329e_HW01

August 30, 2021

1 Assignment 1 - Melbourne Housing Dataset

Practice Loading in Data (20 Points)

Please see the CANVAS system for the due date of this assignment.

1.1 Add YOUR NAME/S HERE !

- Student Name:
- Student UT EID:
- Partner Name:
- Partner UT EID:

[1]:
```
# Imports first on top
import pandas as pd
import matplotlib.pyplot as plt
```

First of all, we load the data from a CSV file into the manin memory in a Pandas Dataframe format.

[2]:
```
melbourne_data = pd.read_csv('melb_data.csv')

melbourne_data
```

[2]:
```
               Suburb           Address  Rooms Type      Price Method  \
0          Abbotsford      85 Turner St      2    h  1480000.0      S
1          Abbotsford   25 Bloomburg St      2    h  1035000.0      S
2          Abbotsford      5 Charles St      3    h  1465000.0     SP
3          Abbotsford  40 Federation La      3    h   850000.0     PI
4          Abbotsford       55a Park St      4    h  1600000.0     VB
…                 …                …    …  …         …    …
13575   Wheelers Hill      12 Strada Cr      4    h  1245000.0      S
13576    Williamstown     77 Merrett Dr      3    h  1031000.0     SP
13577    Williamstown       83 Power St      3    h  1170000.0      S
13578    Williamstown      96 Verdon St      4    h  2500000.0     PI
13579      Yarraville        6 Agnes St      4    h  1285000.0     SP

        SellerG       Date  Distance  Postcode  …  Bathroom  Car  Landsize  \
0        Biggin   3/12/2016      2.5    3067.0  …       1.0  1.0     202.0
1        Biggin   4/02/2016      2.5    3067.0  …       1.0  0.0     156.0
2        Biggin   4/03/2017      2.5    3067.0  …       2.0  0.0     134.0
3        Biggin   4/03/2017      2.5    3067.0  …       2.0  1.0      94.0
4        Nelson   4/06/2016      2.5    3067.0  …       1.0  2.0     120.0
…           …          …       …      …  …        …  …       …
13575     Barry  26/08/2017     16.7    3150.0  …       2.0  2.0     652.0
13576  Williams  26/08/2017      6.8    3016.0  …       2.0  2.0     333.0
13577     Raine  26/08/2017      6.8    3016.0  …       2.0  4.0     436.0
13578   Sweeney  26/08/2017      6.8    3016.0  …       1.0  5.0     866.0
13579   Village  26/08/2017      6.3    3013.0  …       1.0  1.0     362.0

       BuildingArea  YearBuilt  CouncilArea Lattitude  Longtitude  \
0               NaN        NaN        Yarra -37.79960   144.99840
1              79.0     1900.0        Yarra -37.80790   144.99340
2             150.0     1900.0        Yarra -37.80930   144.99440
3               NaN        NaN        Yarra -37.79690   144.99690
4             142.0     2014.0        Yarra -37.80720   144.99410
…               …         …          …        …          …
13575           NaN     1981.0          NaN -37.90562   145.16761
13576         133.0     1995.0          NaN -37.85927   144.87904
13577           NaN     1997.0          NaN -37.85274   144.88738
13578         157.0     1920.0          NaN -37.85908   144.89299
13579         112.0     1920.0          NaN -37.81188   144.88449

                       Regionname Propertycount
0           Northern Metropolitan        4019.0
1           Northern Metropolitan        4019.0
2           Northern Metropolitan        4019.0
3           Northern Metropolitan        4019.0
4           Northern Metropolitan        4019.0
…                             …            …
13575  South-Eastern Metropolitan        7392.0
13576       Western Metropolitan        6380.0
13577       Western Metropolitan        6380.0
13578       Western Metropolitan        6380.0
13579       Western Metropolitan        6543.0

[13580 rows x 21 columns]
```

1.2 Q1 - How many unique suburbs are there? (2 points)

[3]:

```
# code goes here
```

1.3 Q2 - How many unique properties are there? (2 points)

[4]:

```
# code goes here
```

1.4 Q3- What is the mean price of a property in the Kensington suburb? (2 points)

[5]:

```
# code goes here
```

2 Q3.1 (extra) - What is the median price of a property in the Kensington suburb? (1 extra point)

2.1 Q4 - What percentage of properties contain a YearBuilt value? (2 points)

As we see in the table, we do not have for each house the “YearBuild” value and some of them are NaN and not filled. We want to find out the percentage of homes that we know their build year.

[6]:

```
# code goes here
```

2.2 Q5 - Create a histogram plot that shows the data distribution of the Landsizes using a bin size of 20. (2 points)

Describe the shape of this histogram plot and your interpretations in one paragraph.

[7]:

```
# code goes here
```

2.3 Q6 - Plot a scatter plot of price as a function of BuildingArea using only rows that have a valid BuildingArea value. (2 points)

Create a scatter plot that has BuildingArea as x-axis and Price as y-axis.

[8]:

```
# code goes here
```

2.4 Q7 - Who are the top-10 seller/listing agents? And what percentage of properties they have listed? (2 points)

Seller/Listing agents are identified by the “SellerG” column.

I’m going to assume that the SellerG column is the name of the seller group, so “realtor” in US parlance.
So, that column is interpreted as the seller that listed the property.

[9]:

```
# code goes here
```

2.5 Q8 - Fix a problem. (2 points)

Your boss has told you that the number of rooms was calculated incorrectly systematically across the entire dataset.

Solve this problem without using a python for look, and using a single pandas statemnt.

Show your output dataset!

[10]:

```
# code goes here
```

2.6 Q9 - What is the Address of the earliest built house in this dataset? (2 points)

[11]:

```
# code goes here
```

2.7 Q10 - Save a file (2 points)

The council member for Melbourne has asked for the information for their district. Assume that there are legal restrcitions and we can only provide the council member the Suburb, Price, and Date from the CouncilArea “Melbourne” to the counsil person.

Export the file with only the allowed columns, and do not write and index column.

[12]:

```
# code goes here
```

[ ]:

```

```