7/2/24, 6:46 AM 3. California House Price data.ipynb - Colab

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

https://colab.research.google.com/drive/1Yqf8AHrTik1U2E1t9b-beg6XjzFQLcVz#scrollTo=JJMTmwQF5j8u&printMode=true 1/2

7/2/24, 6:46 AM 3. California House Price data.ipynb - Colab

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

https://colab.research.google.com/drive/1Yqf8AHrTik1U2E1t9b-beg6XjzFQLcVz#scrollTo=JJMTmwQF5j8u&printMode=true 2/2