

Nama : Aldi Pradana
Fakultas : STMIK Methodist Binjai
Tugas : Quiz
Date : 16 Maret 2022

Quiz 1 Minggu ke-4

Jawaban :

- Import Library pandas, numpy, matplotlib, seaborn

```
In [1]: import pandas as pd  
import numpy as np  
import matplotlib.pyplot as plt  
import seaborn as sns
```

- Upload data apy.csv

```
In [2]: data = pd.read_csv('apy.csv')
```

```
In [3]: data
```

Out[3]:

	State_Name	District_Name	Crop_Year	Season	Crop	Area	Production
0	Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Arecanut	1254.0	2000
1	Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Other Kharif pulses	2.0	1
2	Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Rice	102.0	321
3	Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Banana	176.0	641
4	Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Cashewnut	720.0	165
...
246086	West Bengal	PURULIA	2014	Summer	Rice	306.0	801
246087	West Bengal	PURULIA	2014	Summer	Sesamum	627.0	463
246088	West Bengal	PURULIA	2014	Whole Year	Sugarcane	324.0	16250
246089	West Bengal	PURULIA	2014	Winter	Rice	279151.0	597899
246090	West Bengal	PURULIA	2014	Winter	Sesamum	175.0	88

246091 rows × 7 columns

Tercatat bahwa jumlah data berisi 146091 baris dan 7 kolom

Data Preprocessing

In [4]: `data.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 246091 entries, 0 to 246090
Data columns (total 7 columns):
State_Name      246091 non-null object
District_Name   246091 non-null object
Crop_Year       246091 non-null int64
Season          246091 non-null object
Crop            246091 non-null object
Area            246091 non-null float64
Production      246091 non-null object
dtypes: float64(1), int64(1), object(5)
memory usage: 13.1+ MB
```

Terlihat data tidak ada yang null

Describe untuk mendeskripsikan data untuk melihat nilai mean, min, max

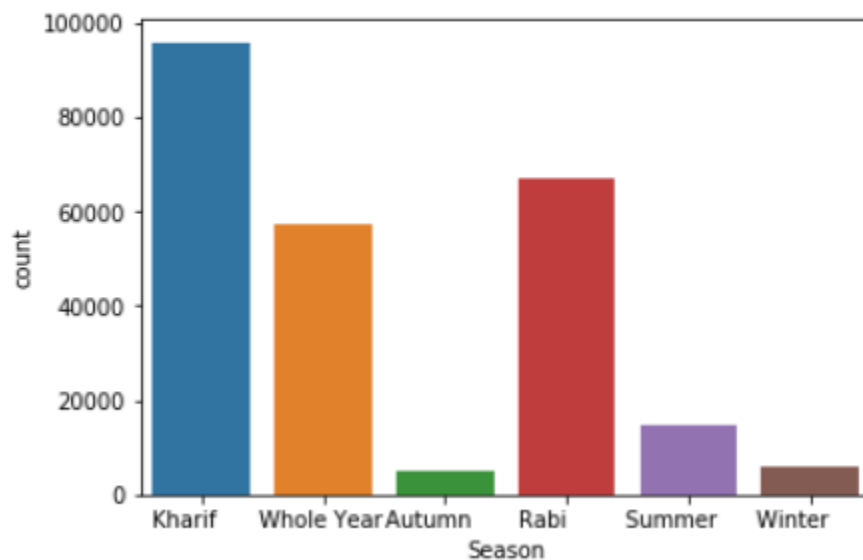
```
In [5]: data.describe()
```

Out[5]:

	Crop_Year	Area
count	246091.000000	2.460910e+05
mean	2005.643018	1.200282e+04
std	4.952164	5.052340e+04
min	1997.000000	4.000000e-02
25%	2002.000000	8.000000e+01
50%	2006.000000	5.820000e+02
75%	2010.000000	4.392000e+03
max	2015.000000	8.580100e+06

Disini saya ingin menghitung jumlah produksi yang dihasilkan berdasarkan Season

```
In [6]: sns.countplot(data['Season'])  
plt.show()
```

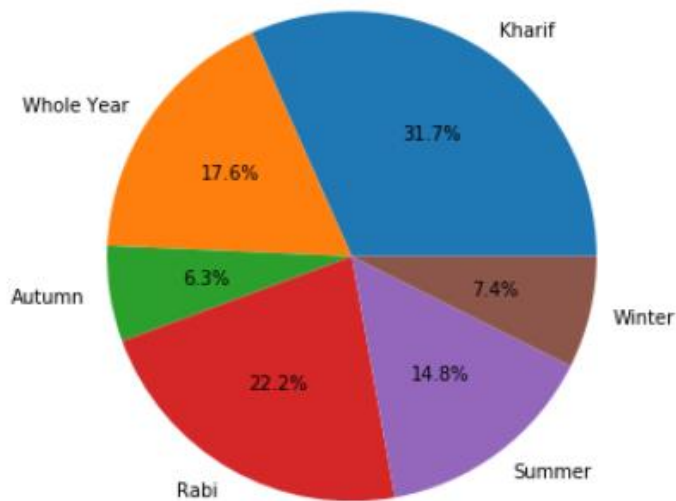


Membuat tampilan dengan menggunakan Chart Pie

```
In [17]: from matplotlib import pyplot as plt
import numpy as np

data1 = ['Kharif','Whole Year', 'Autumn', 'Rabi', 'Summer','Winter',]
data2 = [900, 500, 180, 630,420, 210]

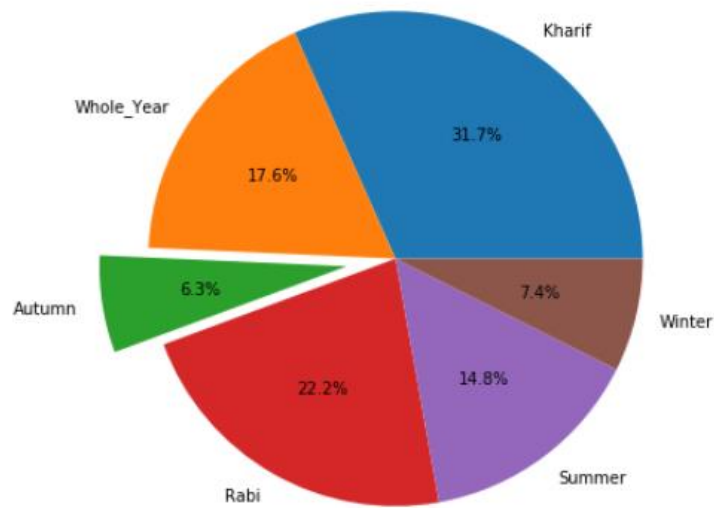
fig = plt.figure(figsize=(8,6))
plt.pie(data2, labels=data1, autopct='%1.1f%%')
plt.show()
```



Karena Autumn merupakan yang terkecil, jadi saya memisahkannya dengan menggunakan explode

```
In [8]: from matplotlib import pyplot as plt
import numpy as np

data1 = ['Kharif', 'Whole Year', 'Autumn', 'Rabi','Summer','Winter']
data2 = [900, 500, 180, 630,420, 210]
fig = plt.figure(figsize=(10,7))
explode=[0,0,0.2,0,0,0]
plt.pie(data2, labels=data1, autopct='%1.1f%%', explode=explode)
plt.show()
```



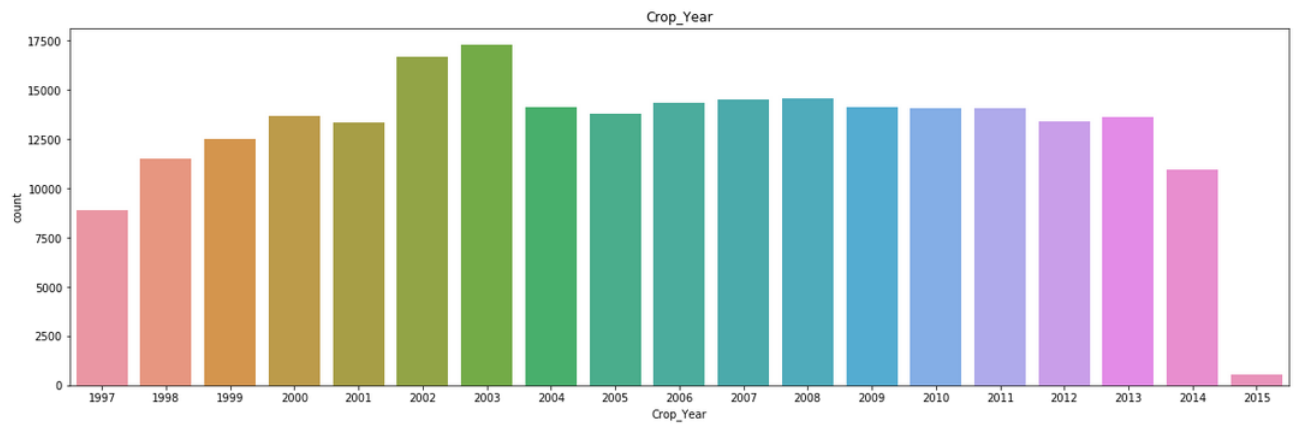
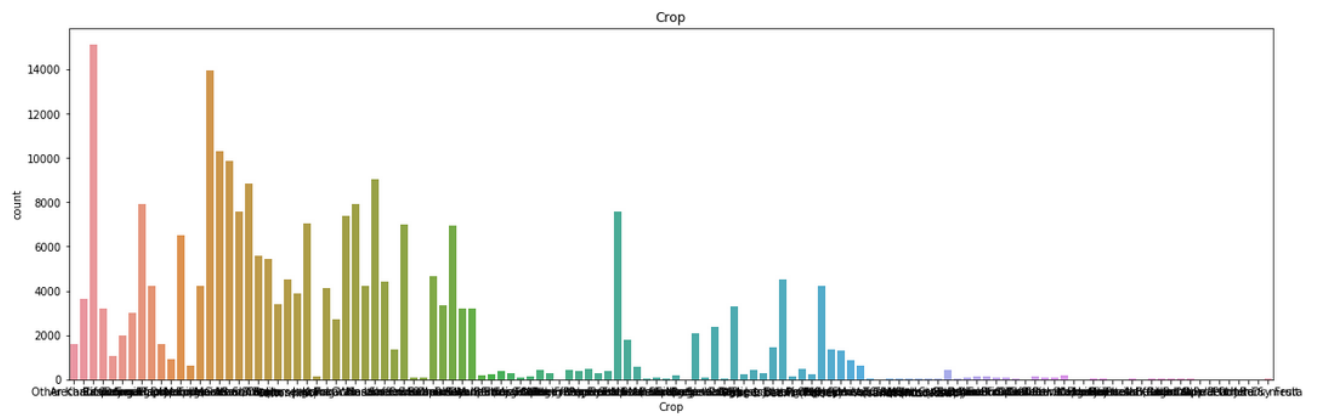
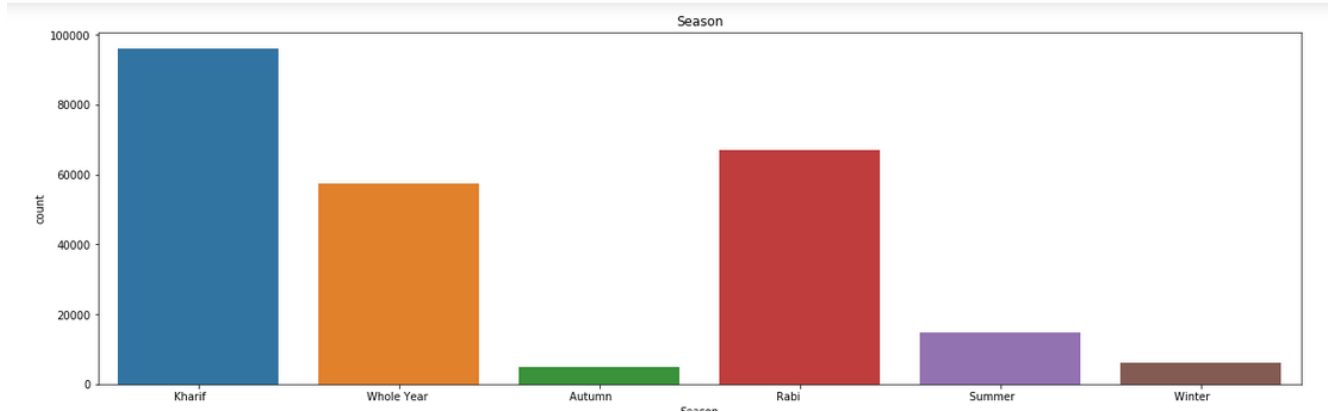
Kesimpulan 1

1. Produksi lebih banyak meningkat pada Season Kharif
2. Autumn merupakan season produksi yang paling kecil

```
In [9]: cat = ['Season', 'Crop', 'Crop_Year']
```

```
In [10]: for column in cat:
plt.figure(figsize=(20,6))
plt.title(f"{column}")
sns.countplot(data[column])
plt.show()
```

Disini saya akan menampilkan insight produksi dengan menggunakan grafik

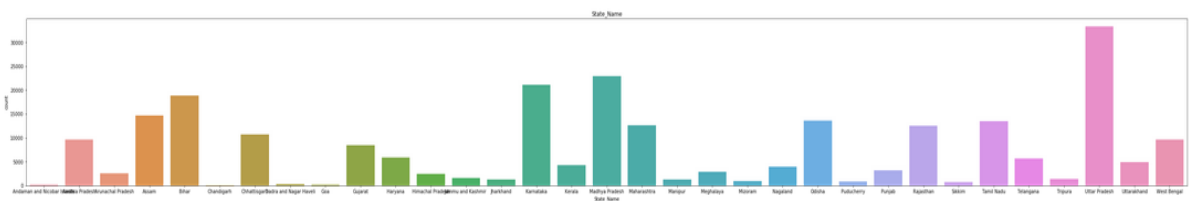


Kesimpulan 2

1. Produksi dari tahun 1997-2000 mengalami peningkatan
2. pada tahun 2003 tercetak sebagai produksi tertinggi
3. produksi tahun 2004-2013 stabil
4. Produksi pada tahun 2015 mengalami penurunan drastis

```
In [11]: Statecat = ['State_Name']
```

```
In [12]: for column in Statecat:  
    plt.figure(figsize=(50,6))  
    plt.title(f"{column}")  
    sns.countplot(data[column])  
    plt.show()
```



Grafik diatas adalah kawasan yang memproduksi kelapa dan lain-lain. Ada yang menurun dan ada yang naik