



Project 4 - Web Scraping Job Postings

- By Carol C

Goals

- Scrape job posting data from Seek.com
- Determine the factors that are most important in predicting salary amount.
- Determine the factors that distinguish job categories and titles from each other.



Tasks Completed

Web Scraping

Apply Requests, BeautifulSoup, CSS path to scrape data from Seek.com

Data Preprocessing

Create feature columns by NLP, and then perform feature dimensions reduction

Data Cleaning

Use Regular Expression to work out the salary number from text

Modeling - Salary

Build both regression and classification models to predict salary amount

EDA

Use visualization to get an overview of the data

Modeling - Job Category

Build classifiers to distinguish job categories

Web Scrapping



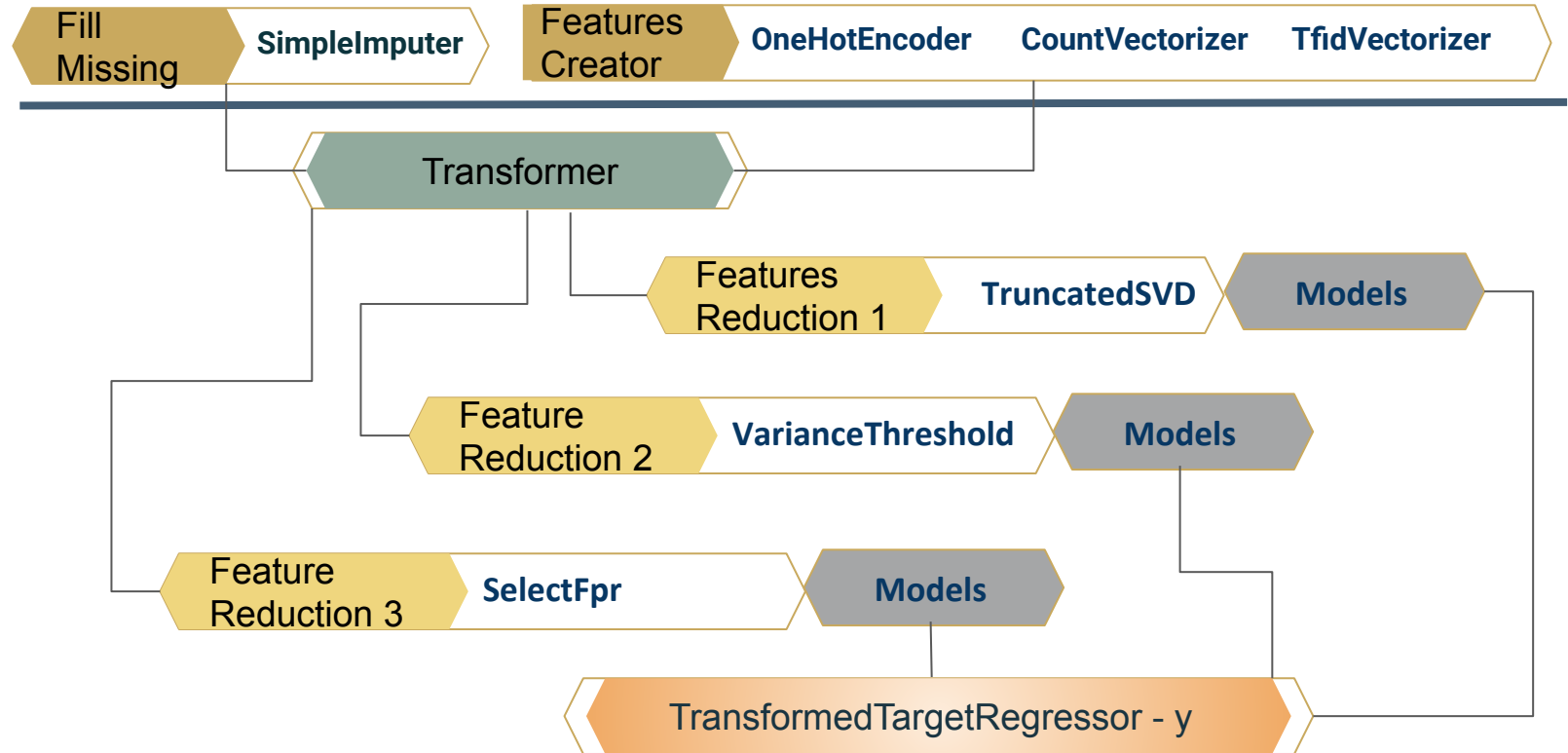
10695 job postings
12 features



1725 job postings contain
salary information

Tools: Requests + BeautifulSoup + CSS path

Data Preprocessing





Models

Predict Salary

- Linear Regression
- Ridge
- DecisionTreeClassifier
- LogisticRegression

Distinguish Job Categories

- SVM
(‘scientist’)
- RandomForestClassifier
(‘Senior’)

Model Performance

- Salary Predictions

Predict Salary Amount	Baseline RMSE	Model RMSE
LinearRegression Ridge	44,332	
		35,217
		34,853

Classify High/Low Pay	Baseline Accuracy	Model Accuracy
DecisionTreeClassifier LogisticRegression	0.52	
		0.70
		0.72



Model Performance

- Job Category Classifications

'Scientist' or not	Baseline Accuracy	Model Accuracy
SVM	0.91	
		0.95

'Senior' or not	Baseline Accuracy	Model Accuracy
RandomForestClassifier	0.82	
		0.88

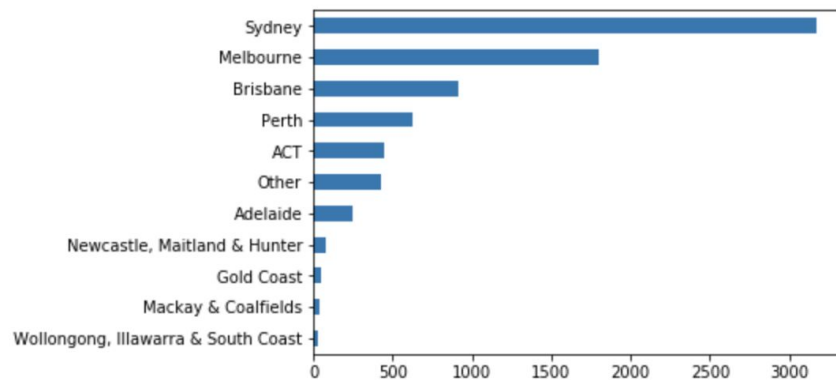




“ What Factors are most important for predicting salary and distinguish job categories? ”

Job Postings Distribution by City

Sydney	41%
Melbourne	23%
Brisbane	12%
Perth	8%
ACT	6%



Job Postings Distribution by Classification

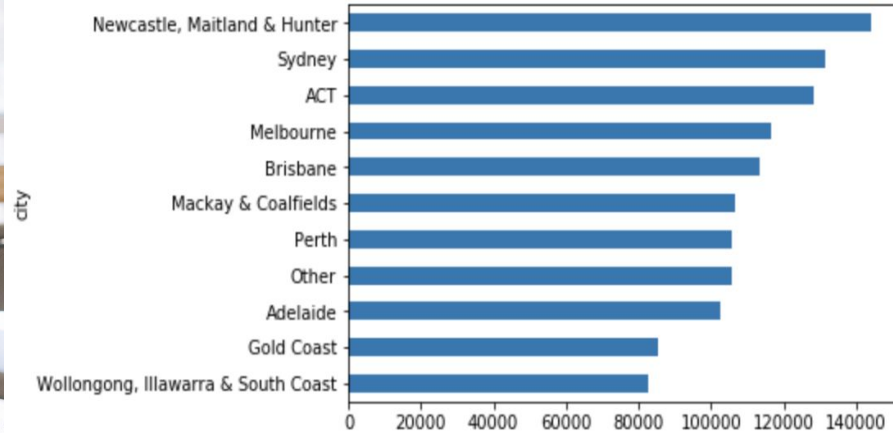
Top 5 classifications:

- Information & Communication Technology
- Accounting
- Engineering
- Banking and Financial Services
- Sales





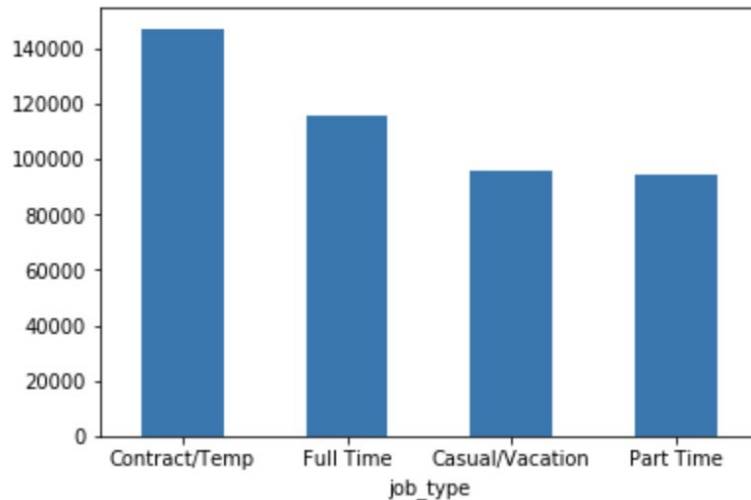
Average Salary by City



Newcastle, Maitland & Hunter 144K
Sydney 131K
ACT 128K
Melbourne 117K
Brisbane 113K



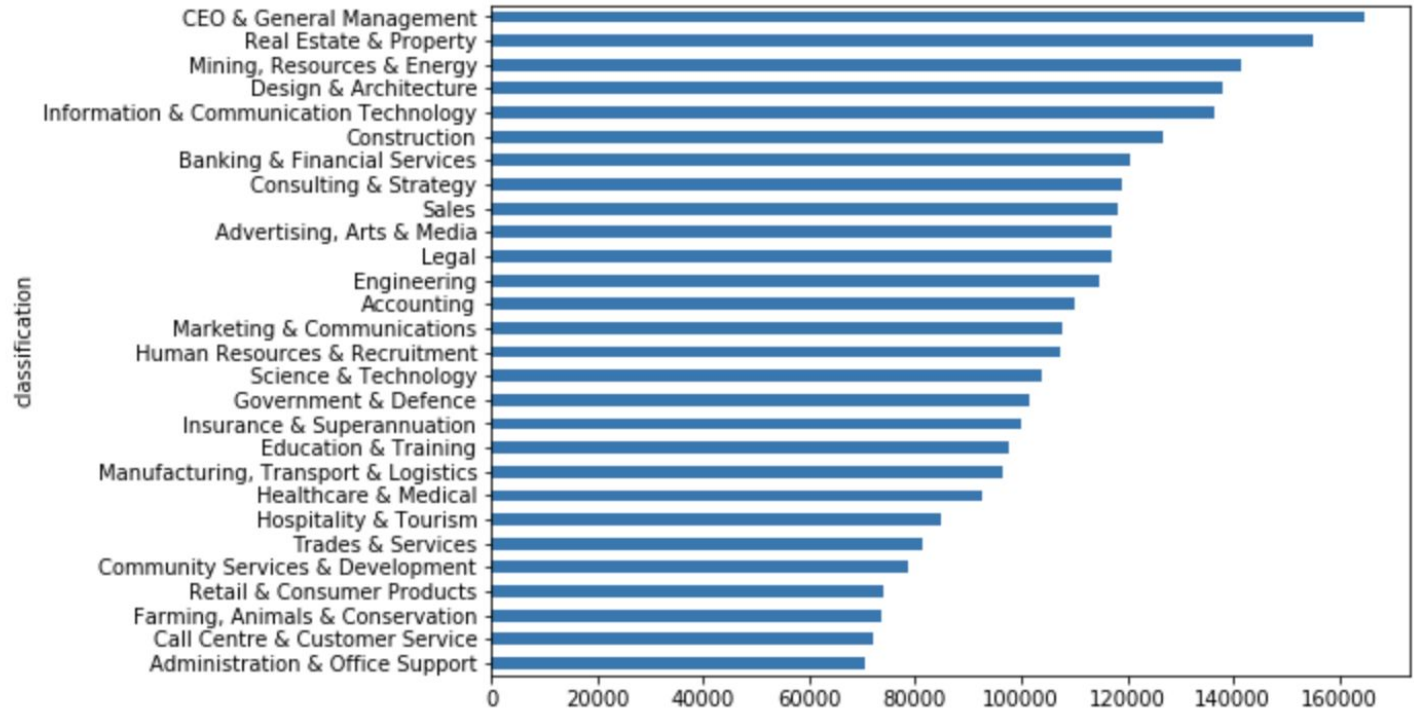
Average Salary by Job Type



Contract/Temp > Full Time > Casual / Part Time

Average Salary by Job Classification

-What industries pay highly to employee ?



What Factors are most important ?

Salary (High/Low)

y=1 top features

Weight?	Feature
+1.977	classification_category__x0_Mining, Resources & Energy
+1.865	job_title_category__head
+1.847	job_title_category__director
+1.785	job_title_category__architect
+1.590	job_title_category__principal
+1.425	job_title_category__team
+1.380	job_title_category__market
+1.361	job_title_category__senior
+1.352	job_title_category__mining
+1.310	classification_category__x0_Information & Communication Technology
+1.294	job_title_category__sap
... 85 more positive ...	
... 72 more negative ...	
-1.239	city_category__x0_Adelaide
-1.244	job_title_category__officer
-1.261	job_title_category__technician
-1.288	job_title_category__operator
-1.300	classification_category__x0_Administration & Office Support
-1.301	job_title_category__research
-1.341	job_title_category__assistant
-1.403	job_title_category__support
-1.874	job_title_category__administrator

Data Scientist ?

y=1 top features

Weight?	Feature
+2.000	classification_category__x0_Science & Technology
+1.000	job_type_category__x0_Contract/Temp
+1.000	job_type_category__x0_Full Time
+0.223	suburb_category__x0_CBD, Inner & Western Suburbs
+0.222	suburb_category__x0_CBD & Inner Suburbs
+0.222	suburb_category__x0_Eastern Suburbs
+0.222	suburb_category__x0_Bayside & South Eastern Suburbs
+0.222	suburb_category__x0_North Shore & Northern Beaches
+0.222	suburb_category__x0_missing
+0.222	suburb_category__x0_CBD, Inner West & Eastern Suburbs
+0.222	suburb_category__x0_Ryde & Macquarie Park
... 4 more positive ...	
... 7 more negative ...	
-0.000	classification_category__x0_Banking & Financial Services
-0.000	city_category__x0_Brisbane
-0.001	classification_category__x0_Insurance & Superannuation
-0.001	classification_category__x0_Accounting
-0.001	classification_category__x0_Healthcare & Medical
-0.001	classification_category__x0_Government & Defence
-0.001	classification_category__x0_Mining, Resources & Energy
-1.777	suburb_category__x0_Fremantle & Southern Suburbs
-2.222	<BIAS>

Thanks!

Does anyone have any questions?

Please contact: carolcheng.au@gmail.com

Credits:

Presentation template by **Slidesgo**

Icons by **Flaticon**

Images & infographics by **Freepik**

