# Data Analyst Project: Business Decision Research

## Chapter 1

### Pendahuluan

#### Proyek dari Kroma

“Aksara, revisi dari Kroma sudah beres? Soalnya kamu dipanggil lagi sekarang ke ruangannya,” ujar Senja.

Aku melangkah menuju ruangan Kroma dengan perasaan campur aduk. Apakah ada revisi lagi? Sesampaiku di sana, Kroma sudah menantiku dengan senyum lebarnya. “Kita akan ada proyek baru. Saya memercayakan kamu untuk menanganinya. Ini sekaligus ujian pertama saya untukmu. Siap?”

Tidak mungkin aku menjawab tidak bukan? Kukumpulkan kepercayaan diriku dan mengangguk mantap.

#### 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.

1. 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.

## Chapter 2

### Theoritical Test

#### Pengantar

Jumat pagi, aku diundang oleh Senja untuk rapat kerja bersama Kroma. Rapat berlangsung cepat, kurang dari satu jam dan aku mendapat kepercayaan dari Kroma untuk menangani proyek market research. Jujur saja, ini tantangan baru buatku. Terlebih pekerjaan ini adalah rikues langsung dari Kroma.

“Sebelumnya, apakah kamu sudah di-brief oleh Senja terkait proyek ini? Apakah Senja sudah sempat membagikan padamu contoh yang pernah dilakukan senior tim terdahulu?” tanya Kroma saat rapat baru berakhir.

Aku mengangguk. “Sudah. Saya juga sudah mempelajari cara untuk coding-nya hingga penyajian visualisasi data di slide presentasi maupun dijadikan dashboard.”

“Baik, sepertinya sudah banyak yang sudah kamu pelajari di sini bersama Senja. Jadi, sudah siap untuk menangani proyek yang tadi saya sebutkan di rapat bukan? Saya ingin kamu melakukan churn analysis terhadap produk di salah satu cabang kita. Harapan saya adalah kamu bisa memberikan rekomendasi dan strategi untuk menurunkan churn dari pelanggan kita,” tukas Kroma seakan menguji kesiapanku.

“Pasti akan saya kerjakan dan berikan yang terbaik,” jawabku mantap.

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

**“**Aksara, tadi saya forward studi kasus untuk proyek market research dari Kroma. Sudah masuk?” tanya Senja padaku begitu aku sampai di meja.

“Iya, notifikasinya sudah masuk. Segera kukerjakan ya.”

Aku pun membuka isi email yang dimaksud:

**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:

Graphical user interface

Description automatically generated

Field yang ada pada data tersebut antara lain:

1. No
2. Row\_Num
3. Customer\_ID
4. Product
5. First\_Transaction
6. Last\_Transaction
7. Average\_Transaction\_Amount
8. Count\_Transaction

#### Market Research and Recommendation and Visualization Technique for Business Decision Making - Part 2

“Sudah kamu baca?” tanya Senja lagi.

Aku mengangguk seraya menyimak kembali isi studi kasus yang perlu kudapatkan solusinya.

“Jadi, 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.”

“Siap!” Dengan saran dari Senja, aku pun membuat catatan dari awal pengerjaan hingga selesai agar lebih sistematis untuk dilaporkan pada Kroma nanti, seperti ini:

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

#### Importing Data dan Inspection

Importlah dataset dari **https://storage.googleapis.com/dqlab-dataset/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:

Text

Description automatically generated

Apa yang dapat terpikirkan terkait dataset yang telah diimport dan diinspeksi?

Graphical user interface, text

Description automatically generated

Table

Description automatically generated

#### 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.

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface

Description automatically generated with medium confidence

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

A picture containing chart

Description automatically generated

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

1. transaksi paling terakhir kapan dilakukan
2. klasifikasikanlah mana customer yang berstatus churn dan mana yang tidak.

Setelah itu cetak lima data teratas dan informasi dataset.

Text

Description automatically generated

Graphical user interface, text

Description automatically generated

Table

Description automatically generated with medium confidence

#### Menghapus kolom yang tidak diperlukan

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

Graphical user interface, text, application

Description automatically generated

Jika metekan tombol  dan kemudian , maka setelah submit akan muncul **KeyError** berikut:

Text

Description automatically generated

Hal ini disebabkan kedua kolom tersebut telah dihapus dari dataframe **df** sewaktu **Run**, ketika **Submit** diminta kembali untuk menghapus kolom yang telah dihapus.

Bisa diabaikan saja, dan melanjutkan ke bagian berikutnya.

## Chapter 3

### 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.

Text

Description automatically generated

Bar chart yang diinginkan:

Chart, bar chart, histogram

Description automatically generated

#### Transaction by year

Visualisasikanlah trend jumlah transaksi per tahunnya dengan menggunakan bar chart.

Graphical user interface, text

Description automatically generated

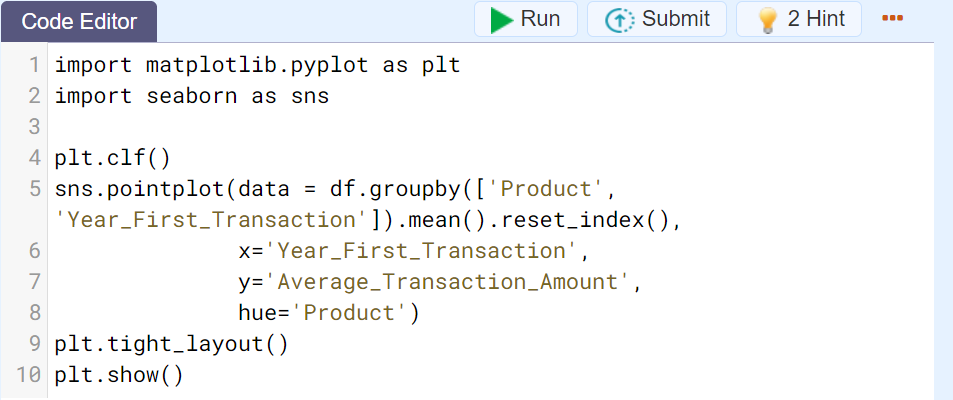
Bar chart yang diinginkan akan seperti ini:

Chart, bar chart

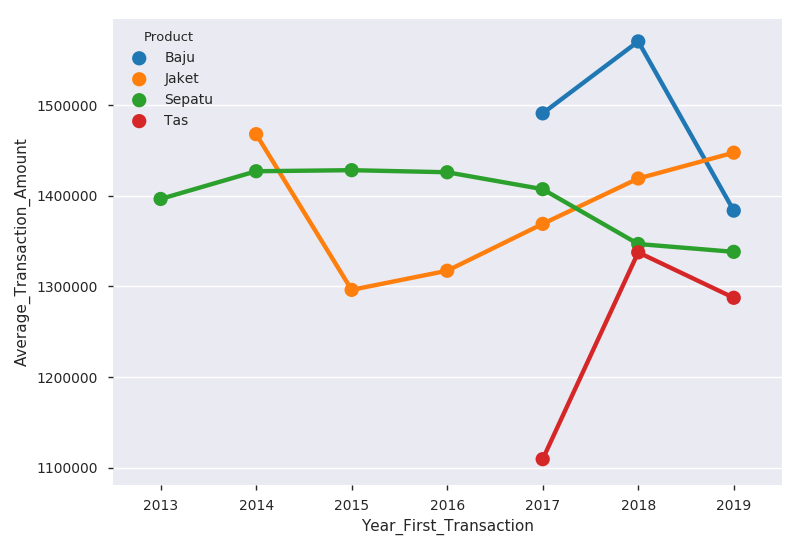
Description automatically generated

#### Average transaction amount by year

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

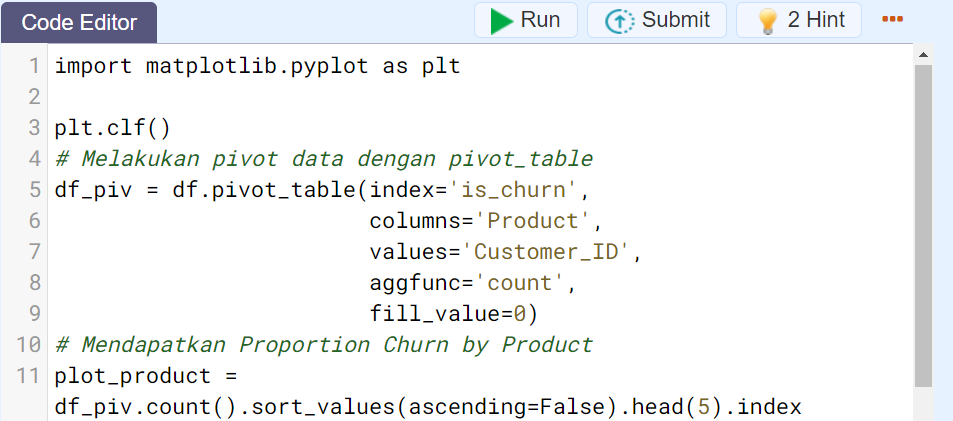


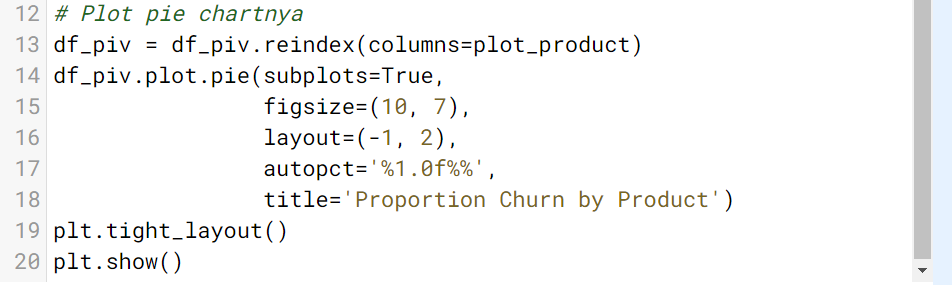
Grafik yang diinginkan akan menjadi seperti ini:



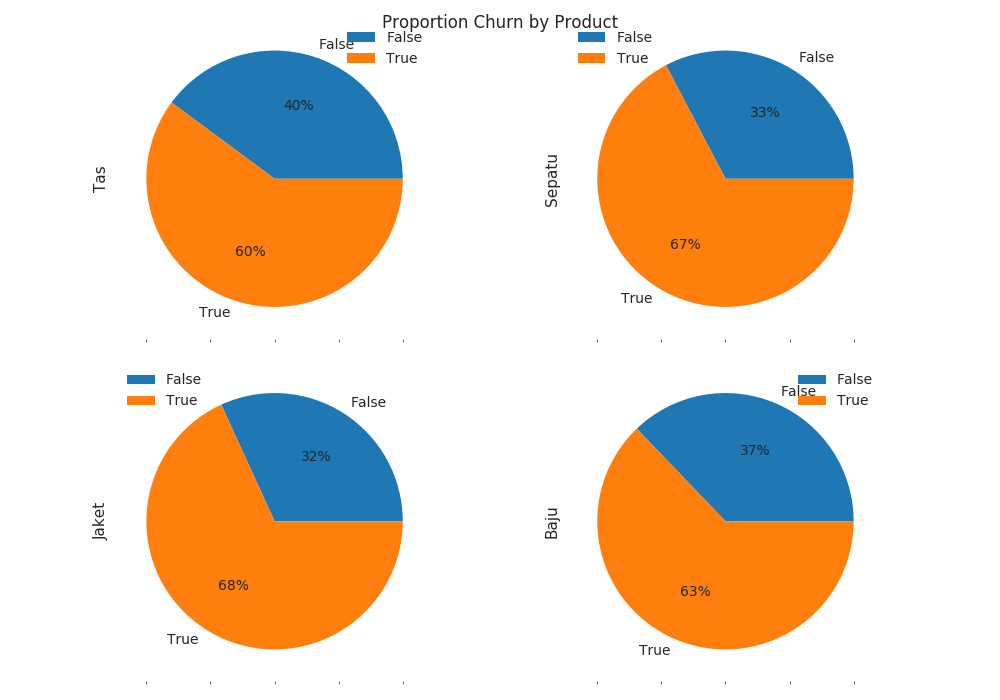
#### Proporsi churned customer untuk setiap produk

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:

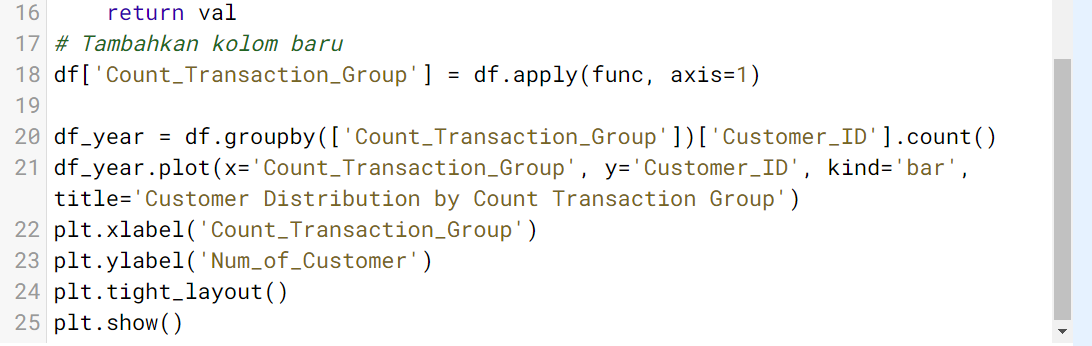
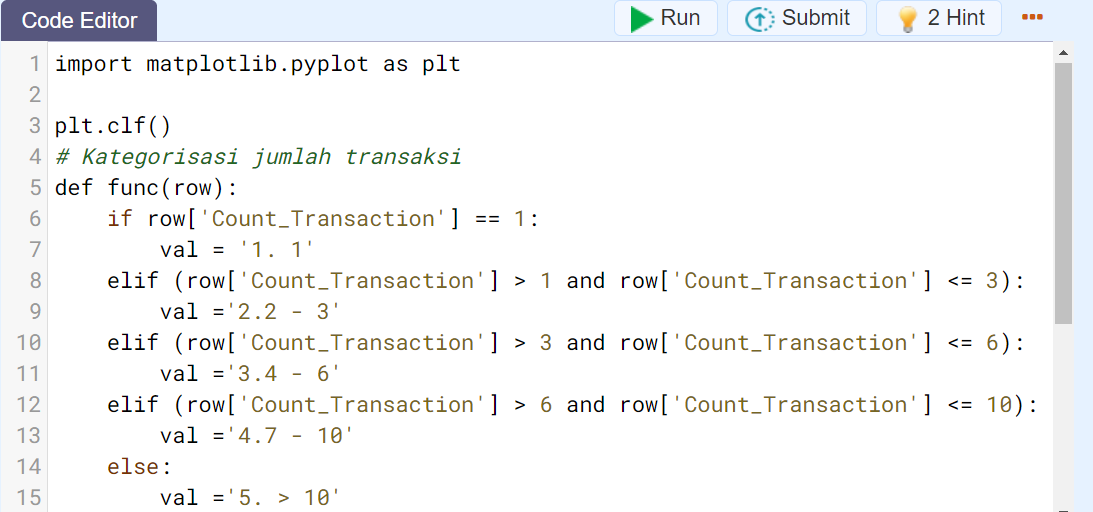


#### Distribusi kategorisasi count transaction

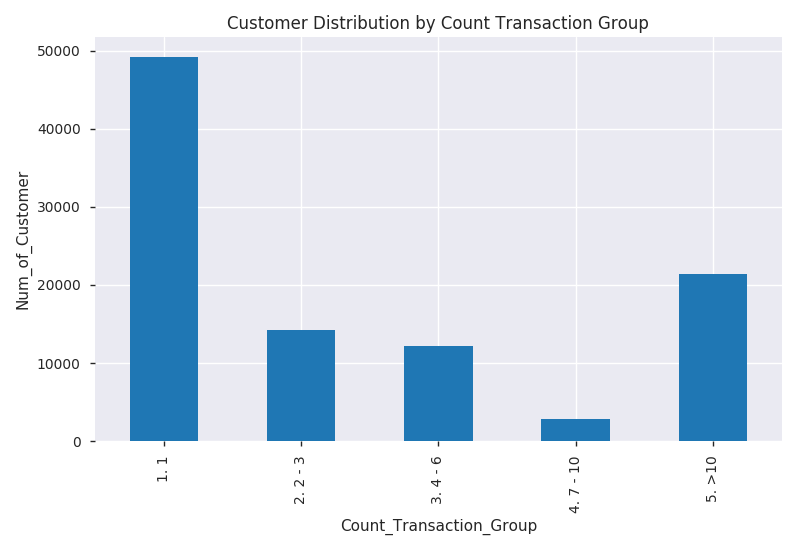
Selanjutnya akan melakukan visualisasi dari distribusi kategorisasi count transaction. Kategorisasi ini dilakukan dengan mengelompokkan jumlah transaksi seperti yang diperlihatkan oleh tabel berikut:

|  |  |
| --- | --- |
| **Rentang jumlah transaksi** | **Kategori** |
| s/d 1 | 1. 1 |
| 2 s/d 3 | 2. 2 - 3 |
| 4 s/d 6 | 3. 4 - 6 |
| 7 s/d 10 | 4. 7 - 10 |
| > 10 | 5. > 10 |

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



Seperti berikut ini bar chart yang dikehendaki:

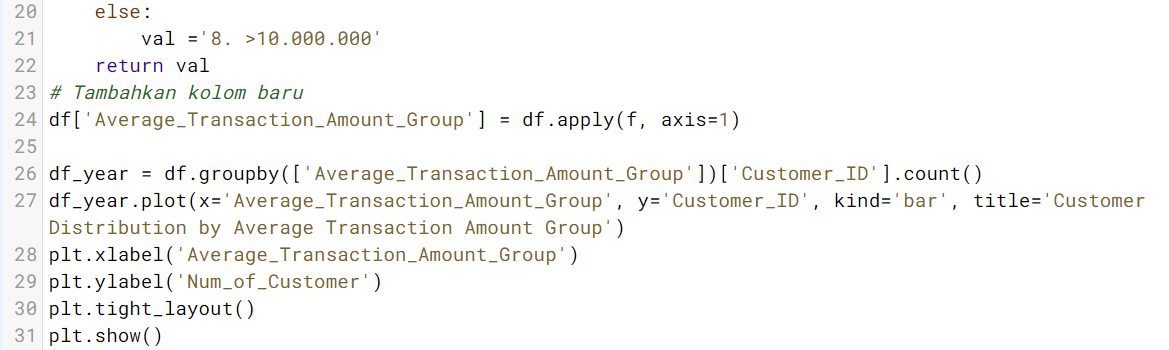
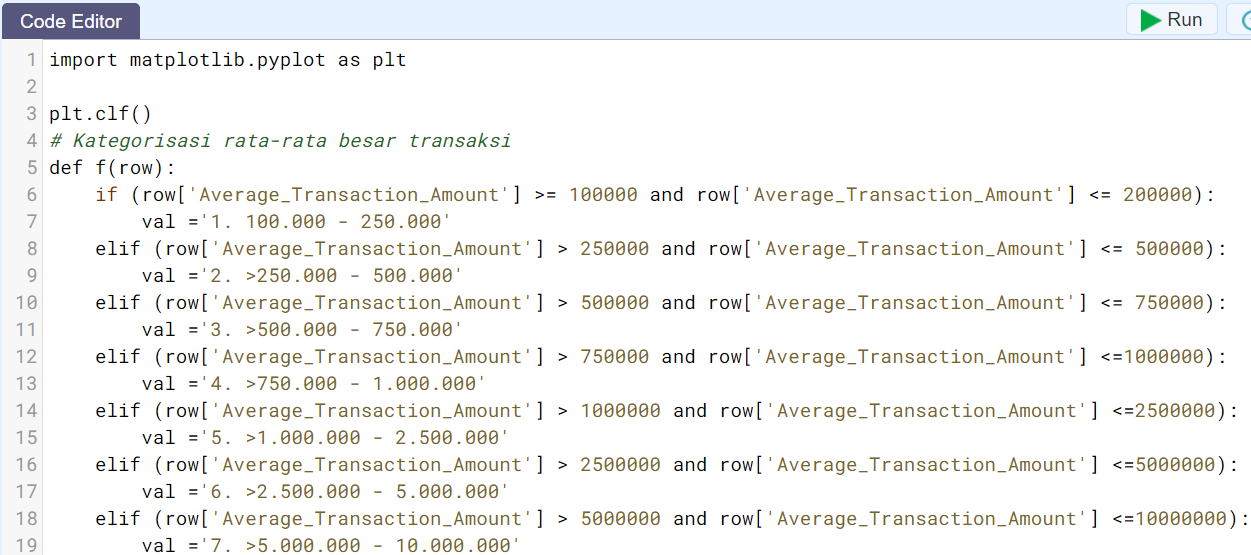


#### Distribusi kategorisasi average transaction amount

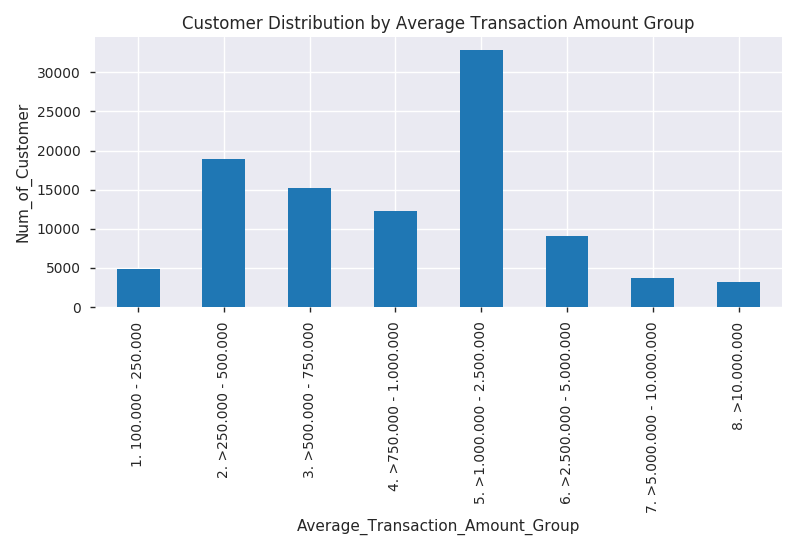
Selanjutnya, akan melakukan visualisasi dari distribusi kategorisasi average transaction amount. Kategorisasi ini dilakukan dengan mengelompokkan rata-rata besar transaksi seperti yang diperlihatkan oleh tabel berikut:

|  |  |
| --- | --- |
| **Rentang rata-rata besar transaksi** | **Kategori** |
| 100.000 s/d 200.000 | 1. 100.000 - 250.000 |
| >250.000 s/d 500.000 | 2. >250.000 - 500.000 |
| >500.000 s/d 750.000 | 3. >500.000 - 750.000 |
| >750.000 s/d 1.000.000 | 4. >750.000 - 1.000.000 |
| >1.000.000 s/d 2.500.000 | 5. >1.000.000 - 2.500.000 |
| >2.500.000 s/d 5.000.000 | 6. >2.500.000 - 5.000.000 |
| >5.000.000 s/d 10.000.000 | 7. >5.000.000 - 10.000.000 |
| >10.000.000 | 8. >10.000.000 |

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:



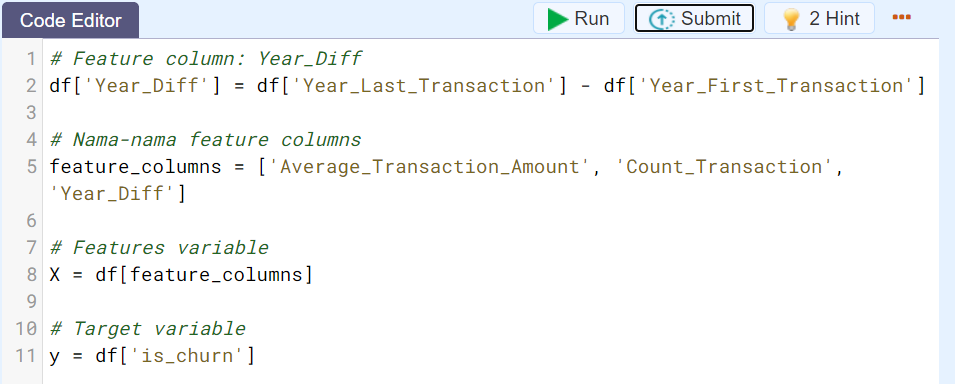
## Chapter 4

### Coding Test: Modelling

#### Feature Columns dan Target

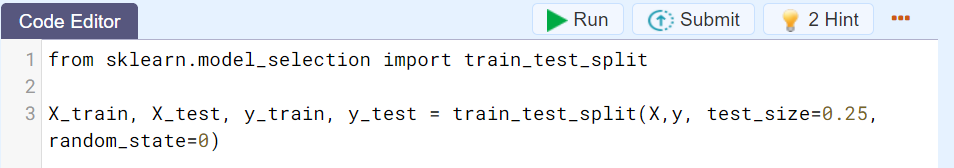
Di bagian ini, selanjutnya akan menentukan feature columns dari dataset yang dimiliki, di sini dipilih kolom Average\_Transaction\_Amount, Count\_Transaction, dan Year\_Diff. Akan tetapi, kolom terakhir belum ada. Silakan dicreate dahulu kolom Year\_Diff ini dan kemudian assign dataset dengan feature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



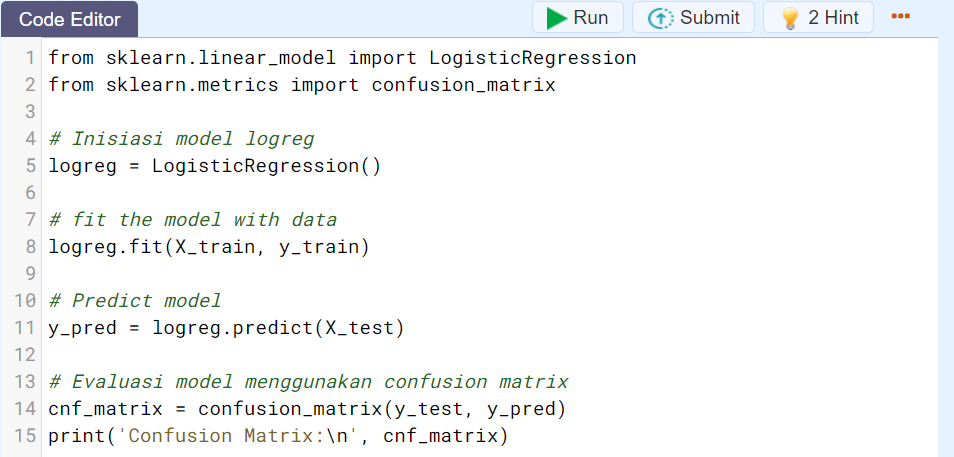
#### Split X dan y ke dalam bagian training dan testing

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.



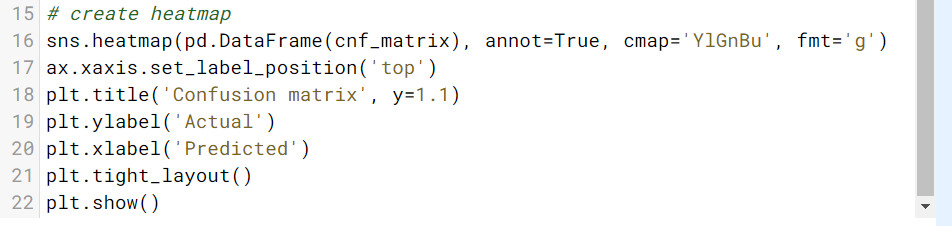
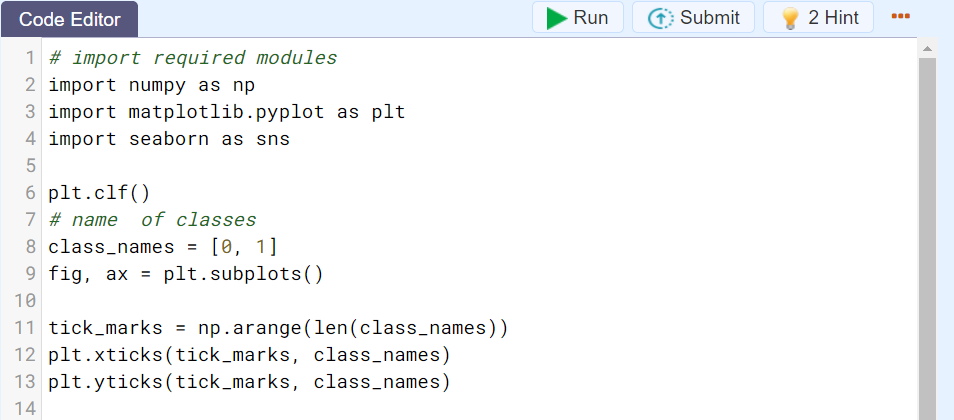
#### Train, predict dan evaluate

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

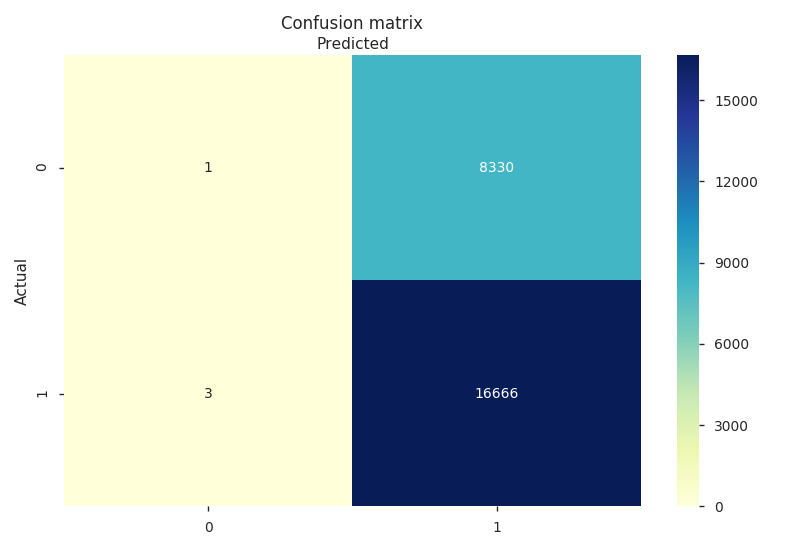


#### Visualisasi Confusion Matrix

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

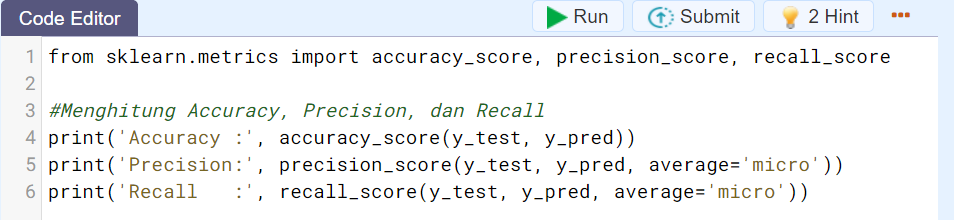


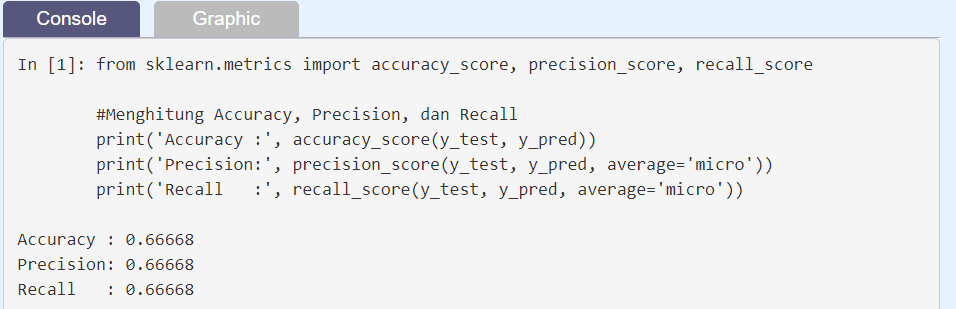
Tampilan dari confusion matrix yang diinginkan:



#### Accuracy, Precision, dan Recall

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





## Chapter 6

### Penutup

#### Ingatan akan perkataan Senja

“Fiuh! Selesai juga untuk studi kasus proyek ini. Bener juga kata Senja, dengan bikin catatan proses kerja dari awal sampai akhir seperti ini, aku jadi bisa meninjau ulang pekerjaanku juga.”

Aku pun merapikan catatanku untuk kuperlihatkan dulu kepada Senja sebelum menghadap Kroma. Sejauh ini sih, lancar dan sukses!

#### 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

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