

Bell Curve - Normal Distribution



What we did:

In last class we learned about analysing data through visualization.

In this class we learned about the bell curve - normal distribution and how to plot it.

How we did it:

1. We visited random.org site and rolled the dice multiple times and recorded the different sums in a file.



2. We wrote python code to get the random dice numbers 100 times and store in a list.

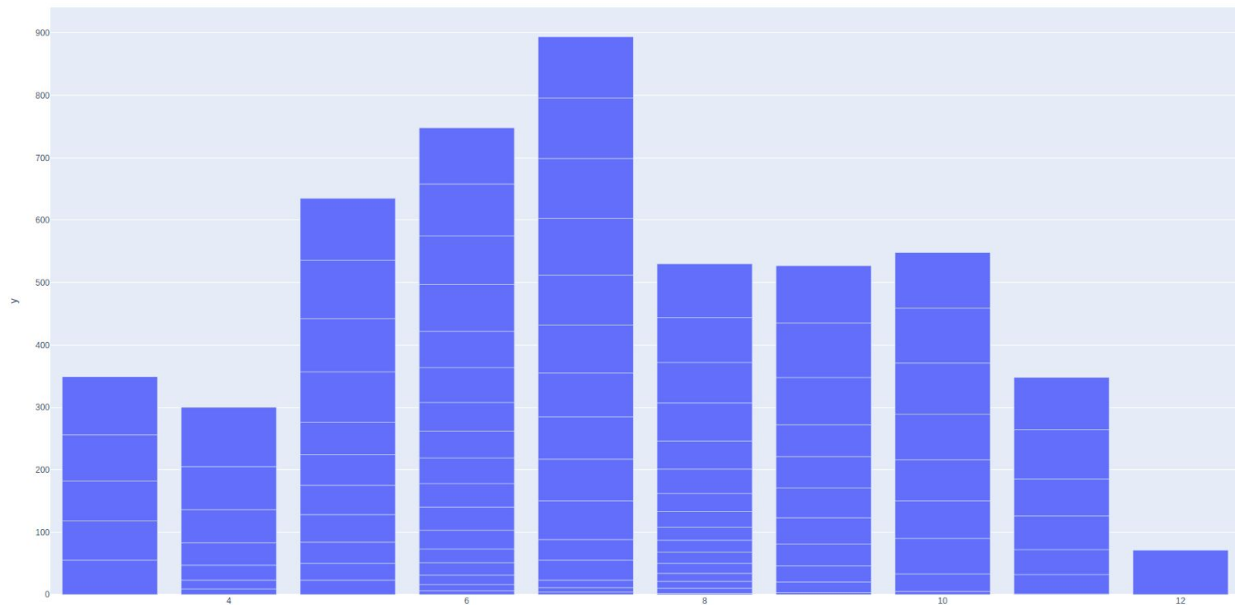
```
import random

dice_result = []
for i in range(0, 100):
    dice1 = random.randint(1, 6)
    dice2 = random.randint(1, 6)
    dice_result.append(dice1 + dice2)
```

3. Using plotly express we plotted the list on the bar graph.

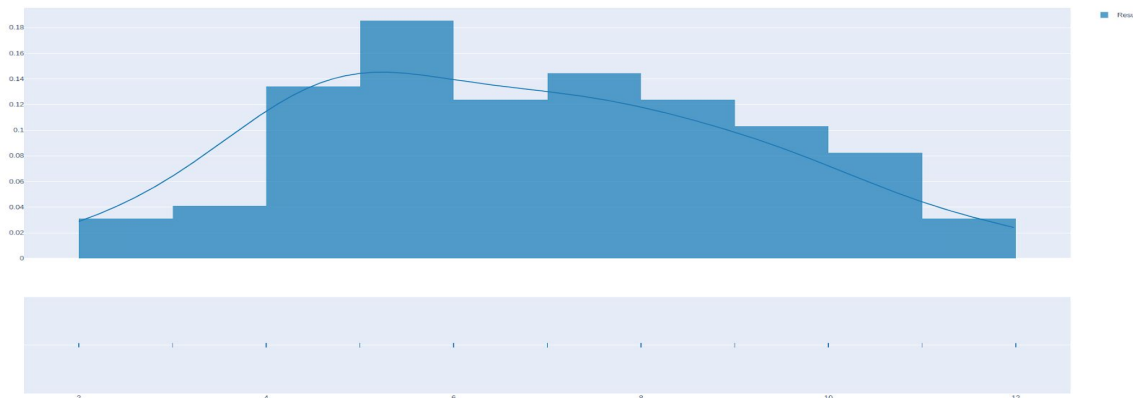
```
import random
import plotly.express as px
count = []
dice_result = []
for i in range(0, 100):
    dice1 = random.randint(1, 6)
    dice2 = random.randint(1, 6)
    dice_result.append(dice1 + dice2)
    count.append(i)

fig = px.bar(x=dice_result, y=count)
fig.show()
```



4. Then we drew the distribution plot using plotly's figure_factory module.

```
dice.py
1  import plotly.figure_factory as ff
2  import random
3
4  dice_result = []
5  for i in range(0, 100):
6      dice1 = random.randint(1, 6)
7      dice2 = random.randint(1, 6)
8      dice_result.append(dice1 + dice2)
9
10 fig = ff.create_distplot([dice_result], ["Result"])
11 fig.show()
12
13
```



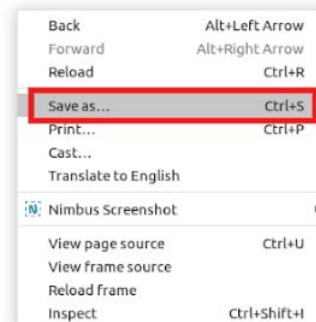
5. Then using the height and weight data that we used in earlier classes we plotted a distribution plot.

6. We got the data and stored it in the data.csv file.



Index,Height(Inches),Weight(Pounds)

```
1,65.78331,112.9925
2,71.51521,136.4873
3,69.39874,153.0269
4,68.2166,142.3354
5,67.78781,144.2971
6,68.69784,123.3024
7,69.80204,141.4947
8,70.01472,136.4623
9,67.90265,112.3723
10,66.78236,120.6672
11,66.48769,127.4516
12,67.62333,114.143
13,68.30248,125.6107
14,67.11656,122.4618
15,68.27967,116.0866
16,71.0916,139.9975
17,66.461,129.5023
18,68.64927,142.9733
19,71.23033,137.9025
20,67.13118,124.0449
21,67.83379,141.2807
22,68.87881,143.5392
23,63.48115,97.90191
24,68.42187,129.5027
25,67.62804,141.8501
26,67.20864,129.7244
27,70.84235,142.4235
28,67.49434,131.5502
```



7. Using pandas we read the data.

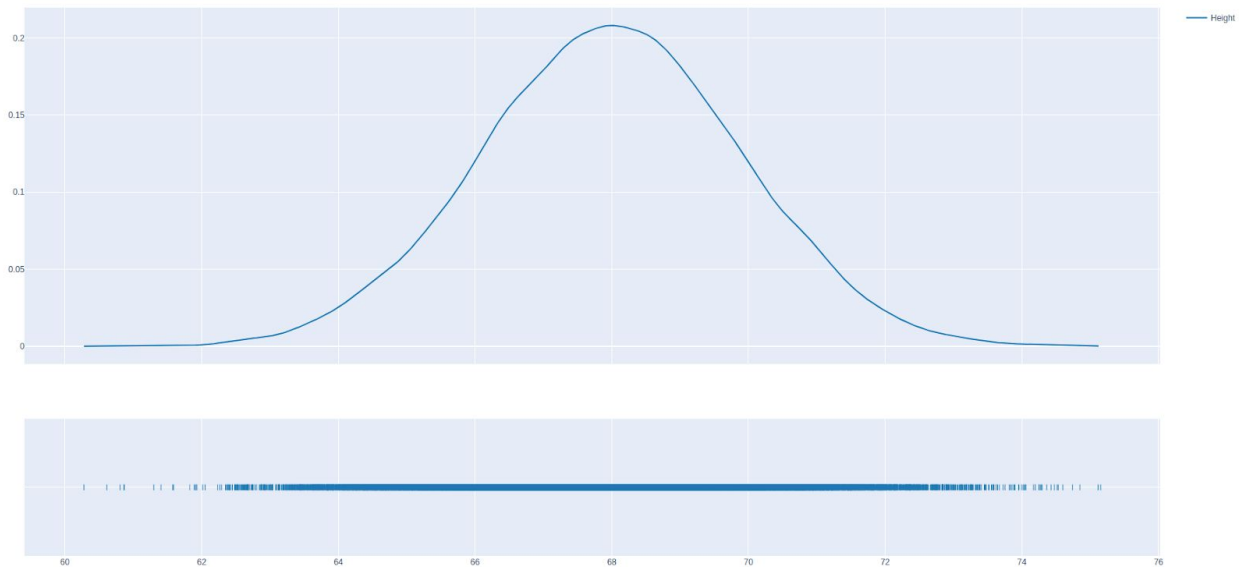
```
import pandas as pd
import csv

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

8. We plotted the height distribution plot.

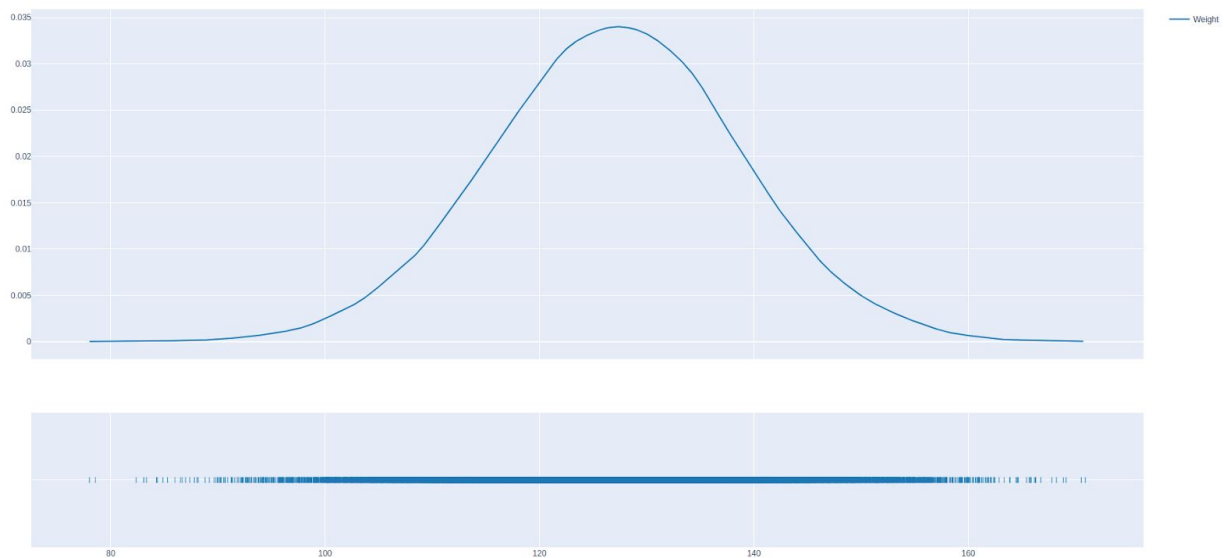
```
import plotly.figure_factory as ff
import pandas as pd
import csv

df = pd.read_csv("data.csv")
fig = ff.create_distplot([df["Height(Inches)"].tolist()], ["Height"], show_hist=False)
fig.show()
```



9. We also plotted the weight distribution plot.

```
1 import plotly.figure_factory as ff
2 import pandas as pd
3 import csv
4
5
6 df = pd.read_csv("data.csv")
7 fig = ff.create_distplot([df["Weight(Pounds)"].tolist()], ["Weight"], show_hist=False)
8 fig.show()
9
```



What's next?

In the next class, we will learn more about normal distribution.