**Tugas Kelompok 4 – Kelompok 4**

Persentase Anggota kelompok:

* Mahatmaditya Favian Rachman Syarief – 2440103546 – **20%**
* Muh. Rizal Sakti Djatmieka – 24400099946 – **20%**
* Ramadhani Al Amin – 2440103546 – **20%**
* Gede Dyava Savitra – 2440100683 – **20%**
* Bachatsa Taqiyya – 2440071733 – **20%**

**Penerapan K-Means**

Dataset: Online Retail

Link GitHub:

Penjelasan teknik penerapan K-means:

1. Melakukan Import Library yang dibutuhkan

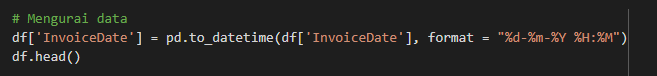
Text

Description automatically generated

Pertama yaitu mengimport *warnings* untuk mengabaikan/menekan semua peringatan yang akan keluar ketika kita menjalankan codenya. Lalu, terdapat *numpy* dan *pandas* yaitu untuk perhitungan segala komputasi, memproses data, melakukan pembersihan, dan lain-lain. Bagian ketigas yaitu *matplotlib.pyplot* dan *seaborn* untuk melakukan visualisasi untuk pemrosesan akhir. Dan terakhir yaitu terdapat *sklearn.preprocessing dan sklearn.cluster* untuk menskalakan data dan untuk melakukan *K-Means clustering.*

1. Pembersihan data / *data cleaning*

* Dilakukannya penguraian data dan pergantian format.



* Pemeriksaan apakah terdapat nilai *null* atau tidak.

Text

Description automatically generated with medium confidence

Ternyata terdapat nilai *null* pada *dataset*.

* Melihat ada berapa nilai null yang terkandung dan kemudian menghapusnya.

Graphical user interface, text

Description automatically generated

* Cek kembali apakah masih terdapat nilai *null*.

Graphical user interface, text

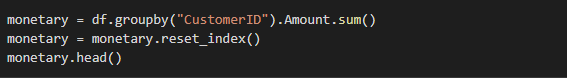
Description automatically generated

1. Melakukan Customer Segmentation dengan menggunakan RFM untuk menargetkan kelompok customer tertentu dengan komunikasi yuang jauh lebih relevean untuk perilaku khusus mereka. Dengan car aini dapat menghasilkan tingkat respons yang jauh lebih ditinggi dan menambahkan peningkatan loyalitas dan nilai seumur hidup customer.

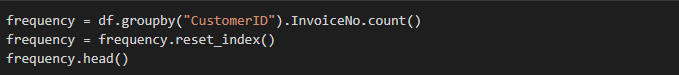
* Mengekstraksi jumlah dengan mengalikan kuantitas dan harga satuan dan menyimpan data ke dalam variabel jumlah.



* **Monetery** value - Menemukan Total jumlah yang dihabiskan setiap customer



* Mendapatkan jumlah pesanan yang dibuat oleh setiap pelanggan berdasarkan ID customer.



* Membuat main dataset untuk menggapungkan kolom Amount dan Frequency.



* Mencari data terbesar



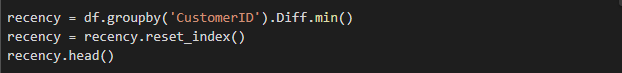
Lalu, menambahkan satu hari pada data *maximum* agar memiliki selisih bernilai satu.



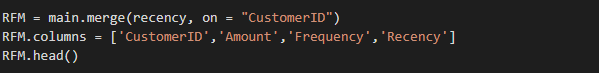
Melakukan pengurangan *maximum* dengan kolom *invoiceDate*.



* Menggabungkan *Dataframe* berdasarkan *recency*



* Menggabungkan *Dataframe* dari semua parameter *recency, frequency,* dan *monetary*.



1. Menskalakan data RFM dengan menggunakan *library StandardScaler*.

Text

Description automatically generated

1. Menerapkan *K-Means Algorithm*

* Menggunakan *silhouette scores* untuk mengevaluasi seberapa bai sampel dikelompokkan dengan sampel lain yang mirip satu sama lain.

Text

Description automatically generated

* Menganalisa cluster yang sudah terbentuk.

Text

Description automatically generated

A screenshot of a computer

Description automatically generated with medium confidence

1. Memvisualisasikan hasil K-Means dengan menggunakan *barplot*.

Text

Description automatically generated

Results:

Chart, bar chart

Description automatically generated