Day 19

DIY

Q1. Problem Statement: Conditional Probability

Load the "*kerala.csv*" data into a DataFrame and perform the following tasks:

1. Explore the DataFrame using `info()` and `describe()` functions
2. June and July are the peak months of rainfall. Consider that if it rains more than 500mm, then the chances of flood become more; create a Datarame with columns –*"YEAR"*, *"JUN_GT_500"* (Contains a boolean value to show whether it rained more than 500 mm in June), *"JUL_GT_500"* (Contains a boolean value to show whether it rained more than 500 mm in July), and *"FLOODS"* (Contains a boolean value to show whether it flooded that year)
3. Calculate the probability of flood given it rained more than 500 mm in June (P(A|B))
4. Calculate the probability of rain more than 500 mm in June, given it flooded that year (P(B|A))
5. Probability of flood given it rained more than 500 mm in July
6. Probability of rain of more than 500 mm in July given it flooded that year (P(B|A))

Dataset:

	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL RAINFALL	FLOODS
0	KERALA	1901	28.7	44.7	51.6	160.0	174.7	824.6	743.0	357.5	197.7	266.9	350.8	48.4	3248.6	YES
1	KERALA	1902	6.7	2.6	57.3	83.9	134.5	390.9	1205.0	315.8	491.6	358.4	158.3	121.5	3326.6	YES
2	KERALA	1903	3.2	18.6	3.1	83.6	249.7	558.6	1022.5	420.2	341.8	354.1	157.0	59.0	3271.2	YES
3	KERALA	1904	23.7	3.0	32.2	71.5	235.7	1098.2	725.5	351.8	222.7	328.1	33.9	3.3	3129.7	YES
4	KERALA	1905	1.2	22.3	9.4	105.9	263.3	850.2	520.5	293.6	217.2	383.5	74.4	0.2	2741.6	NO

Sample Output:

1. June and July are the peak months of rainfall. Consider that if it rains more than 500mm, then the chances of flood become more; create a Datarame with columns – *"YEAR", "JUN_GT_500", "JUL_GT_500",* and *"FLOODS*

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

2. Calculate the probability of flood given it rained more than 500 mm in June (P(A|B)

```
Probailitity of flood given it rained more than 500 mm in June (P(A|B)):
P(Flood|June): 0.5806451612903226
```

3. Calculate the probability of rain more than 500 mm in June, given it flooded that year (P(B|A))

```
Probability of rain more than 500 mm in June given it flooded that year (P(B|A)):
P(June|Flood): 0.9000000000000001
```

4. Probability of flood given it rained more than 500 mm in July

```
Probabilitity of flood given it rained more than 500 mm in July:
P(Flood|July): 0.59375
```

5. Probability of rain of more than 500 mm in July given it flooded that year (P(B|A))

```
Probability of rain more than 500 mm in July given it flooded that year (P(B|A)):
P(July|Flood): 0.9500000000000002
```