### DATA VISUALISASI

CODE :

# Import library yang dibutuhkan

import datetime

import pandas as pd

import matplotlib.pyplot as plt

# Baca dataset https://storage.googleapis.com/dqlab-dataset/retail\_raw\_reduced.csv

dataset = pd.read\_csv('https://storage.googleapis.com/dqlab-dataset/retail\_raw\_reduced.csv')

# Buat kolom order\_month

dataset['order\_month'] = dataset['order\_date'].apply(lambda x: datetime.datetime.strptime(x, "%Y-%m-%d").strftime('%Y-%m'))

# Buat kolom gmv

dataset['gmv'] = dataset['item\_price'] \* dataset['quantity']

# Plot grafik sesuai dengan instruksi

plt.figure(figsize=(10, 5))

dataset[dataset['order\_month']=='2019-12'].groupby(['order\_date'])['customer\_id'].nunique().plot(color='red', marker='.', linewidth=2)

plt.title('Daily Number of Customers - December 2019', loc='left', pad=20, fontsize=20, color='orange')

plt.xlabel('Order Date', fontsize=15, color='blue')

plt.ylabel('Number of Customers', fontsize=15, color='blue')

plt.grid(color='darkgray', linestyle=':', linewidth=0.5)

plt.ylim(ymin=0)

plt.show()

### EKSPLORASI DATA

Aksara, bisa tolong bantu mengurus beberapa data penjualan dari dataset order.csv? Saya sedang rapat dan bahan ini ditunggu dalam pembahasan cabang supermarket kita. Berikut ya detailnya:

1. Median price yang dibayar customer dari masing-masing metode pembayaran.
2. Tentukan metode pembayaran yang memiliki basket size (rataan median price) terbesar.
3. Ubah freight\_value menjadi shipping\_cost dan cari shipping\_cost termahal dari data penjualan tersebut menggunakan sort.
4. Untuk setiap product\_category\_name, berapa rata-rata weight produk tersebut dan standar deviasi mana yang terkecil dari weight tersebut,
5. Buat histogram quantity penjualan dari dataset tersebut untuk melihat persebaran quantity penjualan tersebut dengan bins = 5 dan figsize= (4,5)

Khusus poin 4, tolong diperhatikan lebih ya, Aksara karena hasil analisisnya akan digunakan kepala cabang dalam menyusun strategi free ongkir.

Kubalas email itu segera, *OK! Hasilnya akan selesai sebelum makan siang ya. You can count on me, hehehe.*

*Perhatian*: Semua string dinyatakan dalam kutipan "...".

CODE :

import pandas as pd

import matplotlib.pyplot as plt

order\_df = pd.read\_csv("https://storage.googleapis.com/dqlab-dataset/order.csv")

# Median price yang dibayar customer dari masing-masing metode pembayaran.

median\_price = order\_df["price"].groupby(order\_df["payment\_type"]).median()

print(median\_price)

# Ubah freight\_value menjadi shipping\_cost dan cari shipping\_cost

# termahal dari data penjualan tersebut menggunakan sort.

order\_df.rename(columns={"freight\_value": "shipping\_cost"}, inplace=True)

sort\_value = order\_df.sort\_values(by="shipping\_cost", ascending=0)

print(sort\_value)

# Untuk product\_category\_name, berapa rata-rata weight produk tersebut

# dan standar deviasi mana yang terkecil dari weight tersebut,

mean\_value = order\_df["product\_weight\_gram"].groupby(order\_df["product\_category\_name"]).mean()

print(mean\_value.sort\_values())

std\_value = order\_df["product\_weight\_gram"].groupby(order\_df["product\_category\_name"]).std()

print(std\_value.sort\_values())

# Buat histogram quantity penjualan dari dataset tersebutuntuk melihat persebaran quantity

# penjualan tersebut dengan bins = 5 dan figsize= (4,5)

order\_df[["quantity"]].hist(figsize=(4, 5), bins=5)

plt.show()