Southampton

Coursework Group 1

Group Members

Olivia Wilson

Tianhua Mu

Zhen Wang

Weijian Yao

Yuhe Chen

FoDS(COMP6235)

Github Repo: https://github.com/academic-jobs/academic-jobs



Justification

- An application which helps people find trends in academic jobs data.
- Sources: jobs hosted on www.jobs.ac.uk and REF results from 2014
- Potential impact of REF on job hiring
- Could provide an insight into the inner workings of university hiring



Data collection - Job scraping

- scrapy Python module
- This was used because of the broad built-in functionality
- Sub-pipeline within scrapy:
 - 'Spiders' all relevant links from www.jobs.ac.uk
 - If already downloaded then skip
 - Otherwise, download page and scrape specified details
 - Clean data
 - Parse datetime
 - Clean-up whitespace
 - Field names etc
 - Insert into MySQL database



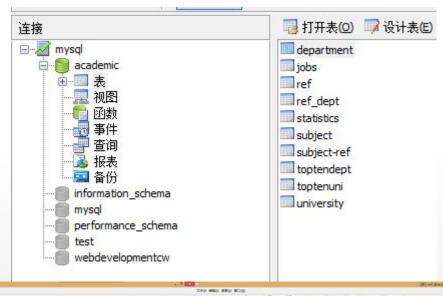
Data Collection - REF

- Research Excellence Framework
- Controls funding provided to universities basically proxy for 'prosperity'/'quality' of a department
- Provided by HEFCE at <u>www.ref.ac.uk</u>

- Simple CSV file
- 'Quality scores' are provided per university department

Southampton Southampton

MySQL Database



let 8	Steam	loan, aut., of	lan.	to lead to the	Inst	ne W	The same of the sa	. at	(wt.dest.id	Person	Street,	1	Service .	- January	Sectionsfeet	
e, Selec Roman ad Novanian	4/307-04		-	The Direction of Earlier is Australia, 20 th first property and the analysis applied regulation for markets and research equality. Authorized in National Association	magning Sub-	10 P				5		10			11.3	46
d Front Ferson	130-04		84 76, 8	falsely, late the highester Office of the Information for one is country for a properly and in the particular and the false of the fact in			-			TT.	3	46			16.7	
France Andreaders Andread	1,000,04		48	The case of this is a chart turned partie, produpt receive an our to the student and out if an agreement force for featured	. Three	eter "						26			7.0	
and fulto, flantage (Charles fulfor)	1.804-018		all from	ers and family made. Family for participation for comprehensing for family find a family of face of family for family for the participation of family for family family for family for family family for family family for family	wante last							-	-			
new method school and	1,805-04			terry and the female for next threat serges, the continuous or excellent apportunity for the exceeds conducts to not account the exceeds for the						46		100	-			
or Australia Streets (CD Oxfor Contr. of Numbers of Constitu	12000			Marie, Nation (Type (Nate (Habited Amount) (Amounted) orbited frank (Habited)	200		-			40	45	-	- 41	- 1	175 8	
more of Chame base' trains Remote frequi	129910			enate and it. Hales Consiling unless an independent behind consideral, deburing recording and and effective solutions selline unit finally hashers or linear	and their	March States	- 2			40			- 10		18.7	
American Company of the Company of t	1,80,04		40	The in an internal promotion on home or security for a 11.1 feature and promotion are flavour trans. The profession of promotion plane (female	246		-					-				
Industrial Park Int.	20 20 20 20		-	The self is responsible for providing features requires in the support and replacementation of a consensity units illustrate featuresing the Professional or Name	many Van	-				44					10.70	
Marting of Locales Trans, December and States	VI 815-016		The Land	age (that has the index easi) studied with the first of t	Trans.	200			-	-						
oth Street	50,000,000			to be been fire better of beautiful policy, a see of the contribute was sharing one count within other activities, as followed in the	wagers have							-	- 4		44 1	
Rames trains	1200		67	(pulse Nationalise (America are of the Sensor provides of principalises, party shapes, probabilises of continued and the Sensor provides of the Sensor provides					-	- 5						
Charles trace: Charleston a partier: Influent Defined festivat dover with a page.	17 210 124			plan format i PC and Ponterioral asubur scalable at the CREECT abcompany storic according to the construction of the CREECT abcompany storic according to the construction of the CREECT abcompany storic according to the construction of the constru	-				-	-		- 2	- 2			
scientia e Parcelviare Inagra tritorio ferfalable arrilles	21 20 0 42 0			as and Autors. The informacy furthers of the information of Europe State, cause there are greated measure for the damp that	-	ther H						- 2				
				and the second recording the second of the second of Second State Continues and Comments and Appendix and the state (No. 1 and No. 1 and	-	-	- 10		-	-	-	- 2	- 2		41	
representation on participation for the content and eventual and eventual participation for contents the contents that	90,006.00				740	-			-					- 5	No.	
THEFA SUPPLY GOOD MAN	0.004.04			e and book 5. Year, Klein Villaged Hosep (Inscaller II), Village Rep England								- 2		- 5		
- Princept	27 20 0-04			No, linears. The Milk to the destrict the electric elect. The proble, or all, a 24 hour Million parth and house service except the lift and hell hupbers, or force			- 0			- 5		- 5	- 5			
Mark Shaketrach Statutes	20 20 20 20 20 20 20 20 20 20 20 20 20 2			ses and filting filtrod following an armount are an addition of students obtaining in the Shoot of European at the Shoet		seb.				- 14		- 0	- 2			
nd trager (a extrator	N.HH-04			roberous an hour son and its boundar 2010 to come excondinate the last and hopes or others asset of the boundary for the boun	PM.								-34		2044 8	
Metape	6 30 6 0 8			on Principal face principals, a recovered for the existing approach to higher exhibition, beautiful person, entering construction and conduct state third beautiful principals.	THE 199							-			4.1	
Market Select Programmer	100614			er Resource: 100, Cercer Institute in Protection Cover(1974) Then The appearance of the on 100, check is 1774 Cercer Resource of Action, Cercer T. Cercer	144	da .			No.	40	17	96		*		
res Messisia	£305-01		57,7666	or Sassarout, There are 1 worked prote in 5th time personnel past and a part time, their laws part for 1 months (other decided from international Characters).	2/9	led .	11			14	- 24			-	AL .	
Miles	20 20 9 0 9		91	Laterate Part (mod LC, CR to an LCR, CR per among an indicating an Arth and experience. Subsy progression focused from the Copies.	Tes/de	Nighter .	28.		N°	Ph.		90		- 40.	063 6	
tre (Plue: 30A) (Fee below)	4/20/02			Notating, Sale S. 1775: Plant Brighout CO, 600: red SE, 775 year van periodusies of Landon Alburance (pending 1987) and Latter (1974) bright code and Co. Co.	389	de .	34			76		16			243	
Autor (New York Cont.)	B(304-04		60,696	rate and a final final at experience and extremely retroted a per-the-rienal audit have at the arresport. There is Audit cover at the a Parliament in their			.88		er .	**	0			40	50.1	
Augures (Mox	W 200-04		107	2004 Carputation are entering to record where of the control of the enterpretation for an extending the convention Color Andreward or their	wagen tiple	en-ter			P.	W.	.00	40	- 1		. 27	
e Season di Associanti	9/30/6/3/9		64	The enceting in apport or experienced and enfounces qualified activistic, with decreastable experience of postograph couples and Christ	Pers		37		N.	40	26	461			46.6	
ulerlybe in Dregs Surious Aubbys and Pryrine	608898			and Digner Substance are rotal to 1990 statements as part of the Design National Codings of Aughterings' Howels (N), 1994 Design PAC	Ind	Herse	.00		M.	91.	. 01	Pi .	III.	AL.	- 34	
Liferality in Plantar Chemistry	10,269-01		IN Prof.	Stemaubidg, and SrPacousterates in Notice by remoter Sobal Supris Charl based in the Noticeberlands. The Remount Force as persistant later has Mill	New		- 20		Nº.	**	W	W	- 11		41 4	
special below Therescholated Corbs for Hollesper the bedien	0.8944		. 10 New	rati transisti. Ne titest af transering et under fordt fant int-entry £362 in his application from existit qualifier undertec for a financial fill	140	der .	36		er.	36	10			16	911 4	
partition tragalation of the recruitment of eulocobia is one of other:	94 (ED) - 1,6 E		71 PM	are professional to the commercial collection from and wall field to industry consumbers but the poster is a continue to report to the	1,646	40.	31		pr.	Mr.	46	46		A.	36.00	
rie dictional study all the Metherseles Objection Carrier	61,310-124		70 Helin	emption and D. The Nathanaska Discrete Cartin (MC) of conglithrough prompts are interestingly recovered partie of nathanaska teaching, in (MC)	led	Plinted	36		St.	40	40	40	M	10.	94.9	
er / Tarriar Lacitures er Schemadoral Rabatoria	96,3006.020		36	If the property of Delta, people are at the least of contribing on the As a regar, but explains no not only offer one of the region Assistance in Sense	and the		36		0.	84	4.	96		100	96.86	
et-Associate (Nettic s./E)	26,2004-019		R lead	Financia and Figur (Note Industry Constanting) or Malanta of England	sach See	001	34		er.	80	- 67	40			37 6	
et hullion is helbar konse (proyects a helba)	0.000.00		10	meat of the School and meat of Deciglini's Apparetient of the case of the contract lidery food an a plant in the orth Schoolment for Appalein, or School	sen Name		. 16		97	86	.00	46	81		an a	
Intert Associate Professor in Lift Foreign Policy	1,89410		56	The Separtment of Follow and International Relations on Softmake an internationally enternal context of condense for teaching and no Augustic or Research	sed: INA	led .	9.			80		101	- 11	18	10.2	
Perspectative for test earliest and (burt-Propusse)	90,999-04		Di Spuri	a est aspec l'Episses, Tra Assec d'Especiale d'Arreit et profet les de réspect à part properties elle les l'expectes d'appece de l'est de l'expectes de l'especiale de l'expectes de l'e	medi (med		10	-		44		10		40	25.49	
merhatre Early Caree Tellowise (E.id)	2 2 4 4 4 4			ages, cleral. The Facults of this and increasing motion against on Early Corner Policeating in the Unit of Life, to that between a September 27 Academic or Nasco		ach I				47	2	56	44	201	4.0	
day Statute Teating February	0.0000			age, Jens, he bearing of hitsual patential means of rendered makes for resent on a role offering to be unit, Autom a force		and the	10	-		86	. 0	10.	**		(6.17	
in to Orini Improvement Industrial land	# 20 G			a Serva (pr. January 2000)) and feet from contaminative resist frontiers of four following learner safe from one data property is Audience of four		-	*			**	in.	No.	-34	10.	0 1	
ictoral feeting halos in toen and realth storals	9.00.00			on and facility programming a strength in a strength collection of the first program of the prog						de:	10.	60.			14 8	
Name the of Tomog Control Proces Storons for Health Sto Physic	3/20/04			one and Carrier Charmoni. Physical and Computer Sciences for Redocument of PROS Research and Treating Carrier in Physical Computer Sciences for Redocument of Pro-		mpan .	- 4		01		34		35		265	
CP studentifies the reflector of features and training intervention is	2 5 5 5 1 1			ance there the present dictinal project of investigate the inflance of features, for othe pattern and other colors over non-economics (NO	Det		- 41		er.	600		18	107		48 8	
Libraria Pennahal antiting making and sorted of Blast durant	20.20 0 0 0			to and below, the counter on a project that continue both theoretical and equationally research observed. Trianslated in bornellad originating, our this		man I			er.	*1	4	20.	- 60		Jan B	
Lawrence to Delian Security 19	1,00400			one, Proming Therefore (COC), a please to mote of a fit paragraph to the apply for the adjusted a force one sufficient are and of all the	Deb	ed III			and the same of th			W			100	
O'D studentials through further Charge and Through Through	36 (0.00)			parts, financia: white is table down charge, we read to need to reaching charge to our every control who encountry arrow final final	- Dec	100			and the same of th	40.				-	14	
Library - Service Audio	21,208,010			endos (sep. apticulos an invital for a file funda file studentes understay research into benefit audit with the context of a burgan or file)	Charles At	STATE	145			441		-	-	- 2	16.7	
lating the of listings of horses allowables	2 2011			p. Marcia II. In hanny (year Martin research continue and other property of an author qualified and the foreign (MC	Det				and the same of th		100	16	12	-	84 8	
calculate. Screening of colors calculated facilities and impainted paid	27 20 4-014			unions and it is NOP 1000, after place francis your of 2007 60 kg personants formy a report discourse about they offer francis assistance in 1 PAD		tente v			in a	-					84	
The second secon	2,0000			Control of the Contro	-	The second second				-			- 0	- 0	-70	



Demo!

Back End

- jFinal framework
- MVC pattern
- JDBC connect to MySQL database



Demo!

```
Analysis Function
//count numbers of job of each university
private List<Jobs> countjobs(int a)
//count job_types for each university
  private List<Jobs> counttype(int a)
// top10 departments
  private void toptendepartment()
// top10 university
  private void toptenuniversity()
//sum all university stars for each department
  private float [] sumdeptstar(int l)
```



Demo!

Front End

Several web pages to display our consequences:

Characteristics:

- Interface:HTML5 and CSS3
- Logic: Javascript
- Interactive
- Responsive



Responsive

Use bootstrap

Change the size according to the size of window Suitable for more device:

- mobile
- ipad



Demo Shiny Server



Conclusions

- There is a small positive correlation between job counts and REF score
- There are very little universities that have low REF scores with higher job counts

Further Work

- Continue scraping to get a better time series of data
- Clean data so that university name is invariant to "the"



Thanks!