

NAMA : ANAS ZAKIA ARDHAN NIM : G.211.19.0136 MATA KULIAH : KOMPUTER GRAFIS (B1)

```
#import library
import datetime
import pandas as pd
import matplotlib.pyplot as plt
#baca dataset
dataset = pd.read_csv('https://dqlab-dataset.s3-ap-southeast-1.amazonaws.com/retail_raw_reduced.csv')
dataset.head(5)
```

	order_id	order_date	customer_id	city	province
product_id \					
0	1703458	2019-10-17	14004	Jakarta Selatan	DKI Jakarta
P1910					
1	1706815	2019-10-24	17220	Jakarta Selatan	DKI Jakarta
P2934					
2	1710718	2019-11-03	16518	Jakarta Utara	DKI Jakarta
P0908					
3	1683592	2019-08-19	16364	Jakarta Barat	DKI Jakarta
P0128					
4	1702573	2019-10-16	15696	Jakarta Timur	DKI Jakarta
P2968					

	brand	quantity	item_price
0	BRAND_J	10	740000
1	BRAND_R	2	604000
2	BRAND_C	8	1045000
3	BRAND_A	4	205000
4	BRAND_R	2	4475000

```
#case 1
#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))
#membuat dataframe baru, filter hanya di bulan Desember 2019 dan hanya top 5 brands
dataset_top5brand_dec = dataset[(dataset['order_month']=='2019-12') & (dataset['brand'].isin(top_brands['brand'].to_list()))]
# print top brands
print(top_brands)
```

	brand	quantity
8	BRAND_S	2197
6	BRAND_P	2194

```

7  BRAND_R      1508
2  BRAND_C      1365
0  BRAND_A      1315

```

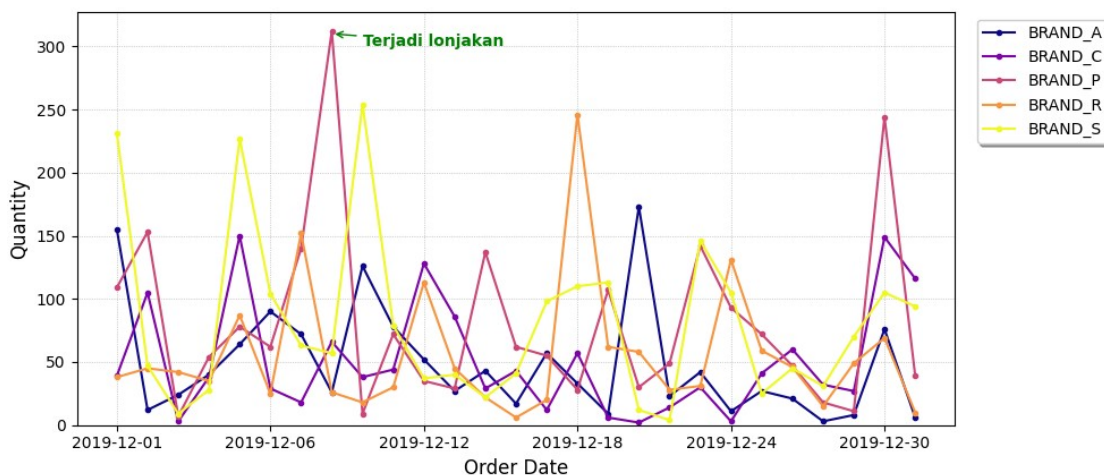
*#case 2*

```

dataset_top5brand_dec.groupby(['order_date','brand'])
['quantity'].sum().unstack().plot(marker='.', cmap='plasma')
plt.title('Daily Sold Quantity Dec 2019-Breakdown by
Brands',loc='center',pad=30, fontsize=15, color='green')
plt.xlabel('Order Date', 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='green',
            arrowprops=dict(arrowstyle='->',
                            connectionstyle="arc3",
                            color='green'))
plt.gcf().set_size_inches(10, 5)
plt.tight_layout()

```

Daily Sold Quantity Dec 2019-Breakdown by Brands



*#case 3*

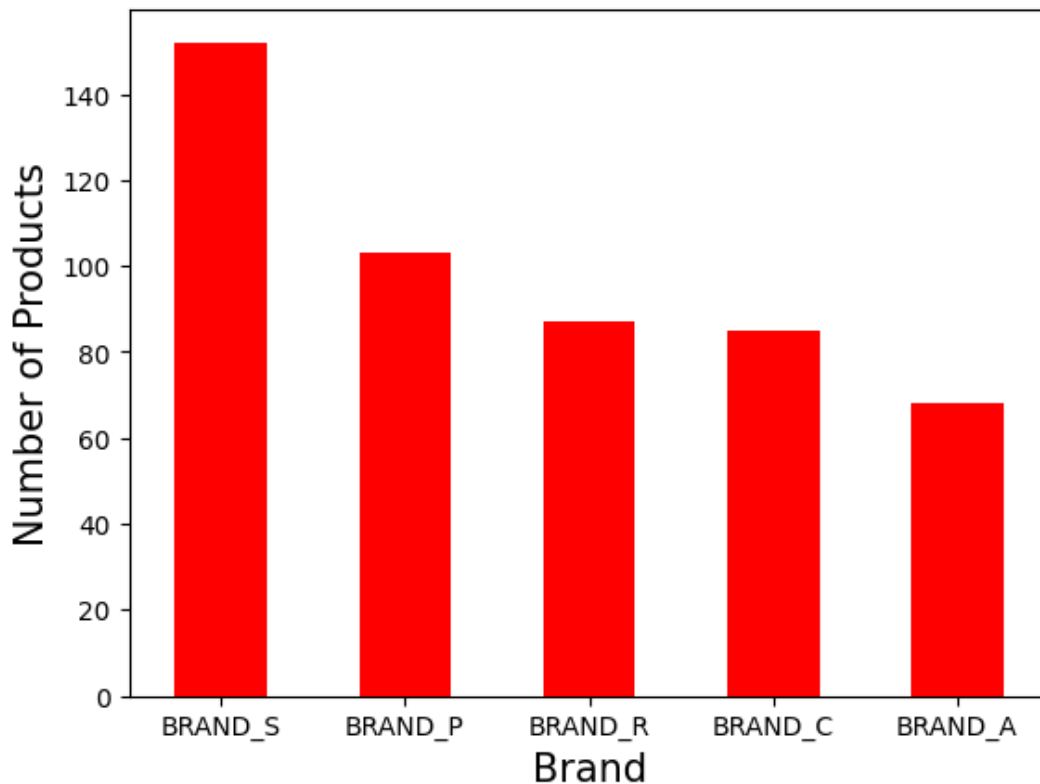
```

plt.clf()
dataset_top5brand_dec.groupby('brand')
['product_id'].nunique().sort_values(ascending=False).plot(kind='bar',
color='red')
plt.title('Number of Sold Products per Brand, December
2019',loc='center',pad=30, fontsize=15, color='green')
plt.xlabel('Brand', fontsize = 15)
plt.ylabel('Number of Products',fontsize = 15)
plt.ylim(ymin=0)
plt.xticks(rotation=0)

```

```
(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')])
```

## Number of Sold Products per Brand, December 2019



```
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)

#case 4
dataset_top5brand_dec_per_product.groupby(['brand', 'quantity_group'])
['product_id'].nunique().reindex(index=s_sort.index,
```

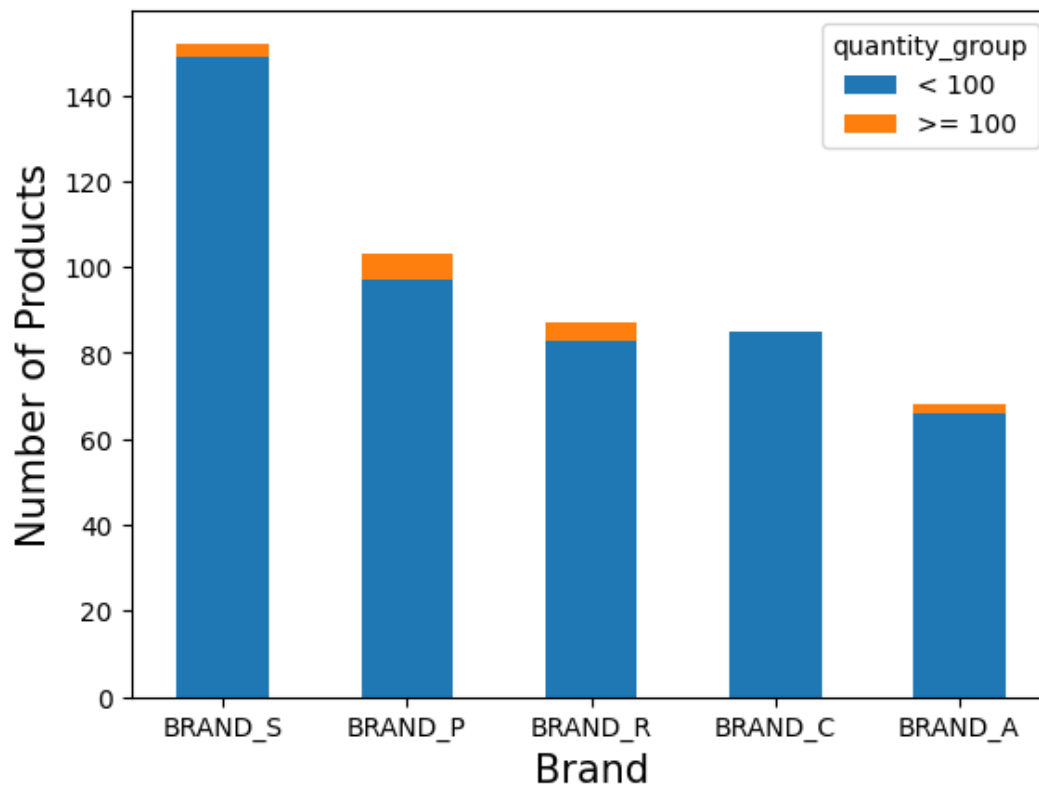
```

level='brand').unstack().plot(kind='bar', stacked=True)
plt.title('Number of Sold Products per Brand, December
2019',loc='center',pad=30, fontsize=15, color='green')
plt.xlabel('Brand', fontsize = 15)
plt.ylabel('Number of Products',fontsize = 15)
plt.ylim(ymin=0)
plt.xticks(rotation=0)

(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')])

```

## Number of Sold Products per Brand, December 2019



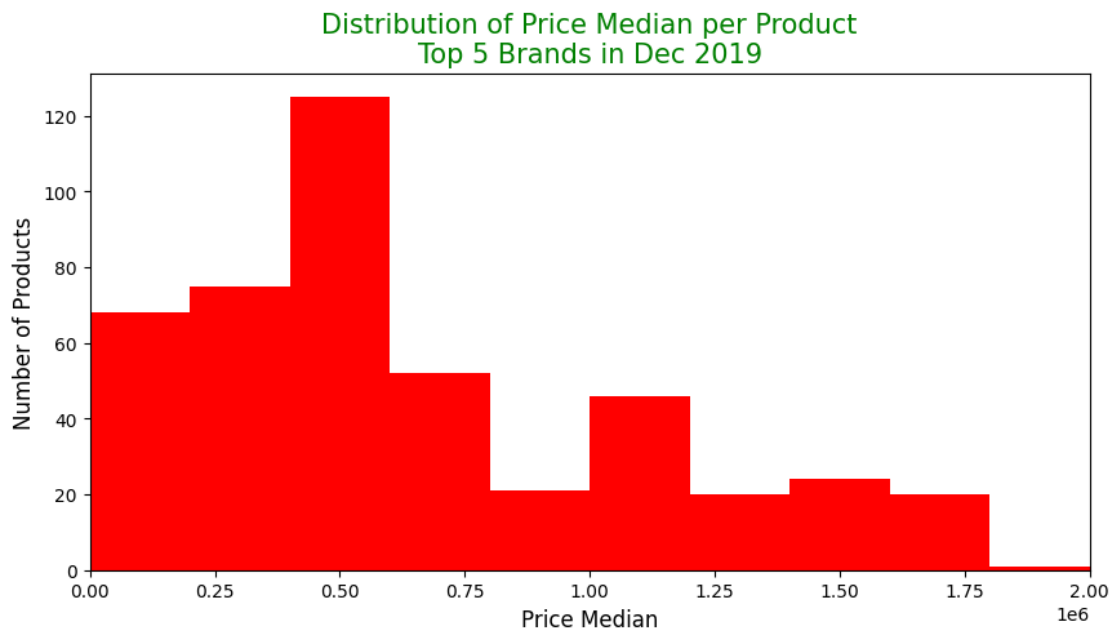
```

#case 5
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')
plt.title('Distribution of Price Median per Product\nTop 5 Brands in
Dec 2019',fontsize=15, color='green')
plt.xlabel('Price Median', fontsize = 12)

```

```
plt.ylabel('Number of Products',fontsize = 12)
plt.xlim(xmin=0,xmax=2000000)

(0.0, 2000000.0)
```



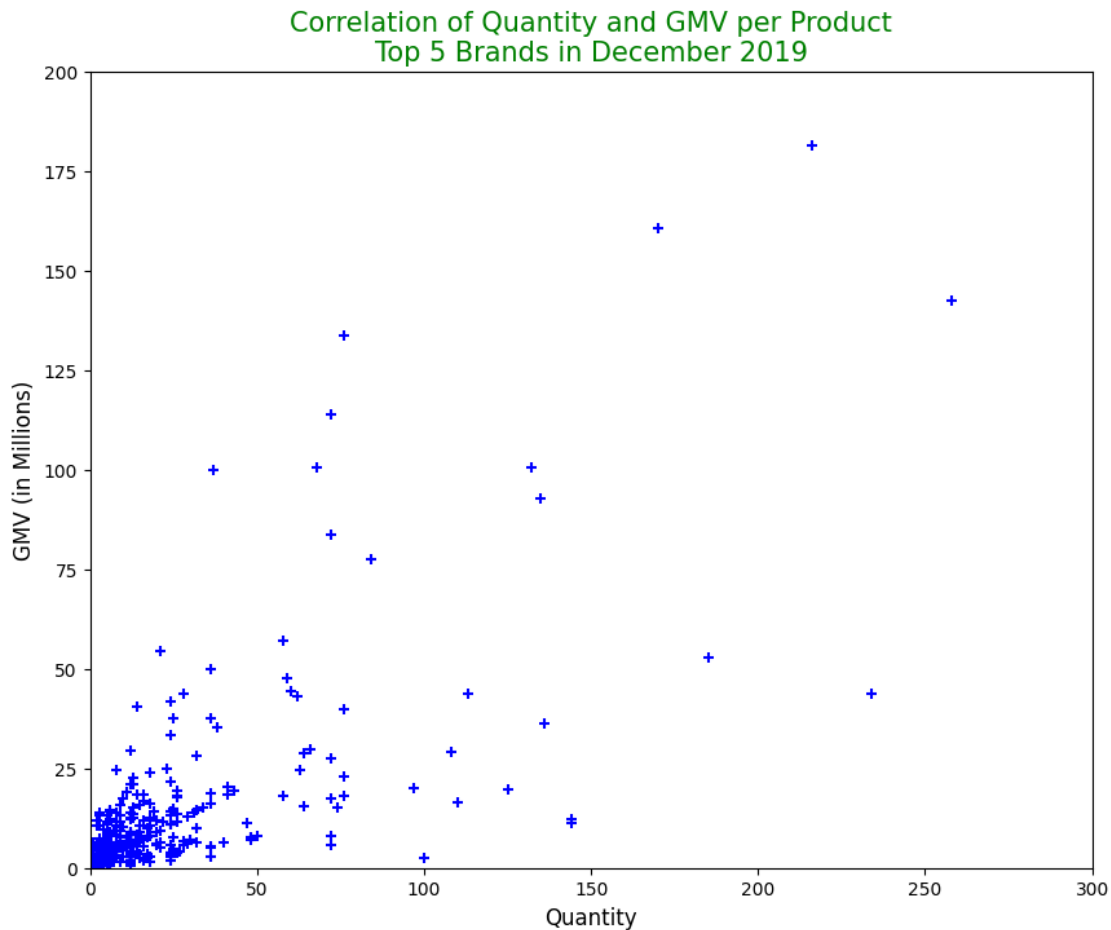
```
#case 6a
#agregat per product
data_per_product_top5brand_dec =
dataset_top5brand_dec.groupby('product_id').agg({'quantity': 'sum',
'gmv': 'sum', 'item_price': 'median'}).reset_index()
#scatter plot
plt.figure(figsize=(10,8))
plt.scatter(data_per_product_top5brand_dec['quantity'],data_per_product_top5brand_dec['gmv'], marker='+', color='blue')
plt.title('Correlation of Quantity and GMV per Product\nTop 5 Brands in December 2019',fontsize=15, color='green')
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))
```

```
([<matplotlib.axis.YTick at 0x7fefc03866b0>,
<matplotlib.axis.YTick at 0x7fefc0384820>,
<matplotlib.axis.YTick at 0x7fefc0384880>,
<matplotlib.axis.YTick at 0x7fefc011ace0>,
<matplotlib.axis.YTick at 0x7fefc01191e0>,
<matplotlib.axis.YTick at 0x7fefc011ad40>,
<matplotlib.axis.YTick at 0x7fefc048ab00>,
<matplotlib.axis.YTick at 0x7fefc0489ea0>],
```

```

<matplotlib.axis.YTick at 0x7fefc048bfa0>],
[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')])

```



```

#case 6b
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_produ
ct_top5brand_dec['quantity'], marker='o', color='green')

```

```
plt.title('Correlation of Quantity and GMV per Product\n Top 5 Brands
in December 2019', fontsize=15, color='blue')
plt.xlabel('Price Median', fontsize = 12)
plt.ylabel('Quantity', fontsize = 12)
plt.xlim(xmin=0, xmax=2000000)
plt.ylim(ymin=0, ymax=250)
```

(0.0, 250.0)

<Figure size 640x480 with 0 Axes>

