

NAME : NERELLA VENKATA RADHAKRISHNA

ID : 190031187

PRACTICAL : 5

In [2]:

```
import pandas as pd
import numpy as np
```

In [4]:

```
df = pd.read_csv('kerala.csv')
df
```

Out[4]:

	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT
0	KERALA	1901	28.7	44.7	51.6	160.0	174.7	824.6	743.0	357.5	197.7	266.9
1	KERALA	1902	6.7	2.6	57.3	83.9	134.5	390.9	1205.0	315.8	491.6	358.4
2	KERALA	1903	3.2	18.6	3.1	83.6	249.7	558.6	1022.5	420.2	341.8	354.1
3	KERALA	1904	23.7	3.0	32.2	71.5	235.7	1098.2	725.5	351.8	222.7	328.1
4	KERALA	1905	1.2	22.3	9.4	105.9	263.3	850.2	520.5	293.6	217.2	383.5
...
113	KERALA	2014	4.6	10.3	17.9	95.7	251.0	454.4	677.8	733.9	298.8	355.5
114	KERALA	2015	3.1	5.8	50.1	214.1	201.8	563.6	406.0	252.2	292.9	308.1
115	KERALA	2016	2.4	3.8	35.9	143.0	186.4	522.2	412.3	325.5	173.2	225.9
116	KERALA	2017	1.9	6.8	8.9	43.6	173.5	498.5	319.6	531.8	209.5	192.4
117	KERALA	2018	29.1	52.1	48.6	116.4	183.8	625.4	1048.5	1398.9	423.6	356.1

118 rows × 16 columns

In [5]:

```
df["FLOODS"] = df["FLOODS"].map({"YES":1, "NO":0})
```

In [7]:

```
df["JUN_GT_500"]=(df["JUN"]>500).astype('int')
df["JUL_GT_500"]=(df["JUL"]>500).astype('int')
df_small=df.loc[:,{"YEAR","JUN_GT_500","JUL_GT_500","FLOODS"}]
df_small["COUNT"]=1
df_small.head()
```

Out[7]:

	YEAR	FLOODS	JUN_GT_500	JUL_GT_500	COUNT
0	1901	1	1	1	1
1	1902	1	0	1	1
2	1903	1	1	1	1
3	1904	1	1	1	1
4	1905	0	1	1	1

In [9]:

```
df_small.shape
```

Out[9]:

(118, 5)

In [11]:

```
pd.crosstab(df_small["FLOODS"],df_small["JUN_GT_500"])
```

Out[11]:

JUN_GT_500		0	1
FLOODS			
0		19	39
1		6	54

In [16]:

```
P_F = (6 + 54)/(6 + 54 + 19 + 39)
P_J = (39 + 54)/(6 + 54 + 19 + 39)
P_F_intersect_J = 54/(6 + 54 + 19 + 39)
print(f"P(F) : {P_F}")
print(f"P(J) : {P_J}")
print(f"P(F AND J) : {P_F_intersect_J}")
```

P(F) : 0.5084745762711864
P(J) : 0.788135593220339
P(F AND J) : 0.4576271186440678

In [17]:

```
# probability of flood given it rained more than 500 mm in June
P_F_J = P_F_intersect_J/P_J
print(f"P(F|J) : {P_F_J}")
```

P(F|J) : 0.5806451612903226

In [19]:

```
#Probability of rain more than 500 mm in June given it flooded that year
# P_J_F = P_F_intersect_J/P_F
P_J_F = P_F_J * P_J / P_F
print(f"P(J|F) : {P_J_F}")
```

P(J|F) : 0.9000000000000001

In [13]:

```
pd.crosstab(df_small["FLOODS"],df_small["JUL_GT_500"])
```

Out[13]:

JUL_GT_500		0	1
FLOODS			
	0	19	39
	1	3	57

In [15]:



```
P_FLOODS = (3 + 57)/(19 + 39 + 3 + 57)
P_JUL = (39 + 57)/(19 + 39 + 3 + 57)
P_F_intersect_JUL = 57/(19 + 39 + 3 + 57)
print(f"P(F) : {P_FLOODS}")
print(f"P(JULY) : {P_JUL}")
print(f"P(F AND JULY) : {P_F_intersect_JUL}")
```

P(F) : 0.5084745762711864
P(JULY) : 0.8135593220338984
P(F AND JULY) : 0.4830508474576271

In [21]:



```
# probability of flood given it rained more than 500 mm in July
P_F_JUL = P_F_intersect_JUL / P_JUL
print(f"P(F|J) : {P_F_JUL}")
```

P(F|J) : 0.59375

In [23]:



```
#Probability of rain more than 500 mm in July given it flooded that year
# P_JUL_F = P_F_intersect_JUL / P_FLOODS
P_JUL_F = P_F_JUL * P_JUL / P_FLOODS
print(f"P(J|F) : {P_JUL_F}")
```

P(J|F) : 0.9500000000000002