

Using Adzuna data to derive an indicator of weekly vacancies: Experimental Statistics

Methodology information for online job advert indices covering the UK job market. These are experimental data created as part of the latest indicators release in response to the coronavirus (COVID-19) pandemic.

Contact:
Tom Evans
vacancy.survey@ons.gov.uk
+44 1633 455277

Release date:
14 June 2021

Next release:
To be announced

Table of contents

1. [Introduction](#)
2. [The data source we are using](#)
3. [How we measure online vacancy data](#)
4. [Applying iterative de-duplication methods](#)
5. [Adverts offering homeworking opportunities](#)
6. [Strengths and limitations](#)
7. [Existing measures of vacancies – ONS Vacancy Survey](#)
8. [Comparison with the ONS Vacancy Survey](#)
9. [Comparison with Institute for Employment Studies](#)
10. [Related links](#)

1 . Introduction

During the coronavirus (COVID-19) pandemic, we have been providing timely indicators of the effect of the disease on the UK economy and society in our [Coronavirus and the latest indicators for the UK economy and society bulletin](#).

These faster indicators now include a set of experimental job advert indices covering the UK job market. These indices are created based upon job adverts provided by [Adzuna](#). These data include information on several million job advert entries live from February 2018 broken down by job category and by region, based on the information included in the job advert. This article sets out the methodology used to derive these indices and provides comparisons with other data sources.

As [Experimental Statistics](#), these data are subject to revisions as our methodology and systems are refined.

We plan to develop these indicators iteratively over the coming weeks and months, taking on user feedback and improving our methodology, and we plan to produce further breakdowns including online vacancies by [Standard Occupational Code](#) and by lower-level geographies such as local authority or Local Enterprise Partnerships.

2 . The data source we are using

[Adzuna](#) is an online job search engine that collates information from thousands of different sources in the UK. These range from direct employers' websites to recruitment software providers to traditional job boards, providing a comprehensive view of current online job adverts.

Adzuna is working in partnership with the Office for National Statistics (ONS) and has made data available for analysis including online advert job descriptions, job titles, job locations, job categories and salary information. The data provided are a point-in-time estimate of all job adverts indexed in Adzuna's job search engine during the point of data extraction.

Prior to our analysis, Adzuna carries out some data cleaning methods such as removing duplicate entries where all information relating to a job advert is the same, because of multiple recruiters advertising at the same time, and applying minimum quality thresholds on some data fields.

Adzuna has a high coverage of all job adverts in the UK but because of the fact that this source is limited to online vacancies, there will be some job adverts missed such as casual work advertised through word-of-mouth and internal vacancies, which are filled using other head-hunting methods.

If you would like more information about Adzuna's data, please contact scottsweden@adzuna.com.

3 . How we measure online vacancy data

We are using online job advert data to create a proxy measure of overall vacancies in the UK.

Allocating job adverts to categories

Adzuna uses a neural network to assign categories to the job adverts. This model uses natural language processing to analyse the text in both the job title and description fields and uses these data to assign the most suitable job category. We use all of these categories as they are defined in the Adzuna data apart from aggregating "Healthcare/nursing" and "Social services/care/work" into "Health and Social Care".

Allocating regions to job adverts

On 18th March 2021, we updated our regional methodology. The online job advert data that we receive from Adzuna contain a free text location field, which is filled in by the company or individual creating the job advert directly. These can contain varying levels of geography information as the advertising company deems fit for the advert, including street names, postcodes, towns, cities or even countries. They may also not fill in a location. To allocate regions to job adverts, we:

- manually review a number of these locations to assign them to a local authority, which accounts for 80% of all unique locations in the dataset
- match this to a postcode Lookup file for adverts that have postcode location information in the raw location data
- text-match the raw location information to a lower-level geography such as wards, local authorities and counties, based on an exact match of the wording used in Office for National Statistics (ONS) Geography names covering the UK
- text-match the raw location information based on partial matches of the wording used in ONS Geography names covering the UK
- choose the lowest geography (that is, ward) as the one to map to region in the case of matching at multiple levels
- only use the postcode information in the case of having both postcode and other raw location information
- map the information from the previous six steps to the right local authority using [ONS Geography Lookup files](#), before aggregating these local authorities to their corresponding NUTS1 regions
- make some manual corrections to re-assign incorrect allocations that are caused by common location names across different regions

Some locations may be too high-level to map to a local authority (for example, counties), but can be mapped to NUTS1 regions using the steps outlined in the first four points outlined.

Using this approach, we can assign regions to the majority of job adverts, but there remain some adverts that are unallocated and hence shown as “Unknown” regions. Some examples of why these adverts have not been assigned a region are:

- the location of a job advert is too granular to be matched to a local authority using our current Lookup files; for example, street-level information is not currently included in our Lookup but may be considered in the future
- the location of a job advert is too high-level to be matched to a region; for example, jobs can be advertised at a “United Kingdom” level
- the location of a job advert cannot be matched to the Lookup file because it is not standard
- the job advert is not assigned a specific location and is instead advertised as a “remote working” or “working from home” opportunity
- the location of a job advert is outside of the UK

We plan to iteratively improve our region allocation process to address these example cases in future.

Presenting the data

We present our total advert, adverts split by category and adverts split by region as index series, which are calculated in the following way:

1. aggregate all live job adverts at a given point in time in a week
2. missing and anomalous values are imputed by linear interpolation, so there is one value for each week
3. calculate the mean of weekly counts of live job adverts listed on Adzuna in February 2020, including the imputed values
4. divide each value in the total time series by this mean value, thus indexing the series so the February 2020 average equals 100
5. round the indexed values to one decimal place

Notes for: How we measure online vacancy data

1. The education industry's total adverts for 21 March was anomalous, and the value has been imputed through linear interpolation.
2. The missing values are one week between 15 and 28 February 2019, three weeks between 31 October and 28 November 2019, two weeks between 5 and 27 December 2019, and one week between 3 and 16 January 2020.

4 . Applying iterative de-duplication methods

From 25 March 2021, we will introduce an additional version of online job advert indices, which have partially accounted for some specific types of duplicate job adverts in the data. The steps taken to identify and remove these duplicates are outlined below:

- apply text-cleaning techniques to the job description and job titles to ensure they are lower case and only contain letters or numbers; that is, special characters are removed
- remove very common words from the job descriptions such as “our”, “a”, “an” and so on
- apply a document similarity detection method to identify sets of job descriptions that are near-identical in their wordings and have the same job title; we then keep a single example for each NUTS 1 region, assigned from the previously mentioned methods, so that a job description from an identical set is found
- the similarity detection is designed to be highly specific, flagging only very similar documents as duplicates; this is because most duplications appear to be repostings with little editing, and because for heavily reworded job descriptions it is often unclear whether the duplication is for the same position or for very similar positions differing slightly in technical skill requirements or seniority
- mark duplicates and drop them from the data

5 . Adverts offering homeworking opportunities

On 14 June 2021, we published a one-off [dataset](#) of online job adverts which offered homeworking opportunities. This shows the trends of jobs referencing some homeworking, as well as trends of the proportion of all jobs referencing the opportunity. A supporting [article](#) explores this.

The steps taken to identify these adverts are:

- combine the job description and job title fields together to one string variable
- apply text-cleaning techniques to the job description and job title to ensure they are lower case
- apply text-matching to identify job adverts which contain key phrases associated with homeworking such as "remote working", "work from home", "home-based" and "telework"
- apply a correction where phrases such as "nursing home" and "care home" are amalgamated to ensure they are not wrongly identified as homeworking.

It is important to note the limitations of the series. These are that:

- the series does not separately identify job adverts that exclusively offer homeworking from those which offer flexible homeworking, such as one day a week at home
- the series does include a small number of incorrect classifications because of adverts listing that the post is "not suitable for homeworking".

We have also produced a deduplicated version of these data. Users should be aware that some trends differ between the non-deduplicated and deduplicated data.

This methodology has been produced by the Office for National Statistics (ONS) in collaboration with our research partners at the University of Warwick. The university team has supported our work in identifying online homeworking adverts.

The University of Warwick team will be publishing an academic paper on their work in June, titled: 'Revolution in progress: the rise of remote work in the UK. CAGE Working Paper (2021)'.

6 . Strengths and limitations

Strengths of Adzuna data

- The data are extremely timely with analysis available to be published six days after the snapshot of adverts has been extracted; this provides an early indication of how the trend of the number of live job adverts is changing in the UK.
- Data are available on a weekly basis, allowing week-to-week comparisons.
- Most adverts in the dataset include detailed information such as potential lower-level geographies, detailed job descriptions and some salary information.

Limitations of Adzuna data

- The number of job adverts being posted is not a direct measure of labour force demand; the number could respond to other changes such as how positions are recruited for (that is, decreased activity from recruitment agencies could lead to decreased duplication of multiple adverts for one post).
- Job adverts may not be removed from online job vacancy boards immediately when the position is filled so the indices may not fully reflect companies who have halted active recruitment. Note that Adzuna perform data cleaning to remove adverts that have not been observed as live for 30 days.
- The data are compiled from multiple job vacancy boards and adverts are considered “live” if the posting is still live on any board, even when it has already been removed from an alternative source.
- The scope of online job adverts does not fully capture the scope of UK economic activity because of differing advertising methods, for example, casual work may be advertised by word-of-mouth or in shop windows as opposed to online.
- There are points in the time series where we know there are increased levels of duplication in the dataset that are resulting in a potentially inflated value of job adverts; where this is the case, footnotes were added to the data tables to identify these anomalies.

7 . Existing measures of vacancies – ONS Vacancy Survey

The Vacancy Survey is a statutory, monthly survey of businesses. The survey asks a single question: how many job vacancies a business had in total (on a specified date) for which they were actively seeking recruits from outside their organisation.

The headline series are based on three-month moving averages, by industry and by size of business. The [Inter-Departmental Business Register \(IDBR\)](#) is used as the sampling frame. The total sample is approximately 6,100 businesses per month, with approximately 1,400 large businesses included every month and the remaining 4,700 consisting of smaller enterprises randomly sampled on a quarterly basis.

The survey covers all sectors of the economy and all industries in England, Scotland and Wales (Great Britain) with the exception of employment agencies (to avoid double-counting of vacancies) and private households, agriculture, forestry and fishing (because of the disproportionate costs involved as these industries mainly consist of very small businesses with few vacancies). Estimates for UK are derived by weighting up the data for Great Britain using employment estimates (Northern Ireland accounts for around 3% of UK employment). Vacancy statistics are not available by region. Northern Ireland businesses are not approached because of the risk of overlap with other surveys conducted by Northern Ireland departments.

The Vacancy Survey reference date falls on the first Friday of the month, unless this is the first day of the month. In this case, the reference date moves to the second Friday of the month. For the May 2020 period, the reference date was Thursday 7 May 2020 because of a change in date of the first May Bank Holiday.

Office for National Statistics (ONS) vacancy statistics are a three-month average measure that are seasonally adjusted and then published in the [Vacancies and jobs in the UK statistical bulletin](#), usually between six and seven weeks after the reference date of the survey.

Further information regarding the ONS Vacancy Survey can be found in the [the Vacancy Survey QMI](#).

8 . Comparison with the ONS Vacancy Survey

When comparing [experimental](#) Adzuna job adverts data with the Office for National Statistics (ONS) Vacancy Survey it is important to be cautious and note the different definition of what each source covers.

Adzuna covers online job adverts listed, which can include multiple job opportunities within one advert or ongoing recruitment campaigns, which do not align directly with one job vacancy; further detail on the limitations of the Adzuna dataset can be found in [Section 3](#).

The ONS Vacancy Survey covers vacancies defined as a vacancy for which businesses are actively seeking recruits from outside their organisation.

However, it may still be sensible to compare both data sources to give an indication of quality. At a higher level, we can see similar movements in Adzuna vacancies compared with ONS vacancies data. The same comparability can be observed with vacancies that are classified at industry level with equivalent [Standard Industrial Classification \(SIC\)](#), such as “Education”, “Healthcare/Social Care” (aggregated from Health care and Social Care), “Retail/Wholesale” and “Catering/Hospitality”. For these groups, we found some correlation with trends in ONS vacancies data.

However, this correlation was not found in other groups because of different methodologies. The vacancies data in some categories assigned by Adzuna such as “Graduates” and “IT/Computing/Software” are linked to both occupations and industries; therefore, it is not appropriate to compare industry-level vacancies directly with the [ONS vacancies estimates](#) which are produced at SIC level.

While the “Healthcare/Social care” category has historically shown a strong correlation with the ONS Vacancy Survey, from April to September 2020, it diverged from the vacancies data.

9 . Comparison with Institute for Employment Studies

[The Institute for Employment Studies](#) (IES) published weekly vacancy analysis from April to the beginning of July 2020 using Adzuna data. These publications provided additional insight including local changes in vacancy levels and changes in vacancies by salary levels. We collaborated with the IES to ensure our methodologies are consistent, but there are some differences that users should be aware of.

Extract date

Adzuna provide point-in-time estimates of job adverts listed at any given time. Although both the Office for National Statistics (ONS) and IES are extracting their data weekly, these extractions take place on different days of the week. This means there are minor differences in our analyses, as there are a different number of job adverts listed on different days.

Location assignment

Some of the job adverts included in the Adzuna data have unidentified locations. IES has removed these adverts from their total advert series, but the ONS has not. This may cause minor differences between the sources at the aggregate level.

Regional breakdowns

The IES and ONS have taken different approaches when assigning regions to job adverts. The ONS is using the raw location variable, but the IES are using longitude and latitude points that are provided by Adzuna. Adzuna use their own algorithm to map the raw location variable to longitude and latitude points.

Additionally, the ONS is carrying out text-matching, but the IES are using the package PostcodesioR to match the longitude and latitude points to local authorities. The IES manually check locations in England because in a very small number of instances, the longitude and latitude points were found to be outside the boundaries of the relevant local authority; locations in Wales, Scotland and Northern Ireland are matched solely by the software package, meaning slightly different methodological approaches. As a result of this, we expect there to be minor differences in the breakdowns.

As of 9 July 2020, the IES are no longer producing these weekly estimates.

10 . Related links

[Coronavirus and the latest indicators for the UK economy and society: 28 May 2020](#)

Bulletin | Released 18 March 2021

Early experimental data on the impact of the coronavirus (COVID-19) on the UK economy and society. These faster indicators are created using rapid response surveys, novel data sources and experimental methods.

[Labour market overview](#)

Bulletin | Released 23 February 2021

Estimates of employment, unemployment, economic inactivity and other employment-related statistics for the UK.

[Vacancies and jobs in the UK](#) Bulletin | Released 23 February 2021 Estimates of the number of vacancies and jobs for the UK.

[Online job advert estimates: May 2020](#)

Dataset | Released 18 March 2021

Experimental job advert indices covering the UK job market.