Central Limit Theorem - Part III

You saw how the **Central Limit Theorem** worked for the sample mean in the earlier concept. However, let's consider another example to see a case where the **Central Limit Theorem** doesn't work...

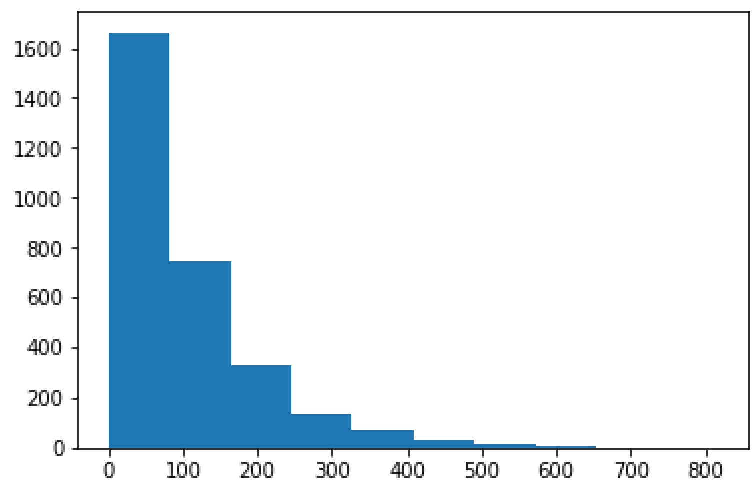
Work through the questions and use the created variables to answer the questions that follow below the notebook.

Run the below cell to get started.

In [1]:
```
import numpy as np
import matplotlib.pyplot as plt

%matplotlib inline
np.random.seed(42)

pop_data = np.random.gamma(1,100,3000)
plt.hist(pop_data);
```



`1.` In order to create the sampling distribution for the variance of 100 draws of this distribution, follow these steps:

`a.` Use numpy's **random.choice** to simulate 100 draws from the `pop_data` array.

`b.` Compute the variance of these 100 draws.

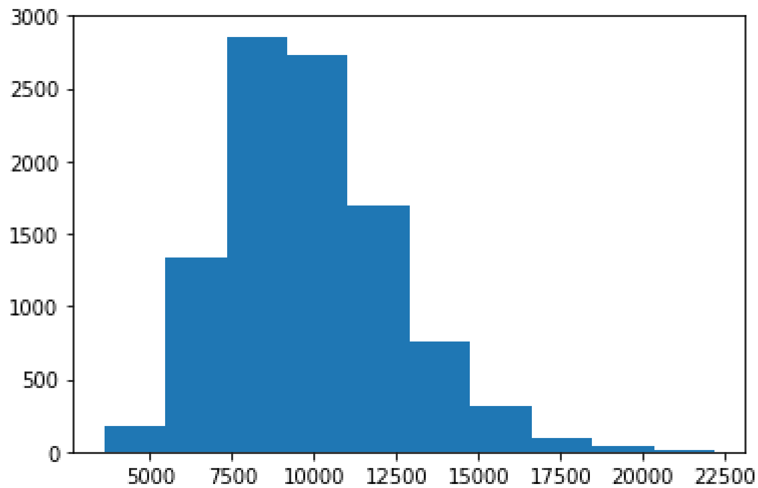
`c.` Write a loop to simulate this process 10,000 times, and store each variance into an array called **var_size_100**.

`d.` Plot a histogram of your sample variances.

`e.` Use **var_size_100** and **pop_data** to answer the quiz questions below.

In [2]:
```
vars_size_100 = []
for _ in range(10000):
    sample = np.random.choice(pop_data, 100)
    vars_size_100.append(sample.var())

plt.hist(vars_size_100);
```



In [6]:
```
pop_data.var() # Variance of the population
```

Out[6]: 9955.7693930654896

In [8]:
```
np.mean(vars_size_100) # The mean of the sampling distribution for the samp
```

Out[8]: 9874.9739450196776

In [9]:
```
np.var(vars_size_100) # The variance of the sampling distribution for the s
```

Out[9]: 6507061.7703391286

Sample variances are actually known to follow a different type of mathematical distribution known as a chi-squared distribution. Which is a more right-skewed distribution than a normal.

In [ ]:
```

```