

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.datasets import make_blobs
from sklearn.cluster import KMeans
```

```
df=pd.read_csv("rainfall.csv")
```

```
year=df["YEAR"]
jf=np.array(df['Jan-Feb'])
mm=np.array(df['Mar-May'])
js=np.array(df['Jun-Sep'])
```

```
od=np.array(df['Oct-Dec'])
ann=np.array(df['ANNUAL'])

annt=ann.reshape(4,4)
print("Matrix of data of annual rainfall\n",annt)
tr = np.transpose(annt)
print("\nTranspose of annual matrix :\n",tr)

MON=df.groupby('YEAR')['ANNUAL'].sum().idxmax()
YEAR=df.groupby('YEAR')['ANNUAL'].sum().max()
print("\nMaximum annual rainfall: ",YEAR)
print("The year of maximum rainfall is: ",MON)
MON1=df.groupby('YEAR')['Jan-Feb'].sum().idxmax()
YEAR1=df.groupby('YEAR')['Jan-Feb'].sum().max()
print("Maximum rainfall in Jan-Feb: ",YEAR1)
print("The year of maximum rainfall is: ",MON1)
```

```
MON2=df.groupby('YEAR')['Mar-  
May'].sum().idxmax()
```

```
YEAR2=df.groupby('YEAR')['Mar-  
May'].sum().max()
```

```
print("Maximum rainfall in Mar-May: ",YEAR2)
```

```
print("The year of maximum rainfall is: ",MON2)
```

```
MON3=df.groupby('YEAR')['Jun-  
Sep'].sum().idxmax()
```

```
YEAR3=df.groupby('YEAR')['Jun-Sep'].sum().max()
```

```
print("Maximum rainfall in Jun-Sep: ",YEAR3)
```

```
print("The year of maximum rainfall is: ",MON3)
```

```
MON4=df.groupby('YEAR')['Oct-  
Dec'].sum().idxmax()
```

```
YEAR4=df.groupby('YEAR')['Oct-Dec'].sum().max()
```

```
print("Maximum rainfall in Oct-Dec: ",YEAR4)
```

```
print("The year of maximum rainfall is: ",MON4)
```

```
print("\n")
```

```
MO=df.groupby('YEAR')['ANNUAL'].sum().idxmin()
```

```
YEA=df.groupby('YEAR')['ANNUAL'].sum().min()
print("Minimum annual rainfall: ",YEA)
print("The year of minimum rainfall is: ",MO)
MO1=df.groupby('YEAR')['Jan-Feb'].sum().idxmin()
YEA1=df.groupby('YEAR')['Jan-Feb'].sum().min()
print("Minimum rainfall in Jan-Feb: ",YEA1)
print("The year of minimum rainfall is: ",MO1)
MO2=df.groupby('YEAR')['Mar-May'].sum().idxmin()
YEA2=df.groupby('YEAR')['Mar-May'].sum().min()
print("Minimum rainfall in Mar-May: ",YEA2)
print("The year of minimum rainfall is: ",MO2)
MO3=df.groupby('YEAR')['Jun-Sep'].sum().idxmin()
YEA3=df.groupby('YEAR')['Jun-Sep'].sum().min()
print("Minimum rainfall in Jun-Sep: ",YEA3)
print("The year of minimum rainfall is: ",MO3)
MO4=df.groupby('YEAR')['Oct-Dec'].sum().idxmin()
YEA4=df.groupby('YEAR')['Oct-Dec'].sum().min()
print("Minimum rainfall in Oct-Dec: ",YEA4)
```

```
print("The year of minimum rainfall is: ",MO4)
```

```
print("\nTotal rainfall in 16 years",ann.sum())
```

```
print("\n")
```

```
print("Average rainfall  
annually",df['ANNUAL'].mean())
```

```
print("Average rainfall in Jan-Feb",df['Jan-  
Feb'].mean())
```

```
print("Average rainfall in Mar-May",df['Mar-  
May'].mean())
```

```
print("Average rainfall in Jun-Sep",df['Jun-  
Sep'].mean())
```

```
print("Average rainfall in Oct-Dec",df['Oct-  
Dec'].mean())
```

```
plt.figure(figsize=(20,10))
```

```
plt.subplot(3,2,1)
```

```
plt.plot(year,jf,color='r')  
plt.xlabel("Year")  
plt.ylabel("Rainfall")  
plt.title("Rainfall in jan-feb")
```

```
plt.subplot(3,2,2)  
plt.bar(year,mm,color='k')  
plt.xlabel("Year")  
plt.ylabel("Rainfall")  
plt.title("Rainfall in Mar-May")
```

```
plt.subplot(3,2,3)  
plt.plot(year,js,color='b')  
plt.xlabel("Year")  
plt.ylabel("Rainfall")  
plt.title("Rainfall in Jun-Sep")
```

```
plt.subplot(3,2,4)  
plt.scatter(year,od,color='g')
```

```
plt.xlabel("Year")  
plt.ylabel("Rainfall")  
plt.title("Rainfall in oct-dec")
```

```
plt.subplot(3,2,5)  
plt.bar(year,ann)  
plt.xlabel("Year")  
plt.ylabel("Rainfall")  
plt.title("Annual Rainfall")
```

```
x=df['ANNUAL']  
x=x.array.reshape(-1, 1)  
y=df['YEAR']
```

```
# Create a K-means clustering model  
kmeans = KMeans(n_clusters=4)
```

```
# Fit the model to the data  
kmeans.fit(x)
```

```
# Predict the clusters for the data points
```

```
y_pred = kmeans.predict(x)
```

```
# Get the coordinates of the cluster centers
```

```
centers = kmeans.cluster_centers_
```

```
# Plot the data points and cluster centers
```

```
plt.subplot(3,2,6)
```

```
plt.scatter(x[:, 0], x[:, 0], c=y_pred, cmap='viridis',  
alpha=0.7)
```

```
plt.scatter(centers[:, 0], centers[:, 0], c='red',  
marker='x')
```

```
plt.xlabel('x')
```

```
plt.ylabel('y')
```

```
plt.title('K-means Clustering')
```