

Mini Project

“Pertemuan 14 Project1 Komputer Grafis”

Muhammad Rizal

G.211.22.0107

Kls B2

Soal !!!



Project 1

Mini Project

Buatlah sebuah line chart dengan menggunakan matplotlib yang menunjukkan jumlah pembeli harian (daily number of customers) selama bulan Desember.

Studi Kasus dari Senja: Daily number of customers on Desember

Dengan menggunakan dataset yang sama ('https://dqlab-dataset.s3-ap-southeast-1.amazonaws.com/retail_raw_reduced.csv'), buatlah sebuah line chart dengan Matplotlib, yang menunjukkan jumlah pembeli harian (daily number of customers) selama bulan Desember.

Beberapa spesifikasi yang harus diperhatikan:

- ✓ Ukuran figure adalah 10x5
- ✓ Sumbu-x adalah tanggal pembelian, dari tanggal 1–31 Desember 2019
- ✓ Sumbu-y adalah jumlah unique customers di tiap tanggal
- ✓ Title dan axis label harus ada, tulisan dan style-nya silakan disesuaikan sendiri

Case 1: Menentukan brand top 5

Lakukan analisis untuk top brands di bulan Desember 2019, di bawah ini adalah beberapa hal yang perlu dilakukan. Silahkan ditambahkan dengan visualisasi-visualisasi lain yang kira-kira dapat memberikan insight tambahan.

Case 1:

Buat variabel baru (e.g. `dataset_top5brand_dec`) yang berisi data penjualan bulan Desember 2019, hanya untuk top 5 brand dengan quantity terjual terbanyak selama bulan Desember 2019. Sebutkan top 5 brands tersebut! Untuk visualisasi-visualisasi selanjutnya, hanya gunakan data frame ini.

Case 2: Multi-line chart daily quantity untuk brand top 5

Buat visualisasi multi-line chart untuk daily quantity terjualnya, breakdown per brand. Maka, akan terlihat 1 tanggal di mana ada salah satu brand yang mengalami lonjakan (quantity lebih tinggi dari tanggal-tanggal lain). Beri anotasi untuk titik lonjakan tersebut.

Case 3:

Cari tahu jumlah product untuk masing-masing brand yang laku selama bulan Desember 2019. Gunakan barchart untuk visualisasinya, urutkan dengan yang kiri adalah brand dengan product lebih banyak.

Case 4: Penjualan produk diatas 100 dan dibawah 100 selama Desember 2019

Gunakan stacked chart, untuk breakdown barchart yang di Case 3, antara product yang terjual ≥ 100 dan < 100 di bulan Desember 2019. Apakah ada pola yang menarik?

Case 5: Murah atau mahalkah harga produk brand top 5

Gunakan histogram untuk melihat distribusi harga product-product yang ada di top 5 brand tersebut (untuk tiap product_id, ambil median harganya). Bagaimana persebaran harga product nya? Cenderung banyak yang murah atau yang mahal?

Case 6a: Korelasi quantity vs GMV

Untuk setiap product_id, cek scatterplot antara quantity dan GMV, apakah ada korelasi? Bagaimana dengan median harga vs quantity? Apakah product yang murah cenderung dibeli lebih banyak?

Case 6b: Korelasi median harga vs quantity

Untuk setiap product_id, cek scatterplot antara quantity dan GMV sudah kamu lakukan pada Case 6a? Untuk Case 6b ini, bagaimanakah dengan median harga vs quantity? Apakah product yang murah cenderung dibeli lebih banyak?

Kesimpulan Case Study:

Dapat disimpulkan bahwa dari kelima top brand, brand S memiliki kuantitas top brands yang paling besar sebesar 2.197. Kemudian, Lonjakan yang cukup drastis ditunjukkan oleh brand P pada tanggal 6 Desember 2019 sebesar 300. Dimana lonjakan ini lebih tinggi sehingga perlu diberikan anotasi. Kemudian, dilihat dari sisi kuantitas produk yang laku terjual, brand S lebih banyak kuantitasnya. Berdasarkan stacked bar-chart top 5 brands, terdapat pola yang menarik. Brand S lebih banyak menjual kuantitas produknya dibawah 100. Sedangkan, brand P lebih banyak menjual kuantitas lebih dari sama dengan 100. Terdapat hubungan positif antara kuantitas dengan GMV yang artinya semakin banyak customer yang membeli produk dalam jumlah yang besar, dapat disimpulkan GMV yang dihasilkan juga akan semakin besar. Terdapat hubungan pula antara median harga dengan kuantitas produk. Hal tersebut berkorelasi positif yang artinya bahwa semakin murah harga suatu produk maka customer berpotensi besar untuk membeli jumlah produk dengan jumlah yang banyak.

Ketentuan:

1. Buatlah dalam 1 file python dengan google colab dan hasilnya disimpan dalam file gambar dengan format jpg, png dll
2. Buatlah akun di github (github.com), upload semua coding dan dataset di github dengan nama MiniProject_NIM.ipynb, MiniProject_NIM.jpg dan file datasetnya
3. Upload link github Sdr. di elearning dibagian Pertemuan 14. Project 1

Jawab!!!

```
#Import Library Terlebih dahulu
import datetime
import pandas as pd
import matplotlib.pyplot as plt
```

```
#Membaca Dari Dataset
ngambil_data = pd.read_csv('https://dqlab-dataset.s3-ap-southeast-
1.amazonaws.com/retail_raw_reduced.csv')
```

```
#Membuat kolom order month
ngambil_data['order_month'] = ngambil_data['order_date'].apply(lambda x:
datetime.datetime.strptime(x, "%Y-%m-%d").strftime('%Y-%m'))
```

```
#Membuat kolom gmv
ngambil_data['gmv'] = ngambil_data['item_price'] *
ngambil_data['quantity']
```

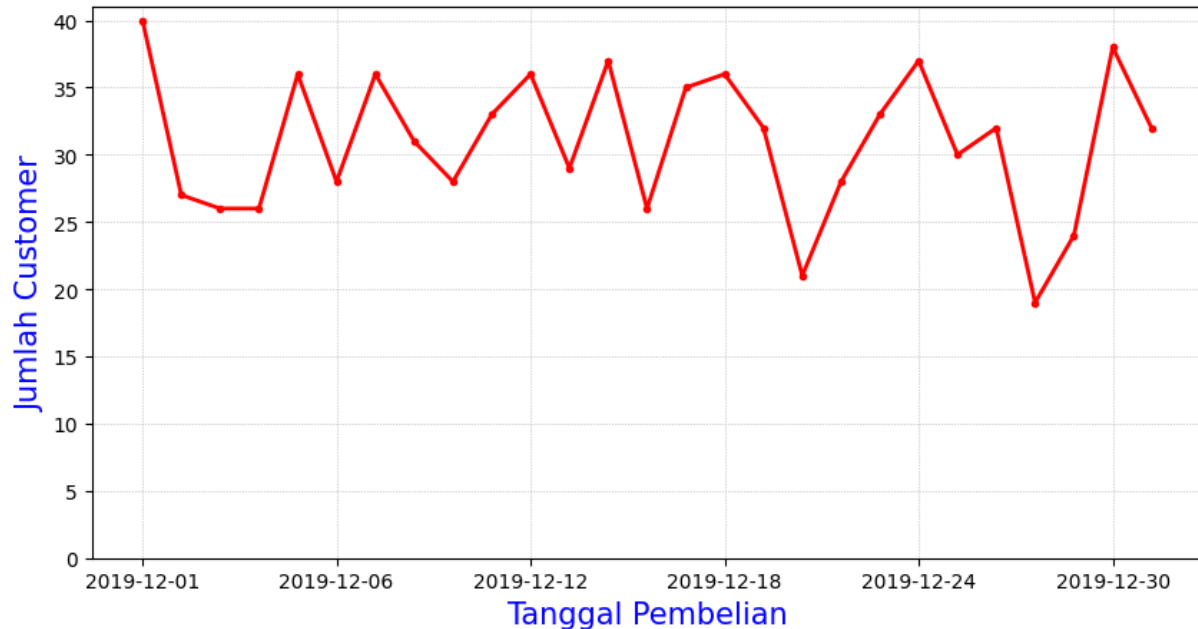
Input :

```
#Membuat plot grafik, sesuai intruksi di soal
plt.figure(figsize=(10, 5))
ngambil_data[ngambil_data['order_month']=='2019-
12'].groupby(['order_date'])['customer_id'].unique().plot(color='red',
marker='.', linewidth=2)
plt.title('daily number of customers - December 2019', loc='left',
pad=30, fontsize=20, color='orange')
plt.xlabel('Tanggal Pembelian', fontsize=15, color='blue')
plt.ylabel('Jumlah Customer', fontsize=15, color='blue')
plt.grid(color='darkgray', linestyle=':', linewidth=0.5)
plt.ylim(ymin=0)
```

Output :

(0.0, 41.05)

daily number of customers - December 2019



Case 1

Buat variabel baru (e.g. dataset_top5brand_dec) yang berisi data penjualan bulan Desember 2019, hanya untuk top 5 brand dengan quantity terjual terbanyak selama bulan Desember 2019. Sebutkan top 5 brands tersebut! Untuk visualisasi-visualisasi selanjutnya, hanya gunakan data frame ini.

Input :

```
#mengambil informasi top 5 brands berdasarkan quantity
merek_teratas = (ngambil_data[ngambil_data['order_month']=='2019-12'].groupby('brand')['quantity']
                  .sum()
                  .reset_index()
                  .sort_values(by='quantity',ascending=False)
                  .head(5))

#membuat dataframe baru, filter hanya di bulan Desember 2019 dan hanya top 5 brands
dataset_top5brand_dec =
ngambil_data[(ngambil_data['order_month']=='2019-12') &
              (ngambil_data['brand'].isin(merek_teratas['brand'].to_list()))]
# print top brands
print(merek_teratas)
```

Output :

```
brand  quantity
8  BRAND_S      2197
6  BRAND_P      2194
7  BRAND_R      1508
2  BRAND_C      1365
0  BRAND_A      1315
```

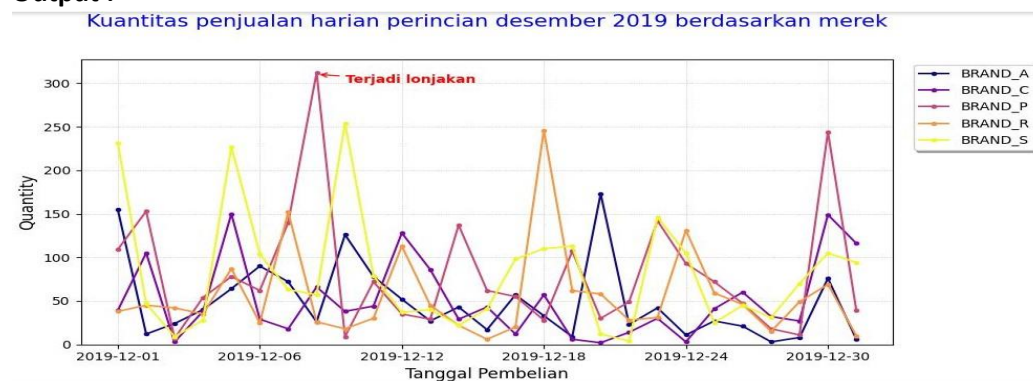
Case 2

Buat visualisasi multi-line chart untuk daily quantity terjualnya, breakdown per brand. Maka, akan terlihat 1 tanggal di mana ada salah satu brand yang mengalami lonjakan (quantity lebih tinggi dari tanggal-tanggal lain). Beri anotasi untuk titik lonjakan tersebut.

Input :

```
dataset_top5brand_dec.groupby(['order_date', 'brand'])['quantity'].sum().
unstack().plot(marker='.', cmap='plasma')
plt.title('Kuantitas penjualan harian perincian desember 2019
berdasarkan merek',loc='center',pad=30, fontsize=15, color='blue')
plt.xlabel('Tanggal Pembelian', fontsize = 12)
plt.ylabel('Quantity',fontsize = 12)
plt.grid(color='darkgray', linestyle=':', linewidth=0.5)
plt.ylim(ymin=0)
plt.legend(loc='upper center', bbox_to_anchor=(1.1, 1), shadow=True,
ncol=1)
plt.annotate('Terjadi lonjakan', xy=(7, 310), xytext=(8, 300),
            weight='bold', color='red',
            arrowprops=dict(arrowstyle='->',
                            connectionstyle="arc3",
                            color='red'))
plt.gcf().set_size_inches(10, 5)
plt.tight_layout()
```

Output :



Case 3

Cari tahu jumlah product untuk masing-masing brand yang laku selama bulan Desember 2019. Gunakan barchart untuk visualisasinya, urutkan dengan yang kiri adalah brand dengan product lebih banyak.

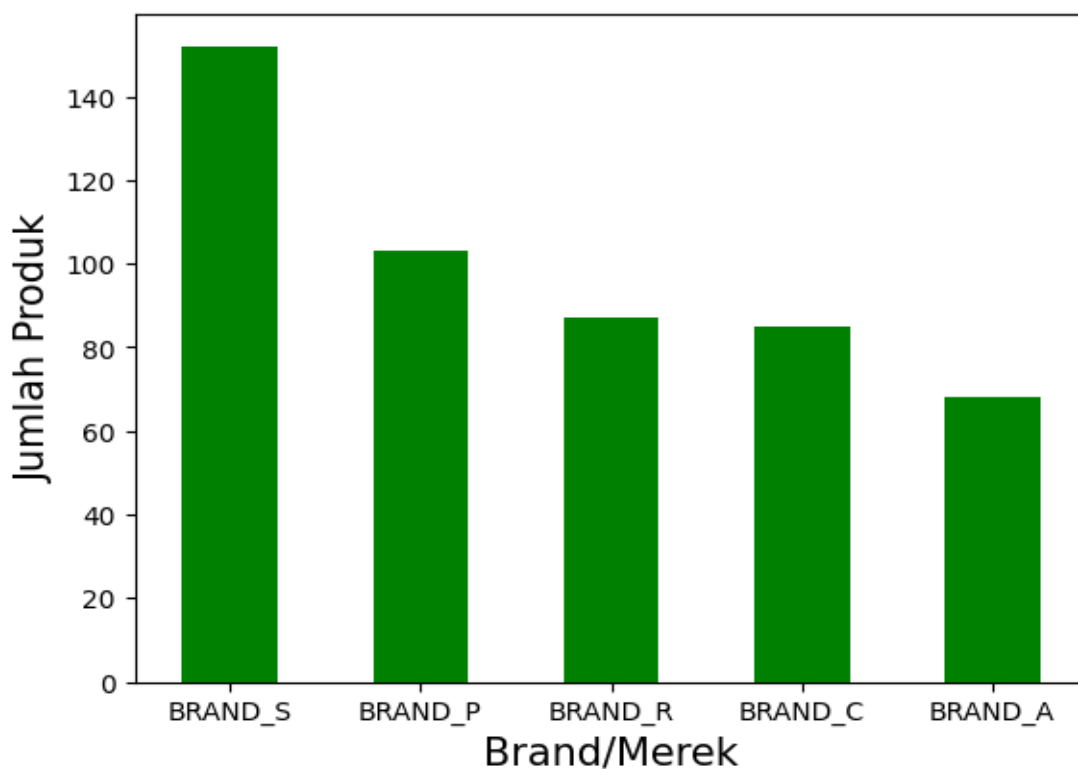
Input :

```
plt.clf()
dataset_top5brand_dec.groupby('brand')['product_id'].nunique().sort_values(ascending=False).plot(kind='bar', color='green')
plt.title('Jumlah produk terjual per merek, December 2019', loc='center', pad=30, fontsize=15, color='blue')
plt.xlabel('Brand/Merek', fontsize = 15)
plt.ylabel('Jumlah Produk', fontsize = 15)
plt.ylim(ymin=0)
plt.xticks(rotation=0)
```

Output :

```
(array([0, 1, 2, 3, 4]),
 [Text(0, 0, 'BRAND_S'),
  Text(1, 0, 'BRAND_P'),
  Text(2, 0, 'BRAND_R'),
  Text(3, 0, 'BRAND_C'),
  Text(4, 0, 'BRAND_A')])
```

Jumlah produk terjual per merek, December 2019



Case 4

Penjualan produk diatas 100 dan dibawah 100 selama Desember 2019 Gunakan stacked chart, untuk breakdown barchart yang di Case 3, antara product yang terjual ≥ 100 dan < 100 di bulan Desember 2019. Apakah ada pola yang menarik?

Input :

```
#Membuat dataframe baru, untuk agregat jumlah quantity terjual per
product
dataset_top5brand_dec_per_product =
dataset_top5brand_dec.groupby(['brand', 'product_id'])['quantity'].sum().
reset_index()

#beri kolom baru untuk menandai product yang terjual  $\geq 100$  dan  $< 100$ 
dataset_top5brand_dec_per_product['quantity_group'] =
dataset_top5brand_dec_per_product['quantity'].apply(lambda x: '>= 100'
if x>=100 else '< 100')
dataset_top5brand_dec_per_product.sort_values('quantity',ascending=False
,inplace=True)

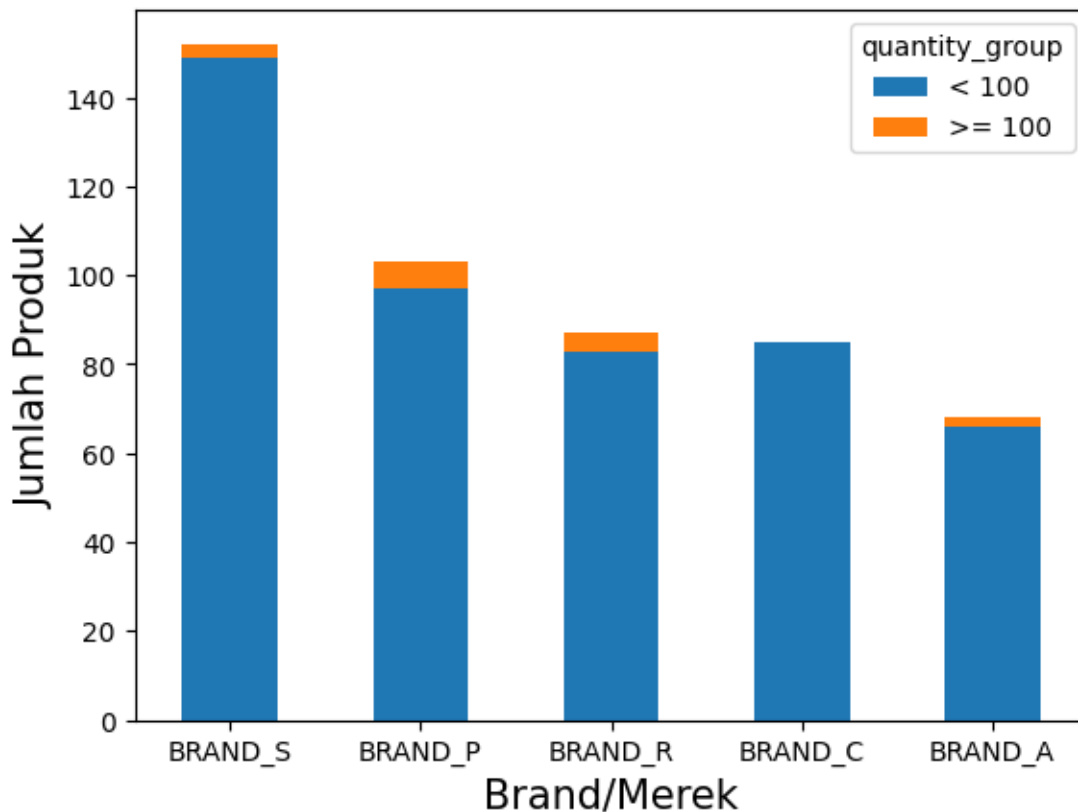
#membuat referensi pengurutan brand berdasarkan banyaknya semua product
s_sort =
dataset_top5brand_dec_per_product.groupby('brand')['product_id'].nunique
().sort_values(ascending=False)

#plot stacked barchart
dataset_top5brand_dec_per_product.groupby(['brand', 'quantity_group'])['p
roduct_id'].nunique().reindex(index=s_sort.index,
level='brand').unstack().plot(kind='bar', stacked=True)
plt.title('Jumlah produk terjual per merek, December
2019',loc='center',pad=30, fontsize=15, color='blue')
plt.xlabel('Brand/Merek', fontsize = 15)
plt.ylabel('Jumlah Produk',fontsize = 15)
plt.ylim(ymin=0)
plt.xticks(rotation=0)
```

Output :

```
(array([0, 1, 2, 3, 4]),
 [Text(0, 0, 'BRAND_S'),
  Text(1, 0, 'BRAND_P'),
  Text(2, 0, 'BRAND_R'),
  Text(3, 0, 'BRAND_C'),
  Text(4, 0, 'BRAND_A')])
```

Jumlah produk terjual per merek, December 2019



Murah atau mahalkah harga produk brand top 5

Case 5

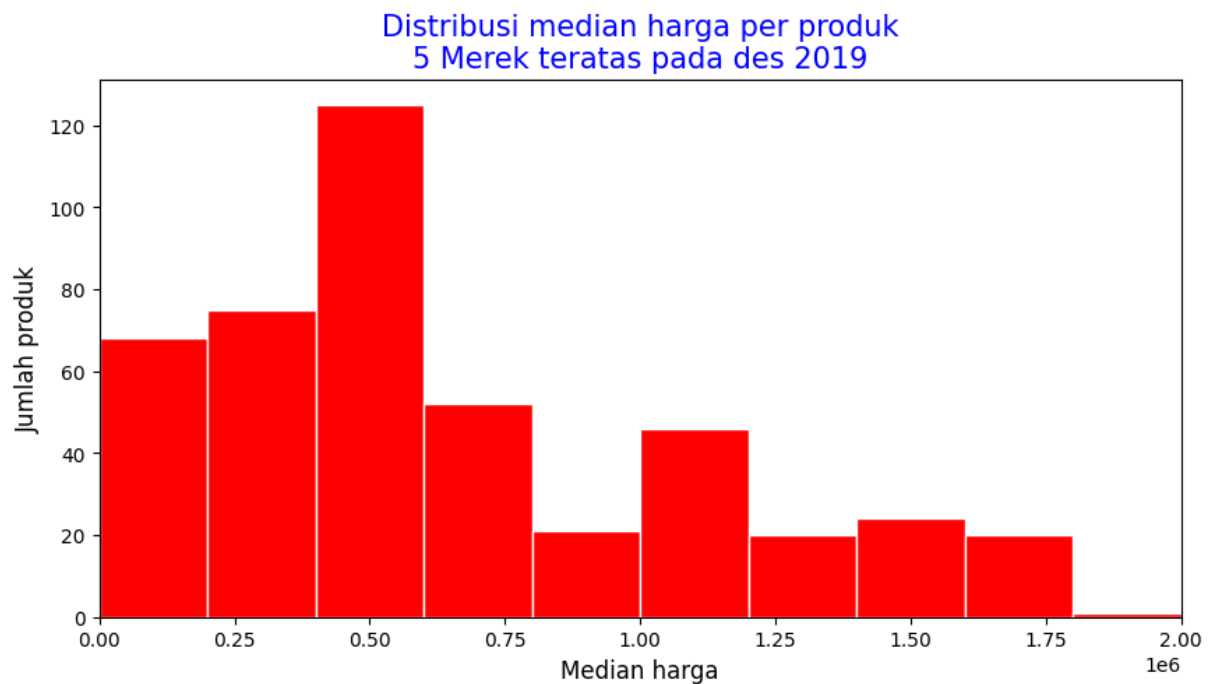
Gunakan histogram untuk melihat distribusi harga product-product yang ada di top 5 brand tersebut (untuk tiap product_id, ambil median harganya). Bagaimana persebaran harga product nya? Cenderung banyak yang murah atau yang mahal?

Input :

```
plt.figure(figsize=(10,5))
plt.hist(dataset_top5brand_dec.groupby('product_id')['item_price'].median(), bins=10, stacked=True, range=(1,2000000), color='red', edgecolor='white')
plt.title('Distribusi median harga per produk\n5 Merek teratas pada des 2019', fontsize=15, color='blue')
plt.xlabel('Median harga', fontsize = 12)
plt.ylabel('Jumlah produk', fontsize = 12)
plt.xlim(xmin=0, xmax=2000000)
```

Output :

(0.0, 2000000.0)



Korelasi Quantity vs GMV

Case 6a

Untuk setiap product_id, cek scatterplot antara quantity dan GMV, apakah ada korelasi? Bagaimana dengan median harga vs quantity? Apakah product yang murah cenderung dibeli lebih banyak?

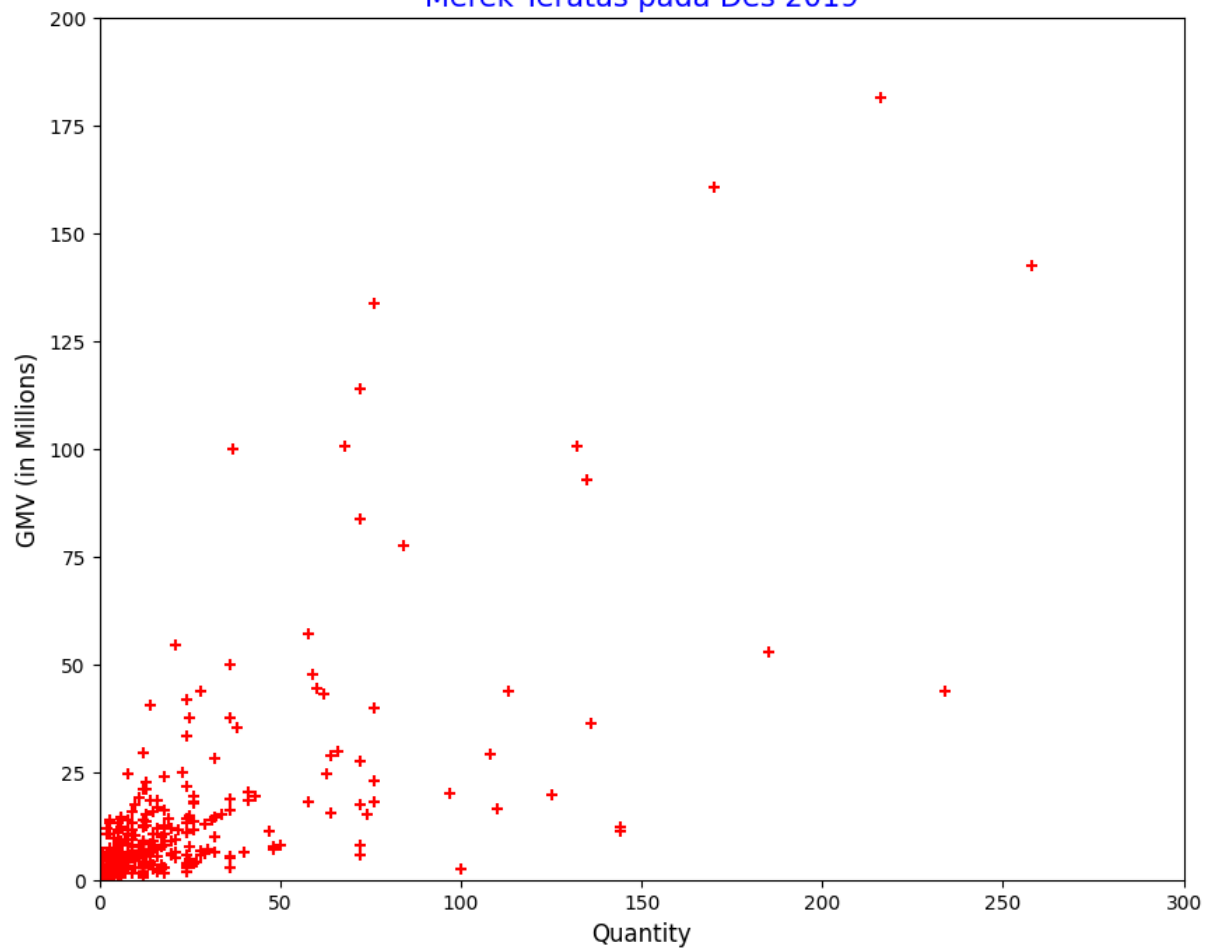
Input :

```
#agregat per product
data_per_product_top5brand_dec =
dataset_top5brand_dec.groupby('product_id').agg({'quantity': 'sum',
'gmv': 'sum', 'item_price': 'median'}).reset_index()
#scatter plot
plt.figure(figsize=(10,8))
plt.scatter(data_per_product_top5brand_dec['quantity'],data_per_product_
top5brand_dec['gmv'], marker='+', color='red')
plt.title('Korelasi Kuantitas dan GMV per Produk\nMerek Teratas pada Des
2019',fontsize=15, color='blue')
plt.xlabel('Quantity', fontsize = 12)
plt.ylabel('GMV (in Millions)',fontsize = 12)
plt.xlim(xmin=0,xmax=300)
plt.ylim(ymin=0,ymax=200000000)
labels, locations = plt.yticks()
plt.yticks(labels, (labels/1000000).astype(int))
```


Output :

```
<matplotlib.axis.YTick at 0x7fa0d6cd0fa0>,  
<matplotlib.axis.YTick at 0x7fa0d6cd1a50>,  
<matplotlib.axis.YTick at 0x7fa0d6cd2500>,  
<matplotlib.axis.YTick at 0x7fa0d6cd2fb0>,  
<matplotlib.axis.YTick at 0x7fa0d6cd1e40>,  
<matplotlib.axis.YTick at 0x7fa0d6cd3a60>],  
[Text(0, 0.0, '0'),  
Text(0, 25000000.0, '25'),  
Text(0, 50000000.0, '50'),  
Text(0, 75000000.0, '75'),  
Text(0, 100000000.0, '100'),  
Text(0, 125000000.0, '125'),  
Text(0, 150000000.0, '150'),  
Text(0, 175000000.0, '175'),  
Text(0, 200000000.0, '200')])
```

Korelasi Kuantitas dan GMV per Produk Merek Teratas pada Des 2019



Korelasi median harga vs quantity

Case 6b

Untuk setiap product_id, cek scatterplot antara quantity dan GMV sudah kamu lakukan pada Case 6a? Untuk Case 6b ini bagaimanakah dengan median harga vs quantity? Apakah product yang murah cenderung dibeli lebih banyak?

Input :

```
import matplotlib.pyplot as plt
plt.clf()
#agregat per product
data_per_product_top5brand_dec =
dataset_top5brand_dec.groupby('product_id').agg({'quantity': 'sum',
'gmv': 'sum', 'item_price': 'median'}).reset_index()
#scatter plot
plt.figure(figsize=(10,8))
plt.scatter(data_per_product_top5brand_dec['item_price'],data_per_product_top5brand_dec['quantity'], marker='o', color='green')
plt.title('Korelasi Kuantitas dan GMV per Produk\n Merek Teratas pada Desember 2019',fontsize=15, color='blue')
plt.xlabel('Median Harga', fontsize = 12)
plt.ylabel('Quantity',fontsize = 12)
plt.xlim(xmin=0,xmax=2000000)
plt.ylim(ymin=0,ymax=250)
```

Output :

(0.0, 250.0)

<Figure size 640x480 with 0 Axes>

