

Nama : Nafhan Haqiqi

Kelas : 1A Informatika

NPM : 2410631170095

## Data Visualization

### 1. Pie Chart

Percobaan Pertama

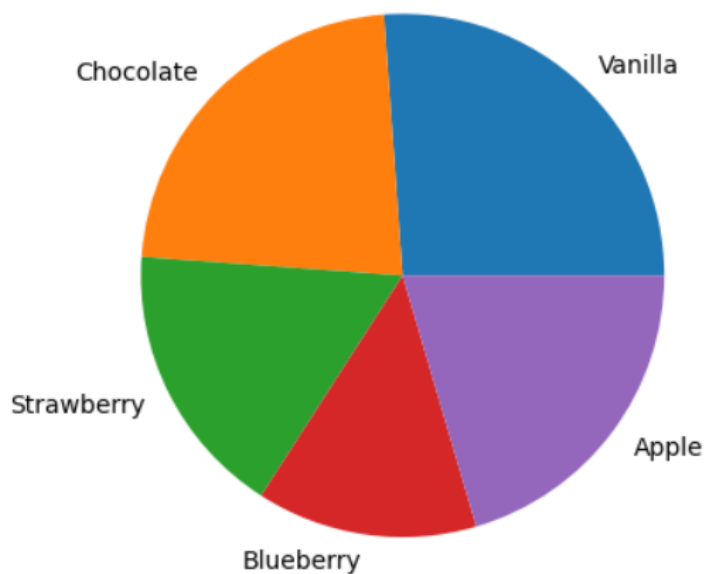
Input

```
import matplotlib.pyplot as plt

flavors = ('Vanilla', 'Chocolate', 'Strawberry', 'Blueberry', 'Apple')
votes = (23, 20, 15, 12, 18)

plt.pie(votes, labels=flavors,)
plt.show()
```

Output



Percobaan Ke Dua

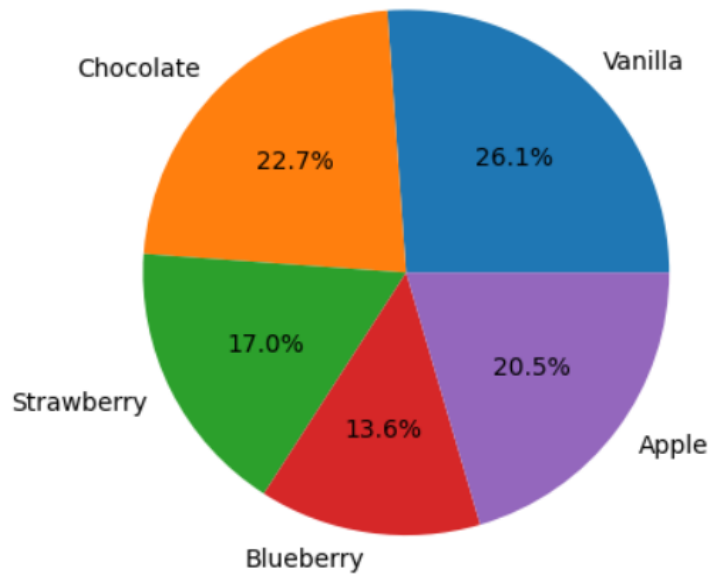
Input

```
import matplotlib.pyplot as plt

flavors = ('Vanilla', 'Chocolate', 'Strawberry', 'Blueberry', 'Apple')
votes = (23, 20, 15, 12, 18)

plt.pie(votes, labels=flavors, autopct = '%1.1f%%',)
plt.show()
```

## Output



## Percobaan Ke Tiga

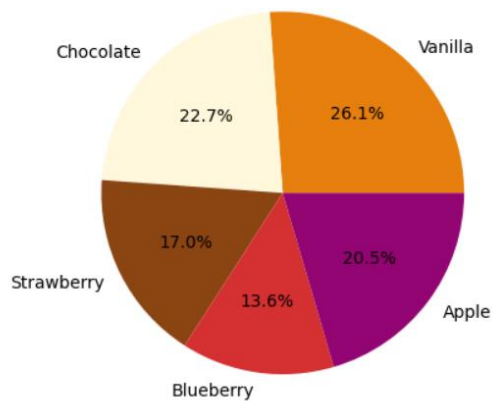
### Input

```
import matplotlib.pyplot as plt

flavors = ('Vanilla', 'Chocolate', 'Strawberry', 'Blueberry', 'Apple')
votes = (23, 20, 15, 12, 18)

warna = ('#E67F0D', '#FF8C00', '#8B4513', '#D53032', '#930572')
plt.pie(votes, labels = flavors, autopct = '%1.1f%%', colors = warna)
plt.show()
```

## Output



## Percobaan Ke Empat

### Input

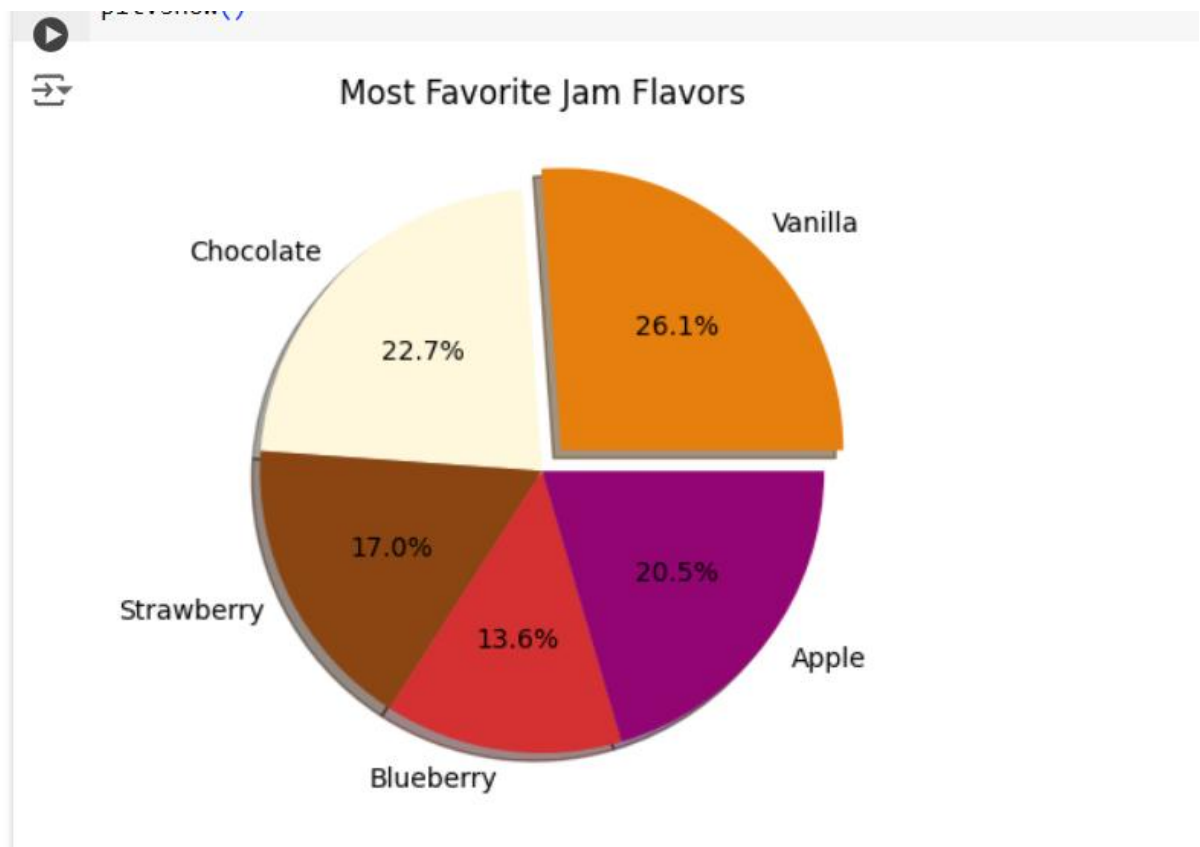
```
import matplotlib.pyplot as plt

flavors = ('Vanilla', 'Chocolate', 'Strawberry', 'Blueberry', 'Apple')
votes = (23, 20, 15, 12, 18)

warna = ('#E67F0D', '#FFF8DC', '#8B4513', '#D53032', '#930572')
explode = (0.1, 0, 0, 0, 0)

plt.title('Most Favorite Jam Flavors')
plt.pie(votes, labels = flavors, autopct = '%1.1f%%', colors = warna, explode=explode, shadow=True)
plt.show()
```

### Output



## 2. Bar Charts

### Percobaan Pertama

### Input

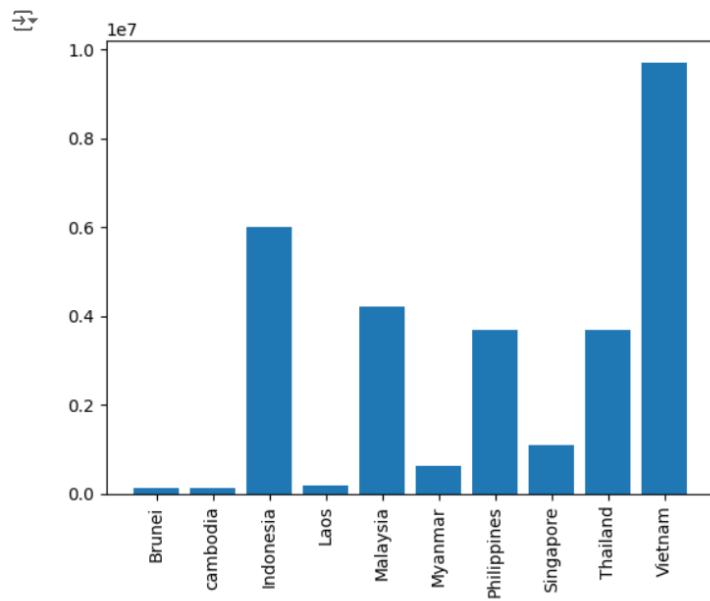
```
import matplotlib.pyplot as plt
import numpy as np

countries = ('Brunei', 'Cambodia', 'Indonesia', 'Laos', 'Malaysia', 'Myanmar', 'Philippines', 'Singapore', 'Thailand', 'Vietnam')
cases = (135430, 135714, 6018048, 181967, 4219395, 611674, 3678968, 1101438, 3684755, 9716282)

x_coords = np.arange(len(countries))

plt.bar(x_coords, cases, tick_label = countries, )
plt.xticks(rotation=90)
plt.show()
```

## Output



## Percobaan Ke Dua

### Input

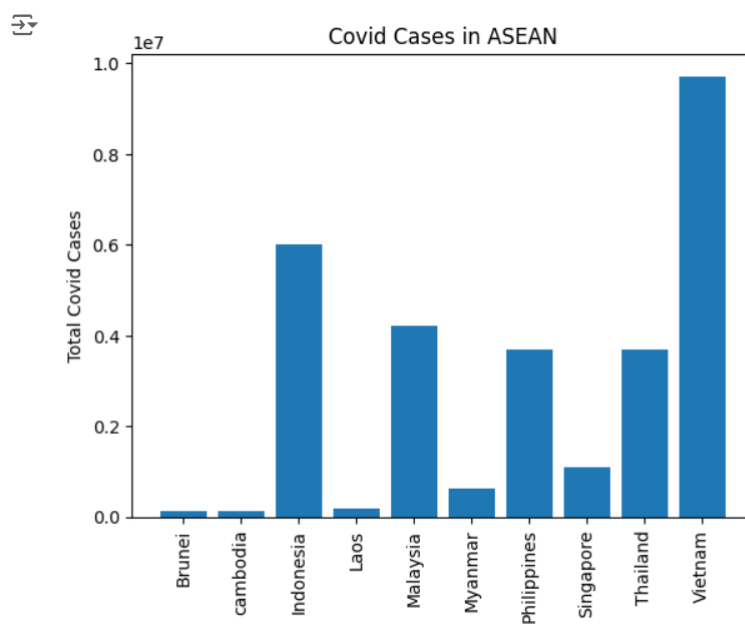
```
import matplotlib.pyplot as plt
import numpy as np

countries = ('Brunei', 'cambodia', 'Indonesia', 'Laos', 'Malaysia', 'Myanmar', 'Philippines', 'Singapore', 'Thailand', 'Vietnam')
cases = (135430, 135714, 6018048, 181967, 4219395, 611674, 3678968, 1101438, 3684755, 9716282)

x_coords = np.arange(len(countries))

plt.bar(x_coords, cases, tick_label = countries, )
plt.xticks(rotation=90)
plt.ylabel('Total Covid Cases')
plt.title('Covid Cases in ASEAN')
plt.show()
```

## Output



## Percobaan Ke Tiga

### Input

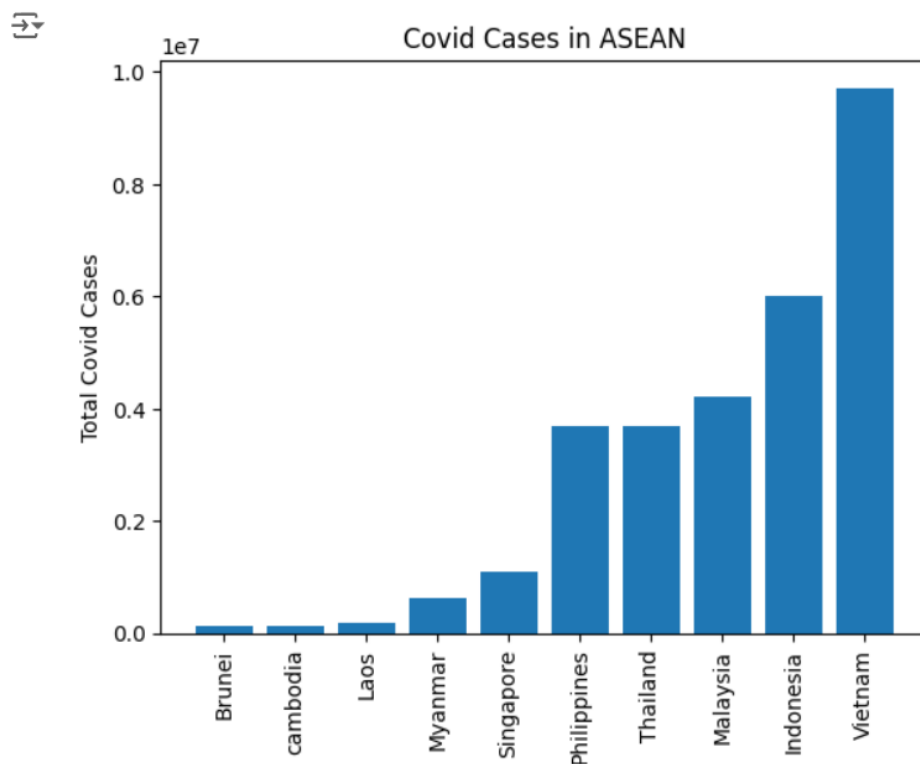
```
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd

countries = ('Brunei', 'cambodia', 'Indonesia', 'Laos', 'Malaysia', 'Myanmar', 'Philippines', 'Singapore', 'Thailand', 'Vietnam')
cases = (135430, 135714, 6018048, 181967, 4219395, 611674, 3678968, 1101438, 3684755, 9716282)

df= pd.DataFrame ({
    'Country':countries,
    'Case': cases,
})
df.sort_values (by='Case', inplace=True)
x_coords = np.arange(len(df))

plt.bar(x_coords,df['Case'], tick_label = df['Country'], )
plt.xticks (rotation=90)
plt.ylabel ('Total Covid Cases')
plt.title('Covid Cases in ASEAN')
plt.show()
```

### Output



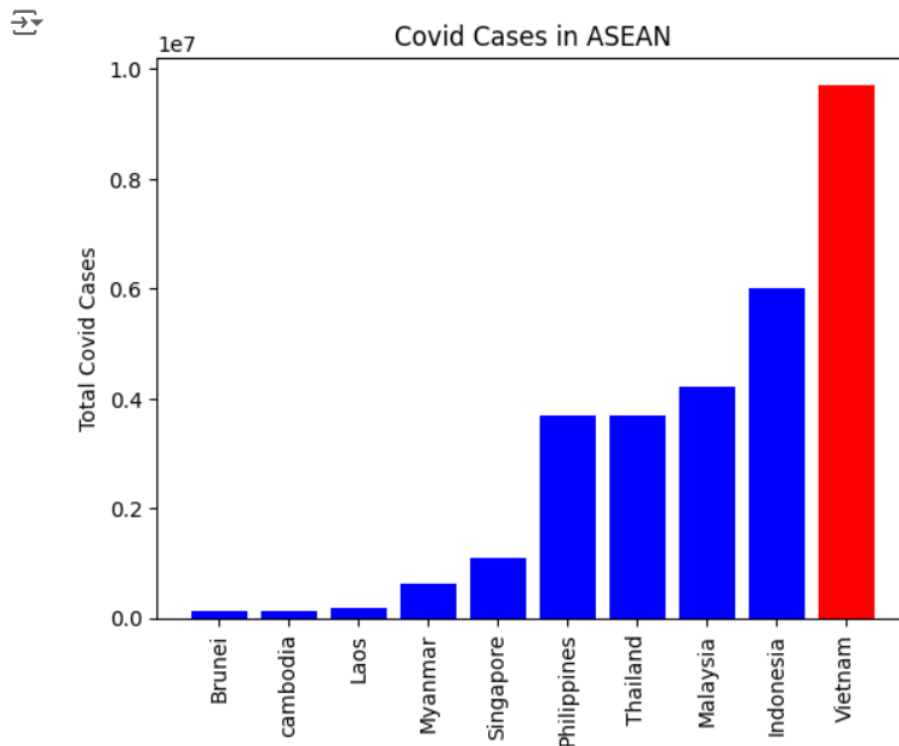
## Percobaan Ke Empat

```
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd

countries = ('Brunei', 'cambodia', 'Indonesia', 'Laos', 'Malaysia', 'Myanmar', 'Philippines', 'Singapore', 'Thailand', 'Vietnam')
cases = (135430, 135714, 6018048, 181967, 4219395, 611674, 3678968, 1101438, 3684755, 9716282)

df = pd.DataFrame ({
    'Country':countries,
    'Case': cases,
})
df.sort_values (by='Case', inplace=True)
x_coords = np.arange(len(df))
colors = ['#0000FF' for _ in range(len(df))]
colors[-1] = '#FF0000'
plt.bar(x_coords, df['Case'], tick_label = df['Country'], color=colors)
plt.xticks (rotation=90)
plt.ylabel ('Total Covid Cases')
plt.title('Covid Cases in ASEAN')
plt.show()
```

## Output



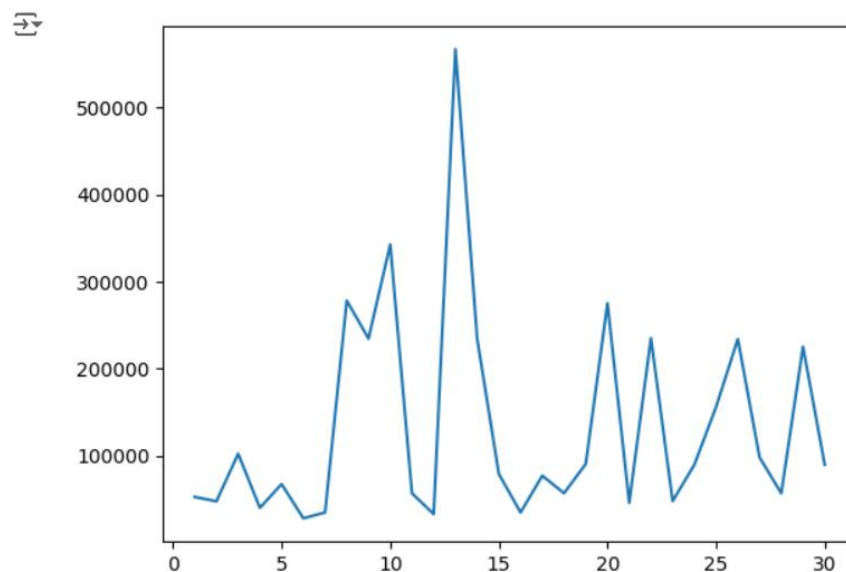
## 3. Line Graphs

### Percobaan Pertamaa

#### Input

```
import matplotlib.pyplot as plt
penj = [52500, 47500, 102200, 40000, 67300, 28000, 34670, 278000, 234500, 342500, 56740, 32780, 567000, 234600, 78900, 34700, 76900, 56800, 90500, 275000, 45800, 235000, 47800, 89600, 156000, 234000, 98000, 56700, 225000, 89500]
hari = [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30]
plt.plot(hari, penj)
plt.show()
```

#### Output

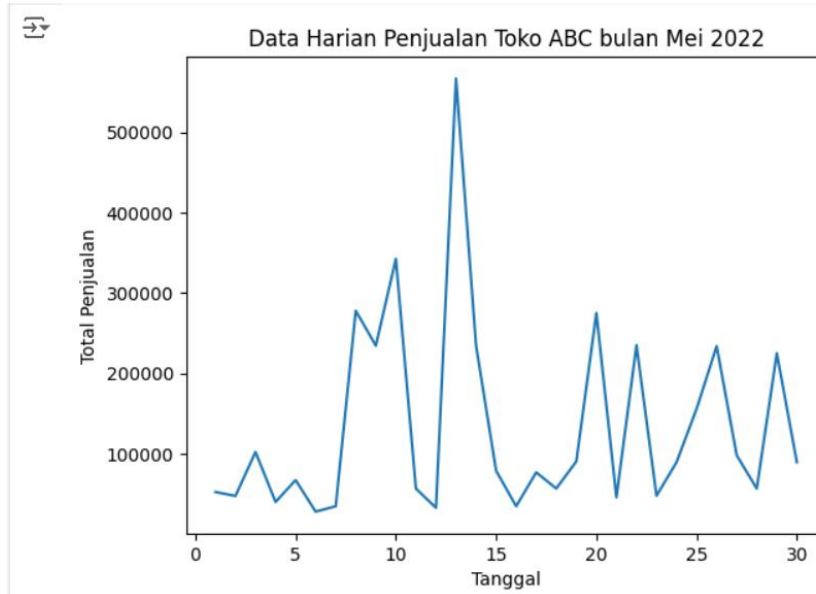


## Percobaan Ke Dua

### Input

```
import matplotlib.pyplot as plt
penj = [52500, 47500, 102200, 40000, 67300, 28000, 34670, 278000, 234500, 342500, 56740, 32780, 567000, 234600,
78900, 34700, 76900, 56800, 90500, 275000, 45800, 235000, 47800, 89600, 156000, 234000, 98000, 56700, 225000, 89500]
hari = [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30]
plt.plot(hari, penj)
plt.title('Data Harian Penjualan Toko ABC bulan Mei 2022')
plt.ylabel('Total Penjualan')
plt.xlabel('Tanggal')
plt.show()
```

### Output

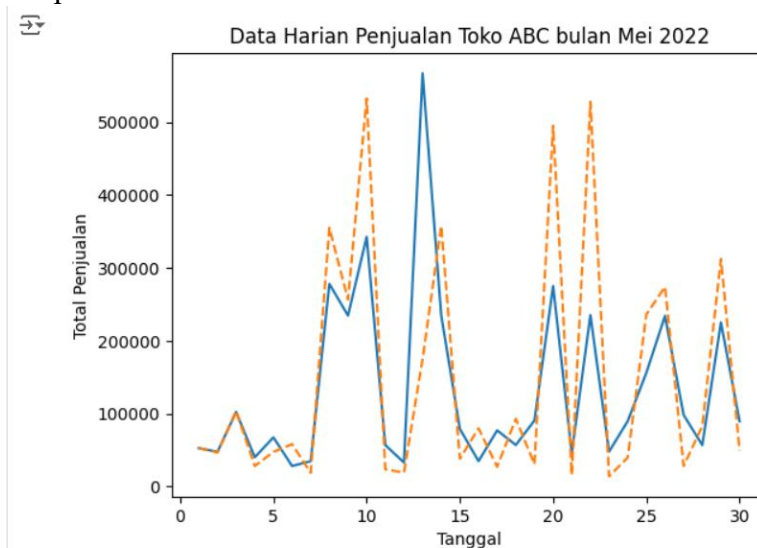


## Percobaan ke Tiga

### Input

```
import matplotlib.pyplot as plt
penj_actual = [52500, 47500, 102200, 40000, 67300, 28000, 34670, 278000, 234500, 342500, 56740, 32780, 567000, 234600,
78900, 34700, 76900, 56800, 90500, 275000, 45800, 235000, 47800, 89600, 156000, 234000, 98000, 56700, 225000, 89500]
penj_predict = [52600, 46500, 103200, 28000, 47300, 58000, 18670, 355000, 256500, 532500, 23540, 18780, 175000, 357600, 38000, 79700,
16900, 92800, 30500, 495000, 15800, 528000, 13800, 39600, 236000, 274000, 28000, 82700, 312000, 49300]
hari = [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30]
plt.plot(hari, penj_actual)
plt.plot(hari, penj_predict, linestyle='--')
plt.title('Data Harian Penjualan Toko ABC bulan Mei 2022')
plt.ylabel('Total Penjualan')
plt.xlabel('Tanggal')
plt.show()
```

### Output



#### 4. Scatter Plot

Percobaan Pertama

Input

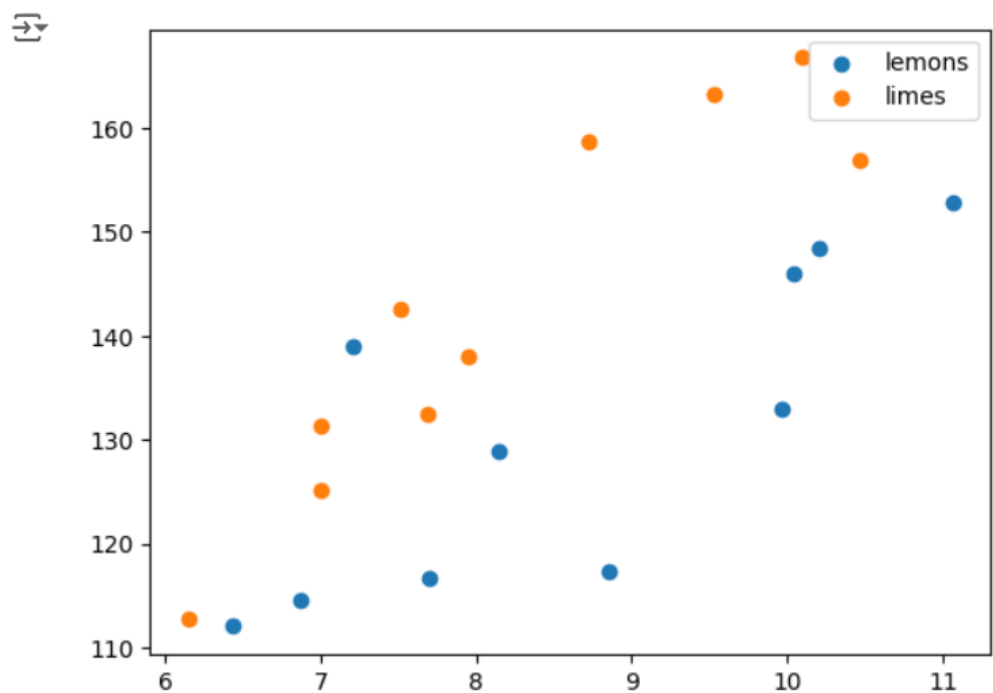
```
[ ] import matplotlib.pyplot as plt

lemon_diameter = [6.44, 6.87, 7.7, 8.85, 8.15, 9.96, 7.21, 10.04, 10.2, 11.06]
lemon_weight = [112.05, 114.58, 116.71, 117.4, 128.93, 132.93, 138.92, 145.98, 148.44, 152.81]

lime_diameter = [6.15, 7.0, 7.0, 7.69, 7.95, 7.51, 10.46, 8.72, 9.53, 10.09]
lime_weight = [112.76, 125.16, 131.36, 132.41, 138.08, 142.55, 156.86, 158.67, 163.28, 166.74]

plt.scatter(lemon_diameter, lemon_weight)
plt.scatter(lime_diameter, lime_weight)
plt.legend(['lemons', 'limes'])
plt.show()
```

Output



Percobaan Ke Dua

Input

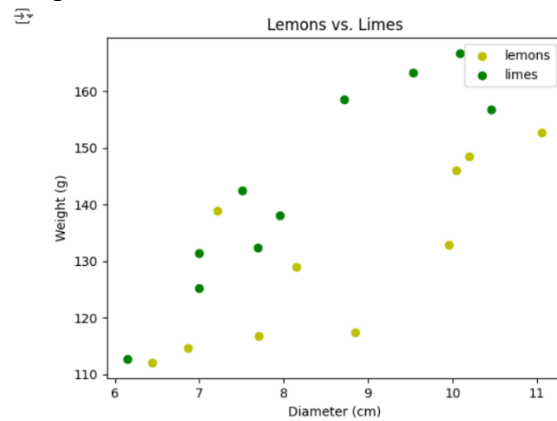
```
import matplotlib.pyplot as plt

lemon_diameter = [6.44, 6.87, 7.7, 8.85, 8.15, 9.96, 7.21, 10.04, 10.2, 11.06]
lemon_weight = [112.05, 114.58, 116.71, 117.4, 128.93, 132.93, 138.92, 145.98, 148.44, 152.81]
lime_diameter = [6.15, 7.0, 7.0, 7.69, 7.95, 7.51, 10.46, 8.72, 9.53, 10.09]
lime_weight = [112.76, 125.16, 131.36, 132.41, 138.08, 142.55, 156.86, 158.67, 163.28, 166.74]

plt.title('Lemons vs. Limes')
plt.xlabel('Diameter (cm)')
plt.ylabel('Weight (g)')
plt.scatter(lemon_diameter, lemon_weight, color='y')
plt.scatter(lime_diameter, lime_weight, color='g')
plt.legend(['lemons', 'limes'])
plt.show()
```



## Output



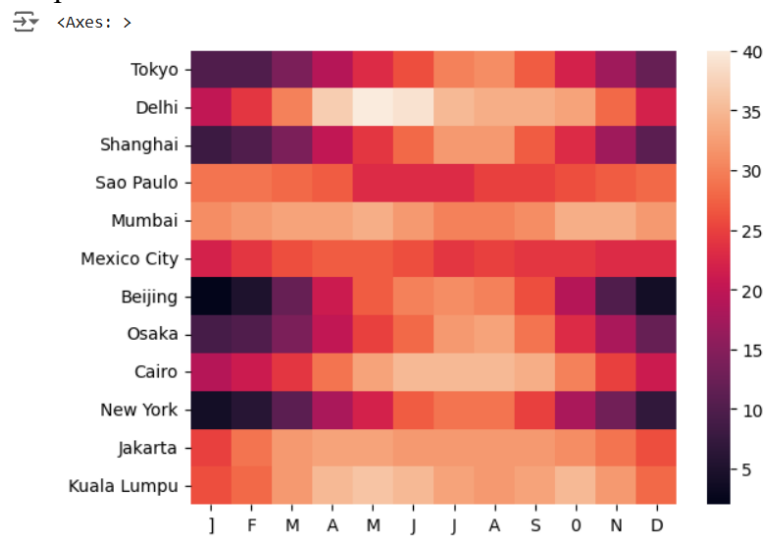
## 5. Heatmap

### Percobaan Pertama

#### Input

```
import seaborn as sns
cities = ['Tokyo', 'Delhi', 'Shanghai', 'Sao Paulo', 'Mumbai',
'Mexico City', 'Beijing', 'Osaka', 'Cairo', 'New York', 'Jakarta', 'Kuala Lumpur']
months = ['J', 'F', 'M', 'A', 'M', 'J', 'J', 'A', 'S', 'O', 'N', 'D']
temperatures = [
[10, 10, 14, 19, 23, 26, 30, 31, 27, 22, 17, 12], # Tokyo
[20, 24, 30, 37, 40, 39, 35, 34, 34, 33, 28, 22], # Delhi
[8, 10, 14, 20, 24, 28, 32, 32, 27, 23, 17, 11], # Shanghai
[29, 29, 28, 27, 23, 23, 23, 25, 25, 26, 27, 28], # Sao Paulo
[31, 32, 33, 33, 34, 32, 30, 30, 31, 34, 34, 32], # Mumbai
[22, 24, 26, 27, 27, 26, 24, 25, 24, 24, 23, 23], # Mexico City
[2, 5, 12, 21, 27, 30, 31, 30, 26, 19, 10, 4], # Beijing
[9, 10, 14, 20, 25, 28, 32, 33, 29, 23, 18, 12], # Osaka
[19, 21, 24, 29, 33, 35, 35, 35, 34, 30, 25, 21], # Cairo
[4, 6, 11, 18, 22, 27, 29, 29, 25, 18, 13, 7], # New York
[25, 29, 32, 33, 33, 32, 32, 32, 32, 31, 29, 26], # Jakarta
[26, 28, 32, 35, 36, 35, 33, 32, 33, 35, 32, 28], # Kuala Lumpur
]
sns.heatmap(temperatures, yticklabels=cities, xticklabels=months)
```

## Output



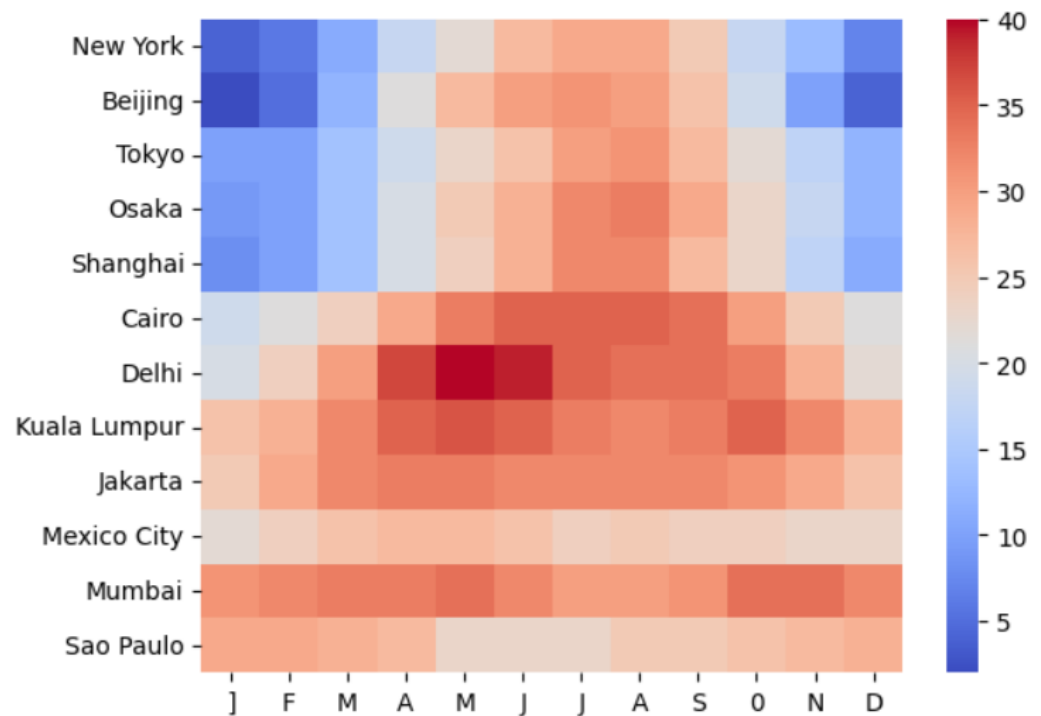
## Percobaan Ke Dua

### Input

```
import seaborn as sns
cities = ['New York', 'Beijing', 'Tokyo', 'Osaka', 'Shanghai', 'Cairo', 'Delhi', 'Kuala Lumpur', 'Jakarta', 'Mexico City', 'Mumbai', 'Sao Paulo']
temperatures = [
    [4, 6, 11, 18, 22, 27, 29, 29, 25, 18, 13, 7], # New York
    [2, 5, 12, 21, 27, 30, 31, 30, 26, 19, 10, 4], # Beijing
    [10, 10, 14, 19, 23, 26, 30, 31, 27, 22, 17, 12], # Tokyo
    [9, 10, 14, 20, 25, 28, 32, 33, 29, 23, 18, 12], # Osaka
    [8, 10, 14, 20, 24, 28, 32, 32, 27, 23, 17, 11], # Shanghai
    [19, 21, 24, 29, 33, 35, 35, 35, 34, 30, 25, 21], # Cairo
    [20, 24, 30, 37, 40, 39, 35, 34, 34, 33, 28, 22], # Delhi
    [26, 28, 32, 35, 36, 35, 33, 32, 33, 35, 32, 28], # Kuala Lumpur
    [25, 29, 32, 33, 33, 32, 32, 32, 32, 31, 29, 26], # Jakarta
    [22, 24, 26, 27, 27, 26, 24, 25, 24, 24, 23, 23], # Mexico City
    [31, 32, 33, 33, 34, 32, 30, 30, 31, 34, 34, 32], # Mumbai
    [29, 29, 28, 27, 23, 23, 23, 25, 25, 26, 27, 28], # Sao Paulo
]
sns.heatmap(
    temperatures,
    yticklabels=cities,
    xticklabels=months,
    cmap='coolwarm',
)
```

### Output

 <Axes: >



## 6. Contoh Kasus

### Kasus Pertama

#### Input

```
import pandas as pd
dataset = pd.read_csv('/content/Sales_Data - Sales_Data.csv')
print('Ukuran Dataset: %d baris dan %d kolom\n' % dataset.shape)
print('Lima data teratas:')
print(dataset.head())
```

## Output



Ukuran Dataset: 25915 baris dan 11 kolom

Lima data teratas:

Unnamed: 0	Order ID	Product	Quantity Ordered	Price Each	\
0	0	295665 Macbook Pro Laptop	1	1700.00	
1	1	295666 LG Washing Machine	1	600.00	
2	2	295667 USB-C Charging Cable	1	11.95	
3	3	295668 27in FHD Monitor	1	149.99	
4	4	295669 USB-C Charging Cable	1	11.95	

	Order Date	Purchase Address	Month	Sales	\
0	2019-12-30 0:01:00	136 Church St, New York City, NY 10001	12.0	1700.0	
1	2019-12-29 7:03:00	562 2nd St, New York City, NY 10001	12.0	600.0	
2	2019-12-12 18:21:00	277 Main St, New York City, NY 10001	12.0	11.95	
3	2019-12-22 15:13:00	410 6th St, San Francisco, CA 94016	12.0	149.99	
4	2019-12-18 12:38:00	43 Hill St, Atlanta, GA 30301	12.0	11.95	

	City	Hour
0	New York City	0.0
1	New York City	7.0
2	New York City	18.0
3	San Francisco	15.0
4	Atlanta	12.0

## Kasus Ke Dua

### Input



```
dataset.describe()
```

### Output



	Unnamed: 0	Order ID	Quantity Ordered	Price Each	Month	Hour
<b>count</b>	25915.000000	25915.000000	25915.000000	25915.000000	25914.000000	25914.000000
<b>mean</b>	12125.955586	302986.455451	1.125140	183.661528	11.700046	14.400170
<b>std</b>	7466.118763	25190.164341	0.444981	332.735800	1.533876	5.413473
<b>min</b>	0.000000	176558.000000	1.000000	2.990000	1.000000	0.000000
<b>25%</b>	5581.500000	300978.500000	1.000000	11.950000	12.000000	11.000000
<b>50%</b>	12093.000000	307198.000000	1.000000	14.950000	12.000000	15.000000
<b>75%</b>	18598.500000	313430.500000	1.000000	150.000000	12.000000	19.000000
<b>max</b>	25116.000000	319670.000000	7.000000	1700.000000	12.000000	23.000000

## Kasus Ke Tiga

### Input



```
dataset.dtypes
```

## Output

0	
Unnamed: 0	int64
Order ID	int64
Product	object
Quantity Ordered	int64
Price Each	float64
Order Date	object
Purchase Address	object
Month	float64
Sales	object
City	object
Hour	float64

dtype: object

Kasus Ke Empat

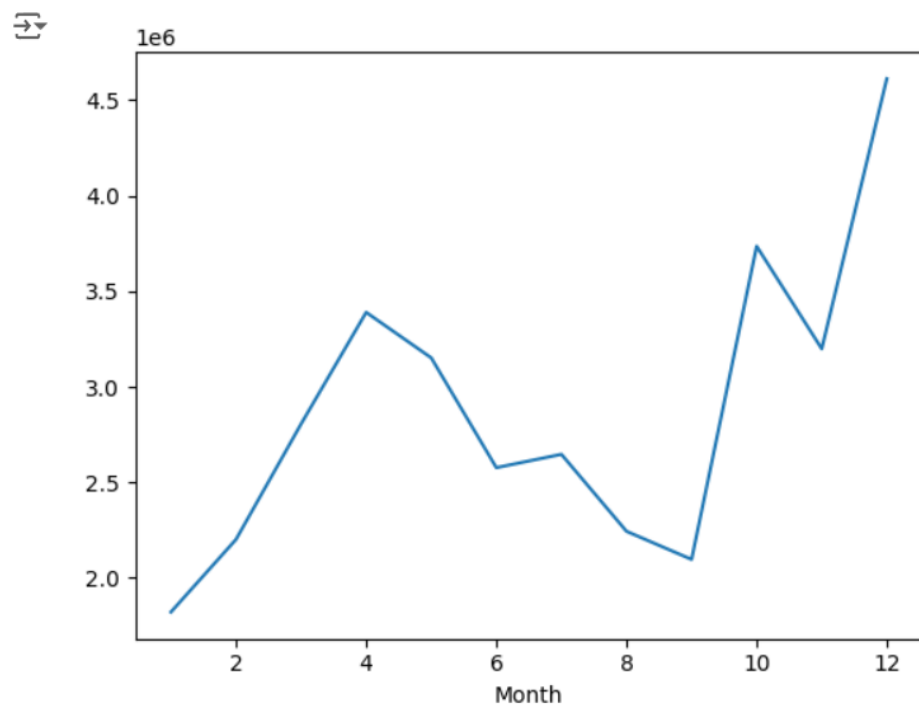
## Input

```
import matplotlib.pyplot as plt

dataset['Sales'] = pd.to_numeric(dataset['Sales'], errors='coerce')
dataset.groupby(['Month'])['Sales'].sum().plot()

plt.show()
```

## Output

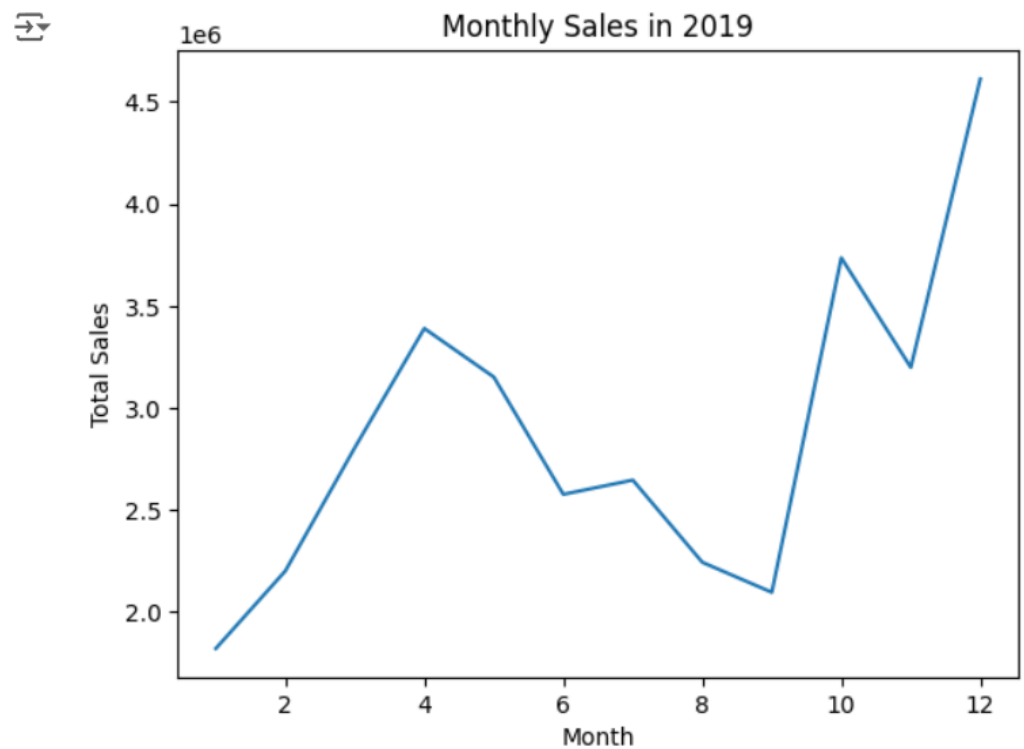


## Kasus Ke Lima

### Input

```
import matplotlib.pyplot as plt
dataset['Sales'] = pd.to_numeric(dataset['Sales'], errors='coerce')
dataset.groupby(['Month'])['Sales'].sum().plot()
plt.title('Monthly Sales in 2019')
plt.xlabel('Month')
plt.ylabel('Total Sales')
plt.show()
```

### Output



## Kasus Ke Enam

### Input

```
top_cities = (dataset.groupby('City')['Sales'].sum().sort_values(ascending=False).head(10))
print(top_cities)
```

### Output

```
City
San Francisco    8258553.17
Los Angeles      5450151.67
New York City    4662897.93
Boston           3660204.51
Atlanta          2794428.58
Dallas           2766776.80
Seattle          2746825.98
Portland         2320012.36
Austin           1818972.00
Name: Sales, dtype: float64
```

## Kasus Ke Tujuh

### Input

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

dataset = pd.read_csv('/content/Sales_Data - Sales_Data.csv')

dataset['cities_top'] = dataset['City'].apply(lambda x: x if x in top_cities.index.to_list() else 'other')
top_cities = dataset['City'].value_counts().nlargest(5)

dataset['Sales'] = pd.to_numeric(dataset['Sales'], errors='coerce')

dataset.groupby(['Month', 'cities_top'])['Sales'].sum().unstack().plot(marker='.', cmap="plasma")

plt.title('Monthly Sales Year 2019 Breakdown by Cities', loc='center', pad=30, fontsize=20, color='blue')
plt.xlabel('Order Month', fontsize=15)
plt.ylabel('Total Sales (in Billions)', fontsize=15)

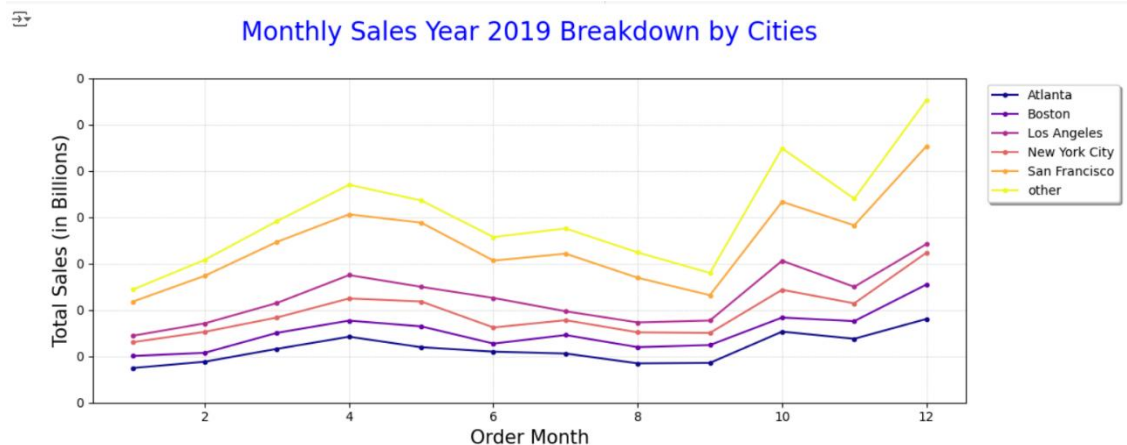
plt.grid(color='darkgray', linestyle=':', linewidth=0.5)
plt.ylim(ymin=0)
labels, locations = plt.yticks()
plt.yticks(labels, (labels / 1000000000).astype(int))

plt.legend(loc='upper center', bbox_to_anchor=(1.1, 1), shadow=True, ncol=1)

plt.gcf().set_size_inches(12, 5)
plt.tight_layout()

plt.show()
```

### Output



## Kasus Ke Delapan

### Input

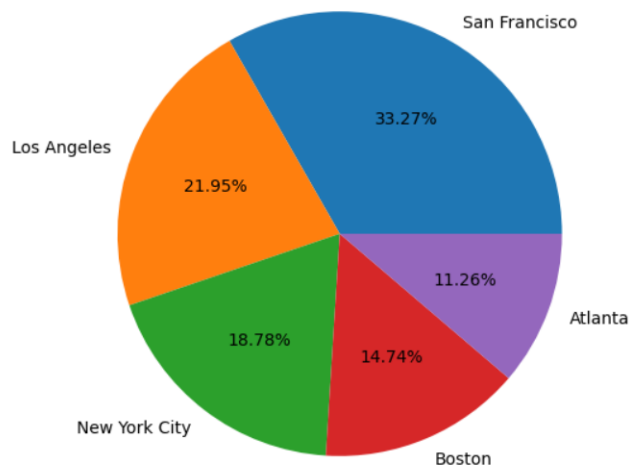
```
import matplotlib.pyplot as plt
top_cities = (dataset.groupby('City') ['Sales']
               .sum()
               .reset_index()
               .sort_values(by='Sales', ascending=False)
               .head())

plt.figure(figsize=(6,6))
plt.pie(top_cities['Sales'], labels = top_cities['City'], autopct='%1.2f%%')
plt.title('Top 5 City with the Biggest Sales in 2019', loc='center', pad=30, fontsize=15, color='blue')
plt.show()
```

## Output



### Top 5 City with the Biggest Sales in 2019



## Kasus Ke Sembilan

### Input

```
import matplotlib.pyplot as plt
top_cities = (dataset.groupby('City') ['Sales']
              .sum()
              .reset_index()
              .sort_values(by='Sales', ascending=False)
              .head())

explode = (0.1, 0, 0, 0, 0)
plt.figure(figsize=(6,6))
plt.pie(top_cities['Sales'], labels = top_cities['City'], autopct='%1.2f%%', explode=explode, shadow=True)
plt.title('Top 5 City with the Biggest Sales in 2019', loc='center', pad=30, fontsize=15, color='blue')
plt.show()
```

### Output



### Top 5 City with the Biggest Sales in 2019

