

## Case Study – London Housing

This project is to find which borough of London has the greatest increase in housing prices, on average, over the two decades (1998-2018) covered by the dataset. The dataset is extracted from the following link:

<https://data.london.gov.uk/download/uk-house-price-index/70ac0766-8902-4eb5-aab5-01951aaed773/UK%20House%20price%20index.xls>

After cleaning, transforming, visualizing, modeling and evaluating the dataset, we find the following:

1. The Hackey Borough has the greatest average increase in house price, about 6.20 times over the two decades among 33 boroughs of London.
2. The top 15 for the greatest average increase are centered around the City of London, with more on the north side of the River Thames while the City of London itself ranks only #7.
3. The most expensive borough is Kensington & Chelsea with average house price \$1,363, 880 in 2018.

The challenges of this project are how to clean and transform the loaded dataset to organize the useful data we need into a clear and tidy format. The original dataset was read into a DataFrame. By observing the DataFrame, we found that a column containing dates has no name, a column containing borough IDs has a name NaT, and other columns are named after the borough names. These issues can be resolved by transposing the DataFrame. As a result, the boroughs of London and the borough IDs become values in columns, and the dates become column labels. Then by melting the transposed DataFrame and renaming

columns in the melted DataFrame, we got a DataFrame with the rows and columns we want.

Furthermore, I want to investigate the average house price increase rates in boroughs of London over 5 years, 10 years, and 15 years to see whether the results are consistent with the result over 20 years. If it is true, we can conclude that the location is the prominent factor for average house price increase.

