Introduction to Adzuna job market data

Transcript from webinar video recording

1
00:00:00,607 --> 00:00:05,742
So, welcome to this webinar on Introduction to Adzuna Data

2
00:00:06,118 --> 00:00:08,597
and how to apply for access.

3
00:00:10,917 --> 00:00:12,101
I'll be taking the session.

4
00:00:12,201 --> 00:00:15,556
My name's Heather Sinclair, Information Services Officer

5
00:00:15,656 --> 00:00:18,061
at the Urban Big Data Centre.

6
00:00:18,656 --> 00:00:22,225
I'll be joined today by Scott Sweden from Adzuna,
Qunshan Zhao, Senior Lecturer in Urban Analytics at UBDC,

Andrew McHugh, Senior Data Science Manager,

and Nadiia Gorash, Data Scientist at UBDC.

I'd like to remind participants that this session is being recorded and will be uploaded on the Urban Big Data Centre website in an accessible format at a later date.

for those who could not attend the session.
Cameras will be turned off and microphones muted to aid privacy and also for bandwidth reasons.

You may wish to adjust your name as it appears on the screen.

Feel free to introduce yourself in the chat box but please don't include any personal information such as phone numbers or emails.

Please also use the chat to ask questions.

These will be collated and responses will be provided
in the Q&A session at the end.

During the webinar, there will be three presentations followed by an opportunity for discussion and a question and answer session where you can submit your questions using the chat function.

Now I'll move on to the first session, which provides an introduction to Adzuna data and how to apply for access.
To start, I'll provide some background information about UBDC.

The Urban Big Data Centre was established in 2014 and it is jointly funded by ESRC and the University of Glasgow.

The aim of UBDC is to promote innovative research methods and the use of big data to improve social, economic, and environmental wellbeing in cities.

UBDC's research centre specialises in areas of transport and mobility, housing and environment,
education and skills,

urban governance, and urban analytics.

As a data service, we achieve this by enabling and supporting the research of our service users.

The objective of UBDC's data service is to build an outstanding data collection for non-commercial research by UK academics which support high quality, impactful research on UK cities.

So, Adzuna data is one of our recent additions to the collection.
Adzuna searches thousands of websites and brings together millions of job advertisements on their website.

The Adzuna data provides information about the job market, job categories, and can be used to analyse vacancies over time.

The data provides snapshots of job ads which were advertised on Adzuna.co.uk at a particular point in time.

UBDC will provide historical data from 2017,
current data, and also future data up to Spring 2022.

So, this is a table with the fields that are included in the Adzuna data.

It includes everything you would expect to see if you were going to a website to view a job vacancy.

So, it has the ID of the job vacancy,

the contracted hours so you could tell if it was full- or part-time,

the type of contract, whether it was permanent or temporary,

the company name,
the date the job listing was added,

a detailed description of the job.

It has location information,

so it includes latitude and longitude co-ordinates

and the region it was in.

So, for example, north-west England or Yorkshire and Humber.

The minimum and maximum salaries.
There's also a field for predictive salaries.

That's for listings without salary information.

The currency of the salary.

We have UK data so in most cases it will be pounds.

The job title and the link to the advert.

I don't expect you to remember all that from the previous slide.

so UBDC has a data catalogue on their website

with information about all the available data sets.
This is a screenshot of the entry for Adzuna data

and by clicking on read more,

you’ll be taken to further information and also a data profile sheet

which contains all the information that was on the last slide.

So, who can apply to use the data?

So, UBDC has a limited number of credits available to sub-license data to PhD level staff at UK universities.
The data is to be used for non-commercial academic research.

UBDC licenses are for academic research

but if anyone on the call is from a government organisation

and would like to access the Adzuna data,

then they could contact Adzuna directly to discuss their licensing options.

Projects should generally be up to two years duration,

although longer qualifying projects may be supported.

Academic staff can use the data for research projects
but it's not to be used for teaching use.

Applicants are expected to be ready to use the data straight away once they receive it.

The data can be used for various purposes.

So, I've included a few here.

Analysis of new vacancies.

Overall vacancy levels.
Vacancies by particular region.

Or by a particular job type.

So, the first call for expressions of interest is now closed.

but UBDC will soon be launching a second call.

And so, once it's launched,

you'll find everything you need to know.

on the call for expressions of interest page.

And at the end of this page,
there'll be an online form that you can fill in with a project summary form, which is just brief details about what your project involves and who's going to be working on the project. This gets checked over, just to make sure you're eligible, that it's coming from an academic organisation, that the timescales are relevant. And then, assuming that's all okay, we send out a more detailed Adzuna proposal form.
This asks you in a lot more detail about your proposal.

So, what we’re looking at in the proposal form is these criteria here.

So, you’ll find these sections in the form where you get opportunities to write about each of these things.

So, for example, privacy and ethics.

We’ll be looking at, does the project require ethical approval?
Feasibility.

The kind of things we're looking at here are,

can the project be completed within the timescales or using the methods that were outlined in the proposal?

The scientific contribution.

This could be development of a methodology or development of new data products or resources.
It could include publication in an esteemed journal

or a presentation at a conference.

Potential for public benefit impact.

So, the panel will be looking at public benefit impact beyond academic research.

So, the things you could include here are information about planned knowledge exchange or other communications activities.
If you have one, you could include a pathways to impact statement, which is where you outline what you would do to make beneficiaries aware of the research so that impact could be achieved. Outcomes. So, we respond to all applicants and provide their decision. Successful applicants will be emailed licensing agreements and we'll share the data on completion of these agreements.
As a free service funded by the Economic and Social Research Council, we rely on end user feedback to provide evidence of the value of what we do.

It's a condition of the licences that data users inform UBDC of any publications they produce that were a result of having the data, and that researchers also include a citation to acknowledge the data owner, Adzuna, and UBDC.

So, I'll now hand over to Scott Sweden from Adzuna.
who will present the next session

about the background of the Adzuna company

and the reasons for the data creation.

Right.

You should see my title page.

If you're not then that is a problem.

So, thanks very much, Heather.
My hope and expectation really is that this is going to prove, obviously, a great partnership.

So, thank you to the UBDC.

And irrespective, I suppose, of your faculty or your expertise,

I'm going to be very excited to learn about the research and importantly, I suppose, the outcomes of your endeavours in the year to follow.

Very briefly, as it says on the tin here,
I'm the Head of Data Sales, which is fairly self-explanatory.

But I'm going to be talking to you about who we are.

I'm going to assume that none of you have familiarity with the data set.

So, those that are very familiar with job advert data,

just please bear with me and I apologise.

Or go and make a coffee.

I'm going to talk about where the data comes from,
some stats of the coverage,

and what you can expect to extract,

what it’s used for, and then I’m going to end with some real life

public sector applications of our data at the moment.

So, Adzuna, who are we and why are we?

So, we’ve only been making our data available, really,

for the last 18 months,

which was through, essentially, organic demand.
It’s come from a very diverse group of clients.

It’s from venture capital, asset management,

quant funds, to corporate HR.

And of course, as you can imagine, those in

public policy and labour market economists.

Adzuna was founded 10 years ago now

by Doug Monro and Andrew Hunter, ex-eBayers and Gumtree.
And their mission was to make it easier for millions of people to find better jobs.

And that's by listing every job everywhere, as the slogan goes.

We do that by leveraging our proprietary technology and our algos to achieve that goal.

So, for any of those that are in business schools that might find this slightly interesting,
our commercial model works by driving click traffic to job boards.

So, there's an important distinction,

and I'll get to that later,

but they are our clients and we aggregate all the jobs from them in the marketplace into one particular place.

And then through this aggregation model,

it means we now have 20 million monthly visitors across the 16 different countries we operate in
for the data set you'll hopefully get access to.

We treat the UK as one country to us, one market, so that essentially means England, Scotland, Wales, and Northern Ireland are where you'll see data.

And as I mentioned, we've developed leading matching technology and we also provide that to central governments.

So, we won the tender for their "Find A Job" service
and that's something we manage to this day

and it's a public service that you can see.

So, what makes us, for a bit of plugging,

what makes us the number one job search engine

in the countries we operate?

For a start, having more vacancies than any other single provider is

obviously going to be very helpful for our cause.
And that should be great news for the analysts amongst you

or perhaps the economists amongst you as well.

So, how do we do this?

So, we aggregate job adverts and we have three primary sources.

The first is our job boards.

Some of these will be your household names.

You'll be familiar with Reed and Monster, etc.

And they will take jobs from their clients
who will pay them a fee to list their job.

Recruiters. So, on the right-hand column,

you'll see recruiters. So, Michael Page.

They'll manage the whole process for a job seeker.

So, from finding a job to application

to negotiating a contract if they're any good.

But they also have now, of course, their own website

so they'll host all their clients' jobs on there.
And then lastly, there's the middle section, which is employers directly. So, we call that direct business Adzuna. So, we will have Virgin, for example, come and host their jobs directly because they have access to two to three million job seekers a month. So, the best place to find all the jobs so they'll want to come to us directly. And just to make sure
everyone is still with me,

247

00:14:38,464 --> 00:14:42,032
I've thrown a little curveball in there. So, Michael Page, as you'll notice,

248

00:14:42,132 --> 00:14:45,352
some of you will notice that they could be a client of ours.

249

00:14:45,478 --> 00:14:49,486
So, they could be those that they are hiring for more recruiters.

250

00:14:49,600 --> 00:14:51,730
I don't know if that's a great idea in the world but.

251

00:14:51,800 --> 00:14:53,666
I know we need more recruiters. But they could be hiring from

252

00:14:53,766 --> 00:14:55,614
lots more recruiters as a direct client.

253

00:14:55,714 --> 00:14:59,350
But they also, we will be taking in their jobs feed

254
hosting all their clients' jobs as well.

So, just to help with the distinction of where our job sources come from.

So, what does that mean for our job representation?

Which is probably more interesting to you.

From our database point of view, in the UK, on the left-hand side,

we've got 85 million jobs plus in our UK historical database.

The USA, it's about 240 million.

240 million and growing. In Russia
it's a good market for us, it's about half that.

In terms of what you'll be able to access though,

the type of data set, more importantly,

you'll be accessing, is,

we have about 1.15 million live jobs on our data set.

So, you will be getting weekly snapshots.

So, week to week, what were the live jobs on the marketplace at that point in time.
And so, what you're actually getting is the same type of access,

the presentation of data is the same as the ONS are receiving today,

and I'll refer to that at the end.

How representative is the sample is another question we get asked a lot.

By that I mean, how reflective of the online job market is it?

In the UK, that hopefully you'll get access to,
the online job market space.

00:16:18,279 --> 00:16:21,216
We have similar coverage in other countries, like Germany,

00:16:21,316 --> 00:16:22,693
France, the Netherlands, etc.

00:16:23,677 --> 00:16:27,367
And we know this by simply going on our clients' websites,

00:16:28,127 --> 00:16:32,683
but also, we start to make comparisons with national statistic databases

00:16:32,783 --> 00:16:35,586
in the country. So, obviously, in the UK it's the ONS.

00:16:39,769 --> 00:16:41,997
This comes from our Head of Data Science.

00:16:42,097 --> 00:16:44,984
He sees it as, as we all do really,
a hugely valuable goldmine of data

for those who aren’t familiar with job advert data.

So, you know, it’s a rich source.

You can look at hiring trends.

You can look at salaries, you can look at perks.

Obviously, as an economist, you can look at labour demands or industry sector growth.
If you're doing public policy work, you can look at skills insights and you can look at changing work patterns, again, sort of, HR, human capital, how things are moving through post-COVID. How are things changing over time? How much has it sped up? And we can cut this by all sorts of other fields, like, you know, obviously, granular location, we can do it by industry, you can look at it at a granular occupational level,
and so on and so forth.

There's a huge wealth of data.

And one client recently said to me,

an economist said to me, actually,

you know, "Data is more like air."

I thought that might have been a bit extreme

but maybe the data scientists amongst you will think so too.
However, it doesn't come to us all nicely extracted.

We've obviously had to do a lot of crunching.

Just as a bit of fun really, you can see some adverts in the middle.

It's £1 per year if you want to be a Consultant Petroleum Chemist.

I don't advise that.

And then, if you look at something, for example,

this Software Engineer role, which some of you may be or become,

you can see that they're relating to ballet in the job description.
So, it's not great, perhaps, if we're an economist.

However, of course, these are written by human beings and we are prone to error and flights of fancy.

But from a social research perspective, this is a great thing because this is a lot of value in these types of errors and what people are writing.

I'll show you why next.
So, if we just go into a bit more detail.

In the full job description text, which you will be getting,

there is a huge amount of data.

So, whilst we extract what we like to think of as the tip of the iceberg,

job descriptions can extract more than we've already done.

So, for example, seniority.

So, just looking at seniority,

you know, questions that are being asked of our data, you know.
Is the workforce getting older, you know, due to remote working?

Therefore, less succession planning.

Therefore, people are staying in their jobs longer.

This could be true of enterprise businesses and when we look at enterprise businesses, company hiring.

So, in 2020, we tagged about 8000 known companies.
There are probably a lot more that we haven’t tagged to our own database.

340
00:19:29,400 --> 00:19:31,600
But, you know, you could look at how does this relate to

341
00:19:31,658 --> 00:19:32,977
financial reporting.

342
00:19:34,059 --> 00:19:37,096
This a metric that VCs, Venture Capital firms are using

343
00:19:37,196 --> 00:19:40,532
for due diligence, or it could be Asset Managers on industry trading.

344
00:19:41,500 --> 00:19:43,537
If we look at perks, just looking at perks,

345
00:19:43,935 --> 00:19:45,424
researching human capital,

346
00:19:46,492 --> 00:19:47,569
remote working.
How are companies now enticing the next generation of employees?

What are they doing?

Qualifications, you could extract all of the qualifications requirements.

Obviously, there's a big piece here on education.

What are the academic credentials for junior roles? For example, for graduate roles.

Is that demand now greater than ever?

What occupations does this apply to?
And my last point on this, this text parsing area, is

the future of work, inverted commas.

Huge topic.

McKinsey, 800 million jobs at risk

from AI and digitalisation of the economy.

Which occupations are most defensible

based on hiring and firing rates?

What are we seeing out there?
Linguistics analysis. So, the second part.

Diversity and inclusion.

I've already spoken to academics about this type of work.

So, for the heavy data scientists, data crunchers amongst you,

looking at sentiment analysis,

understanding gender and cultural bias in the language is

something that is being heavily looked at
at the moment, of course, because,

371
00:20:55,907 --> 00:20:57,963
obviously, D&I, it's very topical

372
00:20:58,163 --> 00:21:02,052
and we are, as I say, we are 95% of the online jobs market

373
00:21:02,152 --> 00:21:04,820
so there's a lot to be mined from that.

374
00:21:05,289 --> 00:21:07,918
And finally, vector embeddings.

375
00:21:08,657 --> 00:21:12,146
So, when you're looking at, you know,

376
00:21:12,246 --> 00:21:14,574
the data scientists amongst you, clustering techniques,

377
00:21:14,674 --> 00:21:19,433
we are using that, as a company, to build recommendation engines for jobs.
But, of course, there are a lot of other things you can do with vector embeddings,

and I'll talk about that after as a real-world application of our data.

So, finally,

just a quick visual representation.

You can segment our data and aggregate it up to look at occupations and salaries and vacancy counts.

You can translate that into interesting...
Well, we've just used a heat map for this example.

There are lots of other ways.

And then, of course, this happens to be looking at profiling of a particular occupation and it's an extraction of skills in the workplace.

So, just to summarise and characterise the things I've said.

So, it's our opinion
that you won't find a more representative sample to work with in the UK.

Our data science team has already done a lot of the heavy lifting in terms of cleaning and deduping.

And the application set can be applied to many different, from your perspective, I suppose, faculties.

So, data science and technology.
We have a lot of people from technical universities accessing our data.

Finance and trading.

Human capital. Looking at salaries and perks.

To, of course, the obvious one is labour market and public policy research.

There's a huge wealth of applications for our data out there.

So, finally,
I'm just going to end on some public sector applications of our data.

And therefore, to really help you appreciate the importance of what this data is used for.

And, obviously, the academic research essentially underpins a lot of what goes on out there in the public and the private sector.

So, you'll know, of course, the ONS and some of you, the economists amongst you,
might know the Labour Workforce Survey that looks at 5000-6000 companies.

It looks at hiring and firing trends,

attitudinal and behavioural characteristics of this as well.

What they are doing, if you look at the left-hand side,

the ONS are augmenting their surveys,

which are quarterly, with our real time data

to build indices on the economic recovery.
Now, just some quick stuff.

Fortunately, we're, as you can see from this data,

we're over, back to, well, just above where we were

pre-pandemic levels, as average job count,

which is good.

They are doing a lot more granular stuff

privately and internally, but interdepartmentally,
that they can't publish,

but they're doing internally a lot more granular work.

More specifically, what are they doing?

So, some of the things they're doing.

They're analysing a relationship between earnings and skills.

They're doing location analysis interdepartmentally.

They're looking at the impact of COVID on self-employed versus employed.
And they’re also about to publish a report with our data on the impact of COVID on remote working.

So, I’ll just park the ONS there.

On the AI, on the OECD side,

the OECD, obviously funded by the EU,

they represent 27 nations under the EU, I believe, last count.

Their AI policy team is looking at
the make up of AI and IT across various occupations.

So, a data scientist versus a database engineer versus those in research and methodology.

and they're looking specifically at what the skills make up is of those jobs today.

And they're looking at how demand differs across regions.

Which you can do with our data but, of course, they are also covering countries, the various other countries that we use.
And finally, something that is not on here.

We have a think tank that is looking more broadly at skills mismatches across professions.

Where is there the biggest imbalance?

They’re mapping vacancy data.

to, sort of, ONS data.

Things like, to furloughs and unemployment.
And they're also looking at things like economic shocks.

So, of course, we've had, some say, a business siphoning in 12 months that would normally take five years or 10 years. So, they're looking at using this as an example of an economic shock and seeing how that extrapolates and compares, etc., historically with other data sets out there.

So, I've gone through an awful lot quite quickly.
I appreciate that.

However, hopefully, the point of this is to give you real food for thought and just get the juices flowing as to all the different applications and where this might sit within your field of expertise and all sorts of interesting research objectives that you can think about in order to obtain our data.

And, as I said at the beginning,
I look forward to seeing the research and the outcomes of that research as well.

Thank you very much.

Thanks very much for that presentation.

It was really interesting.

We'll take questions at the end.

So, I'll hand over to our third speaker, Qunshan Zhao, who will tell us
how he plans to use the Adzuna data.

Okay.

Share my screen.

Can you hear me okay?

- Yep.
- Alright.

Thank you. Okay.

Thank you, Heather, for the introduction.
So, as she said, my name is Qunshan Zhao and I'm a lecturer in urban analytics and based in the Urban Big Data Centre, and urban studies in the University of Glasgow. And today, what I want to present is one of the collaborative efforts I'm thinking of using this new acquisition of Adzuna data set for to actually look at understanding the impact of AI
and smart technologies on jobs

using online job vacancy data.

So, this is a collaborative effort with my colleague from Leeds, and her name is Jiaqi Ge.

Okay.

So, just to give you some background today.

First thing is that everybody is probably aware that more and more robotics, AI, actually, are starting to
substitute for regular tedious human jobs.

And from the current understanding of ONS is basically...

I mean, not only for ONS.

But, if you look at a general job landscape,
	right now we have 140 million full-time workers worldwide

and it's about 1.5 million jobs in the UK.

But when you’re talking about AI,
on the right-hand side,

00:28:30,426 --> 00:28:35,815
the thing is actually saying that after you have all these robotic technologies,

00:28:35,915 --> 00:28:38,004
you have the AIs, you have the automations,

00:28:38,223 --> 00:28:42,617
you can only create about 20% of previous jobs

00:28:42,717 --> 00:28:46,166
and the remaining 80% will be replaced by those kinds of machines.

00:28:46,754 --> 00:28:50,321
And of that 80%, women, young people,

00:28:50,421 --> 00:28:52,800
and part-time workers, and low-income groups are

00:28:52,900 --> 00:28:55,004
more likely to lose their jobs to AI.
So, in this case, to understand what is a job replacement by AI and what is actually job creation by AI is very important.

From the current literature review, most of the research focuses on AI as a job replacement and there's a famous paper, 2017 Frey and Osborne, talking about the likelihood of the automation of different jobs based on three job characteristics.
And they have this categorisation basically saying

if you have a job which has a high social intelligence,

full of creativity,

and really needs a lot of perception and manipulation,

that's something that will not easily be replaced by AI.

But there are some criticisms around their classifications

which say that they really base them on occupations

rather than job tasks,
but actually, AI is really replacing a job task rather than talking about what kind of occupations you work on. And there's a review of the research that recently extends the Frey and Osborne research by OECD or ONS in the UK. But what makes us, as a team, more interested is about the AI and job creation.
So, there’s really limited research looking at the job creation of AI

and also, because of a lack of data.

So, right now, I mean, before all of this online job listing,

it was actually difficult.

You had to wait to get that kind of job information.

You didn’t really get this real time large volume data set.

So, in the current literature,

few have attempted to classify and quantify job creation by AI
using empirical data.

So, the lack of understanding will sometimes lead us to a non-comprehensive picture of the job market.

And I think a few things we identified...

The first thing is that we overestimate the overall net job loss due to AI.

So, actually, how many jobs we actually lose because of the AI or automation.

But on the other hand,

we may also underestimate the negative impact of job loss to AI
on some groups of people and regions in the UK, or maybe you can say a more global context.

So, this is something that triggered us to think of using the new Adzuna data set to look at this research question.

And for our project...

I mean, our project is still under review so we put together a proposal.

So, today, I will not really go into those kinds of details
but I want to share with you some of our initial ideas.

I mean, the project aims,

we really want to use natural language processing

or, say, some of the machine learning deep learning methods to classify

and quantify jobs created in the AI sector first.

So, we have already the long-term data sets,

including Adzuna and including some other data sources
that I will talk about later.

And we are trying to...

Because all of my collaborators and myself all have geography backgrounds,

we are really interested in the spatial and temporal patterns of

the job location

and the impact of AI on job polarisation

and also, the industrial agglomeration.

So, in terms of right now,
certainly, we are still in the COVID-19 period of time

and we're probably in the middle of, the start of recovery.

And we are, if you look at the economic chain
to look at a long-term influence,

we don't really know.

But using this Adzuna data set's real time,

quick response, quickly collected data set has

really helped us to understand the impact of COVID-19
on AI-related jobs

using near real-time online job advertisement data.

So, as I mentioned, the data sets.

So, the Adzuna data obtained by UBDC from 2017 to 2022.

And we also have a Burning Glass data set for the previous five years.

There's also the ONS AI job replacement data set

that would use the official statistics data we can use.

By combining those data sets,
we can roughly look at the past 10 years and see what we can find out in terms of AI,

AI’s influence to the different jobs in the UK.

I’m not talking about methods but I want to talk about some of the impacts we want to generate from this project.

The first thing I want to say is that we really want to provide a more data science focus and critical new evidence to inform UK government policies to maximise job creation by AI,
which is kind of the next level of jobs,

so that it creates more high-tech jobs

and you can attract more top talents to the UK.

And it's really important to maintain

the UK's leadership in AI and innovation,

something like the Alan Turing Institute

which we all know about.

And the second thing is that we really want to understand inequality.
So, actually, COVID-19 exacerbates the inequality in the UK.

So, if we are talking about identifying and supporting those neighbourhoods, after we find out if they have been hit heavily by COVID,

how to help those regions recover from this global pandemic.

That is something we can find out from this kind of job analysis and also, the spatial and temporal analysis.

The final thing is that we want to talk with
the policymakers and local authorities
to help level up the developments in the AI industry
across regions in the UK.
So, potential project ideas.
If we do this kind of webinar,
we have opportunities to share this data set
for non-commercial academic research, right?
So, the potential project ideas.

I think, certainly, the Adzuna data would be a surely useful data set for analysing the labour market,

which is the direct understanding from a data set.

Also, in terms of economic geography,

that's a really variable data set.

And, I mean, for the COVID-19 recovery and renewal,
that's also an important data set

if you can couple it with the urban mobility data set

and with a few other more official statistic data sets.

I mean, as I mentioned in my previous slides,

inequality is something, certainly, that can be looked at

through this job listing data set

because if you keep seeing the jobs number keep going down,

that will probably reflect something you really want to be careful about,
to think about how to really help a region grow again.

So, there can be many, many other things you can do,

especially where you have different academic backgrounds

and also where you have different data sets.

You can couple these new forms of data together.

So, we certainly welcome new ideas

and we're looking forward to seeing your proposals of

how you want to use this data set.
So, that's pretty much it for today for my quick research proposal overview.

I'm happy to take any questions at the end of all the presentations.

Thank you so much.