oleh Praktikan

Bagian kedua dari *query* ini melakukan agregasi tambahan untuk menghitung jumlah pengguna unik (num_clients_resubscribed) yang melakukan resubscription untuk setiap jenis produk yang telah dipisahkan. Hasilnya adalah daftar produk yang telah di*resubscribe*, di mana setiap produk memiliki jumlah pengguna yang melakukan *resubscription* untuk produk tersebut. *Output* ini kemudian diurutkan berdasarkan jumlah total pengguna yang melakukan *resubscription*, dengan produk yang paling sering di*resubscribe* ditampilkan terlebih dahulu. Berikut contoh output produk yang paling sering di*resubscribe* pada **Gambar 3.9**.

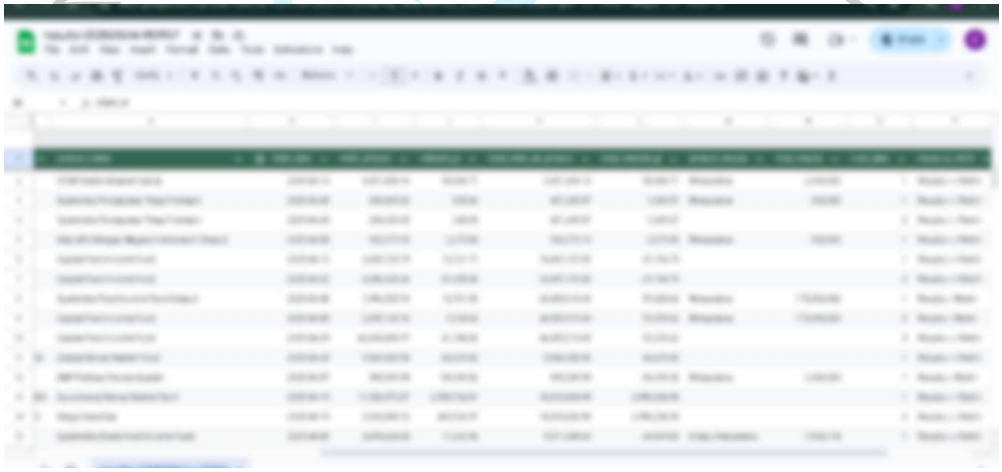



**Gambar 3.9 *Output Trend* Analysis pada BigQuery (b)**
Sumber : Dokumentasi yang Dihasilkan oleh Praktikan

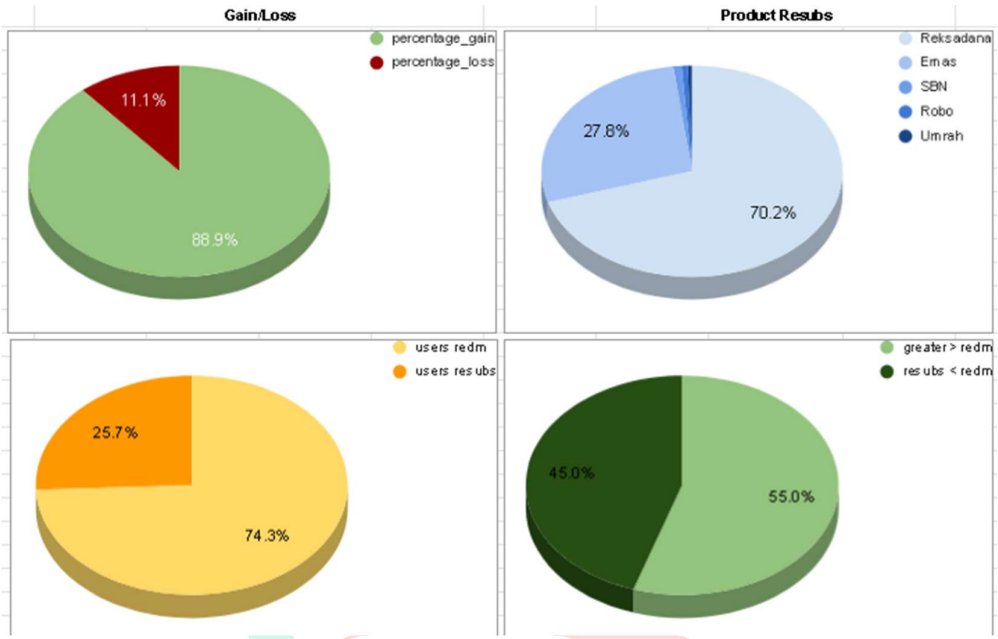
Dengan analisis produk tersebut, dapat menentukan produk mana yang paling sering di*resubscribe* oleh pengguna, serta partisipasi pengguna untuk setiap produk tersebut. *Query* ini membantu dalam mempelajari pola aktivitas *resubscription* pengguna sehubungan dengan produk-produk yang ada dan membantu dalam strategi pemasaran dan pengembangan produk.

### 3. Reporting

Agar hasil dari *query* tersebut dapat tersajikan dengan baik, Praktikan menyajikan hasil *query* ke *google sheet* dan divisualisasikan. Seperti pada **Gambar 3.10** dan **Gambar 3.11**.



**Gambar 3.10 *Reporting* pada *Google Sheets* (a)**
Sumber : Dokumentasi yang Dihasilkan oleh Praktikan




**Gambar 3.11 *Visualisasi* pada *Google Sheets***
Sumber : Dokumentasi yang Dihasilkan oleh Praktikan

Secara keseluruhan, informasi ini menunjukkan bahwa ada lebih banyak pengguna yang melakukan *redemption* dibandingkan dengan yang melakukan *resubscription*. Dalam pandangan yang lebih teragregasi, dapat dilihat bahwa beberapa produk investasi, termasuk reksadana, memiliki tingkat *resubscription* yang secara signifikan lebih rendah atau, pada beberapa kasus, saldo antara *redemption* dan *resubscription* yang saling tumpang tindih. Hal ini juga menunjukkan bahwa persentase klien yang melakukan *resubscription* masih rendah, dengan jumlah pengguna yang mengalami kerugian (*loss*) sebesar 11,09% dan pengguna yang mengalami keuntungan (*gain*) sebesar 88,91%. Pada dasarnya, mayoritas pengguna yang terlibat lebih memilih untuk menarik investasi mereka (*redemption*) saat mengalami keuntungan (*gain*) daripada meningkatkan investasi secara eksponensial dan melakukan *resubscription* pada produk yang sama, yang

menunjukkan ketidakpastian atau kurangnya kepercayaan terhadap kinerja produk-produk tersebut.

Selain itu, distribusi pengguna berdasarkan jenis produk menunjukkan keberagaman yang signifikan, dengan reksadana mendominasi jumlah pengguna yang melakukan *resubscription*. Hal ini tercermin dari 70,12% pengguna yang berpartisipasi dalam *resubscription* dengan produk reksadana, diikuti dengan produk emas, surat berharga negara (SBN), robo, dan umrah. Pengguna yang melakukan *resubscription* menunjukkan permintaan yang kuat terhadap produk investasi yang lebih aman dan terdiversifikasi. Produk-produk yang lebih stabil dan kurang berisiko, seperti emas dan SBN, cenderung menarik investor yang ingin menghindari fluktuasi pasar yang tajam. Di sisi lain, produk robo dan umrah, meskipun jumlahnya lebih sedikit, juga menunjukkan adanya minat dari klien yang mungkin mencari diversifikasi di saluran investasi alternatif lainnya.

### 3.2.3 *Predictive Modelling*

Dalam melaksanakan tugas ini, Praktikan diberikan peran untuk melakukan evaluasi menyeluruh terhadap data pengguna yang berkaitan dengan produk emas Bareksa, dengan kondisi pengguna ini tidak memiliki riwayat transaksi sebelumnya dengan produk emas. Tujuan dari analisis ini adalah untuk mengklasifikasikan pengguna yang berpotensi memiliki keterlibatan tinggi dengan produk emas di masa depan, meskipun mereka belum pernah melakukan transaksi emas. Terkait dengan investasi pada produk emas, Untuk mengetahui kemungkinan minat berinvestasi pengguna, Praktikan menganalisis transaksi produk diluar emas, pola pembelian pengguna, frekuensi dan keteraturan pengguna dalam bertransaksi, serta status portofolio pengguna. Pada fase ini, Praktikan mencoba untuk menemukan aspek-aspek yang dapat mendorong pengguna untuk berinvestasi, tanpa memandang apakah mereka telah membeli produk lain atau tidak memiliki riwayat pembelian emas, dan

menganalisis faktor-faktor tersebut untuk menarik kesimpulan. Analisis ini akan membantu departemen pemasaran untuk mendefinisikan ulang strategi mereka yang ditujukan kepada klien yang berpotensi berinvestasi dalam emas serta meningkatkan efektivitas kampanye pemasaran dan layanan produk emas yang ditawarkan oleh Bareksa.

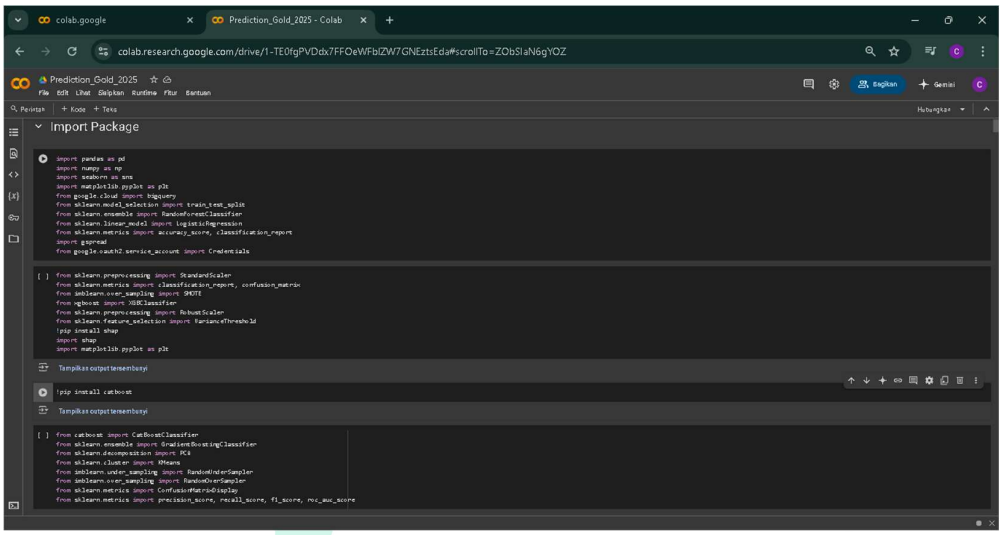
1. ***Data Preparation and Importing Required Package***

Proses ini dimulai dengan mengimpor pustaka dan alat yang diperlukan untuk analisis data dan pemodelan prediktif. Seperti biasa, pustaka pandas dan numpy yang penting sangat krusial untuk pemrosesan dan manipulasi data karena mereka memfasilitasi penanganan kumpulan data besar. Untuk visualisasi data, digunakan matplotlib dan seaborn sebagai alat pemrograman untuk menciptakan representasi visual yang rinci dari data yang tersedia.

Sejumlah pustaka pembelajaran mesin lainnya diterapkan untuk membangun dan mengevaluasi model prediktif. Salah satu sumber daya penting adalah pustaka sklearn, yang memberikan akses ke banyak algoritma pengembangan model seperti RandomForestClassifier, LogisticRegression, dan XGBClassifier yang digunakan dalam berbagai tugas prediktif. Untuk mengevaluasi model pada data yang tidak tersentuh, kumpulan data dibagi menjadi set pelatihan dan pengujian menggunakan fungsi train_test_split yang ditemukan di sklearn.model_selection.

Untuk menghilangkan masalah ketidakseimbangan kelas yang dapat menghambat efektivitas model, pustaka imblearn digunakan. Menggabungkan langkah-langkah ini membantu mencapai generalisasi yang lebih baik untuk semua kelas dalam data yang disediakan. Pustaka-pustaka ini meningkatkan keterbacaan, otomatisasi, dan kepercayaan selama tahap pemodelan yang mengarah pada prediksi yang lebih maju, misalnya, mendeteksi pengguna dengan kecenderungan meningkat untuk berinvestasi

dalam produk emas di tahun-tahun mendatang. Berikut seperti pada **Gambar 3.12**.




**Gambar 3.12** ***Importing Required Packages***
Sumber : Dokumentasi yang Dihasilkan oleh Praktikan

Setelah melakukan *import package*, Praktikan menggunakan data yang berasal dari BigQuery untuk melatih model *machine learning*. *Query* ini bertujuan untuk melakukan analisis mendalam terhadap data pengguna untuk memperkirakan kemungkinan mereka membeli produk emas, tanpa adanya transaksi emas sebelumnya. Langkah pertama dalam proses ini adalah mengidentifikasi pengguna yang tidak pernah membeli emas. Hal ini dilakukan dengan menggunakan *Common Table Expression* (CTE) tbl_user_non_emas, yang mencari pengguna dengan status dan kategori tertentu dalam tabel transaksi sebagai pengguna yang tidak melakukan pembelian emas, disaring berdasarkan status, kategori produk, *channel* transaksi, dan status pengguna. Pada CTE ini, Praktikan menarik pengguna yang melakukan transaksi pada produk reksa dana dengan rentang waktu pembuatan akun 2022 sampai dengan 31 Maret 2025. Setelah pengguna yang tidak membeli emas diidentifikasi, langkah berikutnya adalah menentukan tanggal pertama kali mereka membeli emas. Untuk tujuan ini, CTE

tbl_*first*_emas dibuat untuk mengekstrak data mengenai tanggal pembelian emas pertama dari tabel transaksi yang relevan. Hal ini membantu menetapkan dasar bagi pengguna yang mungkin mempertimbangkan untuk membeli emas di masa depan, meskipun mereka belum pernah melakukan transaksi emas sebelumnya. Proses ini memastikan bahwa setiap pengguna yang sudah bertransaksi emas tercatat dan dipertimbangkan dalam analisis.

Selanjutnya, CTE tbl_*limitation* digunakan untuk menggabungkan informasi dari tbl_user_non_emas dan tbl_*first*_emas untuk menetapkan tanggal batasan atau *date_limitation* bagi setiap pengguna yang terpilih. Bagi pengguna yang belum pernah membeli emas, tanggal pembatasnya ditetapkan pada 31 Maret 2025. Namun, bagi pengguna yang sudah membeli emas, tanggal pembatasnya disesuaikan dengan tanggal pembelian emas pertama. Pembatasan ini memungkinkan analisis lebih terfokus yang berpusat pada transaksi yang terjadi sebelum pembelian emas pertama atau dalam rentang waktu yang relevan. CTE tbl_order digunakan untuk menangkap transaksi lain di luar produk emas, termasuk *subscription* dan *redemption*, untuk menganalisis aktivitas pengguna terkait produk non-emas. Data dari tabel ini dibatasi pada transaksi dengan status selesai (status=4) dan tipe transaksi berupa *subscription* (tipe = 1) atau *redemption* (tipe = 2). Pola informasi ini membantu menganalisis transaksi pengguna terkait produk keuangan lain yang bukan emas, yang dapat membantu mengetahui kemungkinan minat mereka terhadap produk emas di masa depan.

Untuk menganalisis lebih lanjut perilaku transaksi pengguna, metrik transaksi didefinisikan berdasarkan pengguna dan bulan dalam CTE tbl_*order_month*. Di CTE ini, total nilai transaksi bulanan (*net amount*) dan jumlah hari transaksi per bulan dihitung. CTE ini memberikan gambaran pola pembelian yang terstruktur dan rutin, yang sangat berguna untuk segmentasi pengguna berbasis transaksi

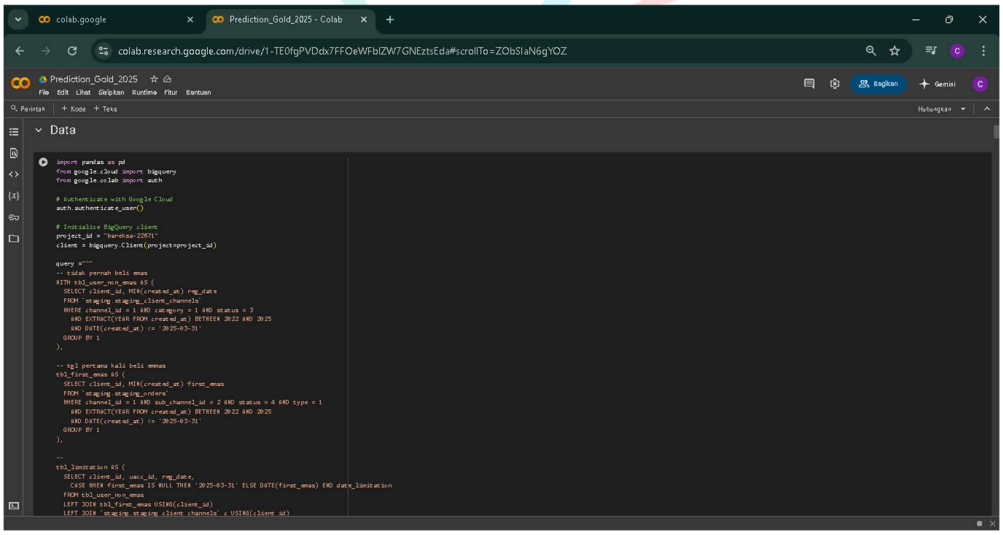
untuk menargetkan pengguna yang berpotensi tertarik pada produk investasi lainnya, termasuk emas, berdasarkan pola transaksi mereka. CTE tbl_order_interval dirancang untuk mempelajari waktu, yang diukur dengan fungsi LAG, antara dua transaksi berturut-turut. Dengan mengukur waktu antara transaksi yang berurutan, analisis ini membantu mengidentifikasi pengguna dengan pola transaksi yang berulang. Pengguna yang memiliki perilaku transaksi yang teratur lebih mungkin untuk melanjutkan investasi pada produk lain, seperti emas, karena mereka memiliki kecenderungan investasi yang konsisten.

Selanjutnya, CTE tbl_rutin menghitung rata-rata interval antara transaksi untuk menganalisis seberapa sering pengguna melakukan transaksi dalam waktu tiga bulan atau kurang. Pengguna yang bertransaksi secara rutin (dan lebih spesifik lagi, yang bertransaksi lebih sering) mendapatkan status transaksi yang telah ditentukan, yang membantu mengukur sejauh mana partisipasi mereka dalam investasi yang stabil. Pengguna dengan pola transaksi yang rutin biasanya lebih cenderung untuk berinvestasi lebih lanjut, termasuk dalam produk emas. CTE tbl_interval_redm memberikan pemahaman mengenai perilaku pengguna dengan menganalisis interval antara redemption dan penarikan dana dari produk investasi mereka. Pengguna yang sering melakukan redemption mungkin juga tertarik pada produk emas untuk diversifikasi portofolio mereka.

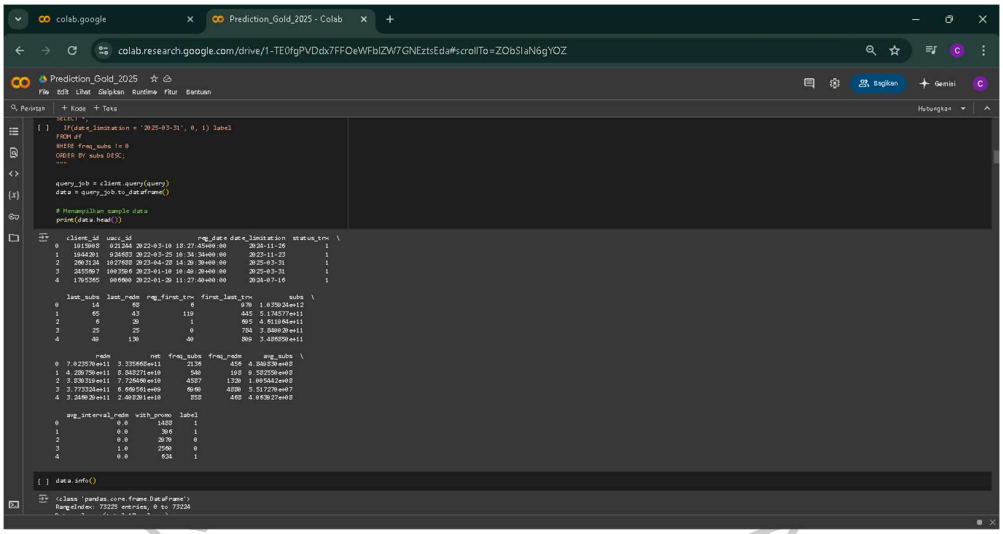
Informasi mengenai transaksi *subscription* dan *redemption* terakhir dari setiap pengguna disimpan dalam CTE tbl_*last_order*. Data ini sangat penting untuk memantau aktivitas terakhir pengguna dalam hal investasi dan memberikan gambaran terbaru tentang keputusan mereka untuk berinvestasi kembali atau menarik dana. Hal ini membantu memahami siklus investasi pengguna dan memprediksi kecenderungan mereka untuk membeli emas di masa depan. Selain itu, CTE tbl_porto mengambil data portofolio

pengguna, yang mencakup total aset yang dikelola (AUM). Dengan mengetahui nilai portofolio dan kemampuan finansial pengguna, kita dapat menilai potensi mereka untuk berinvestasi lebih lanjut dalam produk yang lebih stabil seperti emas. Ini memberikan wawasan penting tentang kemampuan finansial mereka untuk membeli emas.

Terakhir, semua data yang telah dihimpun dari berbagai CTE ini digabungkan dalam CTE utama df, yang menghitung berbagai fitur penting seperti total subscription (subs), redemption (redm), frekuensi transaksi, serta rata-rata interval transaksi dan redemption. Pada tahap ini, label tambahan (label) ditambahkan untuk menandai apakah pengguna layak untuk dianalisis lebih lanjut, dan dataset ini diurutkan berdasarkan total subscription. Dengan menggunakan data ini, analisis mendalam dilakukan untuk mengidentifikasi pengguna yang mungkin tertarik membeli emas di masa depan, bahkan tanpa riwayat pembelian emas sebelumnya. Berikut contoh *code* python yang dijalankan dalam *notebook* pada **Gambar 3.13** dan contoh output dataframe pada **Gambar 3.14**.



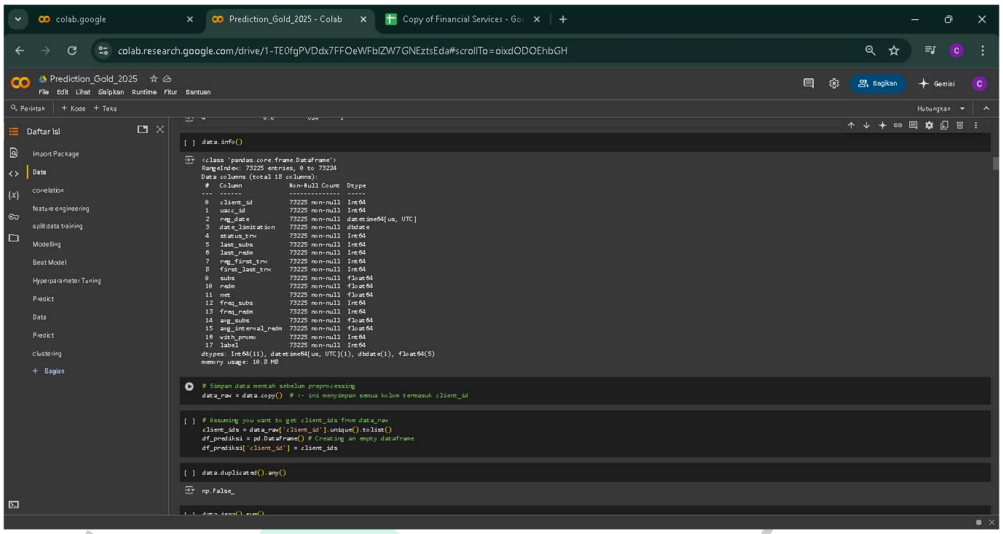

**Gambar 3.13 *Import* Data *dari Bigquery***
Sumber : Dokumentasi yang Dihasilkan oleh Praktikan



**Gambar 3.14 *Output Import* Data *dari Bigquery***
Sumber : Dokumentasi yang Dihasilkan oleh Praktikan

Setelah itu praktikan memproses data dengan memisahkan data yang tidak digunakan untuk prediksi (client_id) ke dalam *data frame* kosong. *Data frame* ini berfungsi sebagai dasar untuk pemrosesan lebih lanjut, dimana prediksi atau analisis terkait lainnya dapat disimpan untuk setiap klien.

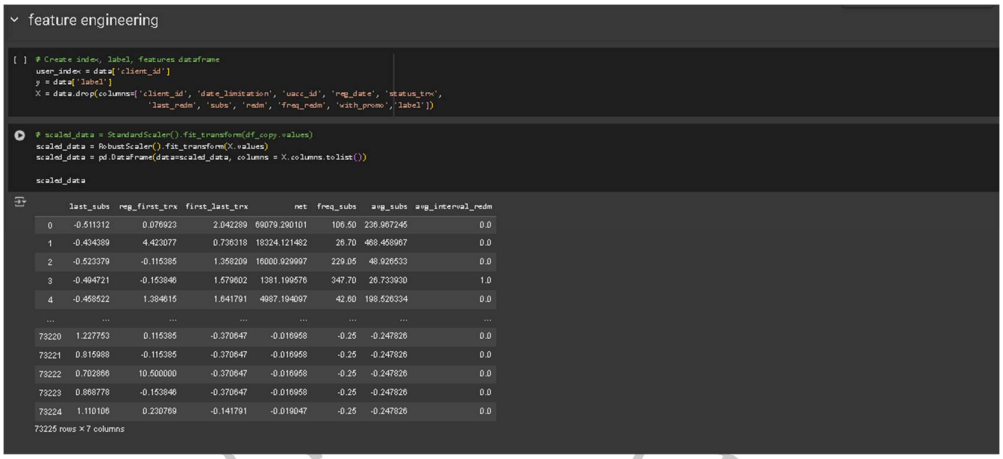
Selanjutnya, mengidentifikasi apakah ada baris yang duplikat dalam dataset. Pemeriksaan ini memastikan bahwa tidak ada data yang terduplikasi yang dimasukkan serta memeriksa apakah ada nilai yang hilang (NaN) dalam dataset. Dengan menghitung jumlah total nilai yang hilang untuk setiap kolom, langkah ini membantu mengidentifikasi masalah potensial dalam data yang perlu ditangani, seperti penghapusan baris yang tidak lengkap, sebelum melanjutkan ke proses analisis atau pembangunan model seperti pada **Gambar 3.15**.



**Gambar 3.15 *Preprocessing* Data dari *Bigquery***
Sumber : Dokumentasi yang Dihasilkan oleh Praktikan

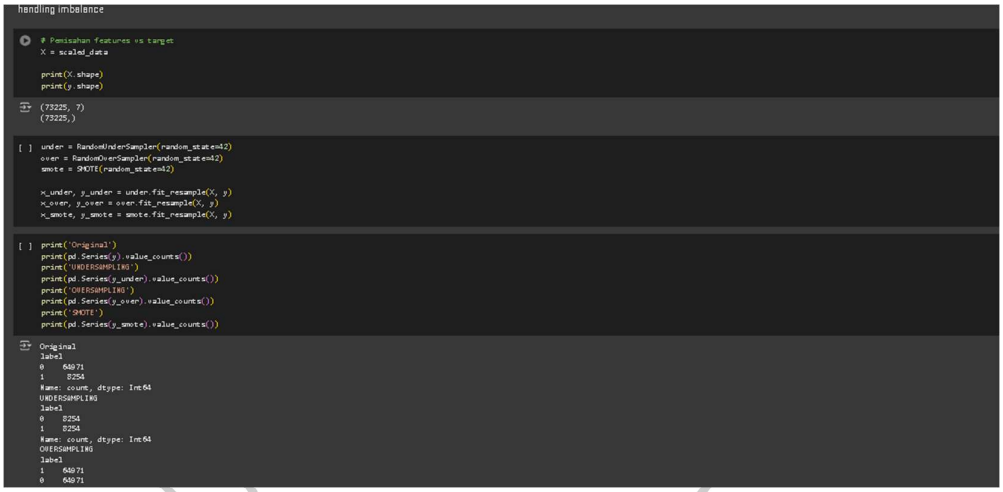
## 2. *Feature Engineering*

Pada fase ini, langkah *feature engineering* dimulai dengan menghapus kolom-kolom yang tidak akan berguna untuk analisis atau pemodelan, yang meliputi, namun tidak terbatas pada, client_id, date_limitation, uacc_id, reg_date, status_trx, dan sebagainya. Data kolom yang tersisa setelah proses ini digunakan untuk membangun fitur model. Data fitur kemudian diskalakan menggunakan *RobustScaler*, yang dianggap lebih tahan terhadap *outlier* dibandingkan dengan *standard scaler.* Hal ini memastikan bahwa fitur memiliki ukuran yang seragam dan model tidak akan terpengaruh oleh perbedaan skala antar fitur. Model kemudian dibangun menggunakan data yang telah diskalakan, dan fitur model lainnya ditambahkan sesuai kebutuhan. Data fitur yang telah ditransformasi kemudian disimpan dalam *data frame* lain dan harus disimpan dengan aman serta dikelompokkan untuk digunakan saat dibutuhkan dalam model prediksi seperti pada **Gambar 3.16**



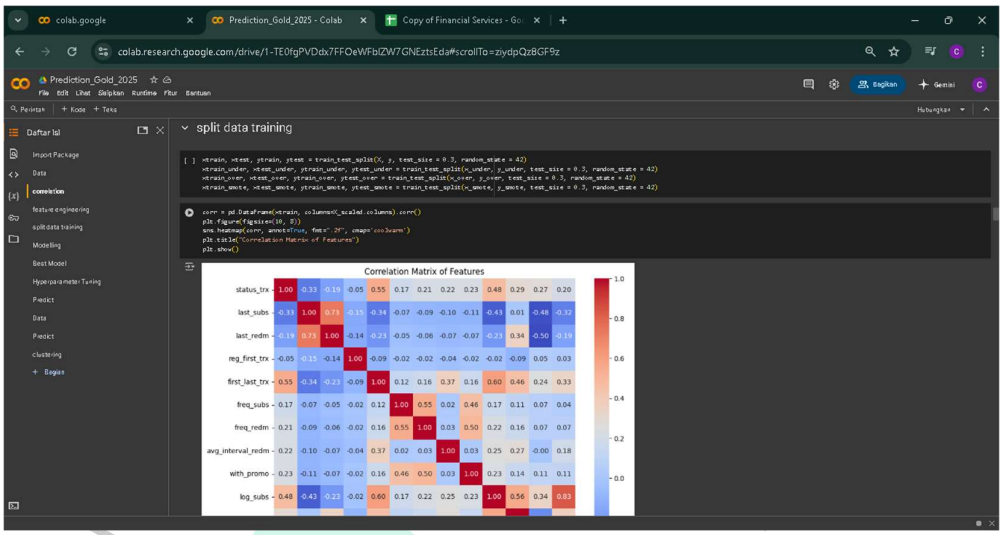

**Gambar 3.16 *Feature Engineering* Data**
Sumber : Dokumentasi yang Dihasilkan oleh Praktikan

Setelah menyelesaikan *feature engineering* dan normalisasi data, perhatian selanjutnya adalah menangani masalah ketidakseimbangan data. Ketidakseimbangan ini akan mempengaruhi kinerja model, terutama ketika satu kelas lebih dominan dibandingkan kelas lainnya. Terdapat tiga metode utama untuk mengatasi masalah ini, yaitu *undersampling*, *oversampling*, dan SMOTE (*Synthetic Minority Over-sampling Technique*). *Undersampling* mengurangi volume data yang tersedia pada kelas mayoritas, sementara *oversampling* meningkatkan volume data pada kelas minoritas. SMOTE meningkatkan keseimbangan data dengan menghasilkan sampel sintetis acak untuk kelas minoritas. Setiap teknik ini diterapkan pada dataset, dan distribusi ketidakseimbangan untuk label setelah penerapan teknik-teknik ini dihitung dan dicetak untuk verifikasi guna memastikan perubahan telah dilakukan seperti pada **Gambar 3.17**.




**Gambar 3.17 *Handling Imbalance* pada Data**
Sumber : Dokumentasi yang Dihasilkan oleh Praktikan

Selanjutnya, kita perlu membagi data menjadi set pelatihan dan pengujian. Menggunakan fungsi train_test_split, data dibagi menjadi dua kelompok dengan alokasi 70% untuk pelatihan dan 30% untuk pengujian. Pembagian ini diterapkan pada empat versi dataset yang telah diproses menggunakan teknik yang berbeda untuk menangani ketidakseimbangan yaitu, dataset asli, dataset yang di-*undersample*, dataset yang di-*oversample*, dan dataset yang diproses dengan SMOTE. Hal ini memungkinkan model untuk dilatih dan diuji pada berbagai dataset untuk menilai efektivitas setiap teknik yang digunakan dalam menangani ketidakseimbangan dan meningkatkan hasilnya seperti pada **Gambar 3.18**.

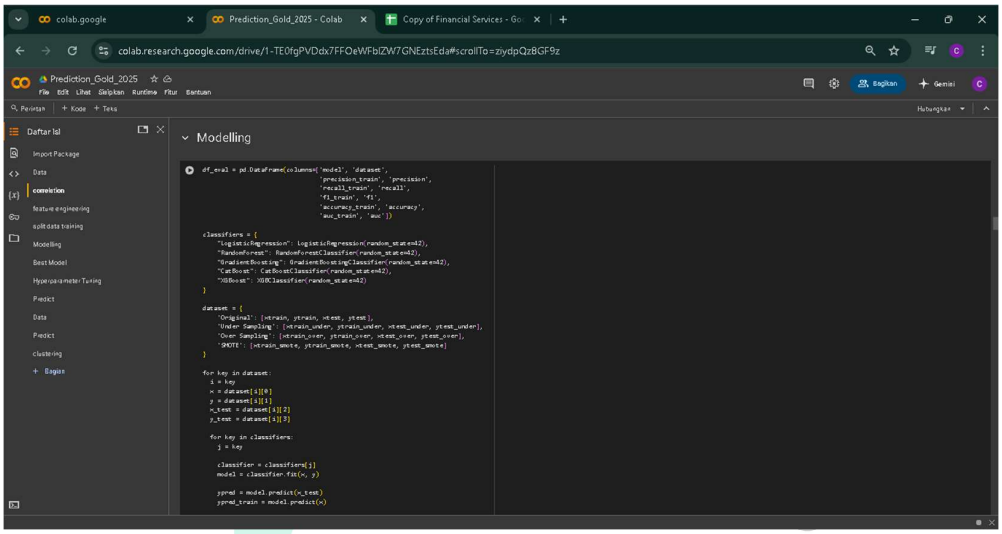



**Gambar 3.18 *Split Data* Menjadi Data *Train* dan Data *Test***
Sumber : Dokumentasi yang Dihasilkan oleh Praktikan

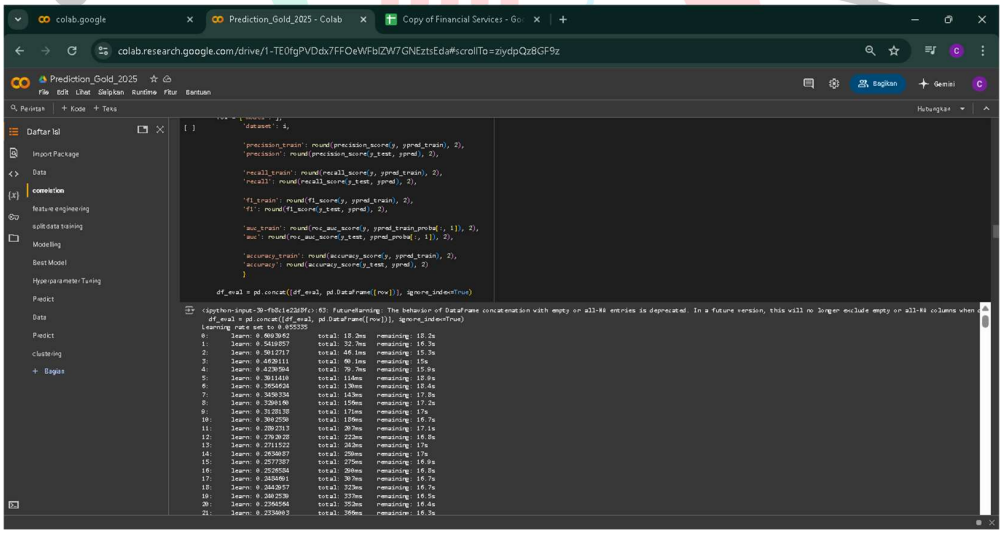
### 3. *Modelling*

Pada bagian ini, Praktikan fokus untuk melatih dan mengevaluasi berbagai model *machine learning* guna menentukan model prediksi terbaik yang dapat memproyeksikan kemungkinan seorang pengguna untuk membeli emas. Langkah pertama adalah menentukan serangkaian *classifier*, yang mencakup *Logistic Regression, Random Forest, Gradient Boosting, CatBoost,* dan *XGBoost* karena mereka banyak digunakan untuk masalah klasifikasi biner. Model-model ini diterapkan pada berbagai versi dataset, yang mencakup dataset asli dan dataset yang telah diproses dengan teknik *undersampling*, *oversampling*, dan SMOTE untuk

memperbaiki ketidakseimbangan yang ada dalam data seperti pada **Gambar 3.19**.



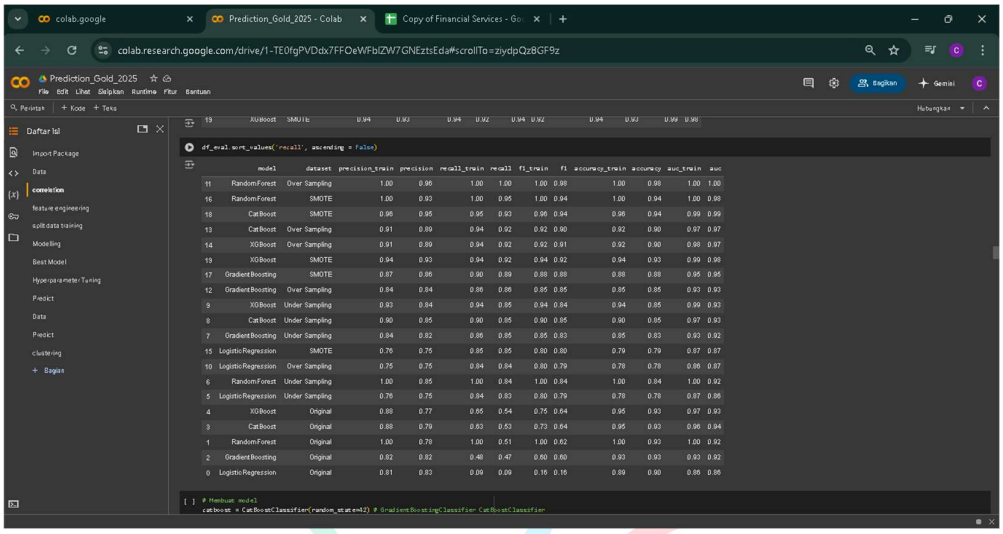
**Gambar 3.19 *Modelling the* Data *dari Bigquery***
Sumber : Dokumentasi yang Dihasilkan oleh Praktikan



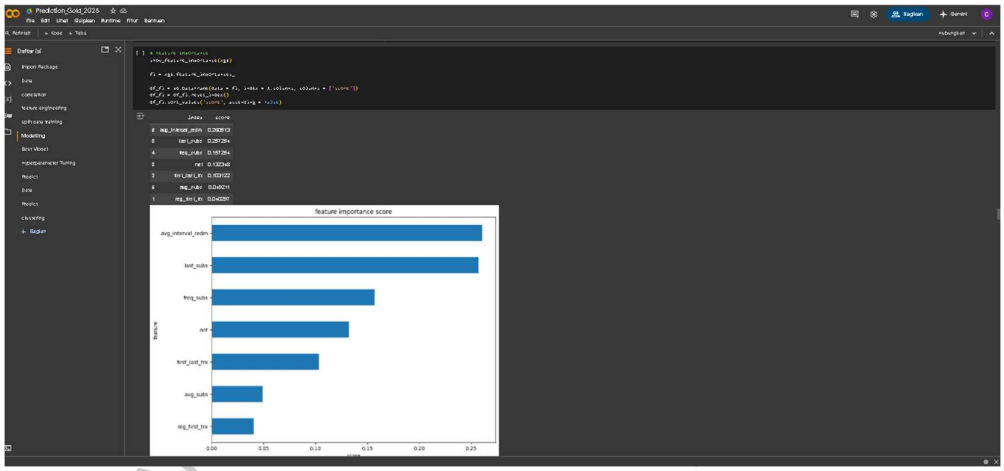
**Gambar 3.20 *Output Modelling* Data *dari Bigquery***
Sumber : Dokumentasi yang Dihasilkan oleh Praktikan

Setiap *classifier* dilatih pada data latih dan dievaluasi pada data uji yang sesuai. Berbagai metrik evaluasi, termasuk *precision*, *recall*, F1-*score*, akurasi, AUC, digunakan untuk menilai performa

model, baik pada fase pelatihan maupun pengujian. Metrik hasil ini mendefinisikan evaluasi terhadap masing-masing model dan disimpan dalam *data frame* (df_eval) untuk analisis lebih lanjut. Prediksi model dibandingkan dengan hasil aktual, sehingga dapat menentukan kredibilitas dan validitas model dalam meramalkan hasil yang diinginkan seperti pada **Gambar 3.21**.

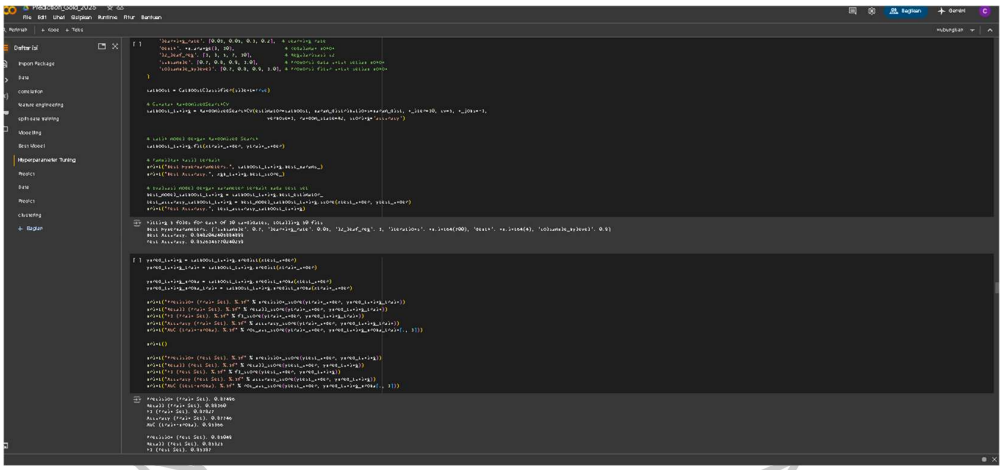



**Gambar 3.21 *Classification Report Modelling* Berdasarkan Model Terbaik**
Sumber : Dokumentasi yang Dihasilkan oleh Praktikan



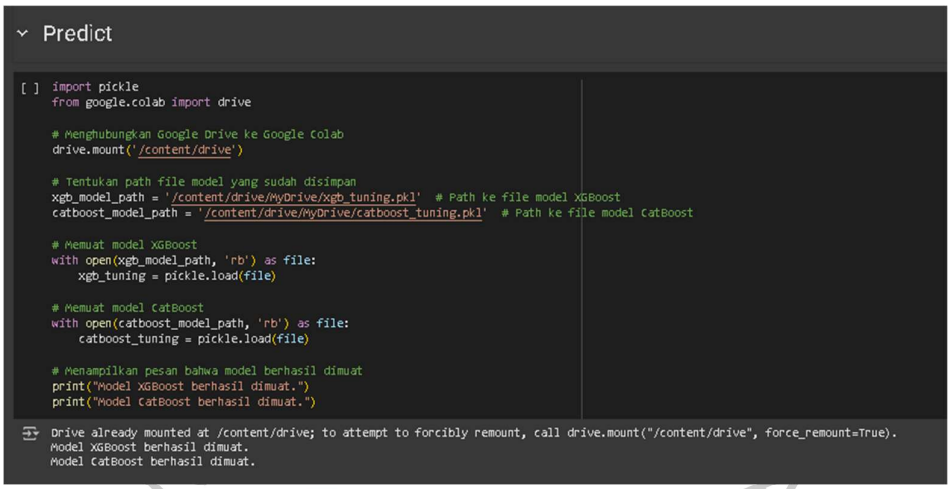

**Gambar 3.22 Mencari *Feature Importance* pada Model Terbaik**
Sumber : Dokumentasi yang Dihasilkan oleh Praktikan

Berdasarkan *classification reports*, model yang terpilih adalah *XGBoost* dan *Catboost* dengan teknik *undersampling* karena dinilai lebih stabil dan tidak *overfitting*. Kemudian, kedua model tersebut dicari *feature importance*nya dan dilakukan tuning dengan optimasi *hyperparameter* menggunakan *RandomizedSearchCV*. Proses ini mengoptimalkan model dengan parameter kunci seperti learning rate, jumlah estimator, kedalaman pohon, dan rasio subsample pada model *XGBoost* dan *CatBoost*. Model yang menghasilkan akurasi tertinggi selama pencarian dipilih untuk optimasi hyperparameter lebih lanjut seperti pada **Gambar 3.22** dan **Gambar 3.33**



**Gambar 3.23 Melakukan *Hyperparameter Tuning* pada Model Terbaik**
Sumber : Dokumentasi yang Dihasilkan oleh Praktikan

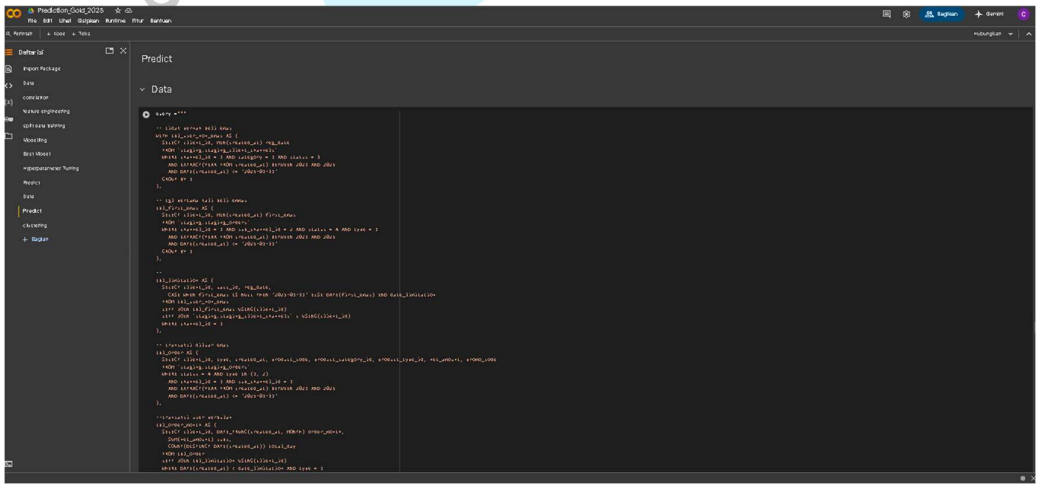
Model yang telah di-*tuning hyperparameter*nya kemudian dievaluasi kembali menggunakan data uji dan menghasilkan laporan klasifikasi yang komprehensif dengan *precision, recall, F1-score*, akurasi, dan AUC. Selain itu, sebuah kurva precision-recall juga dipetakan untuk membandingkan *trade-off* antara *recall* dan *precision* pada berbagai level *threshold*. Grafik ini juga menampilkan feature importance dari model, yang menunjukkan fitur mana yang paling berpengaruh pada hasil prediksi. Prosedur ini memastikan bahwa model disempurnakan untuk memberikan prediksi yang paling akurat mengenai pembelian emas di masa depan. Hasil model yang telah di-*tuning* kemudian disimpan agar dapat digunakan lagi pada data prediksi seperti pada **Gambar 3.24**.



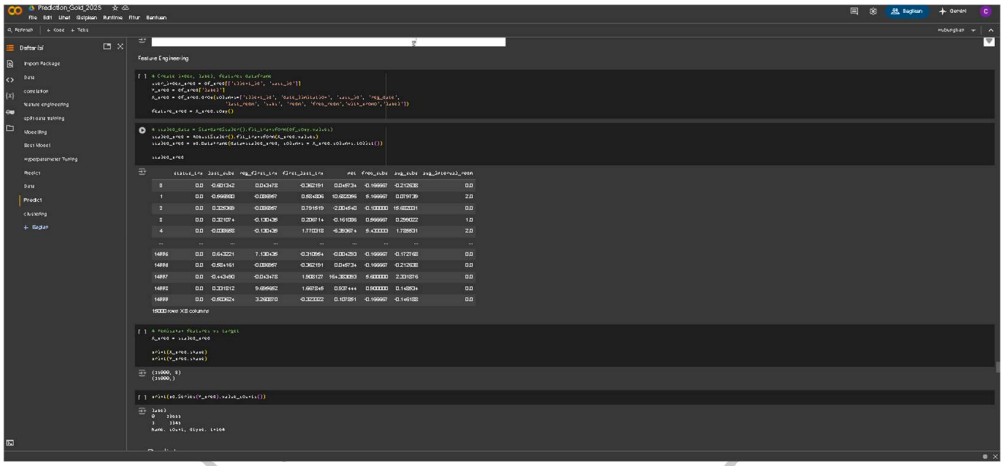

**Gambar 3.24 *Menyimpan Model***
Sumber : Dokumentasi yang Dihasilkan oleh Praktikan

## 4. Predict Data

Ketika model dianggap siap, tugas selanjutnya adalah memprediksi data baru menggunakan model ini. Langkah pertama melibatkan persiapan data yang akan diprediksi, termasuk menerapkan *feature importance* yang sama dengan yang diterapkan pada data pelatihan. Langkah ini diperlukan agar fitur yang ada pada data baru dan yang digunakan pada data pelatihan relevan dan bermakna seperti pada **Gambar 3.25** dan **Gambar 3.26**.




**Gambar 3.25 *Import Data* Baru dari *BigQuery***
Sumber : Dokumentasi yang Dihasilkan oleh Praktikan



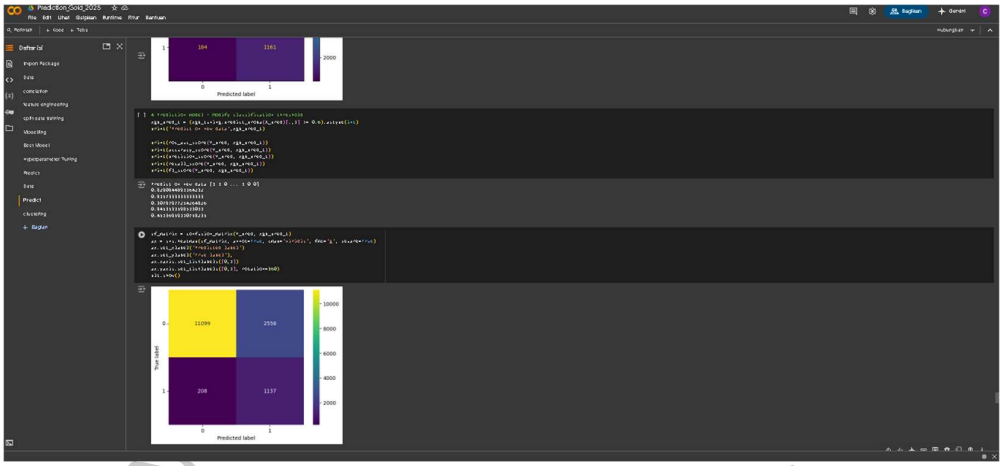
**Gambar 3.26 *Melakukan Feature Engineering* pada Data Baru**
Sumber : Dokumentasi yang Dihasilkan oleh Praktikan

Setelah data diproses dengan benar, langkah pertama adalah memuat model, dan setelah itu, model digunakan untuk membuat prediksi tentang kemungkinan pengguna membeli produk emas seperti pada **Gambar 3.27, Gambar 3.28** dan **Gambar 3.29.**

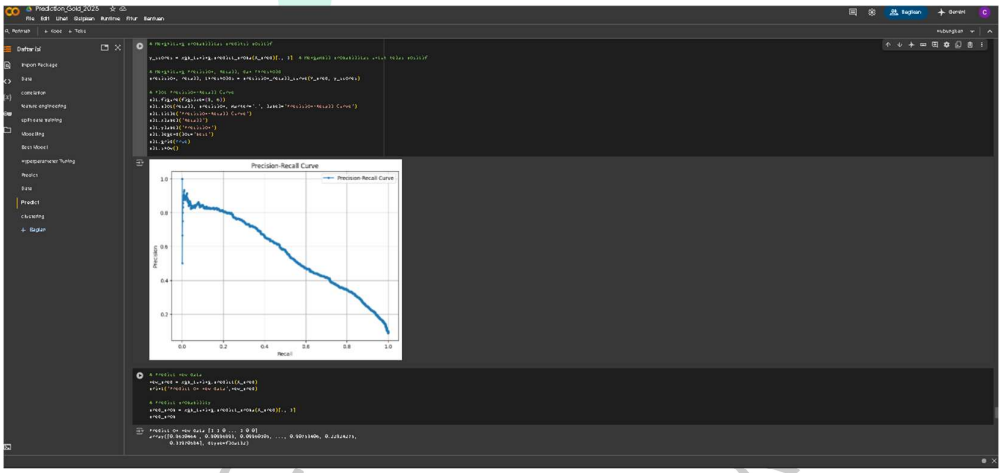
Predict

```
[ ] import pickle
    from google.colab import drive

    # Menghubungkan Google Drive ke Google Colab
    drive.mount('/content/drive')

    # Tentukan path file model yang sudah disimpan
    xgb_model_path = '/content/drive/MyDrive/xgb_tuning.pkl'  # Path ke file model XGBoost
    catboost_model_path = '/content/drive/MyDrive/catboost_tuning.pkl'  # Path ke file model CatBoost

    # Memuat model XGBoost
    with open(xgb_model_path, 'rb') as file:
        xgb_tuning = pickle.load(file)

    # Memuat model CatBoost
    with open(catboost_model_path, 'rb') as file:
        catboost_tuning = pickle.load(file)

    # Menampilkan pesan bahwa model berhasil dimuat
    print("Model XGBoost berhasil dimuat.")
    print("Model CatBoost berhasil dimuat.")

Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).
Model XGBoost berhasil dimuat.
Model CatBoost berhasil dimuat.
```

**Gambar 3.27 *Load Model* yang Sudah Disimpan**
Sumber : Dokumentasi yang Dihasilkan oleh Praktikan



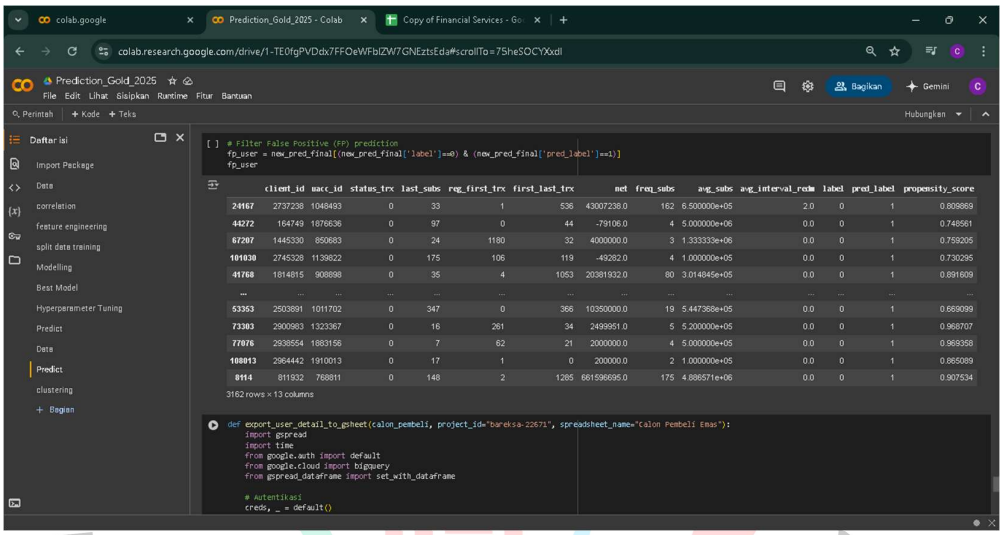
**Gambar 3.28 *Training Model* pada Data Baru (a)**
Sumber : Dokumentasi yang Dihasilkan oleh Praktikan



**Gambar 3.29 *Training Model* pada Data Baru (b)**
Sumber : Dokumentasi yang Dihasilkan oleh Praktikan

Hasil dari prediksi dievaluasi (dinilai), salah satu langkah utama adalah deteksi *false positives* (FP) yang menunjukkan pengguna yang diprediksi tidak akan membeli emas, tetapi sebenarnya mereka diprediksi akan membeli produk tersebut. Mendeteksi *false positives* berguna karena membantu memberikan panduan untuk rencana pemasaran yang lebih baik di masa depan.

Dengan kata lain, *false positives* disini adalah pengguna yang salah diasumsikan oleh model tidak menunjukkan minat untuk membeli emas tapi ternyata tertarik pada produk emas, sehingga tindakan bisa diambil untuk menghindari pemborosan sumber daya dan upaya seperti pada **Gambar 3.30**.




**Gambar 3.30 *Filter False Positive User* di *Bigquery***
Sumber : Dokumentasi yang Dihasilkan oleh Praktikan

Setelah melakukan prediksi dan mengevaluasi dimana *false positives* terjadi, informasi yang relevan diekspor ke *Google Sheets* untuk analisis lebih dalam dan upaya kolaboratif. Informasi yang diekspor mencakup data pengguna seperti client_id, nama, email, dan nomor telfon. Data ini difilter sesuai dengan *propensity score* yang mengukur kemungkinan seorang pengguna membeli emas berdasarkan perhitungan model. Mekanisme ekspor ke *Google Sheets* memungkinkan tim lain seperti pemasaran untuk mengakses dan mengambil tindakan terhadap pengguna yang diprediksi sebagai calon pembeli emas, memfokuskan upaya pemasaran mereka pada pengguna dengan skor propensity tertinggi seperti pada **Gambar 3.31**.

**Gambar 3.31 *Output* di *Google Sheet***
Sumber : Dokumentasi yang Dihasilkan oleh Praktikan

### 3.2.4 Visualisasi Data

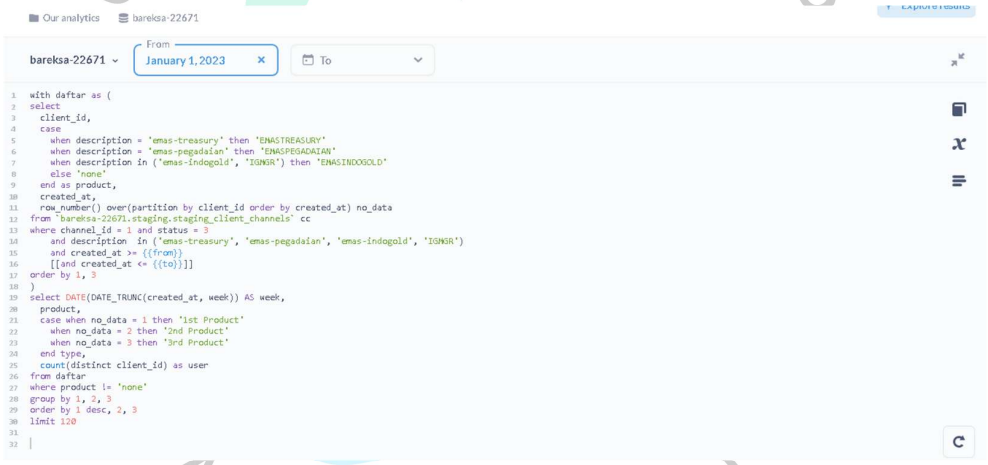
#### 1. Gold Weekly New Registered

Visualisasi data ini bertujuan untuk melacak pengguna yang baru registrasi pada setiap produk emas secara *real-time.* Dalam kueri ini, deskripsi produk seperti emas-*treasury*, emas-pedagiaan, dan emas-indogold dipetakan ke pengenal kategori tertentu yaitu EMASTREASURY, EMASPEDAGAIAN, dan EMASINDOGOLD. Jika ada deskripsi produk yang tidak cocok dengan ini, maka akan diklasifikasikan sebagai 'none.' Klasifikasi ini membantu untuk mengetahui produk mana yang menarik perhatian pengguna untuk keterlibatan lebih lanjut.

Selain itu, dalam *query* ini, kategori pengguna diterapkan dengan menggunakan fungsi row_number(), yang dipartisi berdasarkan client_id dan diurutkan berdasarkan tanggal created_at, untuk melacak transaksi pengguna secara kronologis. Selanjutnya, *query* ini membatasi data hanya untuk mencakup catatan yang memenuhi kriteria channel_id = 1 yang mendeklarasikan produk emas dan status = 3 yang mendeklarasikan pengguna berhasil terdaftar. Selain itu, tanggal transaksi di*filter* untuk memastikan bahwa hanya transaksi yang terjadi setelah tanggal tertentu {{*from*}}

yang diambil, yang dimaksudkan agar visualisasi data ini fokus pada aktivitas terbaru.

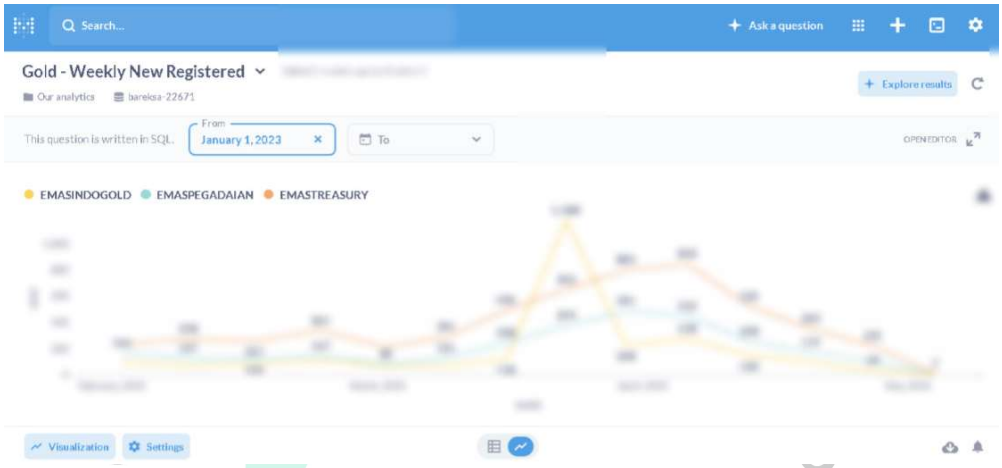
Pada bagian kedua *query*, jumlah pengguna yang terdaftar untuk setiap jenis produk dihitung dan dikelompokkan berdasarkan jenis produk (product). Pengguna diklasifikasikan sebagai terdaftar untuk 1*st Product*, 2*nd Product*, atau 3*rd Product*, tergantung pada minat mereka terhadap produk tersebut, yang membantu untuk melihat produk mana yang lebih banyak diminati oleh pengguna baru. Kueri ini diurutkan berdasarkan tanggal pendaftaran produk untuk memprioritaskan pendaftaran terbaru, dan hasilnya dibatasi hingga 120 catatan untuk memfokuskan hasil dataset seperti pada **Gambar 3.32**.




**Gambar 3.32 Proses *Query* di *Metabase***
Sumber : Dokumentasi yang Dihasilkan oleh Praktikan

Analisis ini memberikan gambaran mengenai tren pendaftaran pengguna, distribusi minat terhadap produk yang berbeda, dan tingkat aktivitas pengguna baru dalam periode waktu tertentu. Bisnis dapat melacak pengguna berdasarkan aktivitas pendaftaran dan minat produk, yang memberikan pemahaman yang lebih baik tentang interaksi pelanggan, mendeteksi peluang baru, dan memungkinkan perusahaan untuk mengembangkan rencana

yang lebih terarah untuk menarik perhatian pelanggan potensial seperti pada **Gambar 3.33**



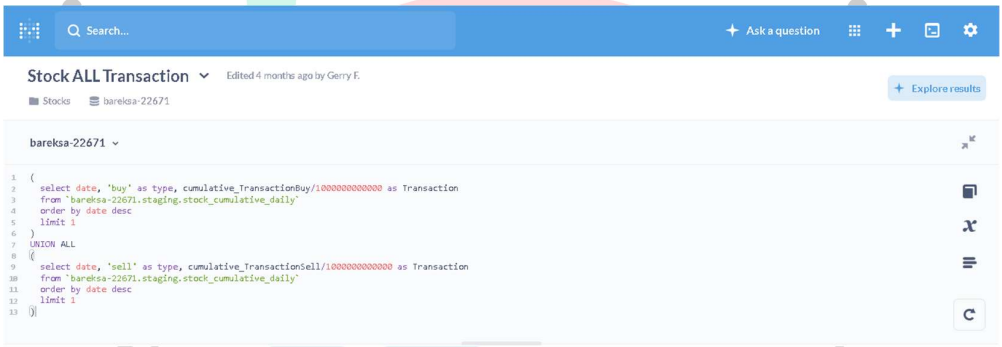

**Gambar 3.33 *Output Weekly Registered* di *Metabase***
Sumber : Dokumentasi yang Dihasilkan oleh Praktikan

### 2. Transaksi Stock

Visualisasi data transaksi *stock* bertujuan untuk melacak transaksi penjualan dan pembelian dalam saham secara *real-time. Query* mengambil data transaksi beli dan jual saham terbaru dari tabel *stock_cumulative_daily*, yang berisi transaksi saham harian secara kumulatif. *Query* ini menggunakan operator *UNION ALL* untuk menggabungkan kedua set transaksi beli dan jual. Pada bagian pertama, *query* mengambil data yang berkaitan dengan transaksi pembelian saham. Kolom *cumulative_TransactionBuy* diambil dari tabel *stock_cumulative_daily* dan untuk memudahkan pemahaman, nilai transaksi pembelian kumulatif yang diambil dibagi dengan 1 triliun (1000000000000). Kolom *date* memberikan informasi terkait tanggal transaksi, sementara kolom *type* diisi dengan *buy*, yang berarti transaksi ini adalah pembelian. Hasilnya kemudian diurutkan berdasarkan kolom *date* dalam urutan menurun, sehingga transaksi terbaru akan muncul pertama kali. *Query* ini

menggunakan klausa *LIMIT* 1 untuk memastikan hanya transaksi pembelian terbaru yang diambil.

Bagian kedua dari *query* menjalankan operasi yang serupa, namun kali ini untuk transaksi penjualan saham. Data yang diambil berasal dari kolom *cumulative_TransactionSell*, yang berisi jumlah kumulatif nilai transaksi penjualan saham. Nilai di kolom ini juga dibagi dengan satu triliun untuk memudahkan pemahaman. Kolom *date* menunjukkan tanggal transaksi, sementara kolom *type* diisi dengan nilai *sell*, yang berarti transaksi ini mencakup penjualan. Seperti pada bagian pertama, data diurutkan berdasarkan kolom tanggal dalam urutan menurun, dan hanya satu baris terbaru yang diambil menggunakan perintah *LIMIT* 1 seperti pada **Gambar 3.34**.




**Gambar 3.34 Proses *Query Stock* di *Metabase***
Sumber : Dokumentasi yang Dihasilkan oleh Praktikan

Kedua set data ini digabungkan menjadi satu hasil menggunakan klausa *UNION ALL*. Dengan menggunakan *UNION ALL*, hasil transaksi beli dan transaksi jual akan disertakan tanpa menghapus baris yang duplikat. Hasil akhirnya adalah dua baris; satu berisi transaksi pembelian terbaru dengan tipe *buy*, dan satu lagi berisi transaksi penjualan terbaru dengan tipe *sell*. Dengan cara ini, jumlah transaksi pembelian dan penjualan saham yang tersedia dapat ditangkap secara *real-time* dan dapat digunakan untuk analisis

lebih lanjut atau pelaporan terkait dengan aktivitas pasar saham seperti terlihat pada **Gambar 3.35**.


**Gambar 3.35 *Output Stock Transaction* pada *Metabase***
Sumber : Dokumentasi yang Dihasilkan oleh Praktikan

## 3.3 Kendala Yang Dihadapi

Selama menjalankan tugas kerja praktik, praktikan menghadapi beberapa kendala yang cukup menghambat kelancaran pekerjaan diantaranya adalah

1. Ketidakpahaman atas *database* yang dimiliki perusahaan sehingga pada saat proses pengumpulan data mentah yang dilakukan secara mandiri tanpa adanya penjelasan yang jelas mengenai makna dan konteks data tersebut, membuat praktikan kesulitan dalam menginterpretasikan data secara tepat dan mengarahkannya pada analisis yang relevan.
2. Miskomunikasi antar divisi, yang mengakibatkan duplikasi dalam pekerjaan dan memperlambat progres proyek. Misalnya, beberapa tim yang bekerja pada tugas yang sama tidak saling berkomunikasi

dengan baik, sehingga satu pekerjaan bisa dikerjakan oleh dua divisi tanpa sepengetahuan satu sama lain. Hal ini menyebabkan pemborosan waktu dan sumber daya yang seharusnya dapat dialokasikan untuk tugas lain.

3. Data mengalami perbaikan atau perubahan yang dilakukan oleh tim *backend engineer*. Ketidakpastian data mengganggu kestabilan hasil analisis yang telah dilakukan sebelumnya, sehingga praktikan harus melakukan penyesuaian dan perbaikan yang memakan waktu ekstra. Kesulitan-kesulitan ini memperlambat proses dokumentasi dan pemrosesan data yang sudah terkumpul.

## 3.4 Cara Mengatasi Kendala

Untuk mengatasi berbagai kendala yang dihadapi selama masa kerja praktik, praktikan mengambil beberapa langkah yang dinilai efektif untuk mengurangi dampak dari masalah-masalah tersebut yaitu

1. Praktikan melakukan diskusi lebih lanjut dengan mentor untuk membahas klasifikasi dan pemahaman tentang data yang digunakan. Ini membantu praktikan dalam mengklarifikasi makna dan tujuan dari data yang dikumpulkan, serta memastikan bahwa pengolahan data dilakukan dengan cara yang benar dan efektif.

2. Mengenai masalah miskomunikasi antar divisi, praktikan secara proaktif melakukan konfirmasi dengan mentor terkait *request* atau permintaan dari divisi lain yang langsung diteruskan kepada praktikan. Ini membantu mengurangi risiko kesalahpahaman

dan memastikan bahwa setiap permintaan yang diterima sudah jelas dan terarah.

3. Pada masalah perbaikan data dari tim *backend engineer*, praktikan mengkonfirmasi dengan mentor untuk mendapatkan kejelasan data yang digunakan, serta menunggu validasi data yang diperlukan untuk analisis lebih lanjut serta menetapkan batas waktu pengumpulan data agar analisis dilakukan pada data yang sudah terkumpul secara final.

## 3.5 Pembelajaran Yang Diperoleh dari Kerja Profesi

Selama menjalankan kerja praktik di PT Bareksa, praktikan memperoleh berbagai pembelajaran yang sangat berharga, baik dalam hal keterampilan teknis maupun komunikasi. Salah satu pembelajaran utama yang diperoleh adalah peningkatan keterampilan dalam analisis data. Praktikan belajar untuk mengelola data mentah yang besar dan kompleks menggunakan berbagai *tools* dan bahasa pemrograman seperti *Python*, *BigQuery*, dan *Metabase*. Praktikan juga menjadi lebih terampil dalam memvisualisasikan data yang kompleks sehingga dapat disajikan dengan cara yang mudah dipahami oleh berbagai pihak, termasuk manajer produk dan tim pemasaran. Selain itu, praktikan memperoleh pemahaman yang lebih dalam mengenai produk investasi dan bagaimana peran data sangat krusial dalam pengembangan produk dan strategi bisnis perusahaan. Praktikan juga memahami bagaimana teknik *machine learning* dapat digunakan untuk memprediksi perilaku pengguna dan meningkatkan efektivitas pemasaran produk. Serta dapat menerapkan bidang ilmu yang telah dipelajari pada Mata Kuliah *Business Intelligence* dan *Data Warehouse*.

Selain keterampilan teknis, praktikan juga mengembangkan keterampilan komunikasi yang sangat penting dalam dunia kerja.Praktikan belajar untuk berkomunikasi dengan jelas dan tepat

waktu dengan berbagai divisi dalam perusahaan untuk memastikan bahwa setiap pekerjaan berjalan dengan baik dan sesuai dengan tujuan yang telah ditetapkan. Praktikan juga melakukan eksplorasi terhadap tools yang sebelumnya belum digunakan, seperti *BigQuery* dan *Metabase*, yang meningkatkan kemampuan praktikan dalam menangani data skala besar dan memberikan *insight* yang lebih mendalam. Secara keseluruhan, pengalaman kerja praktik ini memberikan pembelajaran yang sangat berharga yang dapat diterapkan dalam dunia kerja yang lebih profesional dan teknis.