

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,

7

00:00:22,730 --> 00:00:27,485

Qunshan Zhao, Senior Lecturer in Urban Analytics at UBDC,

8

00:00:27,585 --> 00:00:30,453

Andrew McHugh, Senior Data Science Manager,

9

00:00:30,871 --> 00:00:34,631

and Nadiia Gorash, Data Scientist at UBDC.

10

00:00:35,200 --> 00:00:39,210

I'd like to remind participants that this session is being recorded

11

00:00:39,529 --> 00:00:43,379

and will be uploaded on the Urban Big Data Centre website

12

00:00:43,479 --> 00:00:46,166

in an accessible format at a later date.

13

00:00:46,472 --> 00:00:48,680

for those who could not attend the session.

14

00:00:48,921 --> 00:00:52,830

Cameras will be turned off and microphones muted to aid privacy

15

00:00:52,930 --> 00:00:55,388

and also for bandwidth reasons.

16

00:00:56,091 --> 00:00:59,445

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

17

00:01:00,345 --> 00:01:03,166

Feel free to introduce yourself in the chat box

18

00:01:03,266 --> 00:01:06,791

but please don't include any personal information

19

00:01:06,891 --> 00:01:09,337

such as phone numbers or emails.

20

00:01:10,157 --> 00:01:12,761

Please also use the chat to ask questions.

21

00:01:12,861 --> 00:01:16,709

These will be collated and responses will be provided

22

00:01:16,809 --> 00:01:19,026

in the Q&A session at the end.

23

00:01:19,665 --> 00:01:22,512

During the webinar, there will be three presentations

24

00:01:22,699 --> 00:01:24,926

followed by an opportunity for discussion

25

00:01:25,026 --> 00:01:28,572

and a question and answer session where you can submit your questions

26

00:01:28,672 --> 00:01:30,170

using the chat function.

27

00:01:30,887 --> 00:01:33,115

Now I'll move on to the first session,

28

00:01:33,275 --> 00:01:40,237

which provides an introduction to Adzuna data and how to apply for access.

29

00:01:41,728 --> 00:01:42,744

Okay.

30

00:01:43,262 --> 00:01:47,438

To start, I'll provide some background information about UBDC.

31

00:01:47,876 --> 00:01:52,249

The Urban Big Data Centre was established in 2014

32

00:01:52,479 --> 00:01:57,789

and it is jointly funded by ESRC and the University of Glasgow.

33

00:01:58,678 --> 00:02:03,407

The aim of UBDC is to promote innovative research methods

34

00:02:03,755 --> 00:02:07,084

and the use of big data to improve social, economic,

35

00:02:07,184 --> 00:02:09,793

and environmental wellbeing in cities.

36

00:02:10,719 --> 00:02:16,907

UBDC's research centre specialises in areas of transport and mobility,

37

00:02:17,195 --> 00:02:20,495

housing and environment,

education and skills,

38

00:02:20,595 --> 00:02:23,493

urban governance, and urban analytics.

39

00:02:24,162 --> 00:02:29,399

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

40

00:02:29,499 --> 00:02:31,247

our service users.

41

00:02:32,355 --> 00:02:38,761

The objective of UBDC's data service is to build an outstanding data collection

42

00:02:38,861 --> 00:02:43,556

for non-commercial research by UK academics which support

43

00:02:43,656 --> 00:02:48,134

high quality, impactful research on UK cities.

44

00:02:49,133 --> 00:02:53,812

So, Adzuna data is one of our recent additions to the collection.

45

00:02:56,431 --> 00:02:59,500

Adzuna searches thousands of websites

46

00:02:59,868 --> 00:03:02,689

and brings together millions of job advertisements

47

00:03:03,058 --> 00:03:04,457

on their website.

48

00:03:04,747 --> 00:03:08,735

The Adzuna data provides information about the job market,

49

00:03:08,835 --> 00:03:13,750

job categories, and can be used to analyse vacancies over time.

50

00:03:14,450 --> 00:03:20,778

The data provides snapshots of job ads which were advertised on Adzuna.co.uk

51

00:03:20,878 --> 00:03:23,046

at a particular point in time.

52

00:03:23,496 --> 00:03:28,754

UBDC will provide historical data from 2017,

53

00:03:29,075 --> 00:03:33,743

current data, and also future data up to Spring 2022.

54

00:03:38,783 --> 00:03:45,041

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

55

00:03:46,831 --> 00:03:48,865

It includes everything you would expect to see

56

00:03:48,965 --> 00:03:52,783

if you were going to a website to view a job vacancy.

57

00:03:53,221 --> 00:03:56,112

So, it has the ID of the job vacancy,

58

00:03:56,279 --> 00:04:00,150

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

59

00:04:00,499 --> 00:04:04,550

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

60

00:04:05,509 --> 00:04:06,897

the company name,

61

00:04:07,025 --> 00:04:09,243

the date the job listing was added,

62

00:04:09,999 --> 00:04:12,070

a detailed description of the job.

63

00:04:12,817 --> 00:04:14,647

It has location information,

64

00:04:14,747 --> 00:04:18,496

so it includes latitude and longitude co-ordinates

65

00:04:18,800 --> 00:04:20,179

and the region it was in.

66

00:04:20,279 --> 00:04:23,946

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

67

00:04:24,821 --> 00:04:26,960

The minimum and maximum salaries.

68

00:04:27,927 --> 00:04:30,367

There's also a field for predictive salaries.

69

00:04:30,900 --> 00:04:34,188

That's for listings without salary information.

70

00:04:35,045 --> 00:04:36,696

The currency of the salary.

71

00:04:36,796 --> 00:04:39,834

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

72

00:04:40,583 --> 00:04:43,602

The job title and the link to the advert.

73

00:04:45,651 --> 00:04:49,329

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

74

00:04:50,099 --> 00:04:53,648

so UBDC has a data catalogue on their website

75

00:04:53,748 --> 00:04:57,816

with information about all the available data sets.

76

00:04:58,992 --> 00:05:02,160

This is a screenshot of the entry for Adzuna data

77

00:05:02,478 --> 00:05:04,258

and by clicking on read more,

78

00:05:04,626 --> 00:05:08,823

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

79

00:05:09,081 --> 00:05:12,971

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

80

00:05:18,984 --> 00:05:21,516

So, who can apply to use the data?

81

00:05:22,296 --> 00:05:25,473

So, UBDC has a limited number of credits available

82

00:05:25,573 --> 00:05:29,448

to sub-license data to PhD level students

83

00:05:29,616 --> 00:05:32,512

or staff at UK universities.

84

00:05:35,052 --> 00:05:39,824

The data is to be used for non-commercial academic research.

85

00:05:41,320 --> 00:05:44,205

UBDC licenses are for academic research

86

00:05:44,305 --> 00:05:48,680

but if anyone on the call is from a government organisation

87

00:05:48,780 --> 00:05:51,054

and would like to access the Adzuna data,

88

00:05:51,154 --> 00:05:56,099

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

89

00:05:58,100 --> 00:06:01,325

Projects should generally be up to two years duration,

90

00:06:01,772 --> 00:06:05,299

although longer qualifying projects may be supported.

91

00:06:06,320 --> 00:06:09,974

Academic staff can use the data for research projects

92

00:06:10,074 --> 00:06:12,079

but it's not to be used for teaching use.

93

00:06:13,268 --> 00:06:15,954

Applicants are expected to be ready to use the data

94

00:06:16,054 --> 00:06:18,107

straight away once they receive it.

95

00:06:19,164 --> 00:06:22,159

The data can be used for various purposes.

96

00:06:22,259 --> 00:06:24,418

So, I've included a few here.

97

00:06:25,242 --> 00:06:27,097

Analysis of new vacancies.

98

00:06:27,312 --> 00:06:29,037

Overall vacancy levels.

99

00:06:29,137 --> 00:06:30,993

Vacancies by particular region.

100

00:06:31,338 --> 00:06:33,973

Or by a particular job type.

101

00:06:37,595 --> 00:06:41,623

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

102

00:06:42,038 --> 00:06:45,629

but UBDC will soon be launching a second call.

103

00:06:46,659 --> 00:06:48,806

And so, once it's launched,

104

00:06:49,382 --> 00:06:51,017

you'll find everything you need to know

105

00:06:51,117 --> 00:06:54,003

on the call for expressions of interest page.

106

00:06:56,643 --> 00:06:58,679

And at the end of this page,

107

00:06:58,779 --> 00:07:01,645

there'll be an online form that you can fill in

108

00:07:02,312 --> 00:07:07,337

with a project summary form, which is just brief details about

109

00:07:07,437 --> 00:07:10,722

what your project involves and who's going to be working on the project.

110

00:07:14,316 --> 00:07:17,293

This gets checked over, just to make sure you're eligible,

111

00:07:17,393 --> 00:07:21,722

that it's coming from an academic organisation,

112

00:07:22,071 --> 00:07:24,602

that the timescales are relevant.

113

00:07:25,791 --> 00:07:28,460

And then, assuming that's all okay,

114

00:07:29,417 --> 00:07:33,306

we send out a more detailed Adzuna proposal form.

115

00:07:33,972 --> 00:07:37,722

This asks you in a lot more detail about your proposal.

116

00:07:39,692 --> 00:07:45,684

So, what we're looking at in the proposal form is

117

00:07:46,309 --> 00:07:47,997

these criteria here.

118

00:07:48,097 --> 00:07:50,465

So, you'll find these sections in the form

119

00:07:50,565 --> 00:07:54,153

where you get opportunities to write about each of these things.

120

00:07:54,382 --> 00:07:56,526

So, for example, privacy and ethics.

121

00:07:56,784 --> 00:07:58,115

We'll be looking at,

122

00:07:58,600 --> 00:08:01,811

does the project require ethical approval?

123

00:08:02,940 --> 00:08:04,321

Feasibility.

124

00:08:04,421 --> 00:08:06,479

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

125

00:08:06,579 --> 00:08:10,258

can the project be completed within the timescales

126

00:08:10,358 --> 00:08:13,955

or using the methods that were outlined in the proposal?

127

00:08:15,229 --> 00:08:16,670

The scientific contribution.

128

00:08:16,770 --> 00:08:19,353

This could be development of a methodology

129

00:08:19,895 --> 00:08:23,267

or development of new data products or resources.

130

00:08:23,747 --> 00:08:26,628

It could include publication in an esteemed journal

131

00:08:26,728 --> 00:08:28,899

or a presentation at a conference.

132

00:08:31,552 --> 00:08:34,104

Potential for public benefit impact.

133

00:08:34,204 --> 00:08:35,876

So, the panel will be looking at

134

00:08:35,976 --> 00:08:40,188

public benefit impact beyond academic research.

135

00:08:40,436 --> 00:08:43,308

So, the things you could include here are

136

00:08:43,408 --> 00:08:45,990

information about planned knowledge exchange

137

00:08:46,351 --> 00:08:48,911

or other communications activities.

138

00:08:49,710 --> 00:08:53,201

If you have one, you could include a pathways to impact statement,

139

00:08:53,301 --> 00:08:55,833

which is where you outline what you would do

140

00:08:56,655 --> 00:08:59,452

to make beneficiaries aware of the research

141

00:08:59,552 --> 00:09:01,514

so that impact could be achieved.

142

00:09:04,913 --> 00:09:08,285

Outcomes. So, we respond to all applicants

143

00:09:08,385 --> 00:09:11,325

and provide their decision.

144

00:09:13,318 --> 00:09:17,687

Successful applicants will be emailed licensing agreements

145

00:09:18,108 --> 00:09:21,690

and we'll share the data on completion of these agreements.

146

00:09:26,480 --> 00:09:31,542

As a free service funded by the Economic and Social Research Council,

147

00:09:31,983 --> 00:09:36,674

we rely on end user feedback to provide evidence of the value of what we do.

148

00:09:38,005 --> 00:09:42,167

It's a condition of the licences that data users inform UBDC of

149

00:09:42,267 --> 00:09:44,078

any publications they produce

150

00:09:44,359 --> 00:09:47,249

that were a result of having the data,

151

00:09:47,630 --> 00:09:51,021

and that researchers also include

152

00:09:51,121 --> 00:09:56,237

a citation to acknowledge the data owner, Adzuna, and UBDC.

153

00:09:58,277 --> 00:10:02,138

So, I'll now hand over to Scott Sweden from Adzuna

154

00:10:02,238 --> 00:10:03,946

who will present the next session

155

00:10:04,046 --> 00:10:06,740

about the background of the Adzuna company

156

00:10:06,840 --> 00:10:09,805

and the reasons for the data creation.

157

00:10:18,076 --> 00:10:19,175

Right.

158

00:10:19,295 --> 00:10:23,304

You should see my title page.

159

00:10:23,474 --> 00:10:25,392

If you're not then that is a problem.

160

00:10:27,342 --> 00:10:28,786

So, thanks very much, Heather.

161

00:10:29,656 --> 00:10:31,894

My hope and expectation really is that this is going to

162

00:10:31,994 --> 00:10:33,803

prove, obviously, a great partnership.

163

00:10:33,903 --> 00:10:35,561

So, thank you to the UBDC.

164

00:10:36,971 --> 00:10:41,259

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

165

00:10:42,191 --> 00:10:44,440

I'm going to be very excited to learn about the research

166

00:10:44,500 --> 00:10:47,499

and importantly, I suppose, the outcomes of your endeavours

167

00:10:47,599 --> 00:10:48,787

in the year to follow.

168

00:10:49,467 --> 00:10:52,357

Very briefly, as it says on the tin here,

169

00:10:52,457 --> 00:10:55,224

I'm the Head of Data Sales, which is fairly self-explanatory.

170

00:10:56,205 --> 00:10:58,522

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

171

00:10:58,763 --> 00:11:01,020

I'm going to assume that none of you have

172

00:11:01,120 --> 00:11:02,738

familiarity with the data set.

173

00:11:02,838 --> 00:11:05,505

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

174

00:11:06,253 --> 00:11:07,898

just please bear with me and I apologise.

175

00:11:08,130 --> 00:11:09,846

Or go and make a coffee.

176

00:11:10,100 --> 00:11:12,154

I'm going to talk about where the data comes from,

177

00:11:12,254 --> 00:11:13,402

some stats of the coverage,

178

00:11:13,502 --> 00:11:15,281

and what you can expect to extract,

179

00:11:15,848 --> 00:11:18,596

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

180

00:11:19,185 --> 00:11:21,634

public sector applications of our data at the moment.

181

00:11:22,934 --> 00:11:26,571

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

182

00:11:26,671 --> 00:11:30,396

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

183

00:11:30,496 --> 00:11:32,094

for the last 18 months,

184

00:11:32,693 --> 00:11:34,642

which was through, essentially, organic demand.

185

00:11:34,742 --> 00:11:36,871

It's come from a very diverse group of clients.

186

00:11:36,971 --> 00:11:39,838

It's from venture capital, asset management,

187

00:11:39,938 --> 00:11:42,534

quant funds, to corporate HR.

188

00:11:42,764 --> 00:11:44,513

And of course, as you can imagine, those in

189

00:11:44,613 --> 00:11:46,980

public policy and labour market economists.

190

00:11:47,681 --> 00:11:50,309

Adzuna was founded 10 years ago now

191

00:11:50,409 --> 00:11:55,798

by Doug Monro and Andrew Hunter, ex-eBayers and Gumtree.

192

00:11:56,386 --> 00:11:59,241

And their mission was to make it easier for millions of people

193

00:11:59,341 --> 00:12:00,540

to find better jobs.

194

00:12:00,640 --> 00:12:03,669

And that's by listing every job everywhere,

195

00:12:03,853 --> 00:12:07,134

as the slogan goes.

196

00:12:07,803 --> 00:12:10,793

We do that by leveraging our proprietary technology

197

00:12:11,209 --> 00:12:13,558

and our algos to achieve that goal.

198

00:12:14,300 --> 00:12:16,800

So, for any of those that are in business schools

199

00:12:16,900 --> 00:12:18,603

that might find this slightly interesting,

200

00:12:18,703 --> 00:12:22,896

our commercial model works by driving click traffic to job boards.

201

00:12:23,118 --> 00:12:24,734

So, there's an important distinction,

202

00:12:25,293 --> 00:12:26,413

and I'll get to that later,

203

00:12:26,513 --> 00:12:29,952

but they are our clients and we aggregate all the jobs from them

204

00:12:30,052 --> 00:12:33,606

in the marketplace into one particular place.

205

00:12:34,116 --> 00:12:35,894

And then through this aggregation model,

206

00:12:36,202 --> 00:12:38,763

it means we now have 20 million monthly visitors

207

00:12:38,911 --> 00:12:42,278

across the 16 different countries we operate in

208

00:12:42,886 --> 00:12:45,184

for the data set you'll hopefully get access to.

209

00:12:45,634 --> 00:12:48,902

We treat the UK as one country to us, one market,

210

00:12:49,002 --> 00:12:50,620

so that essentially means England,

211

00:12:50,720 --> 00:12:54,208

Scotland, Wales, and Northern Ireland are

212

00:12:54,308 --> 00:12:56,092

where you'll see data.

213

00:12:56,652 --> 00:13:01,601

And as I mentioned, we've developed leading matching technology

214

00:13:02,463 --> 00:13:05,351

and we also provide that to central governments.

215

00:13:05,451 --> 00:13:08,728

So, we won the tender for their "Find A Job" service

216

00:13:09,508 --> 00:13:11,694

and that's something we manage to this day

217

00:13:11,794 --> 00:13:14,051

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

218

00:13:15,348 --> 00:13:17,916

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

219

00:13:18,016 --> 00:13:20,373

what makes us the number one job search engine

220

00:13:20,473 --> 00:13:21,678

in the countries we operate?

221

00:13:21,778 --> 00:13:27,493

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

222

00:13:27,593 --> 00:13:29,601

obviously going to be very helpful for our cause.

223

00:13:30,301 --> 00:13:34,808

And that should be great news for the analysts amongst you

224

00:13:35,179 --> 00:13:37,156

or perhaps the economists amongst you as well.

225

00:13:37,256 --> 00:13:38,363

So, how do we do this?

226

00:13:38,463 --> 00:13:42,241

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

227

00:13:42,341 --> 00:13:43,839

The first is our job boards.

228

00:13:44,100 --> 00:13:45,854

Some of these will be your household names.

229

00:13:45,954 --> 00:13:49,151

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

230

00:13:49,251 --> 00:13:52,241

And they will take jobs from their clients

231

00:13:52,341 --> 00:13:54,491

who will pay them a fee to list their job.

232

00:13:55,230 --> 00:13:57,440

Recruiters. So, on the right-hand column,

233

00:13:57,540 --> 00:13:59,704

you'll see recruiters. So, Michael Page.

234

00:14:00,074 --> 00:14:02,323

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

235

00:14:02,423 --> 00:14:04,781

So, from finding a job to application

236

00:14:05,428 --> 00:14:07,617

to negotiating a contract if they're any good.

237

00:14:08,464 --> 00:14:10,594

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

238

00:14:10,694 --> 00:14:13,463

so they'll host all their clients' jobs on there.

239

00:14:14,261 --> 00:14:16,530

And then lastly, there's the middle section,

240

00:14:16,630 --> 00:14:20,687

which is employers directly.

241

00:14:20,787 --> 00:14:22,504

So, we call that direct business Adzuna.

242

00:14:22,604 --> 00:14:27,600

So, we will have Virgin, for example, come and host their jobs directly

243

00:14:27,700 --> 00:14:31,899

because they have access to two to three million job seekers a month.

244

00:14:31,999 --> 00:14:33,887

So, the best place to find all the jobs

245

00:14:33,987 --> 00:14:35,655

so they'll want to come to us directly.

246

00:14:36,306 --> 00:14:38,364

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

00:14:59,919 --> 00:15:02,803

hosting all their clients' jobs as well.

255

00:15:02,903 --> 00:15:06,889

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

256

00:15:08,522 --> 00:15:10,594

So, what does that mean for our job representation?

257

00:15:10,694 --> 00:15:12,386

Which is probably more interesting to you.

258

00:15:12,728 --> 00:15:15,642

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

259

00:15:15,742 --> 00:15:20,119

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

260

00:15:20,532 --> 00:15:23,479

The USA, it's about 240 million.

261

00:15:23,579 --> 00:15:26,236

240 million and growing. In Russia

262

00:15:26,796 --> 00:15:28,742

it's a good market for us, it's about half that.

263

00:15:29,653 --> 00:15:31,671

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

264

00:15:31,771 --> 00:15:33,348

the type of data set, more importantly,

265

00:15:33,448 --> 00:15:35,285

you'll be accessing, is,

266

00:15:35,385 --> 00:15:40,974

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

267

00:15:41,074 --> 00:15:42,962

So, you will be getting weekly snapshots.

268

00:15:44,061 --> 00:15:47,950

So, week to week, what were the live jobs on the marketplace

269

00:15:48,050 --> 00:15:49,829

at that point in time.

270

00:15:49,918 --> 00:15:53,606

And so, what you're actually getting is the same type of access,

271

00:15:53,953 --> 00:15:58,041

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

272

00:15:58,141 --> 00:16:00,489

and I'll refer to that at the end.

273

00:16:01,633 --> 00:16:04,993

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

274

00:16:05,443 --> 00:16:08,165

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

275

00:16:08,363 --> 00:16:12,043

In the UK, that hopefully you'll get access to,

276

00:16:12,169 --> 00:16:15,765

we estimate it's about 95% plus representative of

277

00:16:15,865 --> 00:16:17,808

the online job market space.

278

00:16:18,279 --> 00:16:21,216

We have similar coverage in other countries, like Germany,

279

00:16:21,316 --> 00:16:22,693

France, the Netherlands, etc.

280

00:16:23,677 --> 00:16:27,367

And we know this by simply going on our clients' websites,

281

00:16:28,127 --> 00:16:32,683

but also, we start to make comparisons with national statistic databases

282

00:16:32,783 --> 00:16:35,586

in the country. So, obviously, in the UK it's the ONS.

283

00:16:39,769 --> 00:16:41,997

This comes from our Head of Data Science.

284

00:16:42,097 --> 00:16:44,984

He sees it as, as we all do really,

285

00:16:45,084 --> 00:16:47,821

a hugely valuable goldmine of data

286

00:16:47,921 --> 00:16:50,237

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

287

00:16:50,598 --> 00:16:53,184

So, you know, it's a rich source.

288

00:16:53,706 --> 00:16:56,763

You can look at hiring trends.

289

00:16:56,863 --> 00:17:00,420

You can look at, if you're looking at human capital,

290

00:17:00,520 --> 00:17:03,088

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

291

00:17:03,186 --> 00:17:05,694

Obviously, as an economist, you can look at labour demands

292

00:17:05,794 --> 00:17:07,183

or industry sector growth.

293

00:17:07,956 --> 00:17:12,330

If you're doing public policy work, you can look at skills insights

294

00:17:13,000 --> 00:17:15,058

and you can look at changing work patterns, again,

295

00:17:15,596 --> 00:17:17,674

sort of, HR, human capital,

296

00:17:17,920 --> 00:17:21,259

how things are moving through post-COVID.

297

00:17:21,499 --> 00:17:24,448

How are things changing over time? How much has it sped up?

298

00:17:24,548 --> 00:17:27,226

And we can cut this by all sorts of other fields,

299

00:17:27,326 --> 00:17:29,126

like, you know, obviously, granular location,

300

00:17:29,226 --> 00:17:32,350

we can do it by industry, you can look at it at a granular occupational level,

301

00:17:32,400 --> 00:17:33,819

and so on and so forth.

302

00:17:33,938 --> 00:17:35,596

There's a huge wealth of data.

303

00:17:35,696 --> 00:17:37,165

And one client recently said to me,

304

00:17:37,265 --> 00:17:38,643

an economist said to me, actually,

305

00:17:39,101 --> 00:17:40,593

you know, "Data is more like air."

306

00:17:41,312 --> 00:17:43,171

I thought that might have been a bit extreme

307

00:17:43,271 --> 00:17:45,790

but maybe the data scientists amongst you will think so too.

308

00:17:47,989 --> 00:17:53,780

However, it doesn't come to us all nicely extracted.

309

00:17:54,091 --> 00:17:55,942

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

310

00:17:56,434 --> 00:18:00,835

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

311

00:18:01,487 --> 00:18:05,904

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

312

00:18:06,004 --> 00:18:07,166

I don't advise that.

313

00:18:08,007 --> 00:18:09,678

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

314

00:18:09,778 --> 00:18:12,959

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

315

00:18:13,810 --> 00:18:18,435

you can see that they're relating to ballet in the job description.

316

00:18:18,730 --> 00:18:22,071

So, it's not great, perhaps, if we're an economist.

317

00:18:22,171 --> 00:18:24,922

However, of course, these are written by human beings

318

00:18:25,022 --> 00:18:27,254

and we are prone to error

319

00:18:28,039 --> 00:18:29,150

and flights of fancy.

320

00:18:29,800 --> 00:18:32,273

But from a social research perspective,

321

00:18:32,905 --> 00:18:35,036

this is a great thing because this is a lot of value

322

00:18:35,136 --> 00:18:38,674

in these types of errors and what people are writing.

323

00:18:38,912 --> 00:18:40,591

I'll show you why next.

324

00:18:42,231 --> 00:18:44,587

So, if we just go into a bit more detail.

325

00:18:44,687 --> 00:18:47,731

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

326

00:18:48,360 --> 00:18:50,178

there is a huge amount of data.

327

00:18:50,278 --> 00:18:54,186

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

328

00:18:54,286 --> 00:18:57,829

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

329

00:18:58,300 --> 00:19:01,277

So, for example, seniority.

330

00:19:01,664 --> 00:19:03,451

So, just looking at seniority,

331

00:19:03,551 --> 00:19:06,148

you know, questions that are being asked of our data, you know.

332

00:19:06,248 --> 00:19:07,805

Is the workforce getting older,

333

00:19:08,819 --> 00:19:10,887

you know, due to remote working?

334

00:19:11,400 --> 00:19:13,148

Therefore, less succession planning.

335

00:19:13,248 --> 00:19:15,215

Therefore, people are staying in their jobs longer.

336

00:19:15,686 --> 00:19:17,675

This could be true of enterprise businesses

337

00:19:18,939 --> 00:19:21,368

and when we look at enterprise businesses, company hiring.

338

00:19:21,468 --> 00:19:26,225

So, in 2020, we tagged about 8000 known companies.

339

00:19:26,325 --> 00:19:29,382

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.

347

00:19:47,841 --> 00:19:51,200

How are companies now enticing the next generation of employees?

348

00:19:51,300 --> 00:19:52,300

What are they doing?

349

00:19:52,527 --> 00:19:55,842

Qualifications, you could extract all of the qualifications requirements.

350

00:19:56,120 --> 00:19:58,108

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

351

00:19:59,079 --> 00:20:01,326

What are the academic credentials

352

00:20:01,426 --> 00:20:03,895

for junior roles? For example, for graduate roles.

353

00:20:04,004 --> 00:20:05,828

Is that demand now greater than ever?

354

00:20:06,448 --> 00:20:09,036

What occupations does this apply to?

355

00:20:10,023 --> 00:20:14,379

And my last point on this, this text parsing area, is

356

00:20:14,479 --> 00:20:16,873

the future of work, inverted commas.

357

00:20:17,248 --> 00:20:18,278

Huge topic.

358

00:20:18,439 --> 00:20:20,388

McKinsey, 800 million jobs at risk

359

00:20:20,488 --> 00:20:23,127

from AI and digitalisation of the economy.

360

00:20:24,027 --> 00:20:26,704

Which occupations are most defensible

361

00:20:27,476 --> 00:20:29,454

based on hiring and firing rates?

362

00:20:30,096 --> 00:20:31,285

What are we seeing out there?

363

00:20:32,494 --> 00:20:34,713

Linguistics analysis. So, the second part.

364

00:20:35,573 --> 00:20:37,432

Diversity and inclusion.

365

00:20:38,019 --> 00:20:40,388

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

366

00:20:40,482 --> 00:20:43,208

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

367

00:20:43,837 --> 00:20:45,306

looking at sentiment analysis,

368

00:20:45,406 --> 00:20:50,095

understanding gender and cultural bias in the language is

369

00:20:50,195 --> 00:20:52,085

something that is being heavily looked at

370

00:20:52,271 --> 00:20:54,884

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.

378

00:21:19,570 --> 00:21:22,599

But, of course, there are a lot of other things you can do

379

00:21:22,977 --> 00:21:24,023

with vector embeddings,

380

00:21:25,095 --> 00:21:28,433

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

381

00:21:31,733 --> 00:21:33,123

So, finally,

382

00:21:34,040 --> 00:21:35,748

just a quick visual representation.

383

00:21:36,128 --> 00:21:40,719

You can segment our data and aggregate it up to look at occupations

384

00:21:40,819 --> 00:21:42,978

and salaries and vacancy counts.

385

00:21:43,078 --> 00:21:45,365

You can translate that into interesting...

386

00:21:45,547 --> 00:21:47,714

Well, we've just used a heat map for this example.

387

00:21:47,814 --> 00:21:49,102

There are lots of other ways.

388

00:21:49,381 --> 00:21:51,638

And then, of course, this happens to be looking at

389

00:21:51,738 --> 00:21:53,736

profiling of a particular occupation

390

00:21:53,836 --> 00:21:58,274

and it's an extraction of skills in the workplace.

391

00:21:59,913 --> 00:22:01,643

So, just to summarise

392

00:22:02,236 --> 00:22:04,534

and characterise the things I've said.

393

00:22:04,634 --> 00:22:07,402

So, it's our opinion

394

00:22:08,794 --> 00:22:11,531

that you won't find a more representative sample

395

00:22:11,631 --> 00:22:12,998

to work with in the UK.

396

00:22:13,979 --> 00:22:17,398

Our data science team has already done a lot of the heavy lifting

397

00:22:17,498 --> 00:22:19,989

in terms of cleaning and deduping.

398

00:22:21,027 --> 00:22:25,042

And the application set can be applied to many different,

399

00:22:25,460 --> 00:22:27,468

from your perspective, I suppose, faculties.

400

00:22:27,568 --> 00:22:29,627

So, data science and technology.

401

00:22:29,727 --> 00:22:33,987

We have a lot of people from technical universities

402

00:22:34,246 --> 00:22:35,406

accessing our data.

403

00:22:35,974 --> 00:22:37,493

Finance and trading.

404

00:22:38,583 --> 00:22:41,955

Human capital. Looking at salaries and perks.

405

00:22:42,563 --> 00:22:44,932

To, of course, the obvious one is

406

00:22:45,032 --> 00:22:47,551

labour market and public policy research.

407

00:22:48,510 --> 00:22:52,869

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

408

00:22:54,219 --> 00:22:56,206

So, finally,

409

00:22:56,618 --> 00:22:57,867

I'm just going to end on

410

00:22:57,967 --> 00:23:01,975

some public sector applications of our data.

411

00:23:02,766 --> 00:23:05,546

And therefore, to really help you appreciate the importance of

412

00:23:05,746 --> 00:23:07,316

what this data is used for.

413

00:23:08,251 --> 00:23:11,851

And, obviously, the academic research essentially underpins a lot of

414

00:23:11,951 --> 00:23:14,535

what goes on out there in the public and the private sector.

415

00:23:15,403 --> 00:23:18,072

So, you'll know, of course, the ONS

416

00:23:18,172 --> 00:23:20,111

and some of you, the economists amongst you,

417

00:23:20,211 --> 00:23:21,952

might know the Labour Workforce Survey

418

00:23:22,052 --> 00:23:24,747

that looks at 5000-6000 companies.

419

00:23:25,383 --> 00:23:27,164

It looks at hiring and firing trends,

420

00:23:27,390 --> 00:23:31,350

attitudinal and behavioural characteristics of this as well.

421

00:23:32,200 --> 00:23:34,459

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

422

00:23:34,559 --> 00:23:38,718

the ONS are augmenting their surveys,

423

00:23:38,818 --> 00:23:41,156

which are quarterly, with our real time data

424

00:23:41,384 --> 00:23:44,454

to build indices on the economic recovery.

425

00:23:45,593 --> 00:23:47,107

Now, just some quick stuff.

426

00:23:47,207 --> 00:23:49,250

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

427

00:23:49,350 --> 00:23:53,399

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

428

00:23:53,499 --> 00:23:55,947

pre-pandemic levels, as average job count,

429

00:23:56,100 --> 00:23:57,129

which is good.

430

00:23:57,479 --> 00:23:59,416

They are doing a lot more granular stuff

431

00:23:59,516 --> 00:24:01,778

privately and internally, but interdepartmentally,

432

00:24:01,878 --> 00:24:02,956

that they can't publish,

433

00:24:03,056 --> 00:24:05,404

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

434

00:24:06,443 --> 00:24:08,104

More specifically, what are they doing?

435

00:24:08,204 --> 00:24:10,141

So, some of the things they're doing.

436

00:24:10,241 --> 00:24:13,169

They're analysing a relationship between earnings and skills.

437

00:24:14,696 --> 00:24:17,943

They're doing location analysis interdepartmentally.

438

00:24:18,300 --> 00:24:20,897

They're looking at the impact of COVID

439

00:24:21,116 --> 00:24:23,654

on self-employed versus employed.

440

00:24:24,595 --> 00:24:27,574

And they're also about to publish a report with our data

441

00:24:28,182 --> 00:24:31,390

on the impact of COVID on remote working.

442

00:24:32,910 --> 00:24:35,254

So, I'll just park the ONS there.

443

00:24:35,354 --> 00:24:37,752

On the AI, on the OECD side,

444

00:24:38,063 --> 00:24:39,391

on the right-hand side,

445

00:24:40,048 --> 00:24:42,867

the OECD, obviously funded by the EU,

446

00:24:42,967 --> 00:24:47,543

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

447

00:24:48,250 --> 00:24:51,089

Their AI policy team is looking at

448

00:24:51,320 --> 00:24:55,975

the make up of AI and IT across various occupations.

449

00:24:56,292 --> 00:24:59,586

So, a data scientist versus a database engineer

450

00:24:59,686 --> 00:25:02,304

versus those in research and methodology.

451

00:25:02,404 --> 00:25:06,580

and they're looking specifically at what the skills make up is of

452

00:25:06,680 --> 00:25:07,745

those jobs today.

453

00:25:08,073 --> 00:25:12,367

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

454

00:25:12,467 --> 00:25:14,870

Which you can do with our data but, of course, they are also

455

00:25:14,970 --> 00:25:18,209

covering countries, the various other countries that we use.

456

00:25:20,415 --> 00:25:22,304

And finally, something that is not on here.

457

00:25:22,733 --> 00:25:26,555

We have a think tank that is looking more broadly at

458

00:25:26,655 --> 00:25:28,755

skills mismatches across professions.

459

00:25:28,994 --> 00:25:30,521

Where is there the biggest imbalance?

460

00:25:31,108 --> 00:25:33,237

They're mapping vacancy data.

461

00:25:33,337 --> 00:25:34,497

to, sort of, ONS data.

462

00:25:34,597 --> 00:25:38,677

Things like, to furloughs and unemployment.

463

00:25:39,606 --> 00:25:43,115

And they're also looking at things like economic shocks.

464

00:25:43,215 --> 00:25:44,483

So, of course, we've had,

465

00:25:45,997 --> 00:25:48,184

some say, a business siphoning in 12 months

466

00:25:48,284 --> 00:25:50,453

that would normally take five years or 10 years.

467

00:25:50,733 --> 00:25:53,819

So, they're looking at using this as an example of an economic shock

468

00:25:53,919 --> 00:25:57,428

and seeing how that extrapolates and compares, etc., historically

469

00:25:58,248 --> 00:25:59,926

with other data sets out there.

470

00:26:01,267 --> 00:26:04,565

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

471

00:26:04,676 --> 00:26:05,705

I appreciate that.

472

00:26:06,234 --> 00:26:07,584

However, hopefully,

473

00:26:08,153 --> 00:26:10,772

the point of this is to give you real food for thought

474

00:26:10,872 --> 00:26:14,407

and just get the juices flowing as to all the different applications

475

00:26:14,507 --> 00:26:17,205

and where this might sit within your field of expertise

476

00:26:17,305 --> 00:26:20,744

and all sorts of interesting research objectives

477

00:26:20,844 --> 00:26:23,903

that you can think about in order to obtain our data.

478

00:26:24,151 --> 00:26:25,679

And, as I said at the beginning,

479

00:26:25,779 --> 00:26:29,198

I look forward to seeing the research and the outcomes of

480

00:26:29,298 --> 00:26:30,785

that research as well.

481

00:26:31,354 --> 00:26:32,693

Thank you very much.

482

00:26:36,493 --> 00:26:38,650

Thanks very much for that presentation.

483

00:26:38,750 --> 00:26:40,018

It was really interesting.

484

00:26:40,118 --> 00:26:42,156

We'll take questions at the end.

485

00:26:42,956 --> 00:26:46,834

So, I'll hand over to our third speaker,

486

00:26:47,054 --> 00:26:48,735

Qunshan Zhao, who will tell us

487

00:26:48,835 --> 00:26:51,879

how he plans to use the Adzuna data.

488

00:26:58,418 --> 00:26:59,637

Okay.

489

00:26:59,946 --> 00:27:01,426

Share my screen.

490

00:27:01,526 --> 00:27:02,715

Can you hear me okay?

491

00:27:03,343 --> 00:27:05,002

- Yep.

- Alright.

492

00:27:05,102 --> 00:27:06,357

Thank you. Okay.

493

00:27:07,134 --> 00:27:08,854

Thank you, Heather, for the introduction.

494

00:27:09,205 --> 00:27:13,504

So, as she said, my name is Qunshan Zhao

495

00:27:13,604 --> 00:27:15,692

and I'm a lecturer in urban analytics

496

00:27:15,792 --> 00:27:18,334

and based in the Urban Big Data Centre,

497

00:27:18,791 --> 00:27:23,138

and urban studies in the University of Glasgow.

498

00:27:23,260 --> 00:27:25,908

And today, what I want to present is

499

00:27:26,758 --> 00:27:29,766

one of the collaborative efforts

500

00:27:29,866 --> 00:27:34,685

I'm thinking of using this new acquisition of Adzuna data set for

501

00:27:35,605 --> 00:27:40,423

to actually look at understanding the impact of AI

502

00:27:40,523 --> 00:27:42,820

and smart technologies on jobs

503

00:27:42,920 --> 00:27:45,048

using online job vacancy data.

504

00:27:45,199 --> 00:27:49,727

So, this is a collaborative effort with my colleague from Leeds,

505

00:27:49,827 --> 00:27:52,116

and her name is Jiaqi Ge.

506

00:27:52,329 --> 00:27:53,375

Okay.

507

00:27:53,945 --> 00:27:58,194

So, just to give you some background today.

508

00:27:58,463 --> 00:28:01,102

First thing is that everybody is probably aware that

509

00:28:01,202 --> 00:28:04,680

more and more robotics, AI, actually, are starting to

510

00:28:04,780 --> 00:28:09,318

substitute for regular tedious human jobs.

511

00:28:09,694 --> 00:28:13,469

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

512

00:28:13,768 --> 00:28:15,017

I mean, not only for ONS.

513

00:28:15,117 --> 00:28:18,743

But, if you look at a general job landscape,

514

00:28:19,081 --> 00:28:22,379

right now we have 140 million full-time workers worldwide

515

00:28:22,995 --> 00:28:25,623

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

516

00:28:26,220 --> 00:28:28,198

But when you're talking about AI,

517

00:28:28,987 --> 00:28:30,326

on the right-hand side,

518

00:28:30,426 --> 00:28:35,815

the thing is actually saying that after you have all these robotic technologies,

519

00:28:35,915 --> 00:28:38,004

you have the AIs, you have the automations,

520

00:28:38,223 --> 00:28:42,617

you can only create about 20% of previous jobs

521

00:28:42,717 --> 00:28:46,166

and the remaining 80% will be replaced by those kinds of machines.

522

00:28:46,754 --> 00:28:50,321

And of that 80%, women, young people,

523

00:28:50,421 --> 00:28:52,800

and part-time workers, and low-income groups are

524

00:28:52,900 --> 00:28:55,004

more likely to lose their jobs to AI.

525

00:28:55,804 --> 00:28:57,375

So, in this case,

526

00:28:58,072 --> 00:29:03,551

to understand what is a job replacement by AI

527

00:29:03,651 --> 00:29:07,135

and what is actually job creation by AI is very important.

528

00:29:07,909 --> 00:29:09,907

From the current literature review,

529

00:29:10,626 --> 00:29:14,011

most of the research focuses on AI as a job replacement

530

00:29:14,469 --> 00:29:17,336

and there's a famous paper, 2017 Frey and Osborne,

531

00:29:17,436 --> 00:29:21,404

talking about the likelihood of the automation of different jobs

532

00:29:21,504 --> 00:29:23,441

based on three job characteristics.

533

00:29:23,892 --> 00:29:27,920

And they have this categorisation basically saying

534

00:29:28,020 --> 00:29:32,058

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

535

00:29:32,444 --> 00:29:34,132

full of creativity,

536

00:29:34,333 --> 00:29:37,882

and really needs a lot of perception and manipulation,

537

00:29:37,982 --> 00:29:40,810

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

538

00:29:41,057 --> 00:29:44,406

But there are some criticisms around their classifications

539

00:29:44,506 --> 00:29:47,684

which say that they really base them on occupations

540

00:29:47,784 --> 00:29:49,263

rather than job tasks,

541

00:29:50,081 --> 00:29:52,968

but actually, AI is really replacing a job task

542

00:29:53,068 --> 00:29:59,455

rather than talking about what kind of occupations you work on.

543

00:29:59,906 --> 00:30:01,971

And there's a review of the research

544

00:30:02,071 --> 00:30:06,530

that recently extends the Frey and Osborne research

545

00:30:07,688 --> 00:30:11,305

by OECD or ONS in the UK.

546

00:30:12,977 --> 00:30:17,136

But what makes us, as a team, more interested is

547

00:30:17,236 --> 00:30:19,345

about the AI and job creation.

548

00:30:19,795 --> 00:30:26,015

So, there's really limited research looking at the job creation of AI

549

00:30:26,537 --> 00:30:29,077

and also, because of a lack of data.

550

00:30:29,177 --> 00:30:34,864

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

551

00:30:36,303 --> 00:30:37,352

it was actually difficult.

552

00:30:37,452 --> 00:30:40,998

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

553

00:30:41,098 --> 00:30:44,177

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

554

00:30:44,277 --> 00:30:46,185

So, in the current literature,

555

00:30:46,344 --> 00:30:51,283

few have attempted to classify and quantify job creation by AI

556

00:30:51,383 --> 00:30:52,798

using empirical data.

557

00:30:53,020 --> 00:30:56,678

So, the lack of understanding will sometimes lead us to

558

00:31:00,938 --> 00:31:04,107

a non-comprehensive picture of the job market.

559

00:31:04,207 --> 00:31:06,700

And I think a few things we identified...

560

00:31:06,800 --> 00:31:12,068

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

561

00:31:12,266 --> 00:31:18,896

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

562

00:31:19,566 --> 00:31:20,791

But on the other hand,

563

00:31:20,891 --> 00:31:25,739

we may also underestimate the negative impact of job loss to AI

564

00:31:25,839 --> 00:31:28,412

on some groups of people and regions

565

00:31:28,632 --> 00:31:32,938

in the UK, or maybe you can say a more global context.

566

00:31:34,109 --> 00:31:37,618

So, this is something that triggered us to think of using

567

00:31:37,817 --> 00:31:42,004

the new Adzuna data set to look at this research question.

568

00:31:42,454 --> 00:31:43,682

And for our project...

569

00:31:43,782 --> 00:31:45,541

I mean, our project is still under review

570

00:31:45,641 --> 00:31:47,340

so we put together a proposal.

571

00:31:47,440 --> 00:31:50,535

So, today, I will not really go into those kinds of details

572

00:31:50,777 --> 00:31:53,810

but I want to share with you some of our initial ideas.

573

00:31:54,000 --> 00:31:55,109

I mean, the project aims,

574

00:31:55,209 --> 00:31:59,127

we really want to use natural language processing

575

00:31:59,688 --> 00:32:02,938

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

576

00:32:03,038 --> 00:32:06,407

and quantify jobs created in the AI sector first.

577

00:32:07,465 --> 00:32:09,823

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

578

00:32:10,046 --> 00:32:12,805

including Adzuna and including some other data sources

579

00:32:12,905 --> 00:32:14,133

that I will talk about later.

580

00:32:14,502 --> 00:32:16,390

And we are trying to...

581

00:32:16,490 --> 00:32:22,507

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

582

00:32:22,642 --> 00:32:27,011

we are really interested in the spatial and temporal patterns of

583

00:32:27,111 --> 00:32:30,189

the job location

584

00:32:30,438 --> 00:32:34,820

and the impact of AI on job polarisation

585

00:32:34,920 --> 00:32:38,448

and also, the industrial agglomeration.

586

00:32:38,489 --> 00:32:42,128

So, in terms of right now,

587

00:32:42,228 --> 00:32:45,626

certainly, we are still in the COVID-19 period of time

588

00:32:45,726 --> 00:32:50,516

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

589

00:32:50,616 --> 00:32:54,191

And we are, if you look at the economic chain

590

00:32:54,291 --> 00:32:55,730

to look at a long-term influence,

591

00:32:55,830 --> 00:32:57,579

we don't really know.

592

00:32:57,913 --> 00:33:01,717

But using this Adzuna data set's real time,

593

00:33:02,521 --> 00:33:06,451

quick response, quickly collected data set has

594

00:33:06,551 --> 00:33:10,168

really helped us to understand the impact of COVID-19

595

00:33:10,268 --> 00:33:12,026

on AI-related jobs

596

00:33:12,126 --> 00:33:15,254

using near real-time online job advertisement data.

597

00:33:16,900 --> 00:33:19,005

So, as I mentioned, the data sets.

598

00:33:19,105 --> 00:33:26,521

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

599

00:33:26,621 --> 00:33:32,121

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

600

00:33:32,729 --> 00:33:35,683

There's also the ONS AI job replacement data set

601

00:33:35,783 --> 00:33:38,841

that would use the official statistics data we can use.

602

00:33:38,980 --> 00:33:40,345

By combining those data sets,

603

00:33:40,445 --> 00:33:43,444

we can roughly look at the past 10 years

604

00:33:43,919 --> 00:33:48,280

and see what we can find out in terms of AI,

605

00:33:48,839 --> 00:33:52,967

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

606

00:33:55,178 --> 00:34:01,075

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

607

00:34:01,594 --> 00:34:04,891

The first thing I want to say is that we really want to provide

608

00:34:04,991 --> 00:34:08,359

a more data science focus and critical new evidence

609

00:34:08,563 --> 00:34:13,266

to inform UK government policies to maximise job creation by AI,

610

00:34:13,366 --> 00:34:16,055

which is kind of the next level of jobs,

611

00:34:16,964 --> 00:34:22,508

so that it creates more high-tech jobs

612

00:34:22,608 --> 00:34:26,916

and you can attract more top talents to the UK.

613

00:34:27,016 --> 00:34:29,005

And it's really important to maintain

614

00:34:29,105 --> 00:34:31,263

the UK's leadership in AI and innovation,

615

00:34:31,363 --> 00:34:33,262

something like the Alan Turing Institute

616

00:34:33,643 --> 00:34:35,142

which we all know about.

617

00:34:35,382 --> 00:34:38,729

And the second thing is that we really want to understand inequality.

618

00:34:38,898 --> 00:34:42,497

So, actually, COVID-19

619

00:34:44,406 --> 00:34:47,222

exacerbates the inequality in the UK.

620

00:34:47,322 --> 00:34:51,591

So, if we are talking about identifying and supporting those neighbourhoods,

621

00:34:51,968 --> 00:34:54,877

after we find out if they have been hit heavily by COVID,

622

00:34:55,117 --> 00:35:00,413

how to help those regions recover from this global pandemic.

623

00:35:00,513 --> 00:35:04,700

That is something we can find out from this kind of job analysis

624

00:35:05,109 --> 00:35:07,439

and also, the spatial and temporal analysis.

625

00:35:08,046 --> 00:35:10,041

The final thing is that we want to talk with

626

00:35:10,141 --> 00:35:12,349

the policymakers and local authorities

627

00:35:12,449 --> 00:35:15,737

to help level up the developments in the AI industry

628

00:35:15,837 --> 00:35:17,696

across regions in the UK.

629

00:35:19,847 --> 00:35:22,235

So, potential project ideas.

630

00:35:22,845 --> 00:35:25,125

If we do this kind of webinar,

631

00:35:25,225 --> 00:35:29,603

we have opportunities to share this data set

632

00:35:29,703 --> 00:35:31,641

for non-commercial academic research, right?

633

00:35:31,741 --> 00:35:33,720

So, the potential project ideas.

634

00:35:33,820 --> 00:35:36,299

I think, certainly, the Adzuna data would be

635

00:35:36,594 --> 00:35:39,844

a surely useful data set for analysing the labour market,

636

00:35:39,944 --> 00:35:42,402

which is the direct understanding

637

00:35:42,502 --> 00:35:43,781

from a data set.

638

00:35:44,761 --> 00:35:47,046

Also, in terms of economic geography,

639

00:35:47,146 --> 00:35:49,055

that's a really variable data set.

640

00:35:49,420 --> 00:35:53,039

And, I mean, for the COVID-19 recovery and renewal,

641

00:35:53,139 --> 00:35:55,447

that's also an important data set

642

00:35:55,547 --> 00:35:59,307

if you can couple it with the urban mobility data set

643

00:35:59,407 --> 00:36:02,986

and with a few other more official statistic data sets.

644

00:36:03,316 --> 00:36:06,384

I mean, as I mentioned in my previous slides,

645

00:36:06,623 --> 00:36:10,932

inequality is something, certainly, that can be looked at

646

00:36:11,032 --> 00:36:13,421

through this job listing data set

647

00:36:13,521 --> 00:36:20,127

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

648

00:36:20,564 --> 00:36:25,206

that will probably reflect something you really want to be careful about,

649

00:36:25,555 --> 00:36:28,543

to think about how to really help a region grow again.

650

00:36:28,833 --> 00:36:31,510

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

651

00:36:31,610 --> 00:36:35,326

especially where you have different academic backgrounds

652

00:36:35,426 --> 00:36:38,768

and also where you have different data sets.

653

00:36:38,868 --> 00:36:41,436

You can couple these new forms of data together.

654

00:36:41,763 --> 00:36:45,538

So, we certainly welcome new ideas

655

00:36:45,638 --> 00:36:48,536

and we're looking forward to seeing your proposals of

656

00:36:48,636 --> 00:36:50,663

how you want to use this data set.

657

00:36:51,929 --> 00:36:53,417

So, that's pretty much it for today

658

00:36:54,002 --> 00:36:58,639

for my quick research proposal overview.

659

00:36:58,739 --> 00:37:02,406

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

660

00:37:02,506 --> 00:37:03,813

Thank you so much.