
Name:M.Shafeen

Roll No:22P-9278

Subject:BS AI

Intro-to Artificial Intelligence

Assignment #2

1. **Data Collection:** Download a textual data set from **Kaggle**. Write its name, URL, and description in 1-2 lines. [2 marks]

Answer :

Name of the file on website : Canadian house prices for top cities ,

URL of the file

https://www.kaggle.com/datasets/jeremylarcher/canadian-house-prices-for-top-cities/data ,

Description : Listing information from top 45 cities in Canada by population

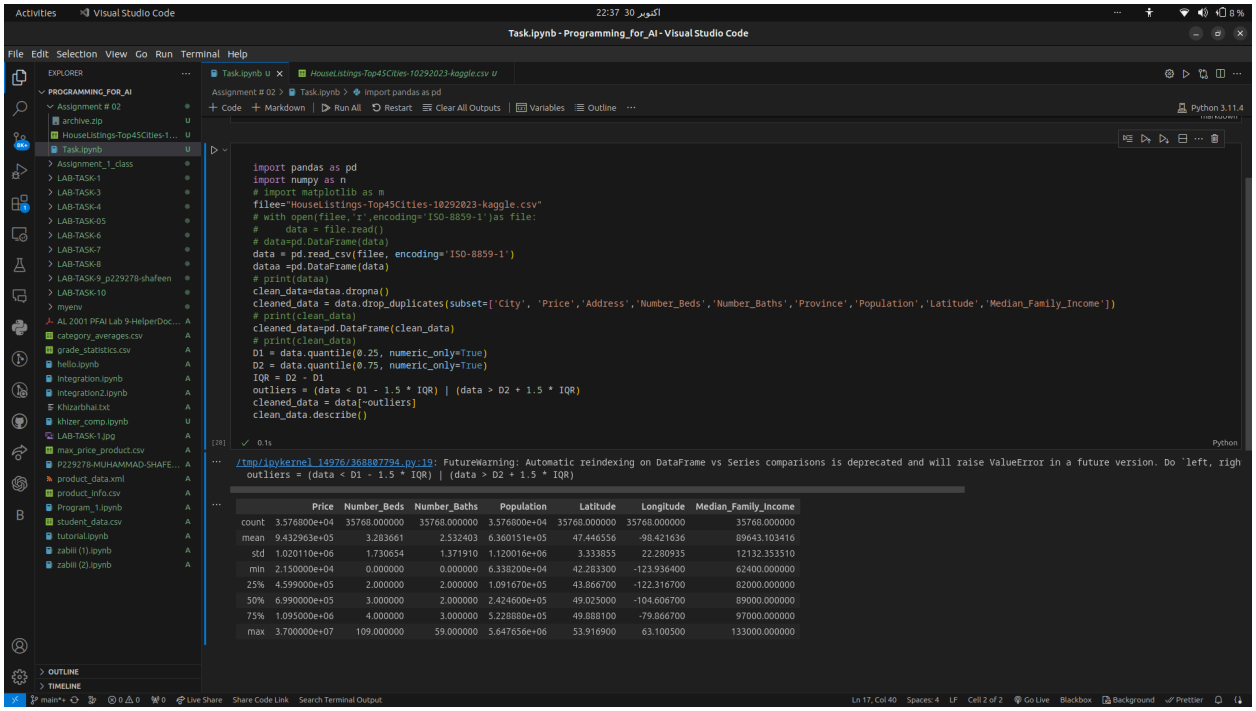
2. **Data Loading:** Load the dataset into a Pandas DataFrame. [2 marks]

```
# import matplotlib as m
filee="HouseListings-Top45Cities-10292023-kaggle.csv"
# with open(filee,'r',encoding='ISO-8859-1')as file:
#     data = file.read()
# data=pd.DataFrame(data)
data = pd.read_csv(filee, encoding='ISO-8859-1')
```

3. **Data Cleaning:** Remove missing values, duplicate records, and outliers from the loaded dataframe. [3 Marks]

```
clean_data=dataa.dropna()
cleaned_data = data.drop_duplicates(subset=['City', 'Price','Address','Number_Beds','Number_Baths','Province','Population','Latitude','Median_Family_Income'])
# print(clean_data)
cleaned_data=pd.DataFrame(clean_data)
# print(clean_data)
D1 = data.quantile(0.25, numeric_only=True)
D2 = data.quantile(0.75, numeric_only=True)
IQR = D2 - D1
outliers = (data < D1 - 1.5 * IQR) | (data > D2 + 1.5 * IQR)
cleaned_data = data[~outliers]
```

4. **Statistical Analysis:** Perform all descriptive statistics functions studied in the class on the DataFrame [3 Marks]



	Price	Number_Beds	Number_Baths	Population	Latitude	Longitude	Median_Family_Income
count	3.576800e+04	35768.000000	35768.000000	3.576800e+04	35768.000000	35768.000000	35768.000000
mean	9.432963e+05	3.283661	2.532403	6.360151e+05	47.446556	-98.421636	89643.103416
std	1.020110e+06	1.730654	1.371910	1.120016e+06	3.333855	22.280935	12132.353510
min	2.150000e+04	0.000000	0.000000	6.338200e+04	42.283300	-123.936400	62400.000000
25%	4.599000e+05	2.000000	2.000000	1.091670e+05	43.866700	-122.316700	82000.000000
50%	6.990000e+05	3.000000	2.000000	2.424600e+05	49.025000	-104.606700	89000.000000
75%	1.095000e+06	4.000000	3.000000	5.228880e+05	49.888100	-79.866700	97000.000000
max	3.700000e+07	109.000000	59.000000	5.647656e+06	53.916900	63.100500	133000.000000