Perform Data Analysis on the California House Price data to answer the following

```
# importing all the necessary libraries
import pandas as pd
import numpy as np
#we need to read the data
data = pd.read_csv("/content/drive/MyDrive/AI Tools Lab/california_housing.csv")
#print top 5 rows
print(data.head())
```

```
   longitude  latitude  housing_median_age  total_rooms  total_bedrooms  \
0    -122.23     37.88                41.0        880.0           129.0
1    -122.22     37.86                21.0       7099.0          1106.0
2    -122.24     37.85                52.0       1467.0           190.0
3    -122.25     37.85                52.0       1274.0           235.0
4    -122.25     37.85                52.0       1627.0           280.0

   population  households  median_income  median_house_value ocean_proximity
0       322.0       126.0         8.3252            452600.0        NEAR BAY
1      2401.0      1138.0         8.3014            358500.0        NEAR BAY
2       496.0       177.0         7.2574            352100.0        NEAR BAY
3       558.0       219.0         5.6431            341300.0        NEAR BAY
4       565.0       259.0         3.8462            342200.0        NEAR BAY
```

a. Data Type of each column and info regarding each column

```
# data information for each column
print(data.info())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 20640 entries, 0 to 20639
Data columns (total 10 columns):
 #   Column              Non-Null Count  Dtype
---  ------              --------------  -----
 0   longitude           20640 non-null  float64
 1   latitude            20640 non-null  float64
 2   housing_median_age  20640 non-null  float64
 3   total_rooms         20640 non-null  float64
 4   total_bedrooms      20433 non-null  float64
 5   population          20640 non-null  float64
 6   households          20640 non-null  float64
 7   median_income       20640 non-null  float64
 8   median_house_value  20640 non-null  float64
 9   ocean_proximity     20640 non-null  object
dtypes: float64(9), object(1)
memory usage: 1.6+ MB
None
```

b. The average age of a house in the data set.

```
# printing average age of house
print(data['housing_median_age'].mean())
```

```
28.639486434108527
```

Determines top 10 localities with the high difference between income and house value. Also, top 10 localities that have the lowest difference

```
#calculating the difference btw House value and income and adding new column 'diff_income_and_house_value' with difference values
data['diff_income_and_house_value'] = data['median_house_value'] - data['median_income']
# sorting the whole dataframe by the difference value in descending order
data.sort_values(by='diff_income_and_house_value', ascending=False,inplace=True)
#printing the top 10 localities with highest difference
print("the top 10 localities with highest difference")
print(data['ocean_proximity'].head(10))
#printing the top 10 localities with lowest difference
print("the top 10 localities with lowest difference")
print(data['ocean_proximity'].tail(10))
```

```
the top 10 localities with highest difference
4861      <1H OCEAN
6688         INLAND
16642    NEAR OCEAN
15661      NEAR BAY
15652      NEAR BAY
6639      <1H OCEAN
459        NEAR BAY
89         NEAR BAY
10448     <1H OCEAN
17819     <1H OCEAN
```

```
Name: ocean_proximity, dtype: object
the top 10 localities with lowest difference
2779         INLAND
16186        INLAND
14326    NEAR OCEAN
1825       NEAR BAY
13889        INLAND
5887      <1H OCEAN
19802        INLAND
2521         INLAND
2799         INLAND
9188         INLAND
Name: ocean_proximity, dtype: object
```

```
data.to_csv("/content/drive/MyDrive/AI Tools Lab/california_housing_2.csv",index=False)
```

What is the ratio of bedrooms to total rooms in the data

```
# total no of rooms
total_rooms = data['total_rooms'].sum()
# total number of bedrooms
total_bedrooms = data['total_bedrooms'].sum()
#printing the ratio of bedrooms to total rooms
print(total_rooms//total_bedrooms)
```

```
4.0
```

e. Determine the average price of a house for each type of ocean_proximity.

```
# average house price for each ocean_proximity type
data.groupby('ocean_proximity')['median_house_value'].median()
```

```
ocean_proximity
<1H OCEAN     214850.0
INLAND        108500.0
ISLAND        414700.0
NEAR BAY      233800.0
NEAR OCEAN    229450.0
Name: median_house_value, dtype: float64
```