

## Problem statement

The number of books sold by a bookseller per day is given in 'bookseller.csv'.

Let

$X$  = Number of books sold by a bookseller per day

$X$  is a Discrete Random variable (because it represents the book count).  
Let's see the distribution of  $X$  and answer the below questions.

1. Find the probability that more than (or equal to) 96 books will be sold on a given day
2. Find the probability that less than (or equal to) 92 books will be sold on a given day

```
import pandas as pd
import numpy as np

df = pd.read_csv('/Users/vishal/Desktop/CSV files/Bookseller.csv')
df.head()
```

	S.No	Date	Number of Books Sold
0	1	01-01-2020	90
1	2	02-01-2020	100
2	3	03-01-2020	100
3	4	04-01-2020	97
4	5	05-01-2020	93

	S.No	Date	Number of Books Sold
<b>0</b>	1	01-01-2020	90
<b>1</b>	2	02-01-2020	100
<b>2</b>	3	03-01-2020	100
<b>3</b>	4	04-01-2020	97
<b>4</b>	5	05-01-2020	93

```
df.tail()
```

	S.No	Date	Number of Books Sold
361	362	27-12-2020	91
362	363	28-12-2020	90
363	364	29-12-2020	92
364	365	30-12-2020	92
365	366	31-12-2020	99

	S.No	Date	Number of Books Sold
<b>361</b>	362	27-12-2020	91
<b>362</b>	363	28-12-2020	90
<b>363</b>	364	29-12-2020	92
<b>364</b>	365	30-12-2020	92
<b>365</b>	366	31-12-2020	99

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 366 entries, 0 to 365
Data columns (total 3 columns):
#   Column                Non-Null Count  Dtype
---  -
0   S.No                  366 non-null   int64
1   Date                  366 non-null   object
2   Number of Books Sold  366 non-null   int64
```

```
dtypes: int64(2), object(1)
memory usage: 8.7+ KB
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 366 entries, 0 to 365
Data columns (total 3 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   S.No                                   366 non-null    int64
1   Date                                   366 non-null    object
2   Number of Books Sold                 366 non-null    int64
dtypes: int64(2), object(1)
memory usage: 8.7+ KB
```

```
df.describe()
```

	S.No	Number of Books Sold
count	366.000000	366.000000
mean	183.500000	94.961749
std	105.799338	3.178465
min	1.000000	90.000000
25%	92.250000	92.000000
50%	183.500000	95.000000
75%	274.750000	98.000000
max	366.000000	100.000000

	S.No	Number of Books Sold
count	366.000000	366.000000
mean	183.500000	94.961749
std	105.799338	3.178465
min	1.000000	90.000000
25%	92.250000	92.000000
50%	183.500000	95.000000
75%	274.750000	98.000000
max	366.000000	100.000000

## Check the distribution Number of Books sold

```
book_distribution = df['Number of Books  
Sold'].value_counts().sort_index()  
prob_distribution = book_distribution / book_distribution.sum()  
prob_distribution
```

```
Number of Books Sold  
90      0.087432  
91      0.095628  
92      0.092896  
93      0.117486  
94      0.068306  
95      0.087432  
96      0.087432  
97      0.084699  
98      0.087432  
99      0.112022  
100     0.079235  
Name: count, dtype: float64
```

```
Number of Books Sold  
90      0.087432  
91      0.095628  
92      0.092896  
93      0.117486  
94      0.068306  
95      0.087432  
96      0.087432  
97      0.084699  
98      0.087432  
99      0.112022  
100     0.079235  
Name: count, dtype: float64
```

---

## Calculate the probability of selling $\geq 96$ books

```
prob_more_equal_96 = prob_distribution[prob_distribution.index >=  
96].sum()  
print(f"Probability of selling  $\geq 96$  books: {prob_more_equal_96}")  
Probability of selling  $\geq 96$  books: 0.4508196721311476
```

Probability of selling  $\geq 96$  books: 0.4508196721311476

## Calculate the probability of selling $\leq 92$ books

```
prob_less_equal_92 = prob_distribution[prob_distribution.index <=
92].sum()
print(f"Probability of selling  $\leq 92$  books: {prob_less_equal_92}")
```

Probability of selling  $\leq 92$  books: 0.27595628415300544

Probability of selling  $\leq 92$  books: 0.27595628415300544

## Optional: Visualize the Probability Distribution

```
import matplotlib.pyplot as plt

# Plot the probability distribution
prob_distribution.plot(kind='bar')
plt.title('Probability Distribution of Number of Books Sold per Day')
plt.xlabel('Number of Books Sold')
plt.ylabel('Probability')
plt.show()
```



