Binomial Distribution

Problem Statement

80% of all the visitors to Lavista Museum end up buying souvenirs from the souvenir shop at the Museum. On the coming Sunday, if a random sample of 10 visitors is picked:

1. Find the probability that every visitor will end up buying from the souvenir shop

2. Find the probability that a maximum of 7 visitors will buy souvenirs from the souvenir shop

Let's check first whether we satisfy the assumptions of the binomial distribution.

• There are only two possible outcomes (success or failure) for each trial. A visitor will buy souvenirs from the souvenir shop or not (yes or no).

• Number of trials (n) is fixed - There are 10 visitors in the sample.

• Each trial is independent of the other trials - It is reasonable to assume that the buying activity of visitors is independent.

• The probability of success (p) is the same for each trial - The probability of success for each visitor is 0.8.

Import the necessary libraies

```
import numpy as np
import pandas as pd
from scipy.stats import binom

df = pd.read_csv('/Users/vishal/Desktop/CSV files/sat_score.csv')

df.head()
```

```
   student_id  score
0           1   1018
1           2   1218
2           3    611
3           4    723
4           5    541
```

	student_id	score
0	1	1018
1	2	1218
2	3	611
3	4	723
4	5	541

```
df.tail()
```

```
     student_id  score
995         996    871
996         997    752
997         998   1087
998         999    987
999        1000   1005
```

	student_id	score
995	996	871
996	997	752
997	998	1087
998	999	987
999	1000	1005

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
```

```
Data columns (total 2 columns):
 #   Column      Non-Null Count  Dtype
---  ------      --------------  -----
 0   student_id  1000 non-null   int64
 1   score       1000 non-null   int64
dtypes: int64(2)
memory usage: 15.8 KB
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 2 columns):
 #   Column      Non-Null Count  Dtype
---  ------      --------------  -----
 0   student_id  1000 non-null   int64
 1   score       1000 non-null   int64
dtypes: int64(2)
memory usage: 15.8 KB
```

```
df.describe()
```

```
        student_id        score
count  1000.000000  1000.000000
mean    500.500000  1007.460000
std     288.819436   204.426007
min       1.000000   288.000000
25%     250.750000   873.000000
50%     500.500000  1010.000000
75%     750.250000  1148.000000
max    1000.000000  1612.000000
```

	student_id	score
count	1000.000000	1000.000000
mean	500.500000	1007.460000
std	288.819436	204.426007
min	1.000000	288.000000
25%	250.750000	873.000000
50%	500.500000	1010.000000
75%	750.250000	1148.000000
max	1000.000000	1612.000000

Probability That Every Visitor Buys Souvenirs

```
# Given parameters
n = 10  # Number of trials (visitors)
p = 0.8  # Probability of success (buying souvenirs)

# Probability that every visitor buys souvenirs (P(X = 10))
prob_all_buy = binom.pmf(10, n, p)
print(f"Probability that all 10 visitors buy souvenirs:
{prob_all_buy}")
```

```
Probability that all 10 visitors buy souvenirs: 0.10737418240000006
```


```
Probability that all 10 visitors buy souvenirs: 0.10737418240000006
```


Probability That a Maximum of 7 Visitors Buy Souvenirs

```
prob_max_7 = binom.cdf(7, n, p)
print(f"Probability that at most 7 visitors buy souvenirs:
{prob_max_7}")
```

```
Probability that at most 7 visitors buy souvenirs: 0.32220047359999987
```

```
Probability that at most 7 visitors buy souvenirs: 0.32220047359999987
```