Data Analytics Test

Data analytics test ini berisi 2 bagian, teori dan test coding yang masing-masing terdiri dari:

1. Teori
   * Konsep Dasar Data Analytics: Tes ini dimaksudkan untuk menguji pemahaman member tentang data analytics.
2. Coding Test
   * Data preparation test: Tes ini dimaksudkan untuk menguji kemampuan member dalam melakukan ETL data.
   * Data visualization test: Tes ini dimaksudkan untuk menguji kemampuan member dalam hal visualisasi data.
   * Basic Stats Method test: Tes ini dimaksudkan untuk menguji kemampuan member dalam melakukan modeling data menggunakan statistika dasar.

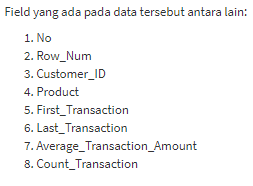
Market Research and Recommendation and Visualization Technique for Business Decision Making - Part 1

DQLab sport center adalah toko yang menjual berbagai **kebutuhan olahraga** seperti **Jaket, Baju, Tas, dan Sepatu.** Toko ini mulai berjualan sejak tahun 2013, sehingga sudah memiliki pelanggan tetap sejak lama, dan tetap berusaha untuk mendapatkan pelanggan baru sampai saat ini.

Di awal tahun 2019, manajer toko tersebut merekrut junior DA untuk membantu memecahkan **masalah yang ada di tokonya**, yaitu **menurunnya pelanggan yang membeli kembali ke tokonya**. Junior DA tersebut pun diberi kepercayaan mengolah data transaksi toko tersebut. Manajer toko mendefinisikan bahwa **customer termasuk sudah bukan disebut pelanggan lagi** (churn) ketika dia sudah **tidak bertransaksi** ke tokonya lagi sampai dengan **6 bulan terakhir** dari **update data terakhir** yang tersedia.

Manajer toko pun memberikan **data transaksi** dari tahun **2013** sampai dengan **2019** dalam bentuk csv (comma separated value) dengan data\_retail.csv dengan jumlah baris **100.000 baris data**.

Berikut tampilan datanya:



Manajer toko dan junior DA di salah satu cabang kita minta bantuan kamu untuk **mengurusi riset pasar mereka dengan data** dan persoalan tadi. Ada baiknya sembari kamu mengerjakan, buat laporan kerjamu **langkah** **per langkah** agar bisa **diketahui proses analisis datanya** untuk mengatasi kasus ini.”

**Langkah Pengerjaannya :**

1. **Data preparation test**

Importing data: Melakukan import data\_retail.csv ke python environment.

Cleansing data: Melakukan pembersihan dan modifikasi data sehingga siap digunakan untuk analisis lebih lanjut.

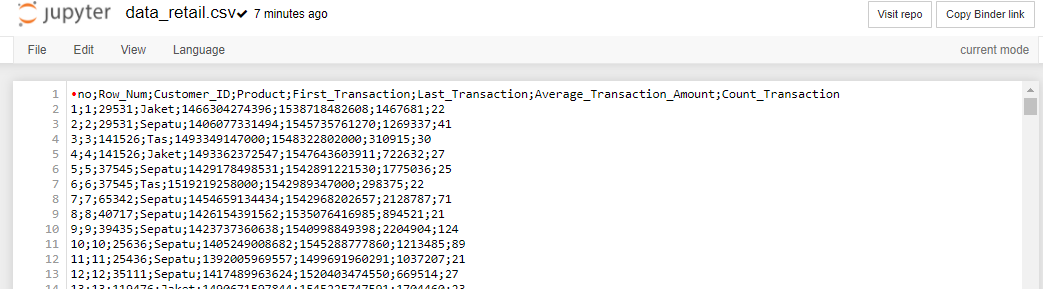
1. **Data visualization test**: Mendapatkan insight dari hasil visualisasi yang telah dibuat.
2. **Basic stats method test**: Mendapatkan insight dari model dan evaluasi model yang sudah dibuat dan diuji.

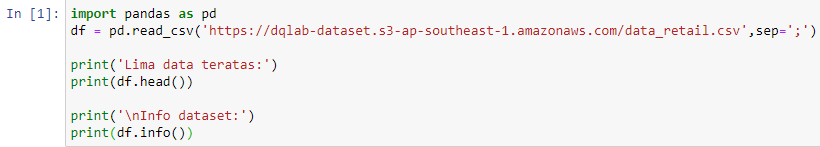
## Importing Data dan Inspection

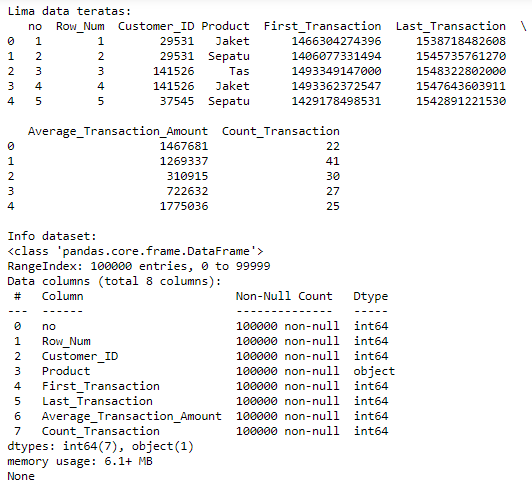
Importlah dataset dari **https://dqlab-dataset.s3-ap-southeast-1.amazonaws.com/data\_retail.csv**dan kemudian inspeksilah dataset tersebut dengan

1. mencetak lima data teratas saja,
2. mencetak info dataset.

Berikut datasetnya yang dibukakan dengan menggunakan Text Editor:







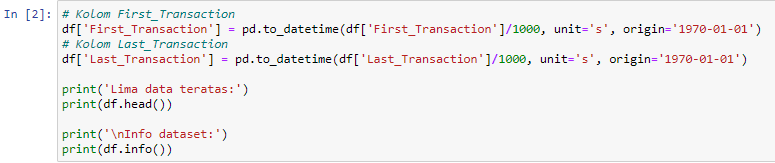
## Data Cleansing

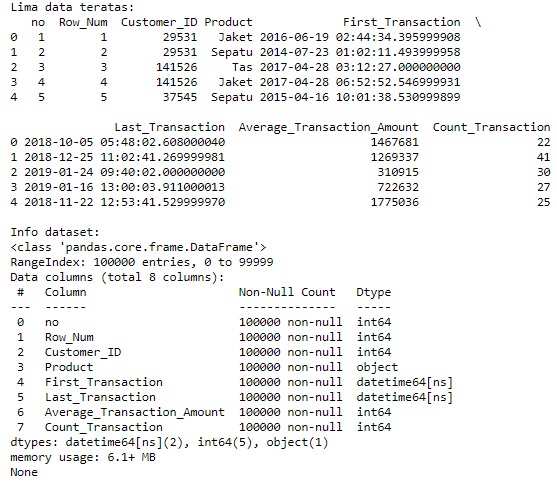
Dua kolom yang menunjukkan terjadinya transaksi tidak bertipe datetime, maka ubahlah kedua kolom tersebut ke tipe data datetime. Kemudian cetaklah kembali 5 data teratas dari dataframe **df** dan juga tipe data masing-masing kolomnya.

 Apabila di  , pada layar console akan menampilkan **TypeError** seperti berikut:

Hal ini disebabkan kedua kolom tersebut telah bertipe Datetime, ketika **Submit** dimeminta kembali untuk menimpa kolom yang telah diubah sewaktu **Run**.

Type Error Message ini dapat diabaikan, karena coding yang diinput sudah benar dan bisa melanjutkan.





## Churn Customers

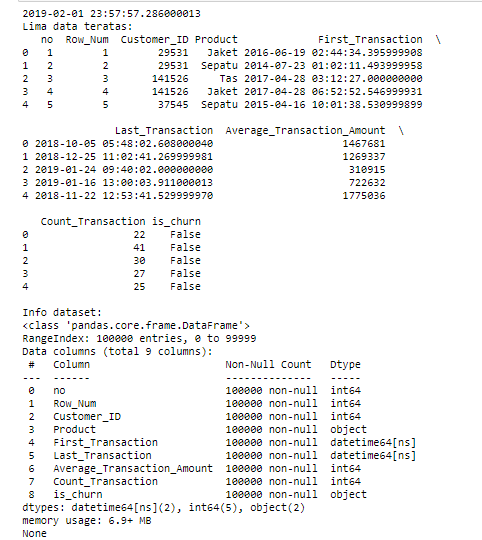
Untuk menentukan churn customers sesuai definisi yang telah diberikan, carilah

transaksi paling terakhir kapan dilakukan

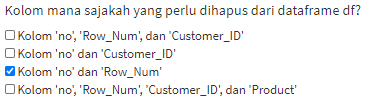
klasifikasikanlah mana customer yang berstatus churn dan mana yang tidak.

Setelah itu cetak lima data teratas dan informasi dataset.



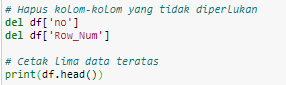


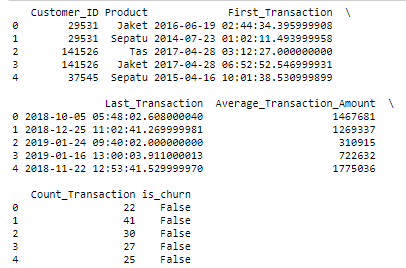
## Kolom yang akan dihapus



## Menghapus kolom yang tidak diperlukan

Sesuai dengan pada jawaban pada quiz sebelumnya, hapuslah kolom-kolom yang dimaksud dengan melengkapi code pada live code editor.



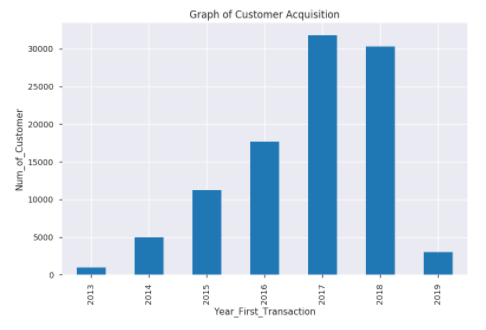


# **Coding Test: Data Visualization**

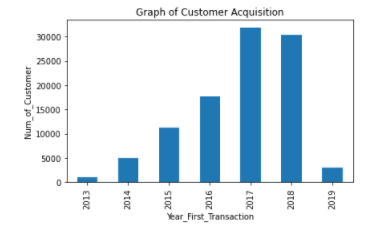
## Customer acquisition by year

Setelah semuanya lancar, langkah berikutnya adalah membuat visualisasi data berupa trend of customer acquisition by year dengan meggunakan bar chart. Untuk itu buatlah feature/kolom tambahan yang merupakan tahun dari First\_Transaction dan tahun dari Last\_Transaction masing-masingnya dengan nama Year\_First\_Transaction dan Year\_Last\_Transaction sebelum melakukan visualisasi.

Bar chart yang diinginkan:

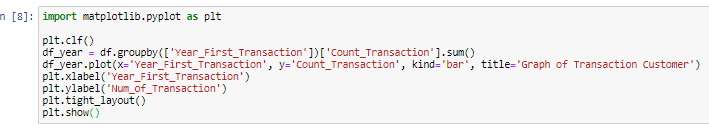


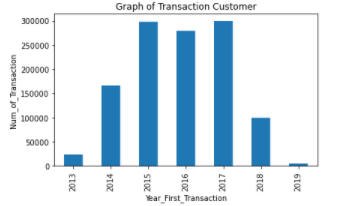




## Transaction by year

Visualisasikanlah trend jumlah transaksi per tahunnya dengan menggunakan bar chart.

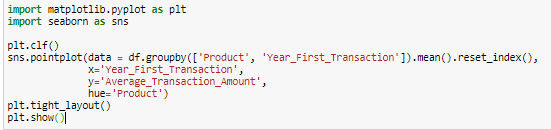


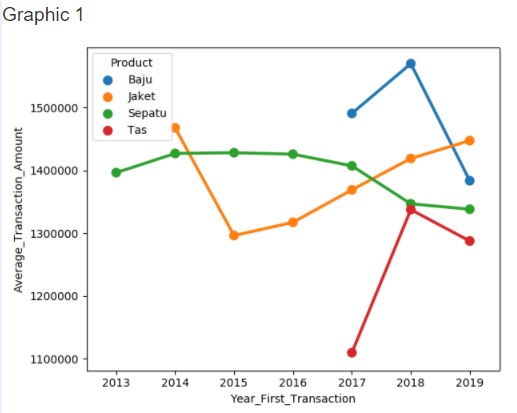


## Average transaction amount by year

Dengan menggunakan seaborn pointplot, visualisasikanlah tren dari tahun ke tahun rata-rata jumlah transaksi untuk tiap-tiap produknya.

Grafik yang diinginkan akan menjadi seperti ini:

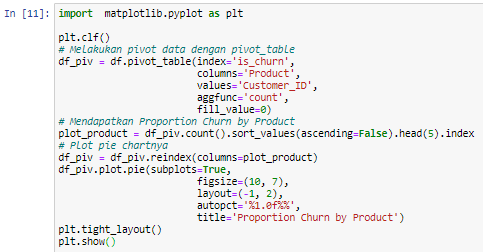
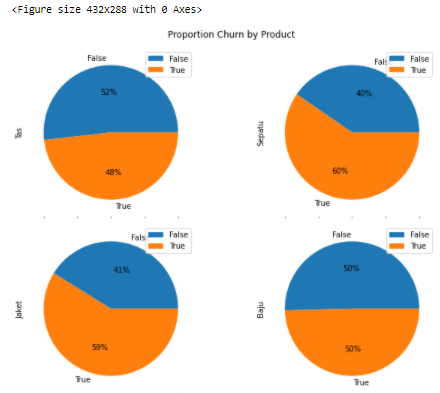




## Proporsi churned customer untuk setiap produk

Dari sisi churned customer, khususnya untuk melihat seberapa besar proporsi churned customer untuk tiap-tiap produk dapat diketahui insight-nya melalui pie chart. Visualisasikan pie chartnya untuk keempat produk yang dimaksudkan.

Pie chart yang diinginkan seperti berikut ini:



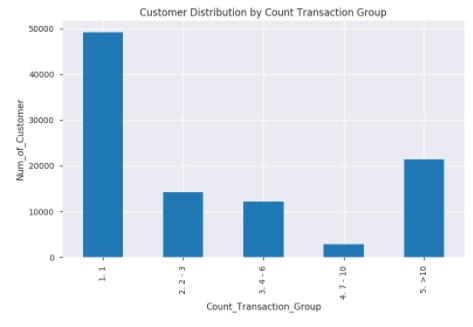
## Distribusi kategorisasi count transaction

Selanjutnya akan melakukan visualisasi dari distribusi kategorisasi count transaction. Kategorisasi ini dilakukan dengan mengelompokkan jumlah transaksi seperti yang diperlihatkan oleh tabel berikut:

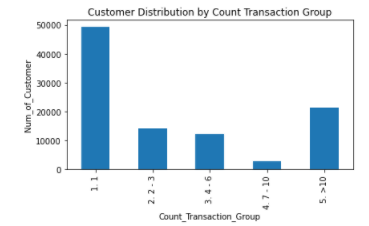


Setelah menambahkan kolom baru untuk kategori ini dengan nama Count\_Transaction\_Group, maka visualisasikanlah dengan bar chart.

Seperti berikut ini bar chart yang dikehendaki:







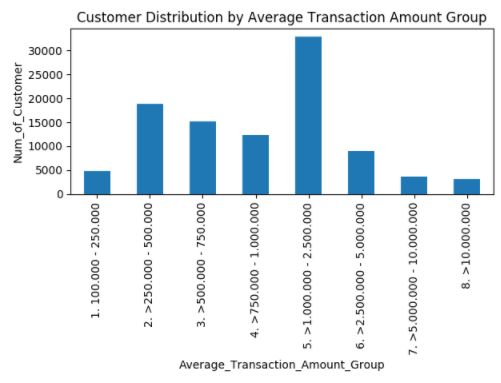
## Distribusi kategorisasi average transaction amount

Selanjutnya, akan melakukan visualisasi dari distribusi kategorisasi average transaction amount. Kategorisasi ini dilakukan dengan mengelompokkan rata-rata besar transaksi seperti yang diperlihatkan oleh tabel berikut:



Setelah ditambahkan kolom baru untuk kategori ini dengan nama Average\_Transaction\_Amount\_Group, maka visualisasikanlah dengan bar chart.

Seperti berikut ini bar chart yang dikehendaki:





# **Coding Test: Modelling**

## Feature Columns dan Target

Di bagian ini, selanjutnya akan menentukan feature columns dari dataset yang dimiliki, di sini dipilih kolom Average\_Transaction\_Amount, Count\_Transaction, dan Year\_Diff. Akan tetapi, kolom terakhir belum ada. Silakan dicreate dahulu kolom Year\_Diff ini dan kemudian assign dataset dengan feature columns ini sebagai variabel independent X.

Untuk target tentunya persoalan costumer dengan kondisi churn atau tidak, assign dataset untuk target ini ke dalam variabe dependent y.



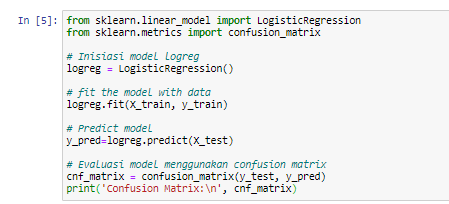
## Split X dan y ke dalam bagian training dan testing

Setelah variabel independent X dan variabel dependent y selesai dilakukan, maka pecahlah X dan y ke dalam bagian training dan testing. Bagian testing 25% dari jumlah entri data.



## Train, predict dan evaluate

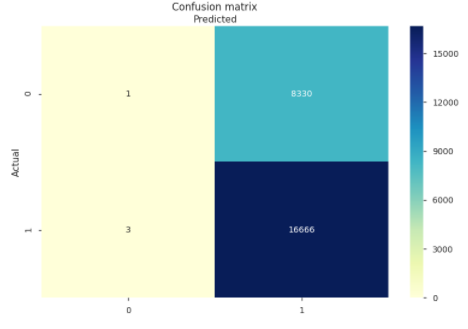
Langkah selanjutnya akan membuat model menggunakan Linear Regression, inisialisasilah model, fit, dan kemudian evaluasi model dengan menggunakan confusion matrix.

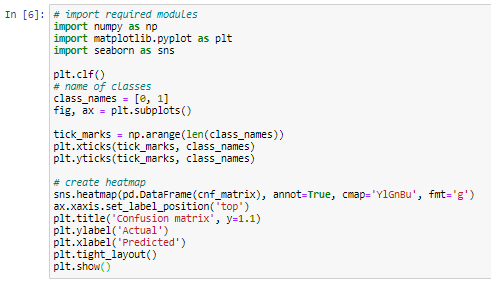


## Visualisasi Confusion Matrix

Confusion matrix yang telah dihitung sebelumnya dapat divisualisasikan dengan menggunakan heatmap dari seaborn. Untuk itu tampilkanlah visualisasi dari confusion matrix ini.

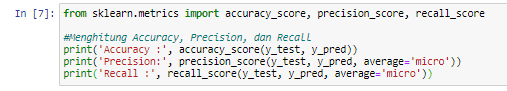
Tampilan dari confusion matrix yang diinginkan:





## Accuracy, Precision, dan Recall

Kemudian, hitunglah nilai accuracy, precission dan recall berdasarkan nilai target sesungguhnya dan nilai target hasil prediksi.





## Hasil belajarku dalam Data Analyst Career Track DQLab

Tidak terasa seluruh materi dalam Data Analyst Career Track telah selesai kupelajari dengan baik. Ini adalah permulaan agar aku dapat melangkah menjadi seorang data analyst.

Aku mengingat apa saja kompetensi yang telah aku capai ya dalam Data Analyst Track ini, yaitu

* Dapat menerapkan bahasa pemograman Python untuk menyelesaikan persoalan untuk analisis data
* Dapat menerapkan SQL dan sistem database relational agar dapat digunakan dalam mengolah data pada perusahaan yang telah menerapkan IT
* Dapat mengolah dataset skala kecil hingga besar
* Dapat menerapkan exploratory data analysis (EDA) untuk persoalan bisnis
* Dapat menghasilkan visualisasi data yang tepat sebagai representasi persoalan bisnis
* Dapat membuat dan menghasilkan model prediktif hingga menguji tingkat akurasi untuk memilih model yang tepat
* Dapat menerapkan teknik investigasi data-data yang memiliki anomali.