TAE Coding

https://edx.org

BerkeleyX

Homework#1

Date Given: June 12, 2019 Due Date:

==========

Problem#1

Analyze the data source in 'kc-house-data.csv' file. This data source is a part of databases available in the public domain. This file contains 21,613 observations of real-estate properties of King county in Washington state. The data for the following 21 variables is provided.

1. id
2. date
3. price
4. bedrooms
5. bathrooms
6. sqft_living
7. sqft_lot
8. floors
9. waterfront
10. view
11. condition
12. grade
13. sqft_above
14. sqft_basement
15. yr_built
16. yr_renovated
17. zipcode
18. latitude
19. longitude
20. sqft_living15
21. sqft_lot15

Write the Python Code:

1. Read Data file ‘kc-house data.csv’.
2. Compute the average house price
3. First sort: Find the highest price house (sort/descending by ‘price’)
4. Store this dataset (sorted by price) in a table ‘sorted_by_price’
5. Second sort: Now sort the ‘sorted_by_price’ table by zip code
6. Find the highest price home of each zip code

id	price	bedrooms	bathrooms	sqft_living	zipcode
3322049005	850000	4	2.75	5440	98001
5409000110	389000	6	4.5	3560	98002
521049227	950000	4	4	5635	98003
9808700762	7.0625e+06	5	4.5	10040	98004
2806300065	1.96e+06	4	4	4430	98005
6065300370	4.208e+06	5	6	7440	98006
6137500310	1.315e+06	5	4	4420	98007
9253900271	3.567e+06	5	4.5	4850	98008
5416300240	935000	4	4.5	5670	98010
1726059134	1.075e+06	3	2.5	2830	98011

... (60 rows omitted)

7. Create a bar chart which shows the highest price house of each zip code