

SUMMARY ON LONDON HOUSING PROJECT

Our project was to determine which boroughs of London have seen the greatest increase in housing prices, on average, over the last two decades. We select the top five which are Hackney, Southwark, Lambeth, City of London and Westminster with a respective increase rate of 7.653565, 6.813154, 6.746656, 6.461971 and 6.380152. We did not really focus on the most expensive borough but what we can at least say is Hackney was the borough with the highest increase rate from 1995 to 2020. We also found that in the dataset given to us there were boroughs that did not belong to London.

These are the steps that we took to come to that conclusion:

1-We imported the data from the website and put it in a panda's data frame. After doing the transposition on our data frame, we used the first row as a column's header.

	London_Borough	ID	1995-01-01 00:00:00	1995-02-01 00:00:00	1995-03-01 00:00:00	1995-04-01 00:00:00	1995-05-01 00:00:00	1995-06-01 00:00:00	1995-07-01 00:00:00	1995-08-01 00:00:00	...	2019-05-01 00:00:00	2019-06-01 00:00:00	2019-07-01 00:00:00	2019-08-01 00:00:00	2
1	City of London	E09000001	91449	82202.8	79120.7	77101.2	84409.1	94900.5	110128	112329	...	718930	761786	738370	795632	
2	Barking & Dagenham	E09000002	50460.2	51085.8	51269	53133.5	53042.2	53700.3	52113.1	52232.2	...	295829	295179	299060	300402	
3	Barnet	E09000003	93284.5	93190.2	92247.5	90762.9	90258	90107.2	91441.2	92361.3	...	504765	512818	515235	528958	
4	Bexley	E09000004	64958.1	64787.9	64367.5	64277.7	63997.1	64252.3	63722.7	64432.6	...	337543	339684	338257	337745	
5	Brent	E09000005	71306.6	72022.3	72015.8	72965.6	73704	74310.5	74127	73547	...	482101	474460	472981	487091	

5 rows × 304 columns

2-We melted our data frame and rename some column of our data frame.

	London_Borough	ID	Month	Average Prices
0	City of London	E09000001	1995-01-01	91449
1	Barking & Dagenham	E09000002	1995-01-01	50460.2
2	Barnet	E09000003	1995-01-01	93284.5
3	Bexley	E09000004	1995-01-01	64958.1
4	Brent	E09000005	1995-01-01	71306.6

3-In order to have the same rows for each column, we removed boroughs that were not part of London and the NULL values.

	London_Borough	ID	Month	Average Prices
0	City of London	E09000001	1995-01-01	91448.98487
1	Barking & Dagenham	E09000002	1995-01-01	50460.22660
2	Barnet	E09000003	1995-01-01	93284.51832
3	Bexley	E09000004	1995-01-01	64958.09036
4	Brent	E09000005	1995-01-01	71306.56698

#Let us check again if the columns have the same number of rows
`df_clean.count()`

```
London_Borough    9966
ID                 9966
Month              9966
Average Prices     9966
dtype: int64
```

4-We created a column called 'Year' that will help us averaging the average price per year for each borough. Then we will calculate the mean for each year for each London borough.

	London_Borough	Year	Average Prices
0	Barking & Dagenham	1995	51817.969390
1	Barking & Dagenham	1996	51718.192690
2	Barking & Dagenham	1997	55974.262309
3	Barking & Dagenham	1998	60285.821083
4	Barking & Dagenham	1999	65320.934441

5-We created another data frame that has 2 columns: Borough (which is the column of London borough) and the increase rate from 1995 to 2020 on average housing price for each borough.

	Borough	Increase Rate
0	Barking & Dagenham	4.814499
1	Barnet	4.688341
2	Bexley	4.265700
3	Brent	5.197119
4	Bromley	4.326400

6-Finally, we sorted in descending order the increase rate and took the top 5 boroughs that have the highest increase rate.

	Borough	Increase Rate
11	Hackney	7.653565
27	Southwark	6.813154
21	Lambeth	6.746656
6	City of London	6.461971
32	Westminster	6.380152

Our main difficulty was to make sure that each column of our data frame has the same number of rows, but we overcame that by deleting the boroughs that was not part of London and by deleting all the NULL values.

An interesting study will be to see if we can create a model predicting the increase rate of each borough from London for 2021.