assignment3-stml

August 23, 2024

1. Find out the maximum likelihood parameters values for california_housing ['latiyude','total_rooms','house_median_age','total_bedrooms'] data, each column individually.

[1]:
```
import pandas as pd
import numpy as np
df=pd.read_csv('/content/sample_data/california_housing_test.csv')
df.head()
```

[1]:

	longitude	latitude	housing_median_age	total_rooms	total_bedrooms	\
0	-122.05	37.37	27.0	3885.0	661.0	
1	-118.30	34.26	43.0	1510.0	310.0	
2	-117.81	33.78	27.0	3589.0	507.0	
3	-118.36	33.82	28.0	67.0	15.0	
4	-119.67	36.33	19.0	1241.0	244.0	

	population	households	median_income	median_house_value
0	1537.0	606.0	6.6085	344700.0
1	809.0	277.0	3.5990	176500.0
2	1484.0	495.0	5.7934	270500.0
3	49.0	11.0	6.1359	330000.0
4	850.0	237.0	2.9375	81700.0

[2]:
```
lt=df['latitude']
tm=df['total_rooms']
tb=df['total_bedrooms']
hma=df['housing_median_age']
print(f"mean of latitude is {lt.mean()} variance of latitude is {lt.var()}")
print(f"mean of total_rooms is {tm.mean()} variance of total_rooms is {tm.
↪var()}")
print(f"mean of total_bedrooms is {tb.mean()} variance of total_bedrooms is {tb.
↪var()}")
print(f"mean of house_median_age is {hma.mean()} variance of house_median_age␣
↪is {hma.var()}")
```

```
mean of latitude is 35.635389999999994 variance of latitude is 4.535492278659553
mean of total_rooms is 2599.578666666667 variance of total_rooms is
4646582.611348672
mean of total_bedrooms is 529.9506666666666 variance of total_bedrooms is
```

```
172768.55375080582
mean of house_median_age is 28.845333333333333 variance of house_median_age is
157.6379575414027
```

2. Find log-likelihood values for 50, 75, and 80. For California_housing [‘latitude’] and find the Max likelihood values.

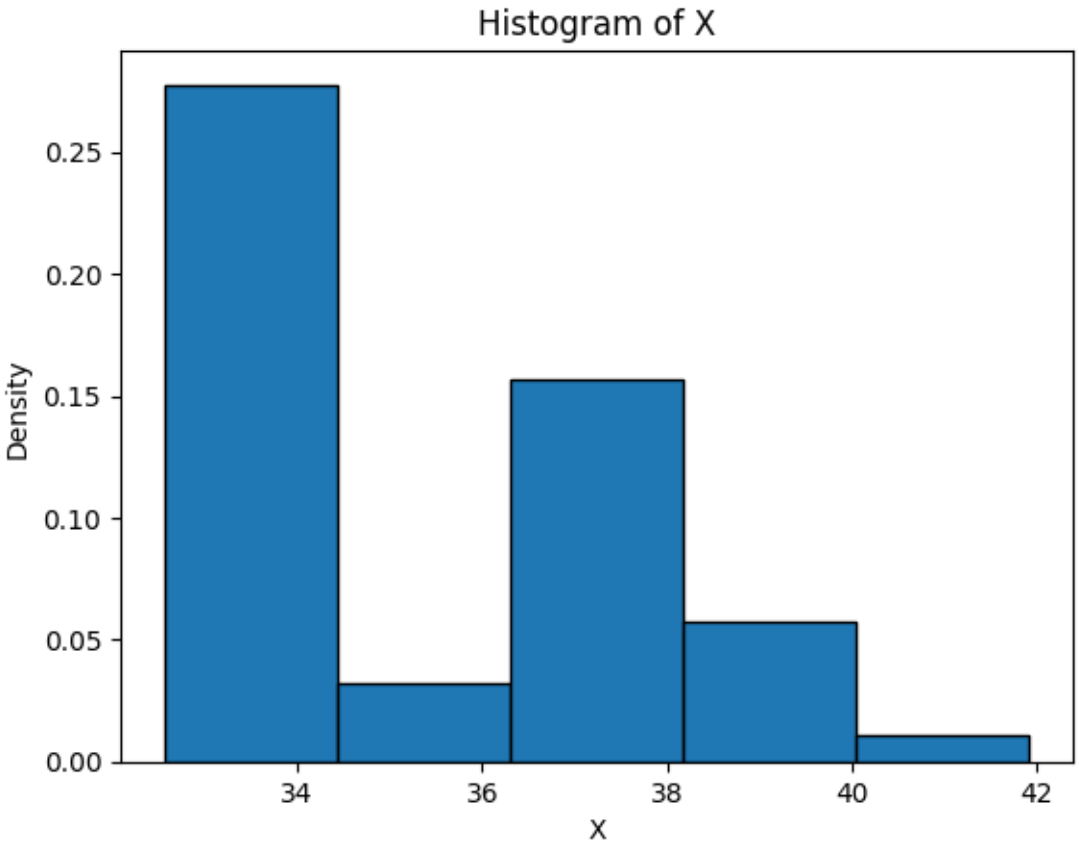
[3]:
```
def log_likelihood(x, mean, var):
    b = -((x - mean) ** 2) / (2 * var)
    log_mle = b - np.log(np.sqrt(2 * np.pi * var))
    return log_mle
x=[50,75,80]
l=df['latitude']
mean=l.mean()
var=l.var()
for i in x:
  print(log_likelihood(i,mean,var))
```

```
-24.422383214813756
-172.50228474173102
-218.6545075081348
```

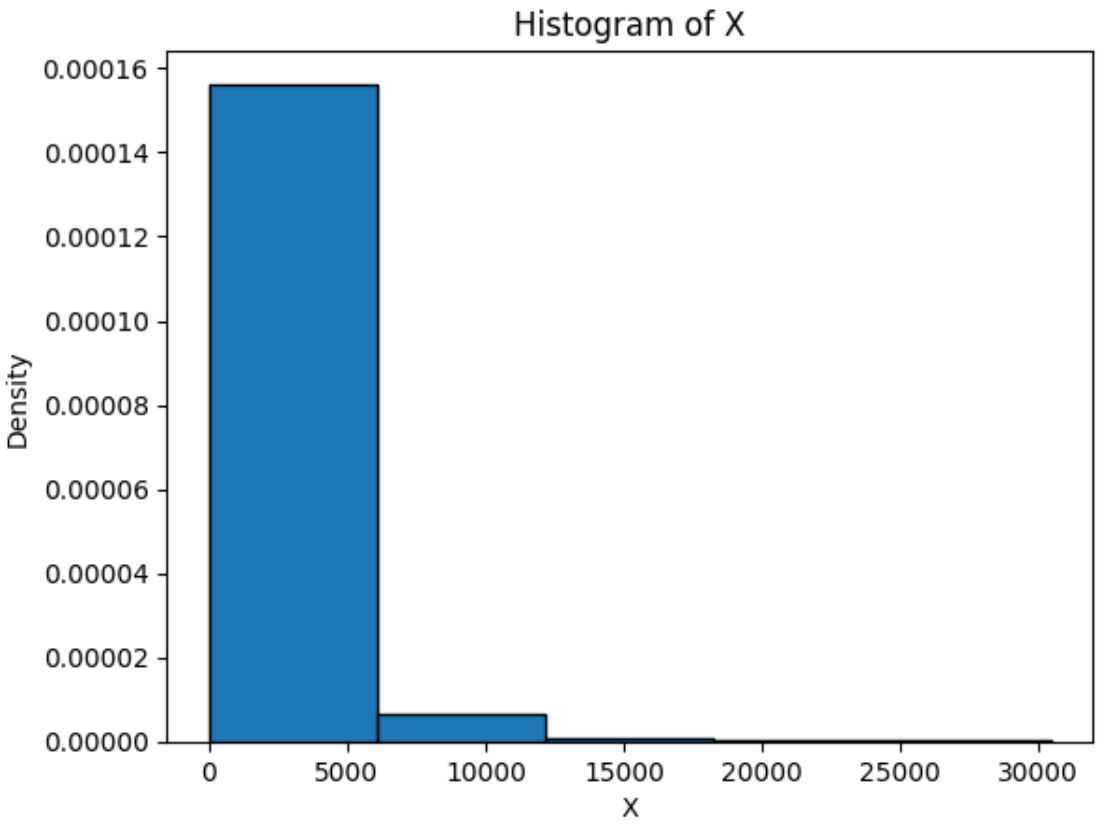
3. Estimate the density of each column of california_housing [‘latitude’,total_rooms’,’house_median_age’,’total_bedrooms’] data, For the bins [5, 10, 15, 20]. And write the observations.

[4]:
```
import matplotlib.pyplot as plt
x1=df['latitude']
x2=df['total_rooms']
x3=df['housing_median_age']
x4=df['total_bedrooms']
def density(x,bins):
  plt.hist(x,bins=bins,density=True,edgecolor='black')
  plt.title('Histogram of X')
  plt.xlabel('X')
  plt.ylabel('Density')
  plt.show()
x=[5,10,15,20]
for i in x:
  print(f"bins are {i}")
  print("latitude")
  print(density(x1,i))
  print("total_rooms")
  print(density(x2,i))
  print("housing_median_age")
  print(density(x3,i))
  print("total_bedrooms")
  print(density(x4,i))
```

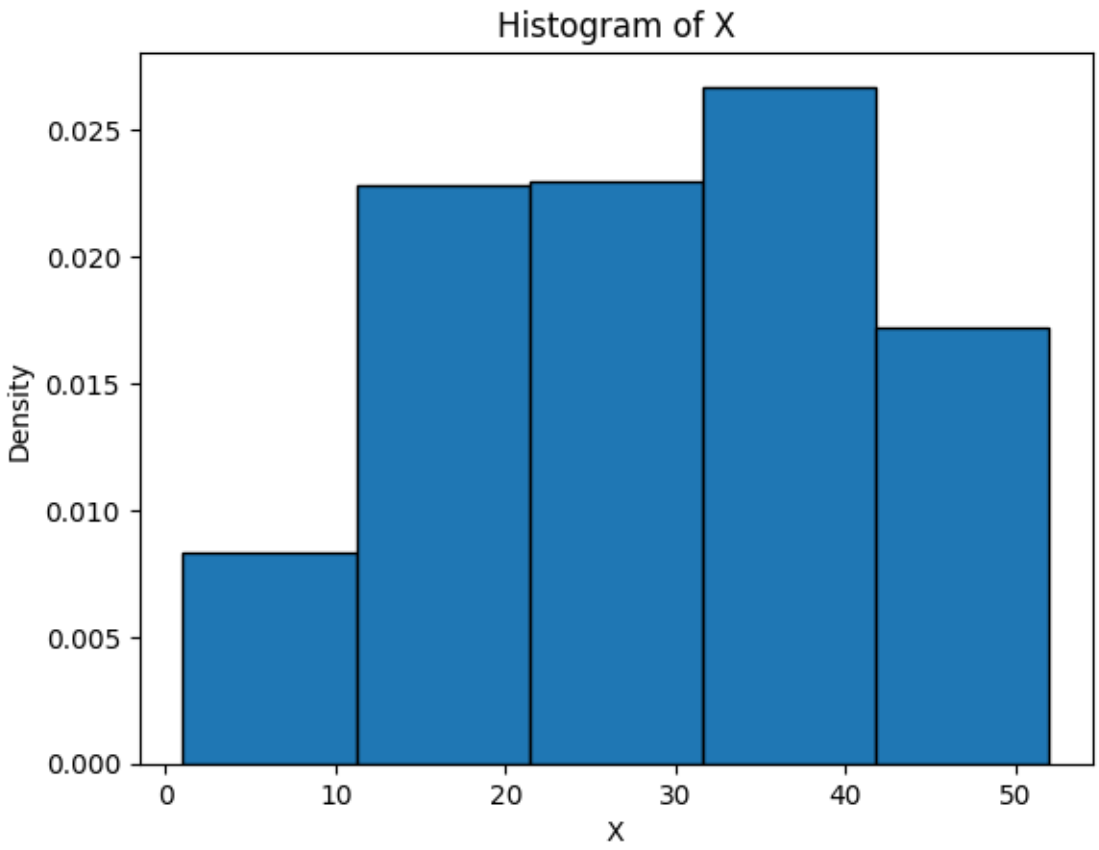
```
bins are 5
latitude
```



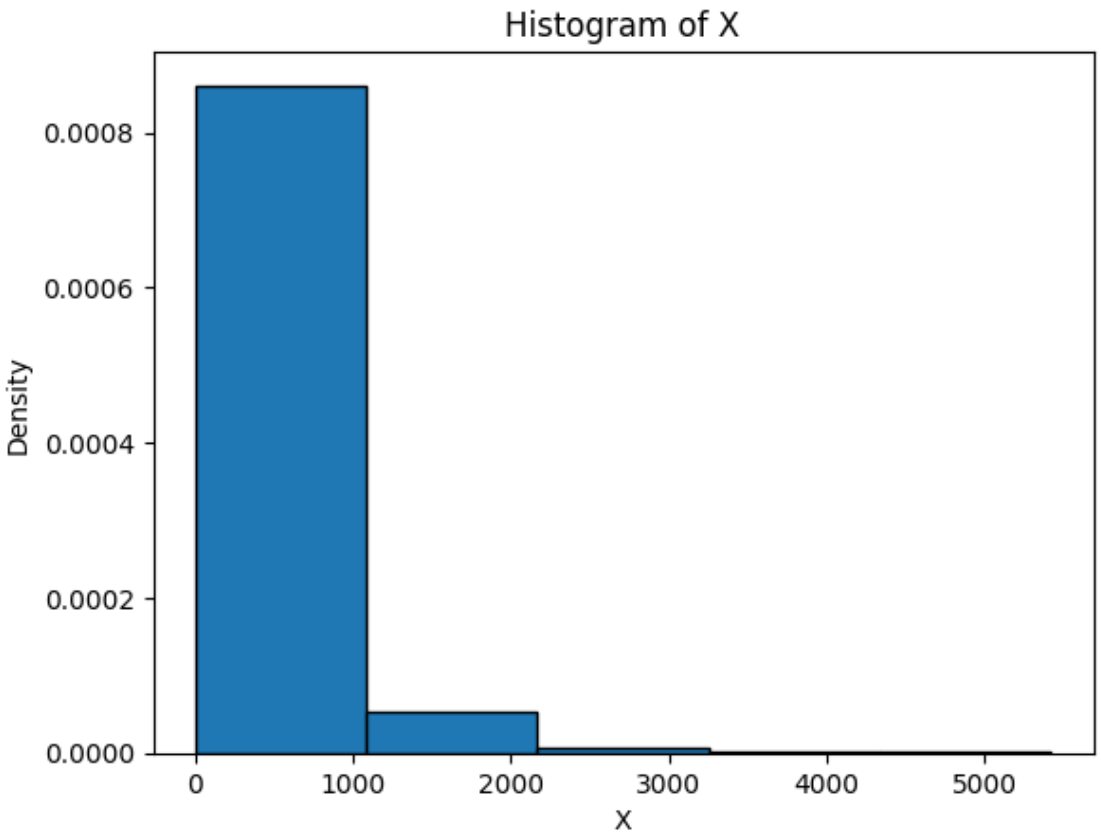

```
None
total_rooms
```



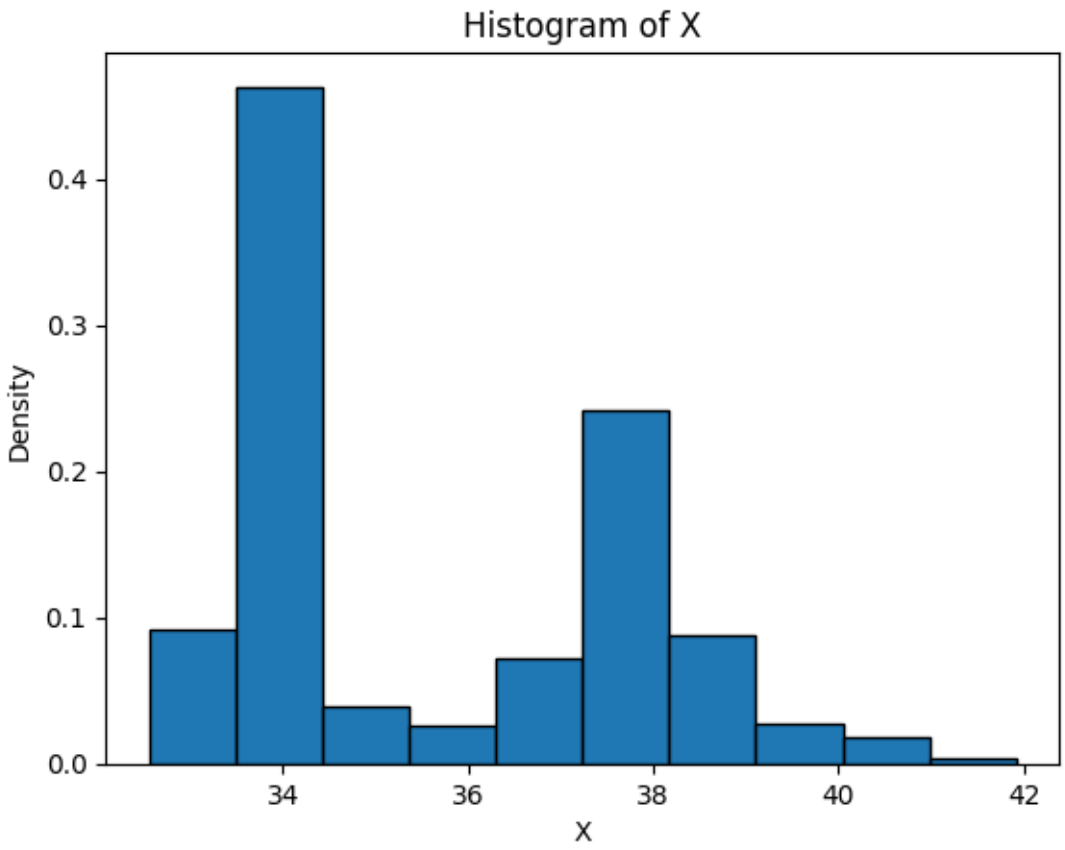

```
None
housing_median_age
```



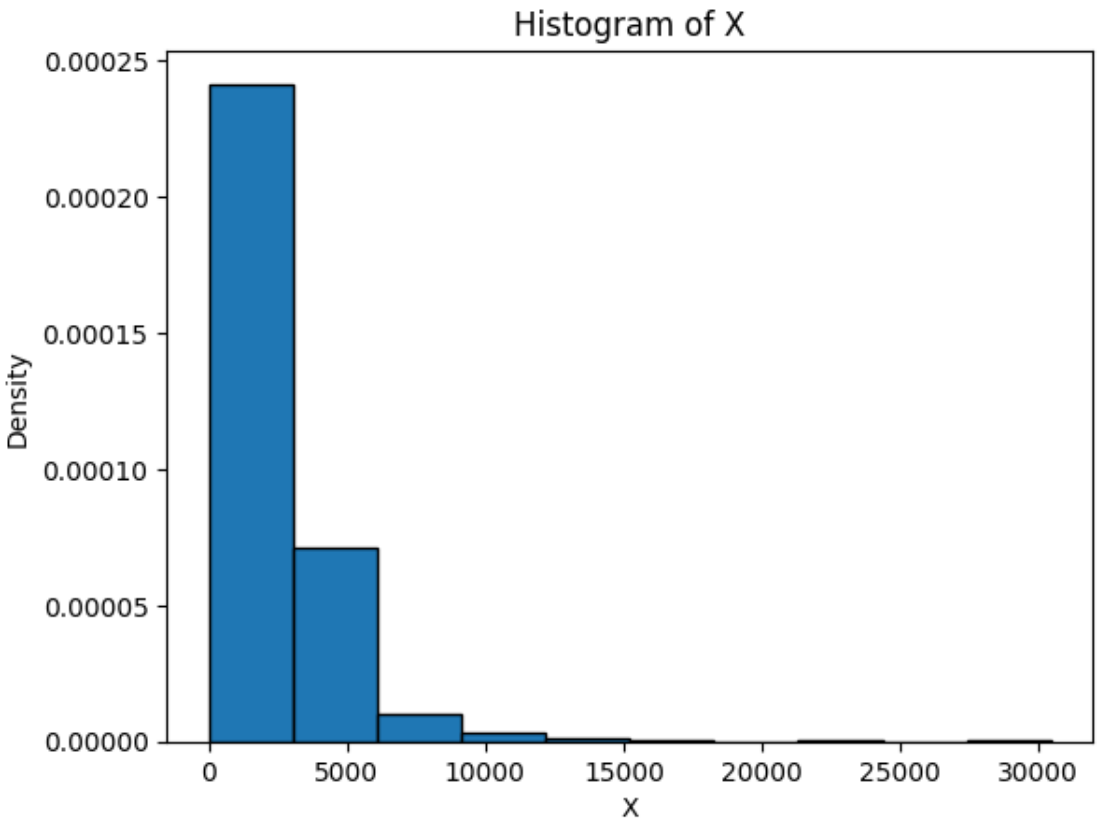

```
None
total_bedrooms
```



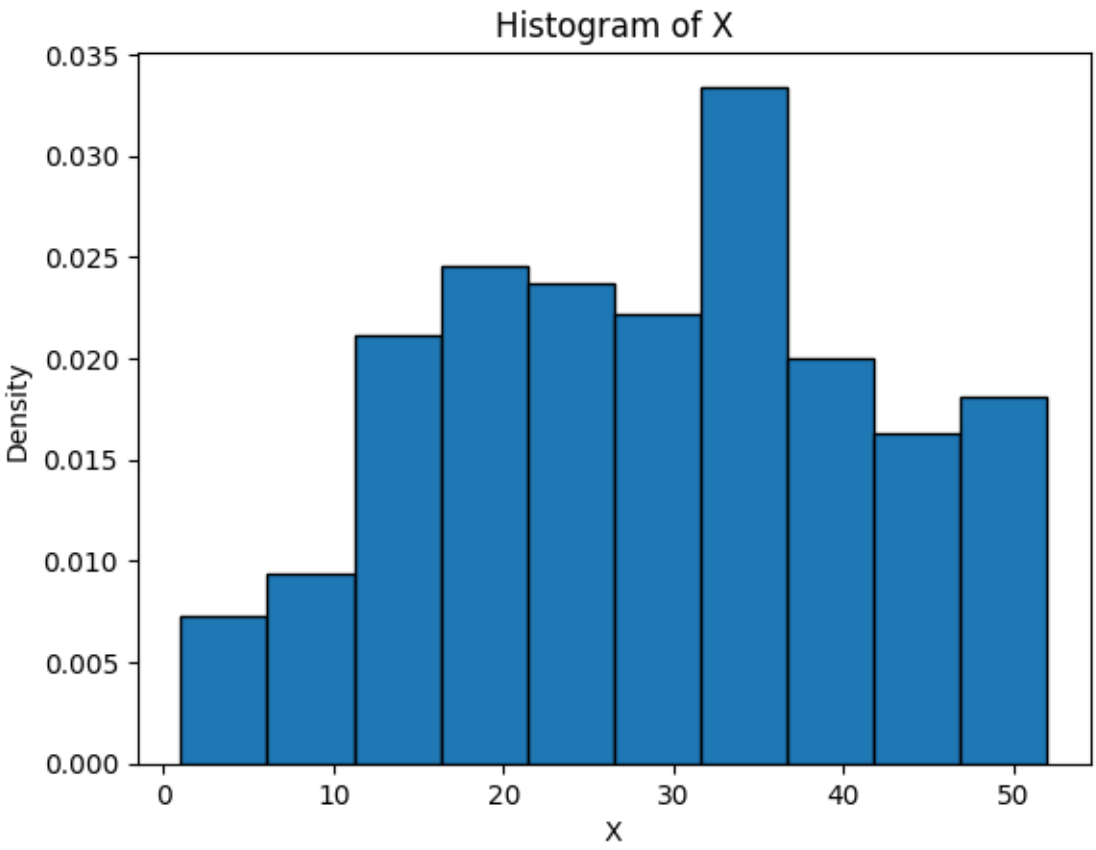

```
None
bins are 10
latitude
```



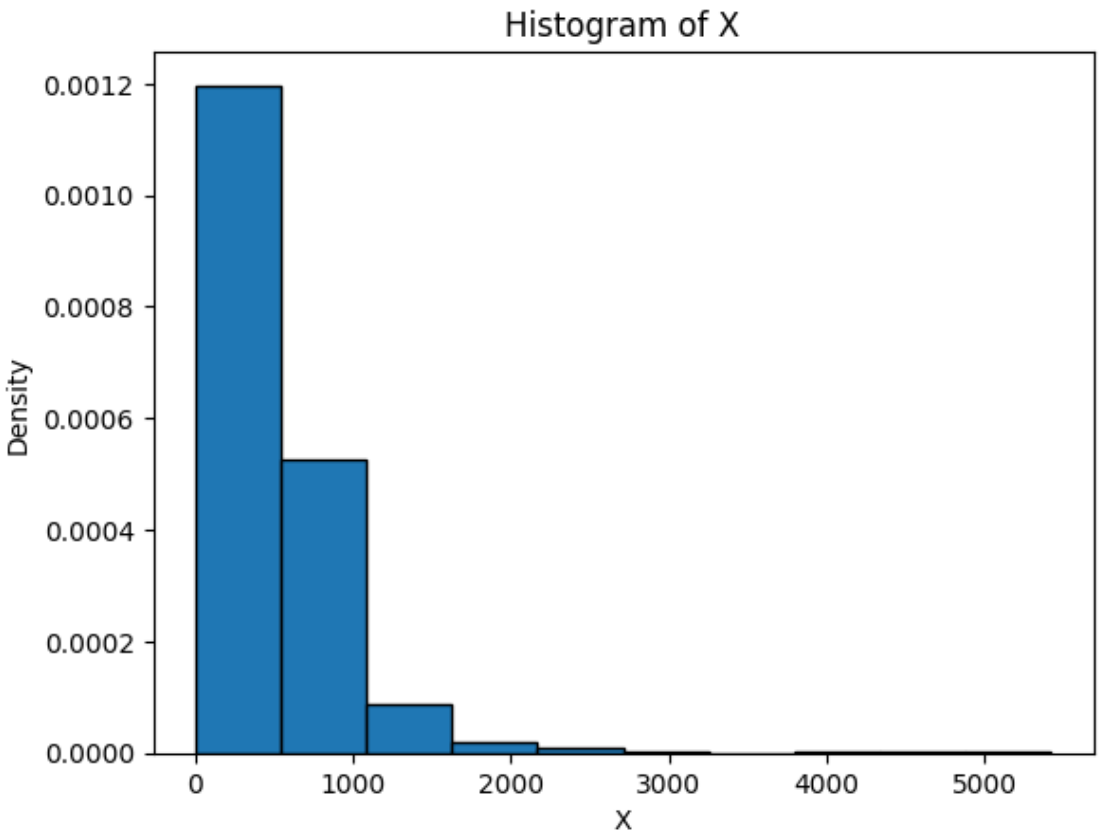

```
None
total_rooms
```



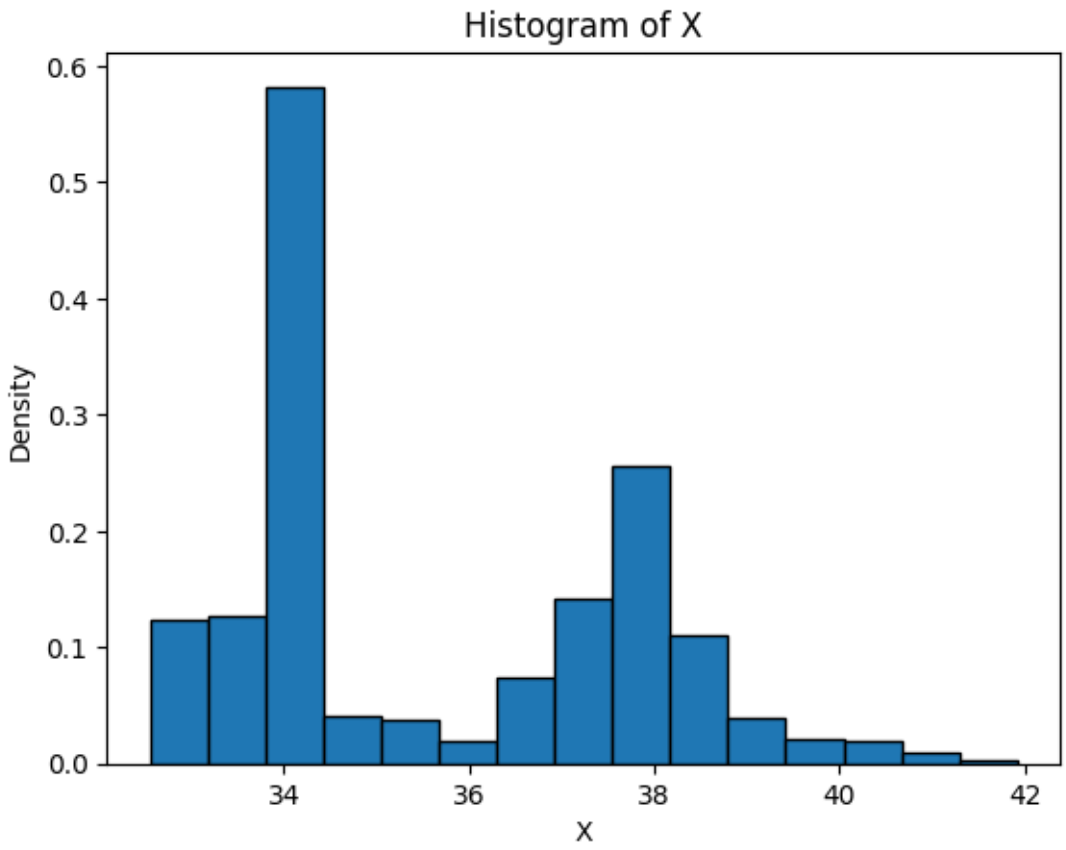

```
None
housing_median_age
```



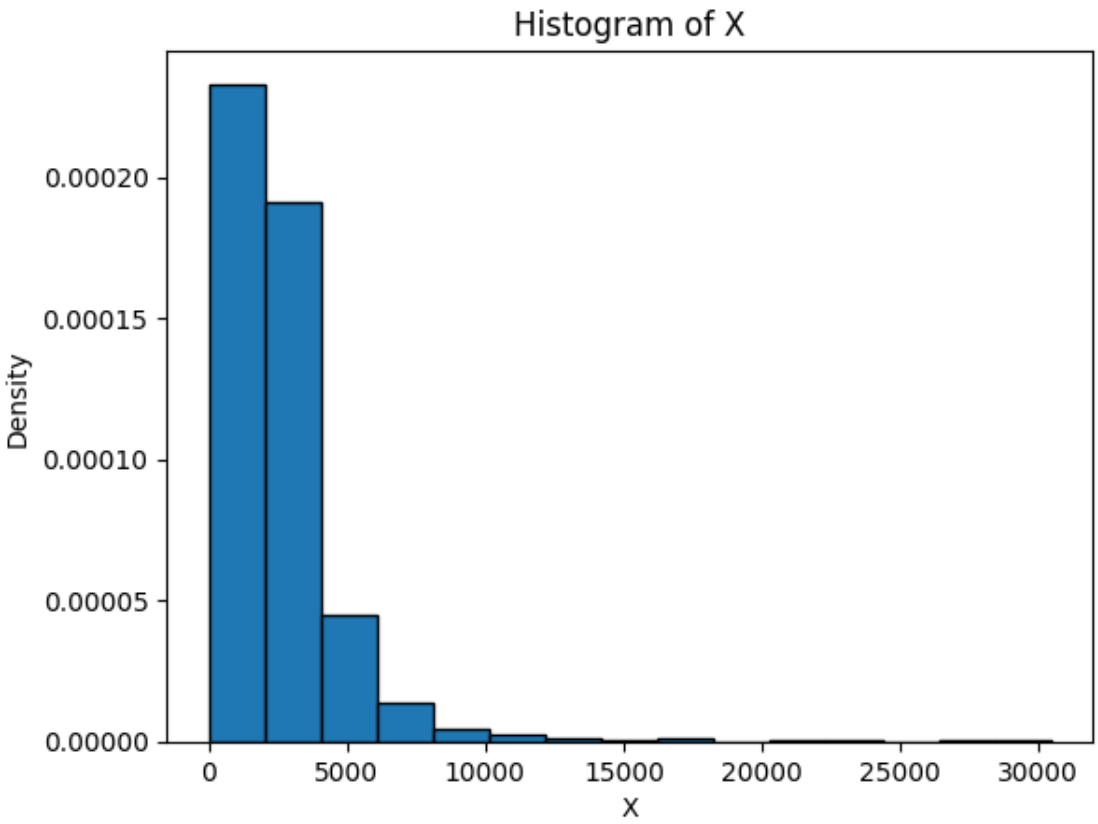

```
None
total_bedrooms
```



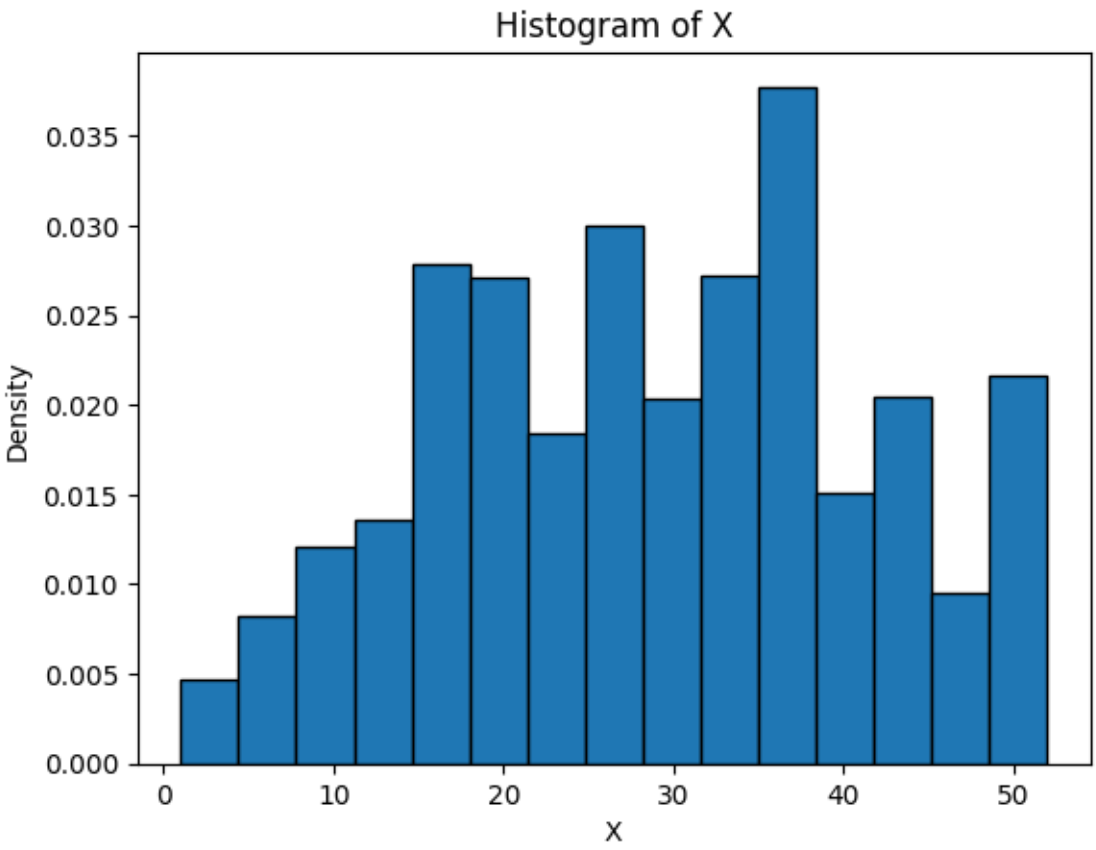

```
None
bins are 15
latitude
```



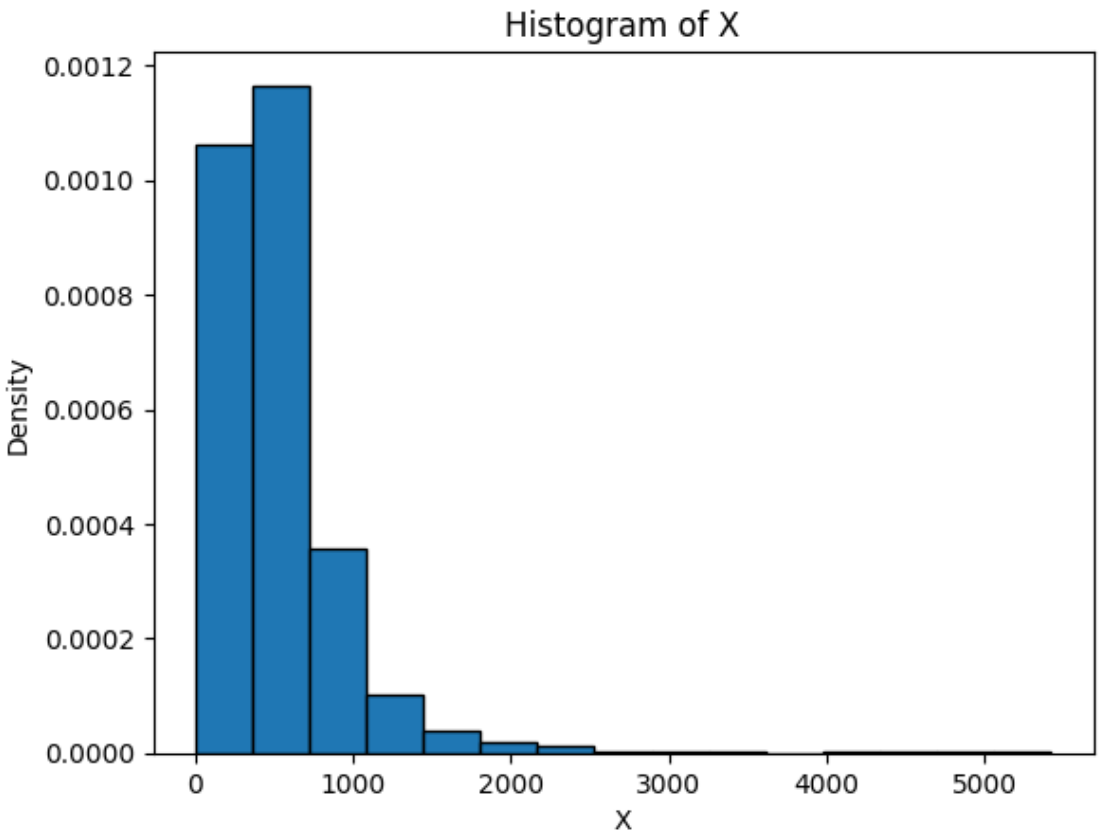

```
None
total_rooms
```



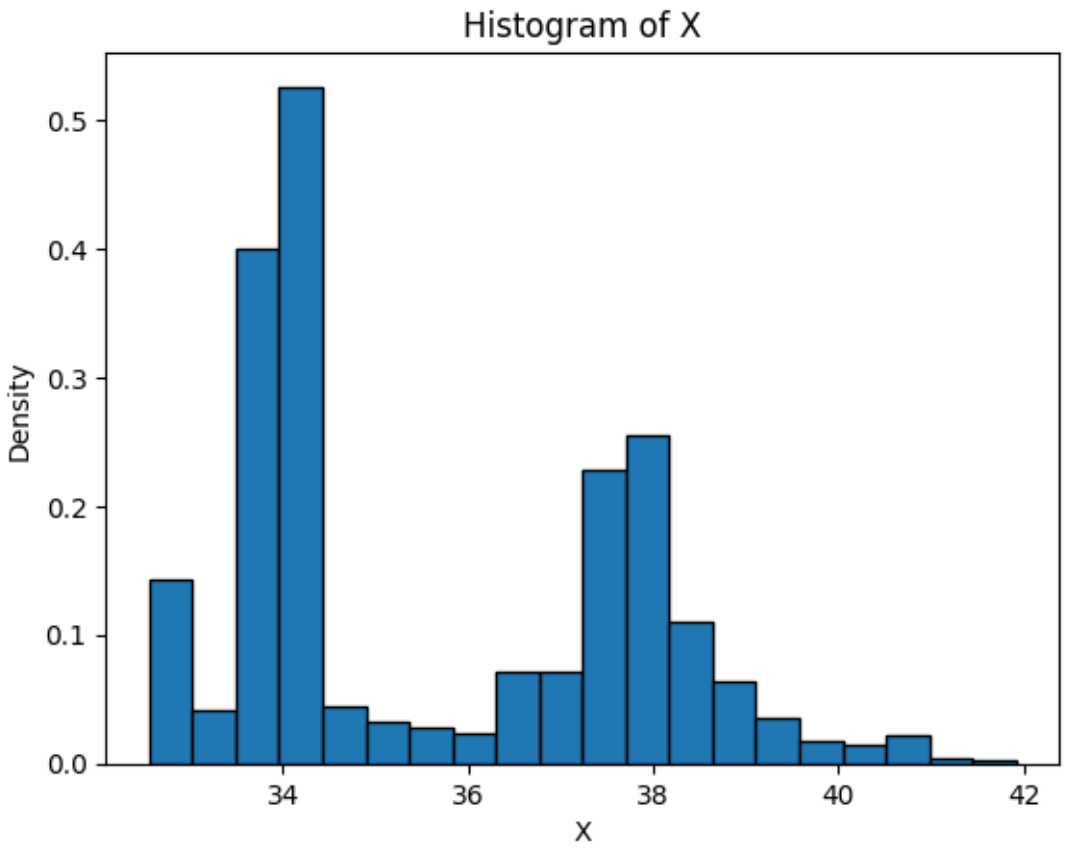

```
None
housing_median_age
```



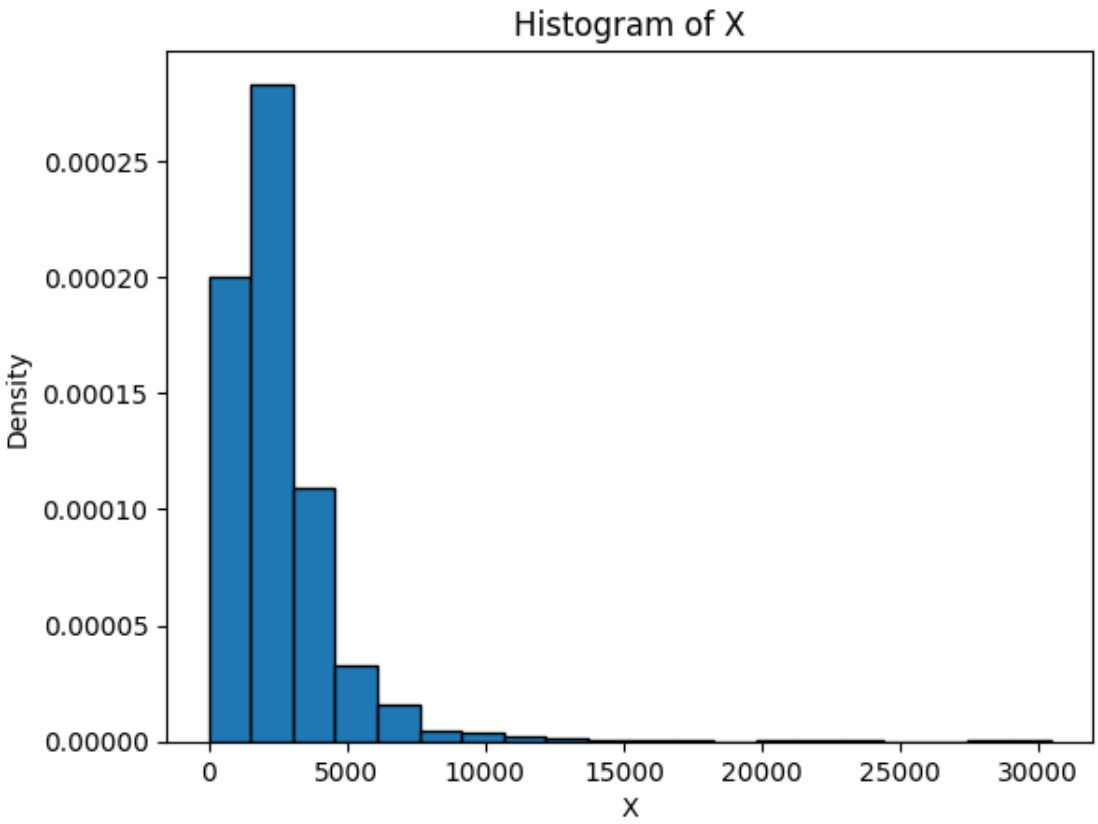

```
None
total_bedrooms
```



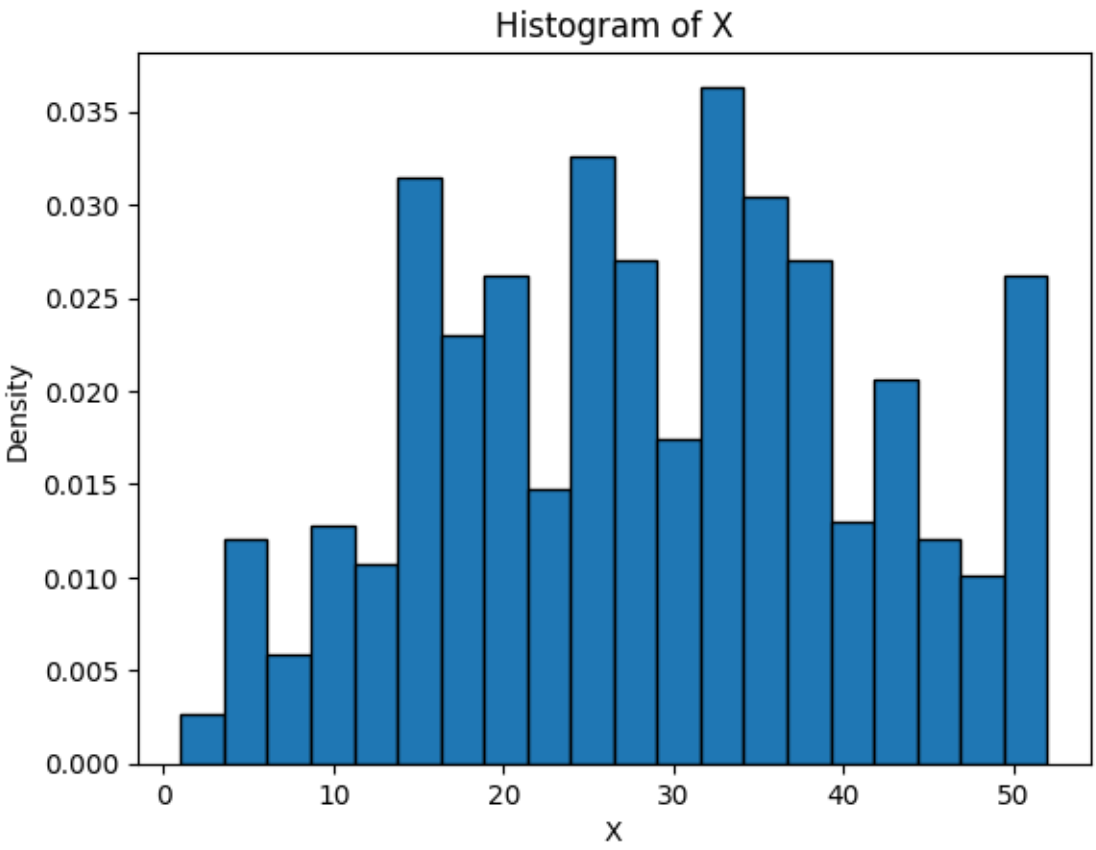

```
None
bins are 20
latitude
```



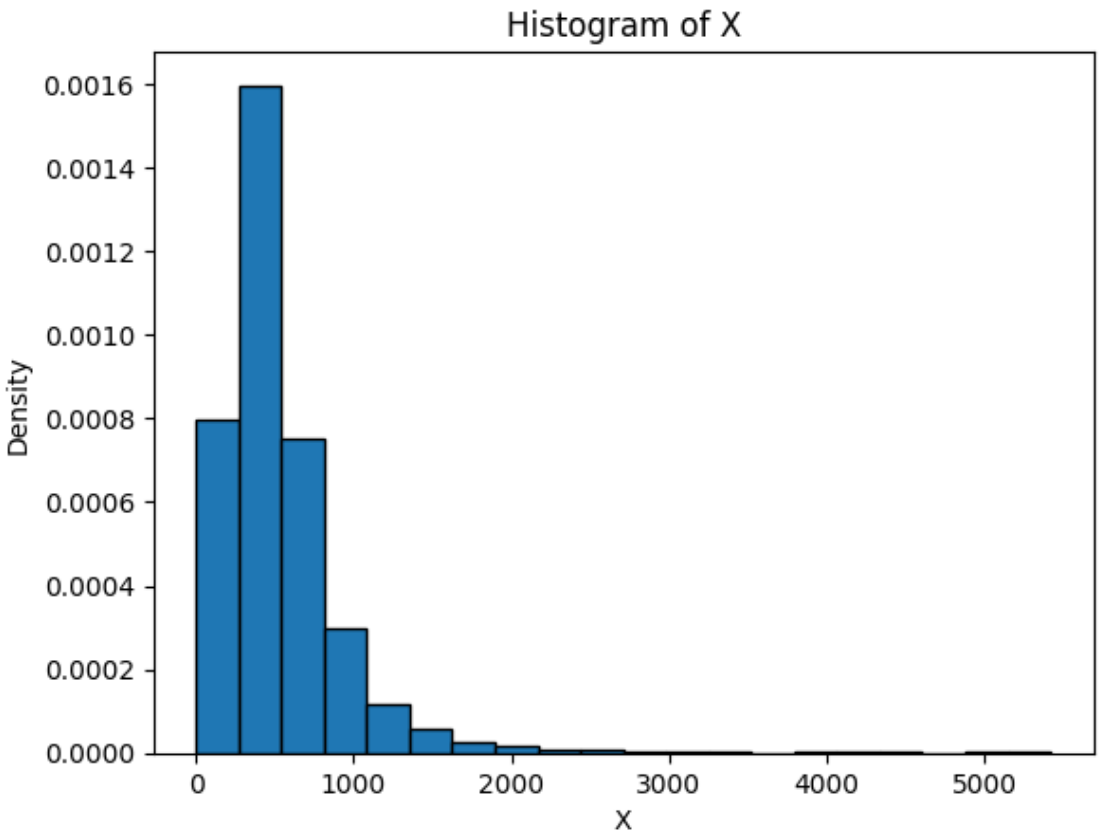

```
None
total_rooms
```




```
None
housing_median_age
```




```
None
total_bedrooms
```


Histogram of X
Density
0.0016
0.0014
0.0012
0.0010
0.0008
0.0006
0.0004
0.0002
0.0000
0
1000
2000
3000
4000
5000
X

```
None
```