

Article

Understanding towns in England and Wales: town characteristics and house prices, 2019

Analysis using machine learning techniques on house prices in towns in England and Wales, with a focus on town characteristics and their relationship to house prices.

Contact:

Joe Davies, Kayleigh Mackay, Lili Bui

integrated.data.analysis@ons.
gov.uk

+44 1633 456639

Release date:

19 April 2022

Next release:

To be announced

Correction

14 July 2022 08:59

The article has been amended to make clear that the “job density” measure used in the model is in fact the “employment to residents ratio”. “Job density” is defined as the number of jobs divided by the working age population (16 to 64) of a town, whereas the “employment to residents ratio” is the number of jobs divided by the total population. No other measures are affected.

Table of contents

1. [Main points](#)
2. [England and Wales analysis](#)
3. [Regional analysis](#)
4. [Future developments](#)
5. [Glossary](#)
6. [Data sources and quality](#)
7. [Related links](#)

1 . Main points

- The three most important characteristics for predicting average property prices in towns were: the distance of the town from London, the types of jobs carried out by the town's resident workers and the level of income deprivation in the town.
- For towns within approximately 200 kilometres of London, each 50 kilometres further away from London was associated with around a £50,000 reduction in average price, after which the effect of distance on price weakened.
- A 10 percentage point increase in the number of people employed within job types with the highest skill levels was associated with around a £25,000 increase in price.
- As the percentage of people experiencing deprivation as a result of low income increased, the average house price decreased.
- Within regions, the types of jobs carried out by the town's resident workers were consistently the most important characteristics, while the importance of income deprivation varied across regions.
- Size of town, age profile, region and distance to nearest city had some influence but were less important, while the presence of a university in a town, the proximity of the town to the coast and travel-to-work area classification had little effect for England and Wales as a whole.

2 . England and Wales analysis

Background

This article forms part of a series providing analysis on towns in England and Wales. The series aims to give new insights into their characteristics and the inequalities that exist.

Recently, in [our article Understanding towns in England and Wales: house price analysis](#), we published new data on the prices of properties in towns.

Approach

In this article we have used a machine learning model, [XGBoost](#), to predict the mean price of residential properties. The model was created using property sales data for towns in England and Wales from 2019. The model was given information about the property, such as size and age, alongside information about the town such as income deprivation level, population and coastal classification.

We then averaged these predictions across towns. Town characteristics, and not property characteristics, were then targeted for further analysis to explore the relationship between these characteristics and prices.

This means that predicted average property prices of towns are presented in this article rather than actual prices. For the latest price data see [our UK House Price Index bulletin](#).

Data on property sales during 2019 were used to avoid atypical volatility in the property market during the coronavirus (COVID-19) pandemic period.

The town characteristics included in this analysis were:

- distance to London
- types of jobs, defined by the percentage of a town's resident workers employed within the skill levels set out in the [UK Standard Occupation Classification](#)
- income deprivation, measured by the percentage of residents experiencing income deprivation
- population of the town in 2019 and population growth from 2009 to 2019
- region of England, or Wales, which the town was in
- age profile of residents of the town
- distance to the nearest city
- employment to residents ratio, measured by the number of jobs in a town as a proportion of the population of the town
- number of large and medium towns, and small towns within 30 kilometres
- coastal town classification
- if a town had high, medium or low numbers of residents with level 4 qualifications or above - this includes first degree or equivalent higher-level qualifications
- employment growth between 2009 and 2019
- travel-to-work area classification
- if a town had high, medium or low numbers of businesses within the production industries
- if a town had one or more universities
- whether the town was an administrative headquarters
- the dependency ratio of the town, defined as the ratio of the economically inactive population to the economically active population, based on age

For more details on methodology, see the [Data sources and quality section](#).

Investigating the most important characteristics

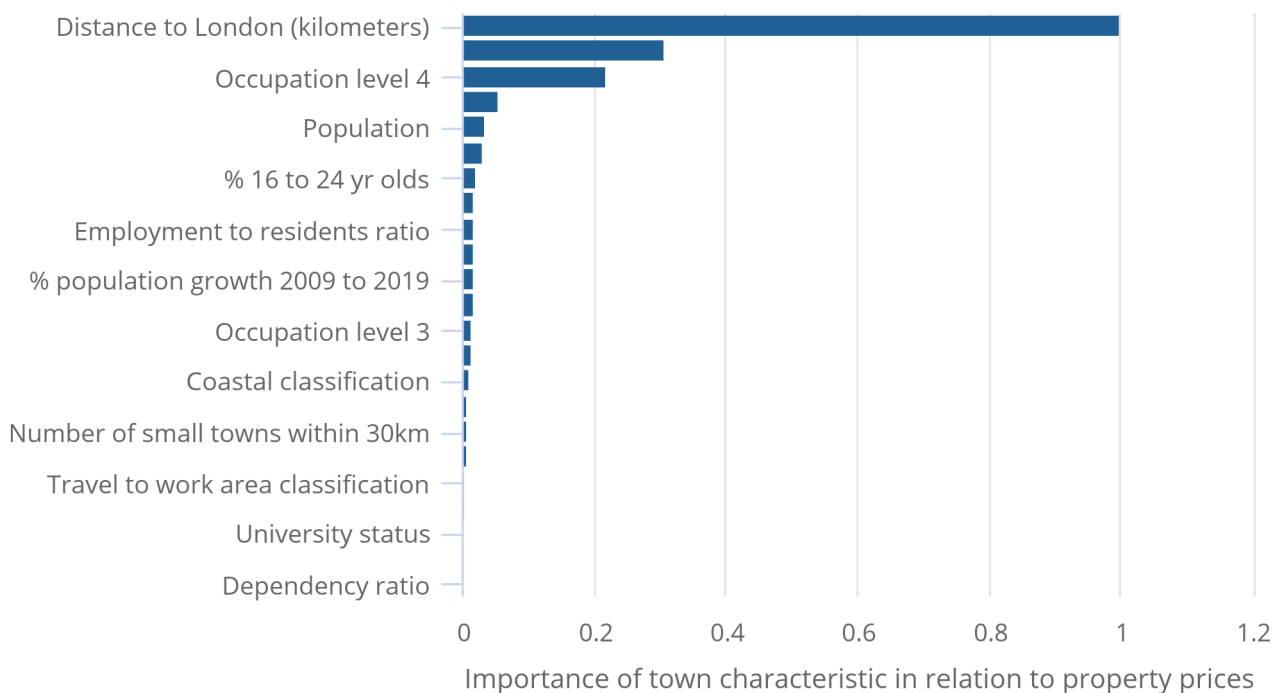
Figure 1 shows the importance of each characteristic in our analysis that predicted average property prices. The importance indicates which characteristics improve accuracy the most. When characteristics were highly correlated, the model predicting the price used the less important characteristic less often or not at all.

Figure 1: A town's distance to London, the types of jobs carried out by the residents, and the level of income deprivation were the most important characteristics for predicting price

Relative importance score, England and Wales, 2019

Figure 1: A town's distance to London, the types of jobs carried out by the residents, and the level of income deprivation were the most important characteristics for predicting price

Relative importance score, England and Wales, 2019



Source: Office for National Statistics using Land Registry Price Paid data

Notes:

1. England and Wales, 2019.
2. Characteristics that are a feature of the property are not included in the importance plot.
3. The x-axis shows the importance of each town characteristic relative to the most important town characteristic.

Distance to London, the types of jobs carried out by the town's resident workers and the level of income deprivation in the town were the most important characteristics. This does not mean that characteristics with smaller importance scores had no association with price. However, the extent to which they explained price differences was lower than the most important characteristics.

Contribution of different characteristics to the average price of each town in England and Wales

We can quantify how much each characteristic is estimated to contribute to the average property price of each town. In Figure 2, any town can be selected and a breakdown shown of the contribution from each characteristic. This can help explain why prices were a certain amount in particular towns.

Figure 2: Impact of different characteristics on the average price in towns, compared to the average town

England and Wales, 2019

Notes:

1. England and Wales, 2019.
2. For each town selected, the chart shows how each characteristic contributes to the average price, over all properties sold in the town. The blue (darker) bars indicate that the particular characteristic contributed positively to price while the red (lighter) bar indicates it contributed negatively.
3. Totals may not sum because of rounding.
4. Property characteristics include floor area, property type, number of rooms, number of bedrooms, and age of property.
5. Other characteristics includes all those shown in Figure 1 that are not already included.
6. Values are predicted mean prices from modelling outputs, not actual mean prices.
7. Values have been rounded to the nearest £100.

[Download the data](#)

Distance to London

Figure 3 shows the relationship between distance to London and average property price.

Being closer to London was associated with higher property prices but only up to a certain point. For towns less than 200 kilometres from London, we saw around a £50,000 reduction in price for every additional 50 kilometres from London. For context, Cardiff and Lincoln are both around 200 kilometres from London.

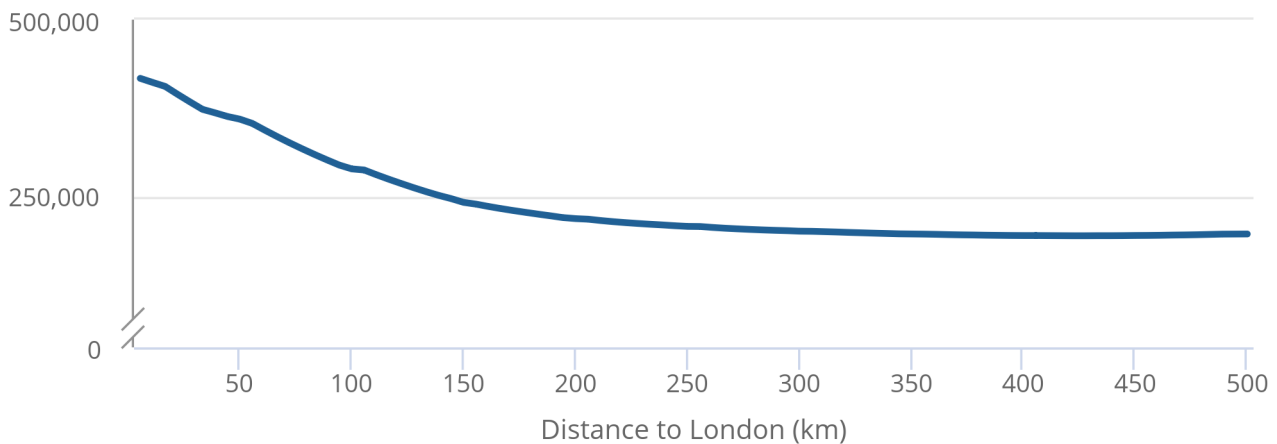
Beyond 200 kilometres from London, the effect weakened, with further increases in distance having little impact on price. This threshold could represent a commutable distance though travel time, distance travelled and cost will not be exactly related to straight-line distance. We did not find strong evidence of a similar effect from other urban areas though this requires further investigation.

Figure 3: Distance of a town from London was highly related with average price, but this weakened after 200 kilometres

Predicted average price in towns (£), England and Wales, 2019

Figure 3: Distance of a town from London was highly related with average price, but this weakened after 200 kilometres

Predicted average price in towns (£), England and Wales, 2019



Source: Office for National Statistics using Land Registry Price Paid data

Notes:

1. England and Wales, 2019.
2. Values are predicted mean prices from modelling outputs, not actual mean prices.
3. Figure has been smoothed in line with common methodology for xgboost outputs.
4. Prices have been rounded to the nearest £100.

Types of jobs

The next most important characteristics were the types of jobs carried out by the town's resident workers. We examined the percentage of people employed within skill levels defined in the [UK Standard Occupation Classification](#). Jobs with higher skill levels generally take a longer time for a person to become fully competent in the performance of the job.

Jobs in skill level 4 relate to "professional" occupations and high-level managerial positions, and normally require a degree or equivalent experience. Skill level 3 includes professions such as science or health associate professionals, skilled construction jobs or other managerial jobs not in level 4. Skill level 2 includes professions such as machine operation, retailing or clerical jobs. Skill level 1 includes elementary trades and administration jobs. Skill levels 2, 3 and 4 were included in the model. The percentage of resident workers employed by skill level is likely highly related to income.

We saw a strong, positive relationship between the percentage of people employed within the highest skill level jobs and house prices. For each 10 percentage point increase, average house prices increased by around £20,000. The effect strengthened in towns with the highest percentage of workers in these jobs.

This contrasts with the percentage of people employed within skill level 2 jobs, where we saw a negative relationship with prices. The percentage of people employed within skill level 2 is highly correlated with the percentage of people employed within skill level 1 jobs. Therefore, we can interpret this chart as showing the relationship between those employed within the lowest-skilled jobs and average price. For each 10 percentage point increase, average house prices decreased by around £20,000 though this relationship weakened when higher numbers of people were employed in these jobs.

There was a weak relationship between the percentage of people employed within skill level 3 and price.

Figure 4: The percentage of people in a town employed in the highest-skilled job types showed a strong, positive relationship with prices

Predicted average price in towns, England and Wales, 2019

Notes:

1. England and Wales, 2019.
2. Values are predicted mean prices from modelling outputs, not actual mean prices.
3. Figure has been smoothed in line with common methodology for xgboost outputs.
4. Prices have been rounded to the nearest £100.

[Download the data](#)

Income deprivation

The income deprivation measure was the next most important characteristic predicting prices. The income deprivation measure used shows the percentage of people in a town experiencing deprivation from low income. This uses the income domains of the English Index of Multiple Deprivation and the Welsh Index of Multiple Deprivation. More information about this measure can be found in [our article Understanding towns in England and Wales: an introduction](#).

Income deprivation was related to the types of jobs workers in a town had. However, because this characteristic has been identified as important in addition to job type, this suggests that it explains additional variation in house prices, which cannot be accounted for entirely by job type. A potential explanation is that income deprivation incorporates additional information about people who are not in employment such as retired people, students or others out of the labour force.

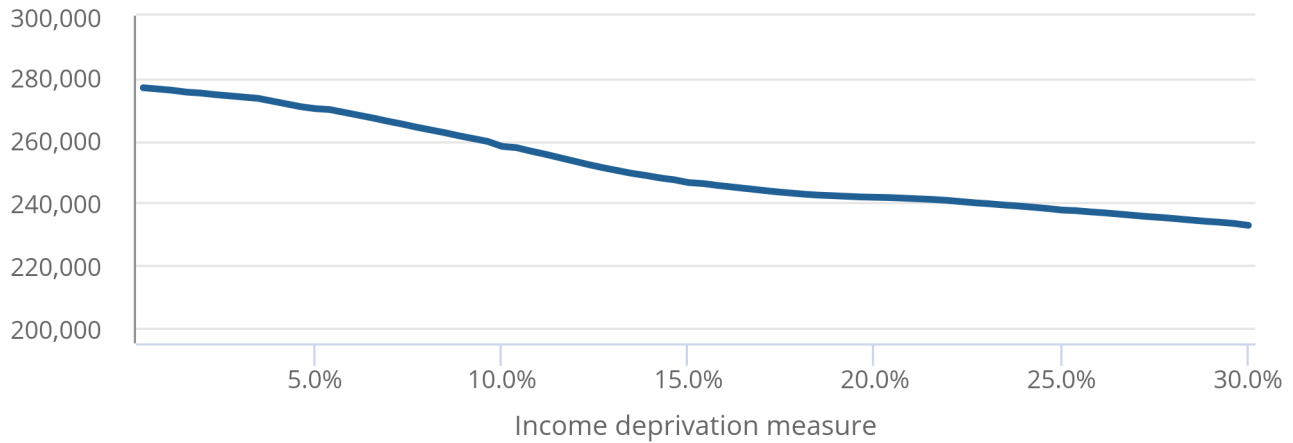
The relationship between income deprivation and price was strong, with the average price reducing by around £10,000 for every 5 percentage point increase in deprivation level. We only consider house prices sold from the Land Registry Price Paid dataset and not any rental market or social housing data. We might expect to see more interaction between income deprivation and the private rental or social housing market, which this analysis does not capture.

Figure 5: Lower income deprivation was associated with higher prices

Predicted average price in towns (£), England and Wales, 2019

Figure 5: Lower income deprivation was associated with higher prices

Predicted average price in towns (£), England and Wales, 2019



Source: Office for National Statistics using Land Registry Price Paid data

Notes:

1. England and Wales, 2019.
2. Values are predicted mean prices from modelling outputs, not actual mean prices.
3. Figure has been smoothed in line with common methodology for xgboost outputs.
4. Prices have been rounded to the nearest £100.

3 . Regional analysis

We investigated the regions of England, and Wales, separately to explore the relationship between characteristics and property prices in those areas. Some characteristics that were not particularly important for England and Wales as a whole were more important within certain regions.

The strength of the positive relationship between numbers employed in the highest skill level jobs and average prices varied across regions. The North West had the strongest association between the percentage of workers employed in the highest skill level jobs and house prices.

In the East Midlands, the most important characteristic was the percentage of a town's resident workers employed in skill level 2 occupations. This includes professions such as machine operation, retailing or clerical jobs. As the percentage of employees in these jobs increased, the average price of properties decreased. This pattern was broadly true for most regions of England, and Wales, though for Yorkshire and The Humber the relationship was weak.

Income deprivation, although the fourth most important characteristic in England and Wales, appears only in the top five characteristics for Yorkshire and The Humber, the East Midlands, West Midlands and the South East. For Yorkshire and The Humber in particular, we saw a strong association, with lower deprivation towns seeing much higher average prices.

Within Wales, the distance to nearest city is the second-strongest characteristic, with prices being higher nearer to cities. The only statistical city (defined as a built-up area with a population greater than 225,000 people) within Wales is Cardiff, though Bristol, Stoke-on-Trent and Liverpool are near to the Welsh border. In the South West the university status of the town was the second-most important characteristic, with seven of the 136 towns having a university.

Distance to London remained consistently in the top five, though it is less important compared with England and Wales as a whole. In further away regions, distance to London had less of an effect.

Figure 6: The relationship between a town's average property price and its occupational levels, income deprivation and distance to London varied between regions

Predicted average price in towns, England and Wales, 2019

Notes:

1. England and Wales, 2019.
2. Values are predicted mean prices from modelling outputs, not actual mean prices.
3. Figure has been smoothed in line with common methodology for xgboost outputs.
4. Prices have been rounded to the nearest £100.

[Download the data](#)

4 . Future developments

This experimental approach was an attempt to explore the characteristics of towns and their relationship to property prices. There is scope to expand and improve this analysis to incorporate additional information about towns including schools, transport infrastructure and other characteristics, which were not immediately available to us. This approach could also be applied to different geographies or the private rental and social housing markets.

We focused on England and Wales only as the data sources for housing sales differ and the town characteristics data have not yet been defined for Scotland and Northern Ireland. However, both Scotland and Northern Ireland have alternative definitions available and were recently incorporated into other articles on towns. Work is ongoing to incorporate Scotland and Northern Ireland into future towns releases.

Further articles on towns are likely to be released in 2022.

5 . Glossary

Towns

In this article a "town" is defined using built-up area subdivision boundaries (or built-up area boundaries where no subdivisions exist). Built-up areas (BUA) and built-up area subdivisions (BUASD) were created as part of the 2011 Census outputs and refer to urban areas defined as "irreversibly urban in character". To be classified as a town, the 2011 Census population of the BUASD (or BUA) had to be between 5,000 and 225,000. Within this context, 1,082 urban settlements in England and 104 in Wales were identified as towns. Further information can be found in [our article Understanding towns in England and Wales: an introduction](#).

XGBoost

Extreme gradient boosting is a machine learning algorithm, which uses decision trees to make predictions. The algorithm works in stages, initially fitting multiple decision trees on a variety of input variables to predict an output variable. After the initial fit, the algorithm finds where its predictions are weakest and aims to correct for these at the next iteration. The process is repeated either a set number of times or until certain error criteria are met.

Cross-validation

Cross-validation is a scheme for splitting data into different parts. In this analysis, we have used five-fold cross-validation. This involves splitting the data into five partitions and fitting the model five times, withholding a partition from each fit to be used for prediction. This reduces the effects of overfitting. Overfitting is when a model is fitted too strongly to input data meaning performance is low on new, unseen data.

Coastal classification

In this article, we have used the same definition and classification of coastal towns used in [our article Coastal towns in England and Wales: October 2020](#).

Travel-to-work area classification

In this article, we have used the same definition and travel-to-work area classifications used in [our article Understanding towns in England and Wales: spatial analysis](#).

Income deprivation

The income deprivation measure for a town indicates the percentage of people in that town that are experiencing deprivation as a result of low income, using the income domains of the English Index of Multiple Deprivation and the Welsh Index of Multiple Deprivation.

6 . Data sources and quality

Data sources

Data from HM Land Registry were used to provide statistics on the price paid for properties sold for value in England and Wales. These data were linked to the Valuation Office Agency property attributes dataset, used for the administration of Council Tax, to provide information on floor area and number of rooms and bedrooms.

These data were then linked to towns data used in previous articles in the Understanding towns series, providing information on the towns in which these houses are located.

Methodology

To predict house prices in towns, a gradient-boosted decision trees model was developed using the [XGBoost](#) package in R. For more details of the method used within the XGBoost package, see [our article Valuing green spaces in urban areas: a hedonic price approach using machine learning techniques](#), which took a similar approach.

Steps in the modelling process were:

1. Data were cleaned, linked, recoded and filtered to only include transactions for towns in 2019.
2. Parameters were optimised using cross-validation to avoid over-fitting. Model performance was evaluated using error metrics of predictions on test data.
3. Predictions of the mean town price, and the mean contributions for each characteristic, were calculated using 2019 transactions within each town.

Table 1 shows the mean absolute error and R-squared values for the England and Wales towns' model.

Table 1: Model performance metrics

	Transaction level	Town level
Mean absolute error (£)	39,100	3,700
R squared value	84.10%	-

Source: Office for National Statistics

Strengths and limitations

The inferences made in this article used a machine learning model. The benefits of this approach compared with standard statistical models such as linear regression were the lack of assumptions needed, the ability to handle non-linear data and the automatic incorporation of complex interactions. For instance, the distance to London variable and some of the job type variables exhibit clear non-linear behaviour, which would violate regression assumptions.

All models involve simplification and there will be characteristics that we have not adequately captured. Findings from the model should be treated with some caution because of the experimental nature of the approach used.

This analysis investigated properties sold and not rental market or social housing. Homeowners will typically differ socioeconomically and demographically from those living in private rentals or social housing so it is likely there were different drivers of prices in these markets.

Some cleaning of erroneous data was conducted prior to linking. Linking was performed using a fuzzy matching technique for linking addresses. Some records were known to be matched incorrectly.

The main set of official statistics for house prices in England and Wales are the [UK House Price Index](#) and [House price statistics for small areas in England and Wales](#).

Unlike the UK House Price Index, the estimates in this publication have not been quality adjusted to account for changes in quality and composition. In line with other Office for National Statistics housing outputs, we have not adjusted prices for inflation.

7 . Related links

[House price statistics for small areas in England and Wales](#)

Bulletin | Quarterly

House prices and number of transactions for property sales in England and Wales, on an annual basis.

[UK House Price Index](#)

Bulletin | Monthly

Monthly house price inflation in the UK, calculated using data from HM Land Registry, Registers of Scotland, and Land and Property Services Northern Ireland.

[Understanding towns in England and Wales: house price analysis](#)

Article | Released 18 October 2021

Data and analysis on house prices in towns in England and Wales. Looks at house price trends in towns up to 2020, and changes over the pandemic period from Jan 2020 to April 2021.

[Understanding towns in England and Wales: an introduction](#)

Article | Released 9 July 2019

The first in a series of articles that provide new data and analysis on towns in England and Wales, to help inform policy.

[Productivity in towns and travel to work areas, UK 2019](#)

Article | Released 30 March 2022

An analysis of labour productivity data for towns and travel to work areas in the UK in 2019.

[Understanding towns: industry analysis](#)

Article | Released 13 December 2021

Analysis of industry in towns and cities in England and Wales, localities in Scotland, and travel to work areas in Great Britain.