Analysis of UK Cities House Prices

based on City Economic Fundamentals

IVAN TODOROVIC

OCTOBER 2017

EXECUTIVE SUMMARY

- This project examines the movement of house prices for 62 major cities in the UK in the period from 2003/4 to 2015/16
- Based on city data that includes economic indicators for each city (such as average house prices, housing stock, population, employment, average wages and other data), the analysis attempts to identify the factors which have a statistical effect on house prices.
- Understanding the behavior of house prices is of huge interest to mortgage lenders, real estate investors, property developers, and the buying public.
- Key questions addressed in this project include:
  - What key factors influence the movement of house prices in major cities the UK? What makes house prices certain cities more expensive than in others?
  - Can the cities be grouped based on their economic fundamentals and risk profile?
  - Can we use any of these factors to predict the movements of house prices by city e.g. one year from now?
  - How is the overall level of house prices in the UK related to the FTSE100 stock market index?

WHAT THE DATA SET INCLUDES

- The Centre for Cities (www.centreforcities.org) provides information for 62 cities in the UK. The period between 2003/4 and 2015/16 has most of the items available on an annual basis. Items included in the data and used in the analysis are:
  - Population
  - Employment
  - Average wages
  - Housing stock
  - Average house price (annual)

- The FTSE 100 stock market index was also added to the data set, as an annual average for each year

- Average house prices represent the labels that the analysis will analyze and try to predict

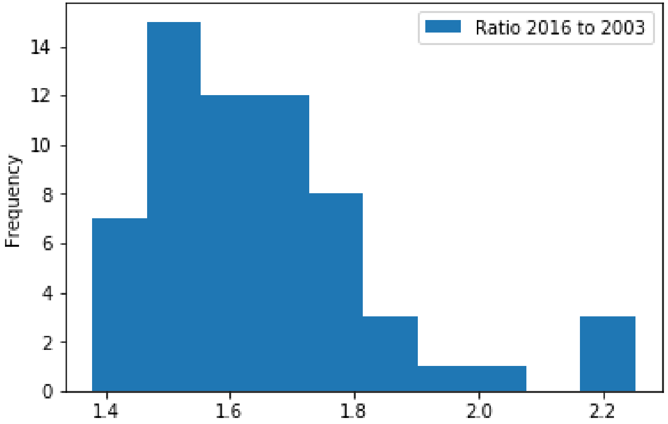
- Remaining data are factors that may influence house prices

ANALYSES PERFORMED

ANALYSIS	DESCRIPTION OF METHODS AND GOALS
1. EXPLORATORY DATA ANALYSIS (EDA)	⦿ EDA is used to visually assess the development of prices in various cities / regions and obtain a high-level understanding of the data.
2. STATISTICAL ANALYSIS	⦿ This section attempts to identify statistically significant correlations between economic factors and house prices.
3. CLUSTERING ANALYSIS	⦿ Unsupervised machine learning algorithms are performed on the data to identify clusters of cities that exhibit similar behaviour.
4. PREDICTION MODELS	⦿ Regression models are used to build prediction models based on historic prices and economic factors. Specifically, the models will show if the economic factors can help in obtaining more accurate predictions than simply using past prices.

1. EDA – DEVELOPMENT OF PRICES

HOW THE PRICES MOVED FROM 2003 TO 2016



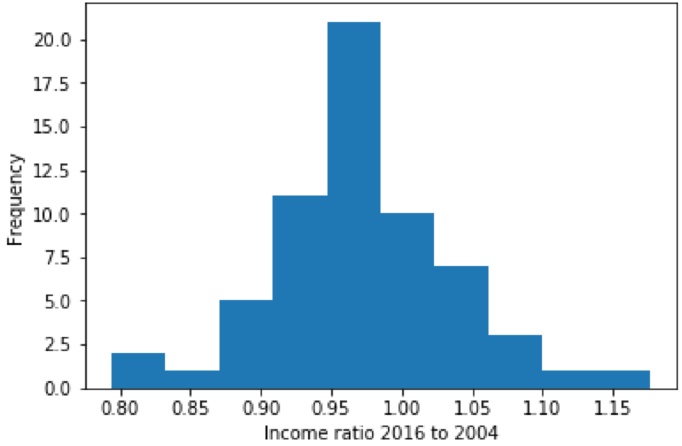
Metric	Ratio 2016 / 2003	Increase
Mean	1.65	65%
Standard deviation	0.19	19%
Min	1.38	38%
Max	2.25	125%

KEY TAKEAWAYS

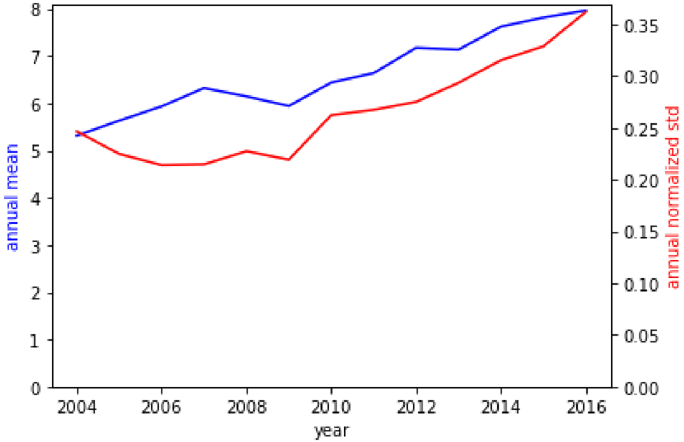
- All 62 cities have experienced an increase in house prices
- The mean increase is 65%, smallest increase is 38% and largest increase is 125%
- The distribution has a negative skew, with the majority of cities showing an increase between 40% and 80%
- There is a group of 3 cities with a significantly greater increase, of approx. 120%

1. EDA – AFFORDABILITY

INCOME RATIO 2016 / 2004



AFFORDABILITY OVER THE PERIOD 2004 - 2016

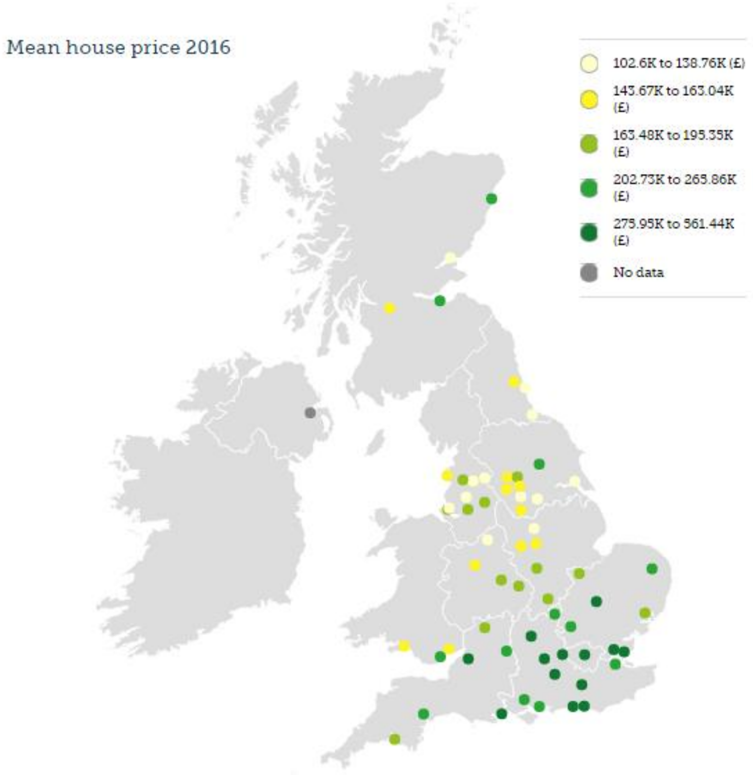


- House prices have increased but wages have (on average) stayed the same between 2004 and 2016: hence it has become increasingly difficult to afford a home:
  - In 2004 5.3 average annual wages were needed to afford an average home
  - In 2016 this ratio increased to 8.0 years
  - The standard deviation has increase as well, meaning that geographical differences in affordability have become more pronounced

1. EDA

Nationwide Overview

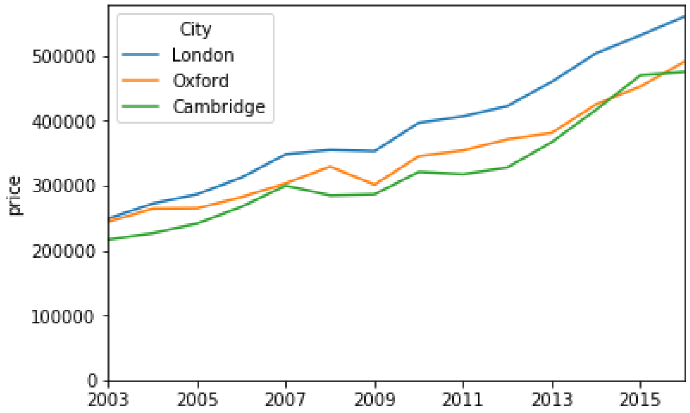
- The map shows mean prices in 2016 for the cities analyzed
- As is widely known, the Southeast has the highest property prices
- The following slides will show a more detailed breakdown by region as well as how prices reached the 2016 levels
- The following sections will then attempt to identify factors which contributed to the observed variations



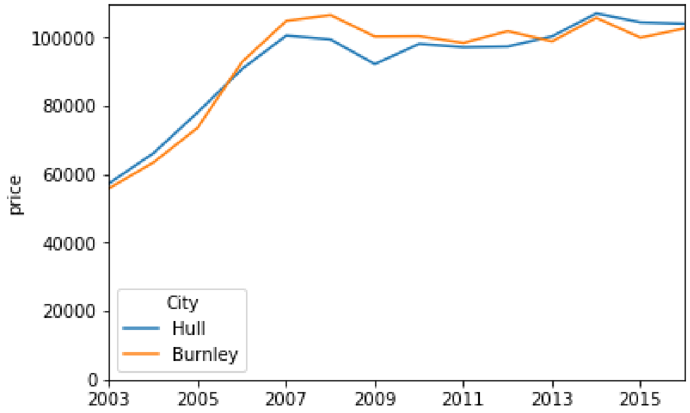
Source: centreforcities.org

1. EDA (CONT'D)

HIGHEST PRICED



LOWEST PRICED

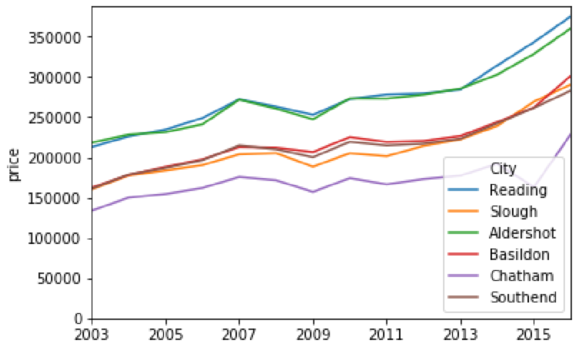


- London, Oxford and Cambridge are the 'outliers' in the data set, with the highest prices (above £450k) and the highest increases – most likely influenced by a different set o factors than the 59 other cities

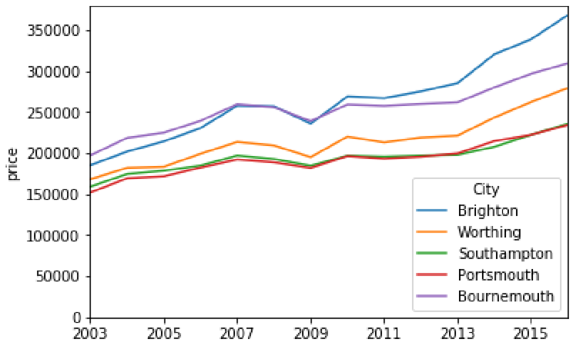
- By contrast, the lowest priced cities (e.g. Hull and Burnley) also have low employment rates, and their prices struggled to increase past the boom ending in 2008, averaging at approx. £100k

1. EDA (CONT'D)

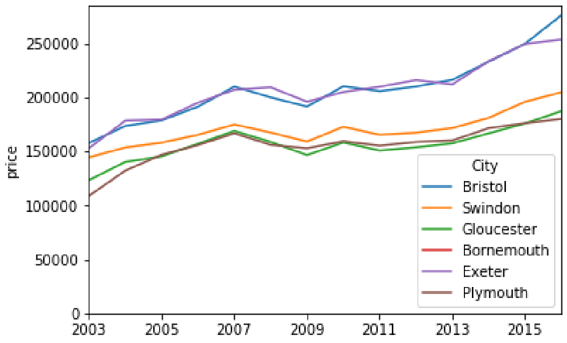
SOUTHEAST



SOUTH COAST



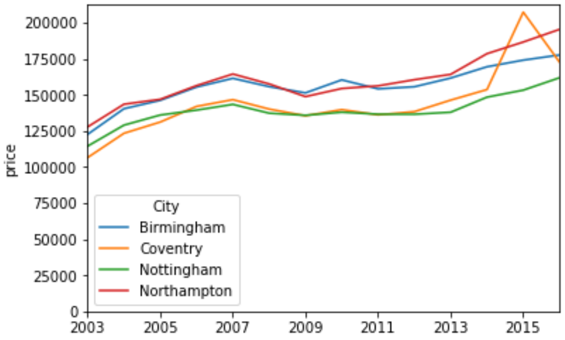
WEST



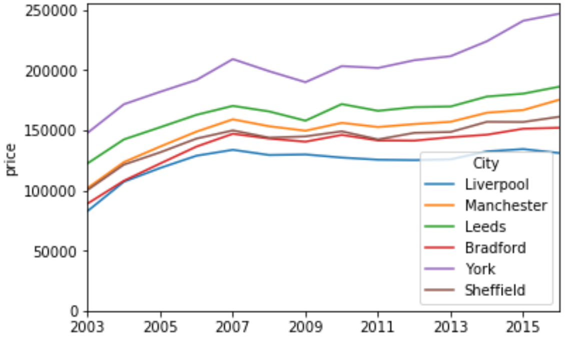
- Places gravitating to London (Reading, Aldershot, Brighton) exhibit some of the highest average prices outside of the 3 outliers, at over £350k
- Cities that are further away from London are more aligned, with prices ranging between approx. £150k - £250k. For these cities the city-specific economic fundamentals may be much more indicative of price movements

1. EDA (CONT’D)

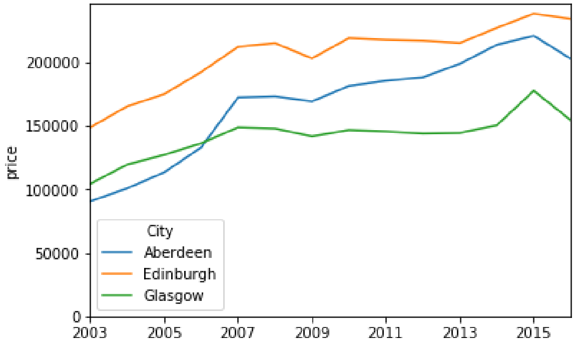
MIDLANDS



NORTH



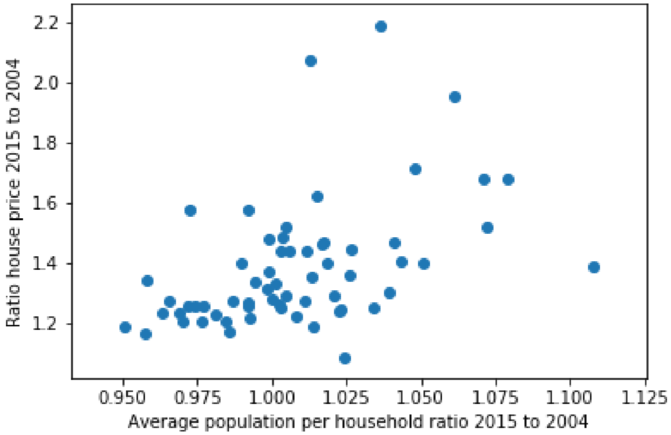
SCOTLAND



- Moving further north, prices tend to decline, with a few notable exceptions:
  - York has historically higher prices
  - Edinburgh has historically high prices as the capital of Scotland
  - Aberdeen has experienced an outstanding increase in prices, almost reaching Edinburgh levels

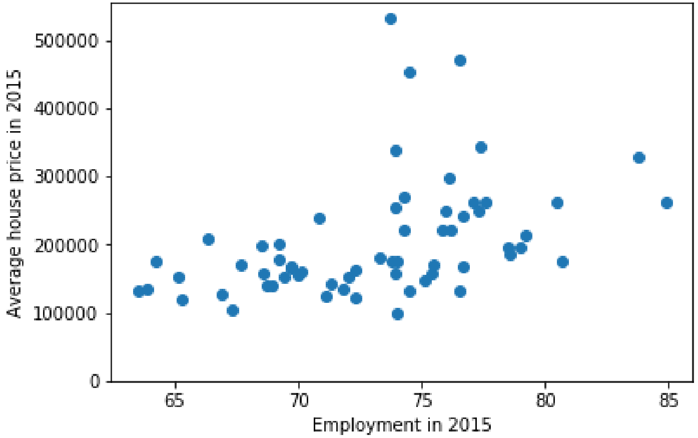
2. STATISTICAL ANALYSIS

RATIO OF POPULATION TO HOUSING STOCK



Metric	Value
R^2	~ 0.22
p-value	~1 x 10^{-4}

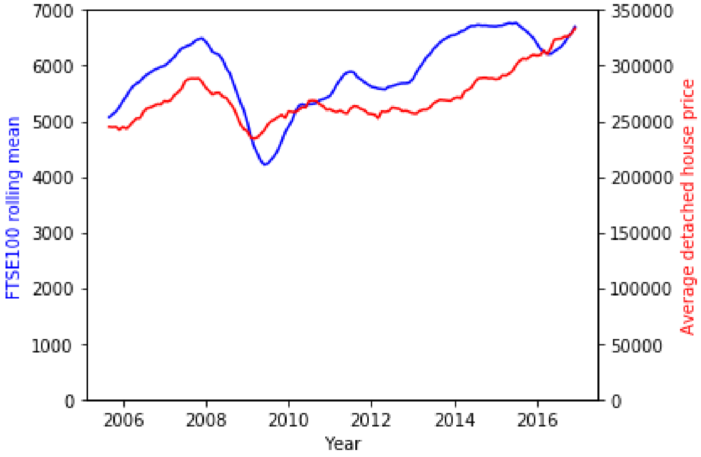
EMPLOYMENT RATIO (%)



Metric	Value
R^2	~ 0.18
p-value	~5 x 10^{-4}

2. STATISTICAL ANALYSIS

CORRELATION OF HOUSE PRICES AND FTSE100*



Metric	Value
R^2	~ 0.55
p-value	~1 x e^{-24}

Source: The FTSE100 9-month moving average is used and the average UK detached house price, both with a monthly frequency

CONCLUSIONS FROM THE STATISTICAL ANALYSIS:

- Three statistically significant relationships have been identified:
  - Ratio of population to housing stock, acting as an indicator of supply (housing stock) vs. demand (population)
  - Employment ratio as an indicator of demand from the income-earning population
  - Value of the FTSE100 index as an indicator of UK investment asset valuations
- Average wages have no statistically significant correlation to house prices – as indicated in the EDA section, houses have simply become less affordable
- The FTSE100 index is the moth closely correlated indicator – suggesting that valuations reflect equity prices more closely than economic fundamentals

3. CLUSTERING

What clustering will try to achieve

- Clustering is an unsupervised machine learning concept, meaning that algorithms used will try to identify patterns in the data, without being given any prior instructions
- The data fed into the clustering algorithms will not include price data, instead the algorithm will group together cities with similar behaviour
- The clusters will then be checked against price performance, to verify if cities with similar patterns of economic fundamentals do indeed have similar property prices
- This would allow e.g. a mortgage lender to view their portfolio exposure in terms of clusters that are expected to exhibit broadly similar price movements

Methods used

K-means

- This algorithm tries to group data points into a user-defined K number of groups
- A key challenge is to find the right value for K, that demonstrates good performance, without fitting too many small clusters

Affinity propagation

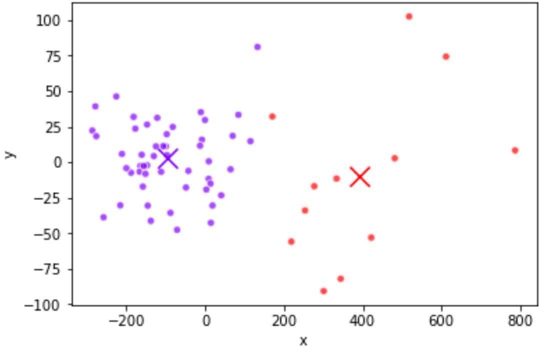
- By contrast, affinity propagation is an algorithm that determines the number of clusters given the data
- A number of parameters need to be tuned, most importantly the ‘damping coefficient’

Metric used

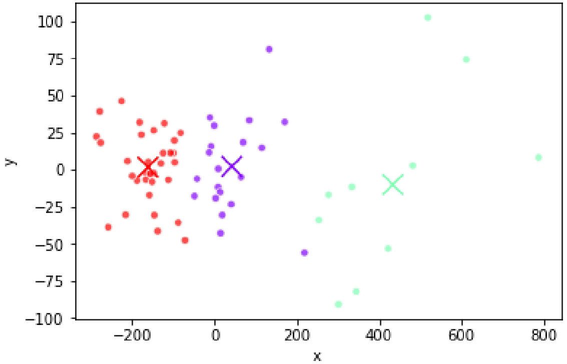
- In both cases the silhouette coefficient is used as the performance metric

3. CLUSTERING: K-MEANS RESULTS

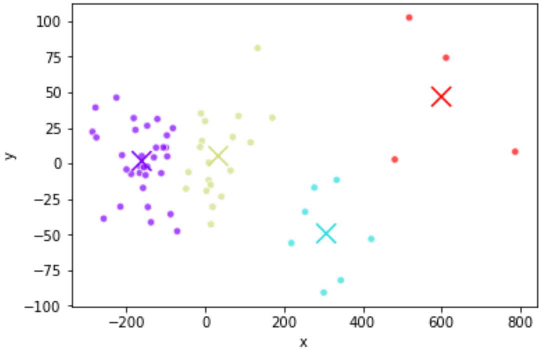
K = 2
S.C.:0.64



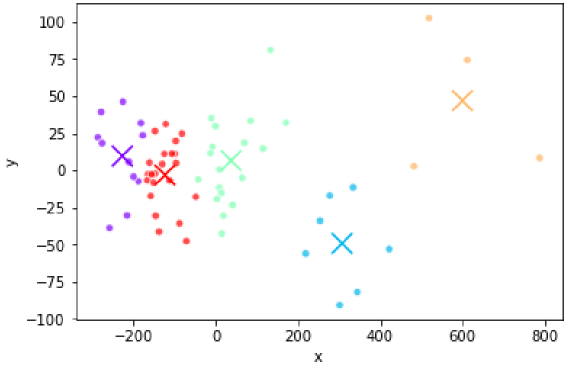
K = 3
S.C.: 0.45



K = 4
S.C.: 0.45

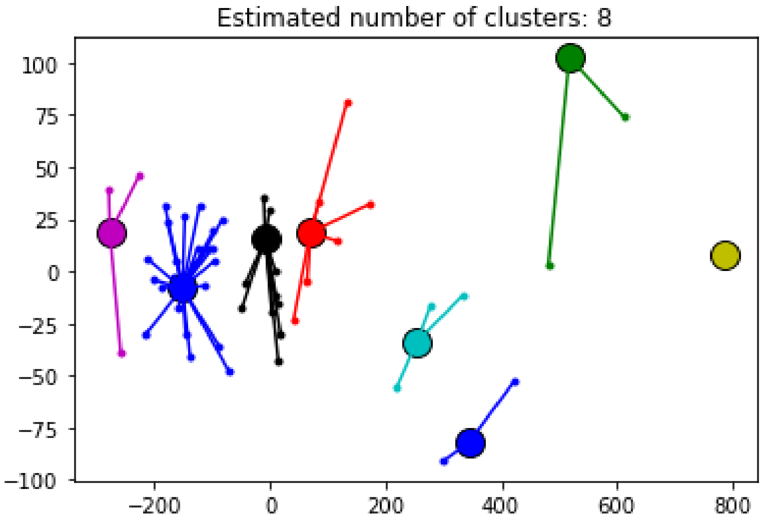


K = 5
S.C.: 0.33



3. CLUSTERING: AFFINITY PROPAGATION

- While K-means worked best with 2-3 clusters, affinity propagation proposes 8 clusters regardless of the damping coefficient parameter
- This is a less intuitive result as some clusters only contain 1-4 members, and one cluster seems to be very large
- For further analysis, the **3-cluster K-means** result will be used



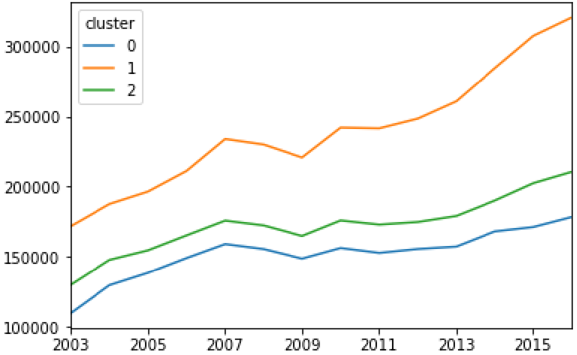
S.C.: 0.28

3. CLUSTERING: RESULTS

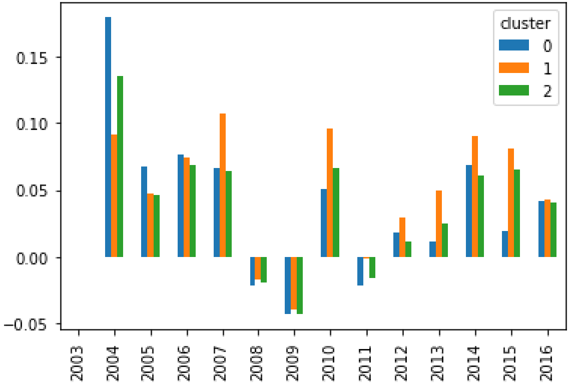
What the clustering analysis tells us:

- **Cluster 0**: low price and low variance cities that has a strong period of growth in 2004-2005 but struggled to recover at the same pace as the other two clusters post-2009
- **Cluster 1**: the high mean and high variance cluster containing the most expensive cities (London and its peripherals plus Cambridge).
- **Cluster 2**: priced above cluster 0 and with a stronger growth rate post-2009 (in particular stronger growth in 2015)

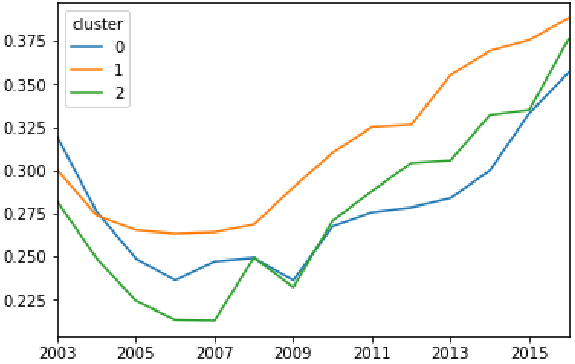
Mean prices



Annual growth



Normalized standard deviation



4. LINEAR REGRESSION

How the test was constructed:

- The purpose of the linear regression model is to try to predict the movement of house prices based on past prices and the shift in economic indicators
- As the EDA and clustering analysis has shown that there is a wide degree of variation between cities, the linear regression is performed cluster by cluster
- Four tests are devised to determine the best predictor
- In each test, the target is to predict prices in year t based on a combination of:
    - Economic data up to year t
    - FTSE100 data up to year t
    - House prices up to year t-1
- The performance metric comparing results between tests is the R^2 correlation coefficient between actual and predicted prices for each year tested

Test	Description
Test 1	Using economic data for year t and prices for year t-1 to predict prices in year t
Test 2	Using economic data for years t and t-1, plus prices for year t-1. The Idea here is to capture the change in conditions from year t-1 to year t
Test 3	Removing all economic fundamentals (wages, employment and population to housing stock) and using only the FTSE value for year t and house prices for year t-1 to fit prices for year t
Test 4	Removing the FTSE value from the features matrix and using only house price data for year t-1 as a predictor for prices in year t

4. LINEAR REGRESSION - RESULTS

Test results

Test	Test 1	Test 2	Test 3	Test 4
Cluster 0	0.9165	0.9270	0.9192	0.9953
Cluster 1	0.9248	0.9267	0.9280	0.9953
Cluster 2	0.9189	0.9100	0.9184	0.9948

- For each cluster, Test 4 provides the most accurate results
- This test only uses the t-1 price to predict the price in year t
- In other words, adding any additional information introduces noise to the data that makes predictions less accurate

CONCLUSIONS

- House prices across the 62 major cities in the UK have risen by an average of 65% from 2003 to 2016
- As wages have on average remained at the same level, owning a house has become increasingly difficult to afford for the average worker, needing 8.0 annual wages to buy a house compared to 5.3 annual wages in 2003.
- The ratio housing stock to population, and the employment ration, are identified as two factors with a (weak) correlation to house prices. The FTSE100 index correlates better to house prices and therefore indicates that houses are regarded more as an investment asset
- In the clustering analysis yields no obvious grouping and the 3-cluster K-means result was chosen as an interpretable solution. Indeed, the clusters do translate into groups that can be distinguished from the perspective of price movement
- The most accurate predictor of future house prices is the last available price. The weak correlation between house prices and indicators such as the FTSE 100 index, employment and the ratio of population to housing stock in a city, adding these features to the linear regression model introduces noise and increases the prediction error
- A more sophisticated algorithm, such as the random forest method, may be more successful in identifying the effect of other features besides the prior year’s price