Ankush Kumar

CSE - P

AP21110011026

```
#Import libraries
import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
import seaborn as sns
import matplotlib.pyplot as plt
from scipy.stats import norm

# Import dataset using a raw string
height_weight = pd.read_csv(r"C:\Users\tiwar\Downloads\SOCR-
HeightWeight.csv")


# Check data using head()
height_weight.head()
```

```
   Index  Height(Inches)  Weight(Pounds)
0      1        65.78331        112.9925
1      2        71.51521        136.4873
2      3        69.39874        153.0269
3      4        68.21660        142.3354
4      5        67.78781        144.2971
```

```
#check info about dataset
height_weight.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 25000 entries, 0 to 24999
Data columns (total 3 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   Index           25000 non-null  int64
 1   Height(Inches)  25000 non-null  float64
 2   Weight(Pounds)  25000 non-null  float64
dtypes: float64(2), int64(1)
memory usage: 586.1 KB
```

```
#cut only 'Height' column for our data
data = height_weight.iloc[:, 1]
print(data)
```

```
0        65.78331
1        71.51521
```

```
2         69.39874
3         68.21660
4         67.78781
             ...
24995     69.50215
24996     64.54826
24997     64.69855
24998     67.52918
24999     68.87761
Name: Height(Inches), Length: 25000, dtype: float64
```

```
mean = data.mean()
print(mean)
```

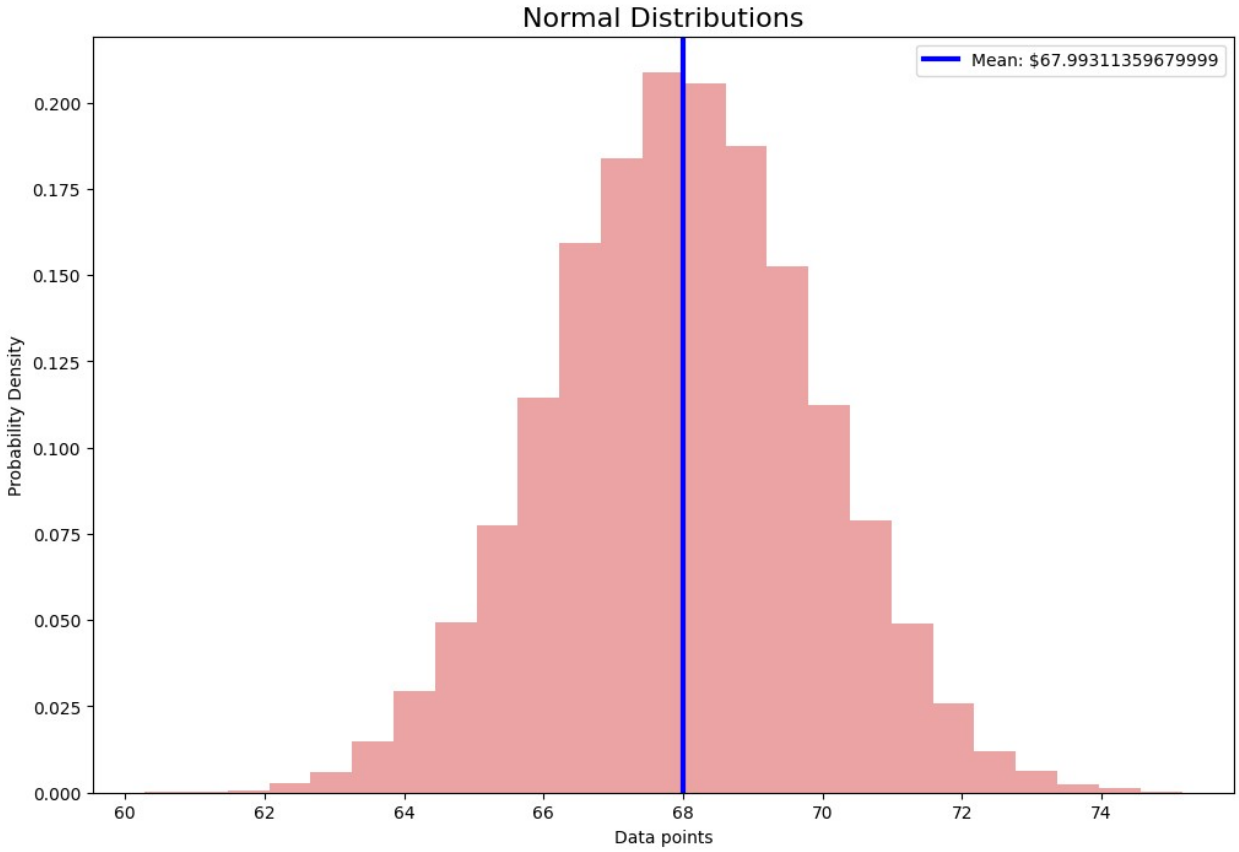
```
67.99311359679999
```

```
plt.figure(figsize=(12,8))
plt.hist(data, bins=25, density=True, alpha=0.6, color='#e06666')

plt.axvline(mean, color='blue',lw=3, label=f'Mean: ${mean}')

plt.xlabel('Data points')
plt.ylabel('Probability Density')
plt.title('Normal Distributions', fontsize=16)
plt.legend()
```

```
<matplotlib.legend.Legend at 0x2483f3a23f0>
```




```
median = data.median()
print(median)
```

```
67.9957
```

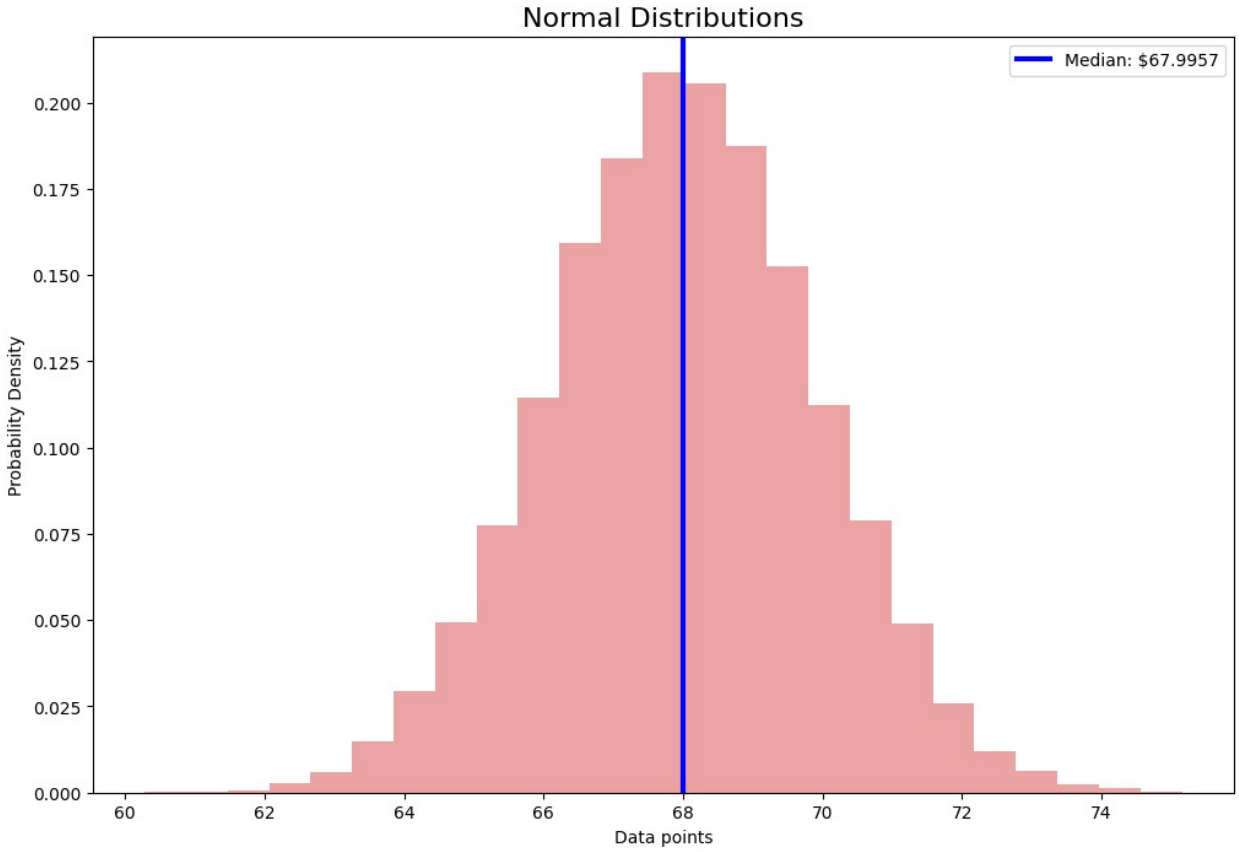
```
#get median
median = data.median()

plt.figure(figsize=(12,8))
plt.hist(data, bins=25, density=True, alpha=0.6, color='#e06666')

plt.axvline(median, color='blue',lw=3, label=f'Median: ${median}')

plt.xlabel('Data points')
plt.ylabel('Probability Density')
plt.title('Normal Distributions', fontsize=16)
plt.legend()
```

```
<matplotlib.legend.Legend at 0x248404719d0>
```




```
mode = data.mode()
print(mode)
```

```
0     65.65796
1     67.34629
2     67.41672
3     67.94061
4     67.97769
5     67.98509
6     68.30287
7     68.92380
8     68.93456
9     68.97830
10    70.04724
Name: Height(Inches), dtype: float64
```

```
#get mode
mode = data.mode()[9] #I took the last one

plt.figure(figsize=(12,8))
plt.hist(data, bins=25, density=True, alpha=0.6, color='#e06666')

plt.axvline(mode, color='blue',lw=3, label=f'Mode: ${mode}')
```

```
plt.xlabel('Data points')
plt.ylabel('Probability Density')
plt.title('Normal Distributions', fontsize=16)
plt.legend()
```

```
<matplotlib.legend.Legend at 0x2484049bad0>
```