1. What is the average median income of the data set and check the distribution of data using appropriate plots. Please explain the distribution of the plot.

import pandas as pd

df=pd.read\_excel('housing+(1).xlsx')

df.head()

import matplotlib.pyplot as plt

plt.hist(df.median\_income,edgecolor='black')

plt.xlabel('median\_income')

plt.ylabel('frequency')

plt.title('income Distribution')

plt.show()

2. Draw an appropriate plot to see the distribution of housing\_median\_age and explain your observations.

plt.hist(df.housing\_median\_age,edgecolor='black',bins=10)

plt.xlabel('housing\_median\_age')

plt.ylabel('frequency')

plt.title('Observations of housing\_age')

plt.show()

3. Show with the help of visualization, how median\_income and median\_house\_values are related?

plt.scatter(df.median\_income,df.median\_house\_value,edgecolors='black')

plt.xlabel('median\_income')

plt.ylabel('median\_house\_value')

plt.title('income vs House\_values')

plt.show()

4. Create a data set by deleting the corresponding examples from the data set for which total\_bedrooms are not available.

dff.dropna()

5. Create a data set by filling the missing data with the mean value of the total\_bedrooms in the original data set.

dff=pd.DataFrame(df.total\_bedrooms)

dff

dff.fillna(value=dff.mean())

dff.isna().sum()

6. Write a programming construct (create a user defined function) to calculate the median value of the data set wherever required.

droped\_df=df.drop(columns=['longitude','latitude','ocean\_proximity'])

def median\_value(n):

  result=n.median()

  return result

median\_value(droped\_df)

7. Plot latitude versus longitude and explain your observations.

plt.plot(df.longitude,c='r',label='Log')

plt.plot(df.latitude,c='g',label='Lat')

plt.title('latitude versus longitude')

plt.legend()

plt.show()

8. Create a data set for which the ocean\_proximity is ‘Near ocean’.

import pandas as pd

Near\_ocean = df.loc[df['ocean\_proximity'] == 'NEAR OCEAN']

print(Near\_ocean)

9. Find the mean and median of the median income for the data set created in question 8.

print('Mean of median\_income is',(Near\_ocean['median\_income'].mean()))

print('Median of median\_income is',(Near\_ocean['median\_income'].median()))

10. Please create a new column named total\_bedroom\_size. If the total bedrooms is 10 or less, it should be quoted as small. If the total bedrooms is 11 or more but less than 1000, it should be medium, otherwise it should be considered large.

df['total\_bedrooms']=pd.cut(df['total\_bedrooms'], bins=[0,10,1000,float('inf')],labels=['small','medium','large'])