

Mini Project
“Pertemuan 14 Project 1 Komputer Grafis”

Nama : Tegar Adimas Nugroho
NIM : G.211.22.0088
Kelas : TI-B Pagi (B1)

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. Silakan 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 berpeluang 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 :

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.

```
import matplotlib.pyplot as plt
import pandas as pd
import datetime
```

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

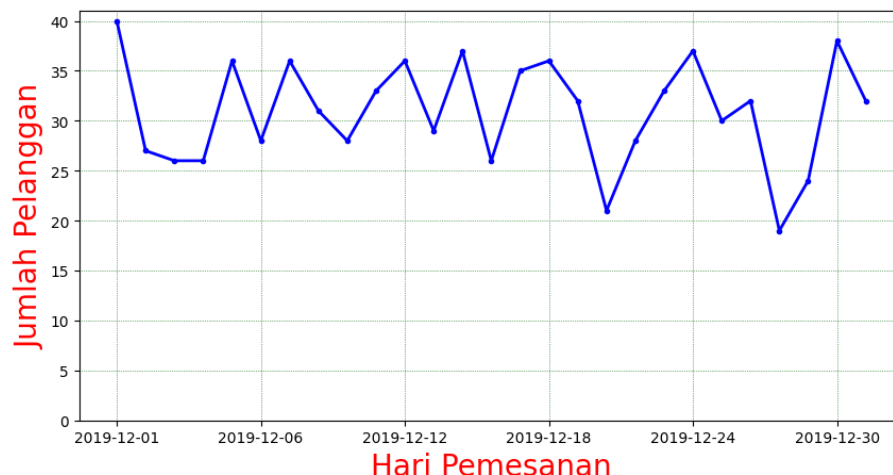
Input:

```
#Case 1 : Menentukan brand top 5
dataset['order_month'] = dataset['order_date'].apply(lambda
x:datetime.datetime.strptime(x, "%Y-%m-%d").strftime('%Y-%m'))
dataset['gmv'] = dataset['item_price'] * dataset['quantity']
plt.figure(figsize=(10, 5))
dataset[dataset['order_month']=='2019-
12'].groupby(['order_date'])['customer_id'].nunique().plot(color='blue',
marker='.', linewidth=2)
plt.title('Top 5 Brand Dengan Quantity Terjual Terbanyak Selama Bulan
Desember 2019', loc='left', pad=30, fontsize=20, color='red')
plt.xlabel('Hari Pemesanan', fontsize=20, color='red')
plt.ylabel('Jumlah Pelanggan', fontsize=20, color='red')
plt.grid(color='darkgreen', linestyle=':', linewidth=0.5)
plt.ylim(ymin=0)
plt.show
```

Output:

```
<function matplotlib.pyplot.show(close=None, block=None)>
```

Top 5 Brand Dengan Quantity Terjual Terbanyak Selama Bulan Desember 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 in

Input :

```
#mengambil informasi top 5 brands berdasarkan quantity
top_brands=(dataset[dataset['order_month']=='2019-12'].groupby('brand')['quantity']
              .sum()
              .reset_index()
              .sort_values(by='quantity', ascending=False)
              .head(5))
#filter hanya bulan desember dan hanya top 5 brands
dataset_topbrand5_dec = dataset[(dataset['order_month']=='2019-12') &
                                (dataset['brand'].isin(top_brands['brand'].to_list()))]
print(top_brands)
```

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: Multi-line chart daily quantity untuk brand top 5

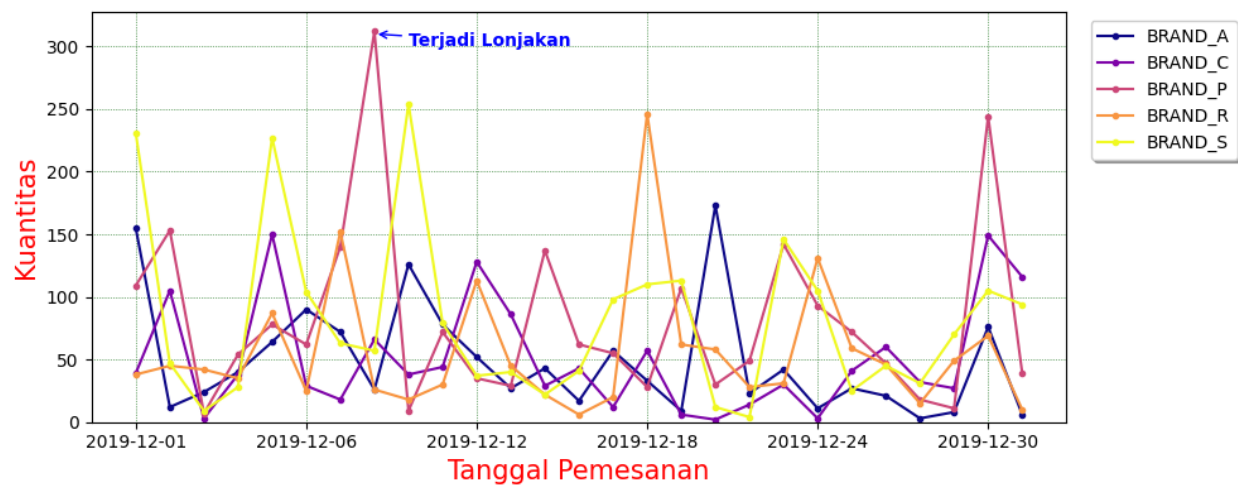
Buat visualisasi multi-line chart untuk daily quantity terjualnya, breakdown per brand. Maka, akan terlihat 1 tanggal di mana ada salah satu brand yang mengalami lonjakan (quantity lebih tinggi dari tanggal-tanggal lain). Beri anotasi untuk titik lonjakan tersebut.

Input :

```
#Case 2
dataset_topbrand5_dec.groupby(['order_date', 'brand'])['quantity'].sum().un
stack().plot(marker='.', cmap='plasma')
plt.xlabel('Tanggal Pemesanan', fontsize=15, color='red')
plt.ylabel('Kuantitas', fontsize=15, color='red')
plt.title('Kuantitas Penjualan Harian Desember 2019-\nPerincian
berdasarkan Merek', loc='center', pad=30, fontsize=20, color='red')
plt.grid(color='darkgreen', 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='blue',
            arrowprops=dict(arrowstyle='->', connectionstyle="arc3",
color='blue'))
plt.gcf().set_size_inches(10, 5)
plt.tight_layout()
```

Output :

Kuantitas Penjualan Harian Desember 2019- Perincian berdasarkan Merek



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.

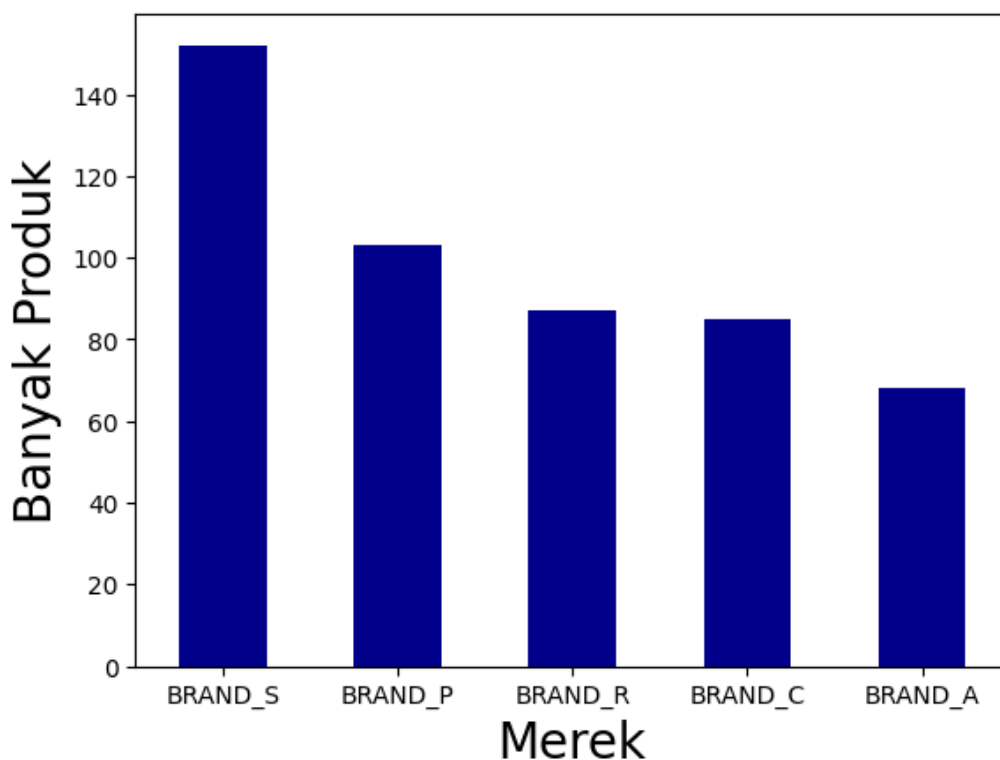
Input :

```
#Case 3
plt.clf()
dataset_topbrand5_dec.groupby('brand')['product_id'].nunique().sort_values(
    ascending=False).plot(kind='bar', color='darkblue')
plt.ylabel('Banyak Produk', fontsize=20)
plt.xlabel('Merek', fontsize=20)
plt.title('Jumlah Produk Terjual per Merek, Desember 2019', loc='center',
    pad=30, fontsize=20, color='red')
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, Desember 2019



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

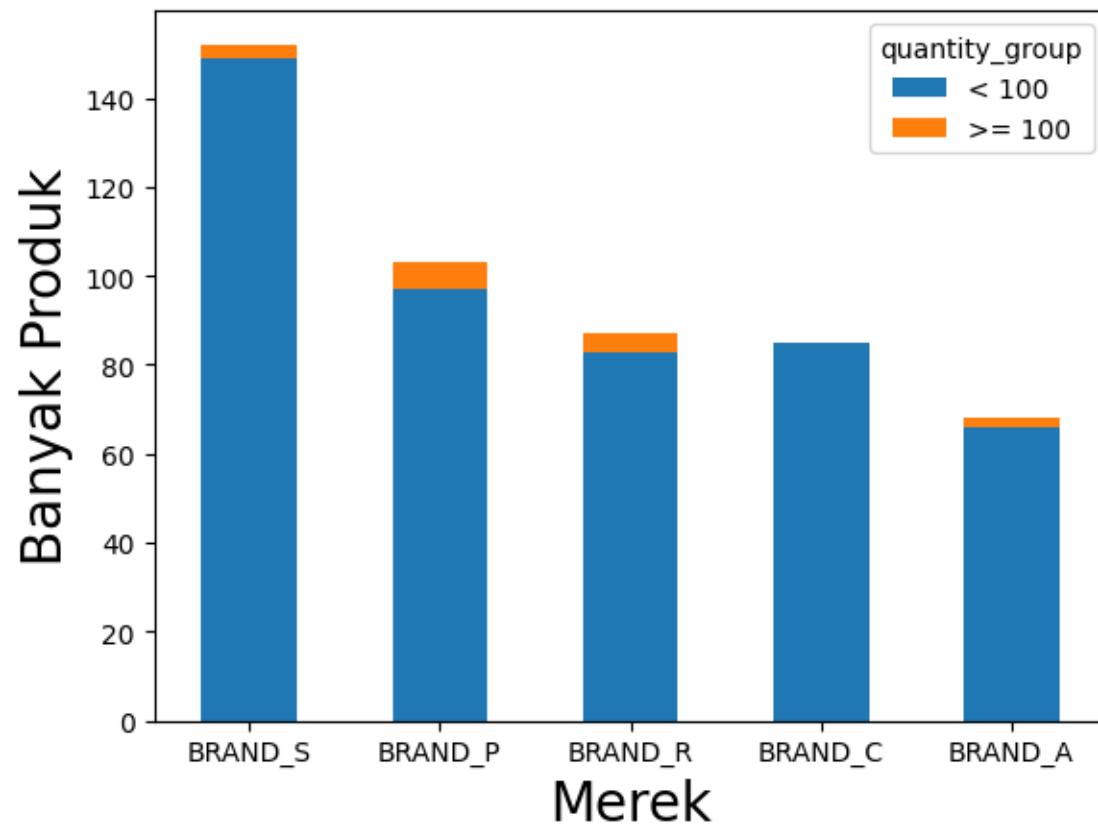
Gunakan stacked chart, untuk breakdown barchart yang di Case 3, antara product yang terjual ≥ 100 dan < 100 di bulan Desember 2019. Apakah ada pola yang menarik?

Input :

```
dataset_top5brand_dec_per_product =
dataset_top5brand_dec.groupby(['brand', 'product_id'])['quantity'].sum().reset_index()
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)
s_sort =
dataset_top5brand_dec_per_product.groupby('brand')['product_id'].nunique().sort_values(ascending=False)
dataset_top5brand_dec_per_product.groupby(['brand',
'quantity_group'])['product_id'].nunique().reindex(index=s_sort.index, level='brand').unstack().plot(kind='bar', stacked=True)
plt.ylabel('Banyak Produk', fontsize=20)
plt.xlabel('Merek', fontsize=20)
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')])
```



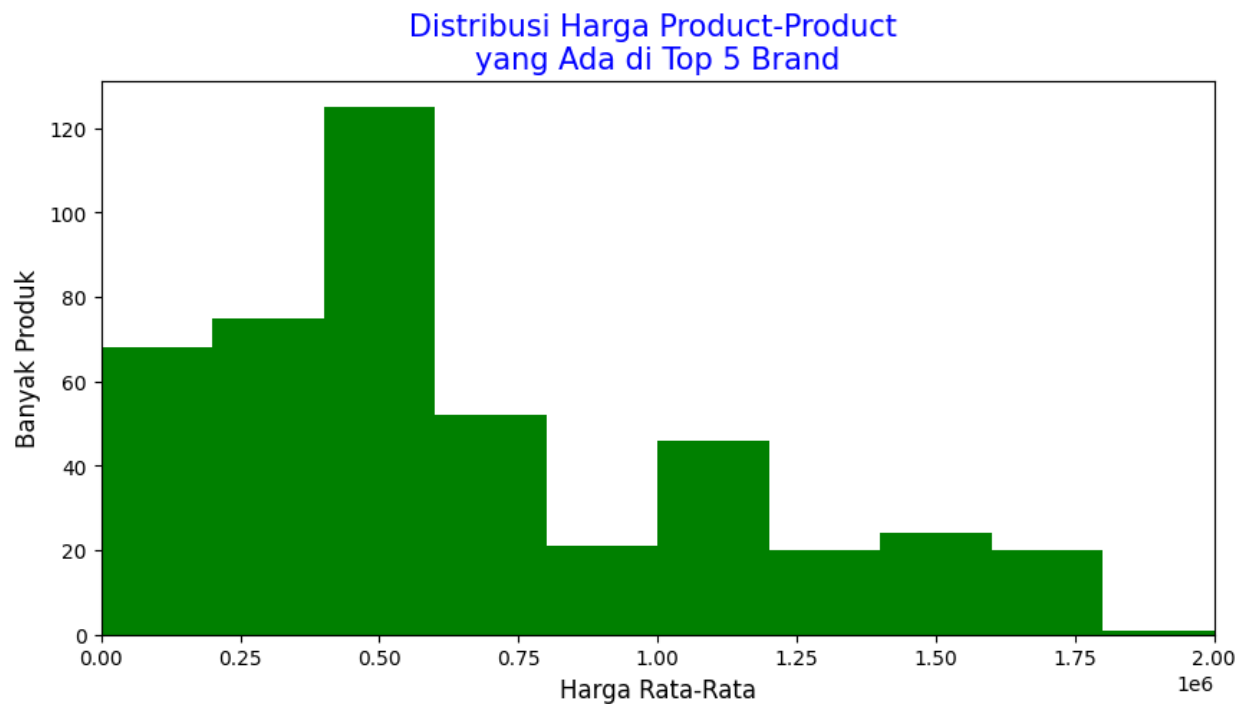
Case 5: Murah atau mahalkah harga produk brand top 5

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?

Input:

```
plt.figure(figsize=(10,5))
plt.hist(dataset_topbrand5_dec.groupby('product_id')['item_price'].median(
), bins=10, stacked=True, range=(1,2000000), color='green')
plt.title('Distribusi Harga Product-Product \nyang Ada di Top 5
Brand',fontsize=15, color='blue')
plt.ylabel('Banyak Produk', fontsize = 12)
plt.xlabel('Harga Rata-Rata',fontsize = 12)
plt.xlim(xmin=0,xmax=2000000)
```

Output :



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?

Input :

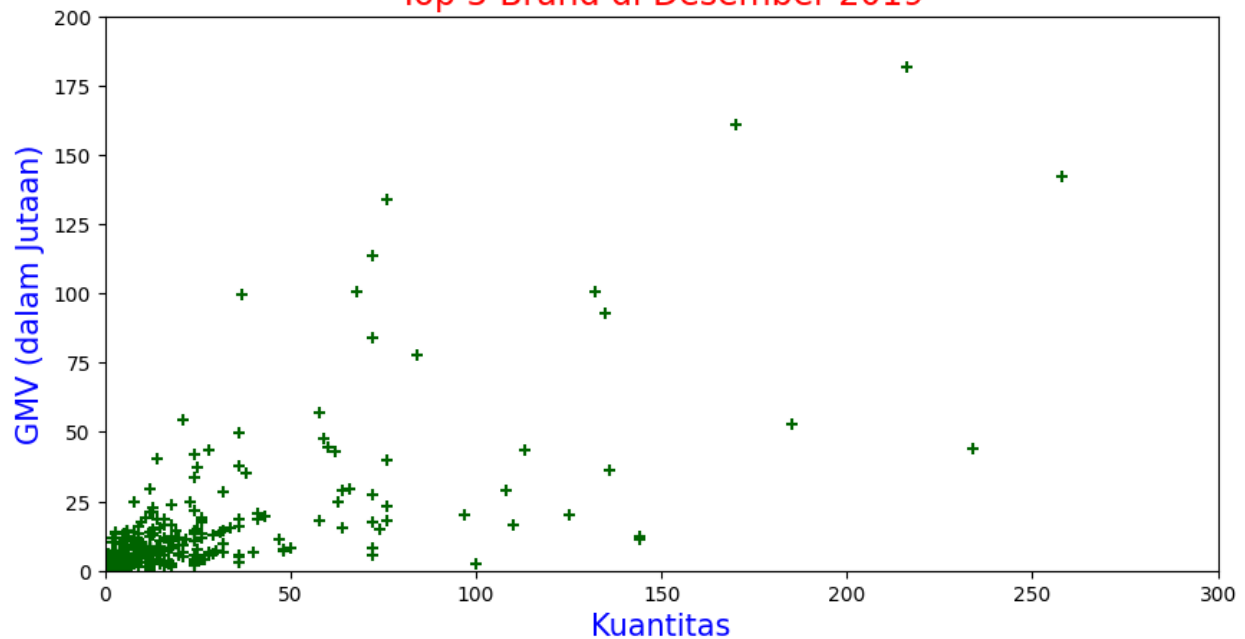
```
#agregasi setiap product
data_per_product_topbrand5_dec =
dataset_topbrand5_dec.groupby('product_id').agg({'quantity': 'sum',
'gmv':'sum', 'item_price':'median'}).reset_index()

#plot scatter
plt.figure(figsize=(10,5))
plt.scatter(data_per_product_topbrand5_dec['quantity'],data_per_product_topbrand5_dec['gmv'], marker='+', color='darkgreen')
plt.xlabel('Kuantitas', fontsize = 15, color='blue')
plt.ylabel('GMV (dalam Jutaan)',fontsize = 15, color='blue')
plt.title('Korelasi dari Kuantitas dan GMV per Produk \nTop 5 Brand di Desember 2019',fontsize=17, color='red')
plt.ylim(ymin=0,ymax=200000000)
plt.xlim(xmin=0,xmax=300)
labels, locations = plt.yticks()
plt.yticks(labels, (labels/1000000).astype(int))
```

Output :

```
([<matplotlib.axis.YTick at 0x7f4535bebcd0>,\n  <matplotlib.axis.YTick at 0x7f4535beab30>,\n  <matplotlib.axis.YTick at 0x7f4536a0f430>,\n  <matplotlib.axis.YTick at 0x7f4536b5d720>,\n  <matplotlib.axis.YTick at 0x7f4536b5ee60>,\n  <matplotlib.axis.YTick at 0x7f4536b5cb80>,\n  <matplotlib.axis.YTick at 0x7f453697a2f0>,\n  <matplotlib.axis.YTick at 0x7f4536b30c40>,\n  <matplotlib.axis.YTick at 0x7f4536b32650>],\n [Text(0, 0.0, '0'),\n  Text(0, 25000000.0, '25'),\n  Text(0, 50000000.0, '50'),\n  Text(0, 75000000.0, '75'),\n  Text(0, 100000000.0, '100'),\n  Text(0, 125000000.0, '125'),\n  Text(0, 150000000.0, '150'),\n  Text(0, 175000000.0, '175'),\n  Text(0, 200000000.0, '200')])
```

Korelasi dari Kuantitas dan GMV per Produk
Top 5 Brand di Desember 2019



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?

Input :

```
#agregat per product
data_per_product_topbrand5_dec =
dataset_topbrand5_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_topbrand5_dec['quantity'],data_per_product_to
pbrand5_dec['gmv'], marker='+', color='orange')
plt.title('Korelasi dari Kuantitas dan GMV per Product\nTop 5 Brand di
Desember 2019',fontsize=15, color='blue')
plt.xlabel('Kuantitas', fontsize = 12)
plt.ylabel('GMV (dalam Jutaan)',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 0x7f4538653400>,  
<matplotlib.axis.YTick at 0x7f4538651a80>,  
<matplotlib.axis.YTick at 0x7f4538652b60>,  
<matplotlib.axis.YTick at 0x7f453a7aaaa0>,  
<matplotlib.axis.YTick at 0x7f4536a774c0>,  
matplotlib.axis.YTick at 0x7f453a7aa140>,  
<matplotlib.axis.YTick at 0x7f453a7a82b0>,  
<matplotlib.axis.YTick at 0x7f4538600400>,  
<matplotlib.axis.YTick at 0x7f453a70f3a0>],  
[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 dari Kuantitas dan GMV per Product
Top 5 Brand di Desember 2019

