# **Web Scraping and Analysing Job Postings**

**Introduction:**

This report focuses on web scraping a recruitment website (<https://www.seek.com.au/>) , analysing the Data Science related jobs and presenting the results.

**Data:**

Using various web tools managed to collect data pertaining to the following job titles:

* Data Scientist
* Data Analyst
* Research Scientist
* Data Engineer
* Big Data

*(Jobs from all over Australia)*

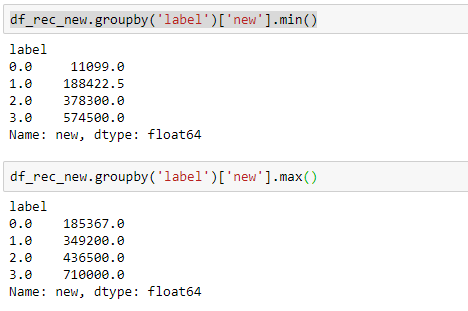
Under the following categories:

* Information & Communication Technology
* Science & Technology
* Government & Defence
* Education & Training
* Consulting & Strategy
* Engineering
* Manufacturing
* Transport & Logistics
* Construction
* Healthcare & Medical
* Advertising
* Advertising, Arts & Media
* Arts & Media
* Marketing & Communications
* Sales
* Human Resources & Recruitment
* Administration & Office Support
* Banking & Financial Services
* Insurance & Superannuation
* Accounting

**Data Cleaning Process:**

Salary Data Cleaning and Binning Process:

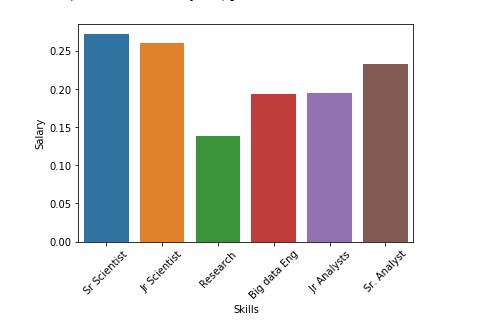
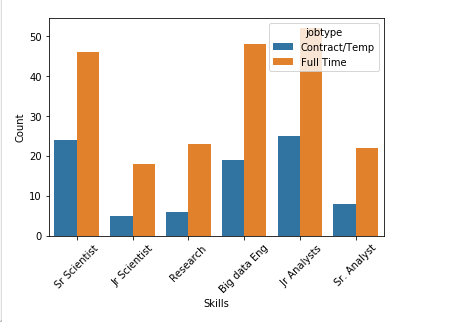




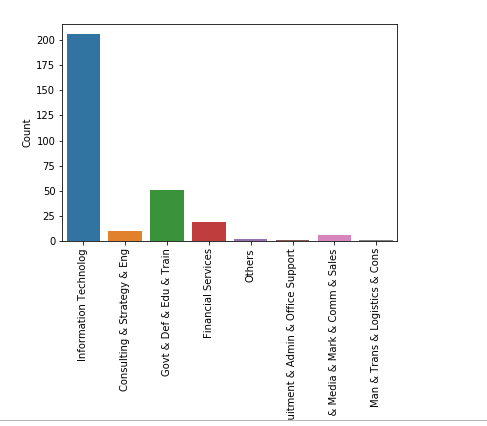
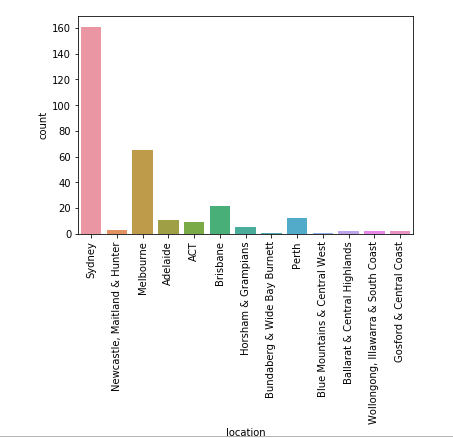
* Salary:
  + Divided the Salary into 4 bins
  + Within each bin grouped on Skills, Location and Job type for missing salary values.
* Skills
  + Used Spacy to extract proper nouns from Title and divide title into
    - Jr. Analysts
    - Sr. Analysts
    - Jr. Scientist
    - Sr. Scientist
    - Big Data Eng
    - Research
* Category:
  + Jobs binned into fewer major category depending on the industry.

**Findings:**

**Job Counts Vs Type: Skills Vs Salary:**

****

**No.of Jobs vs Category : No of jobs vs Location:**

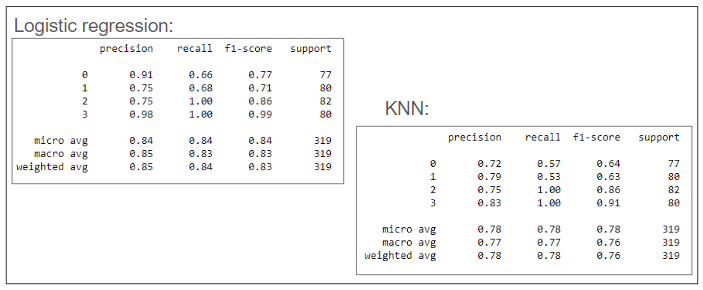
**** ****

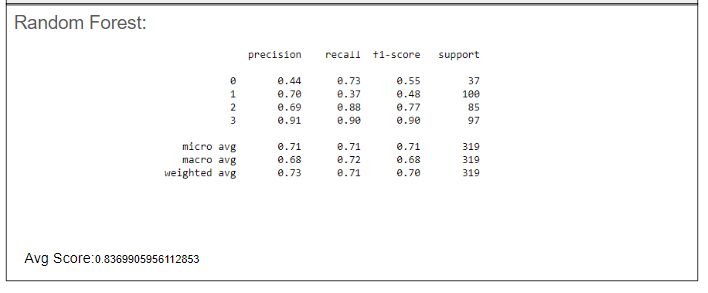
**Modeling & Results:**

Used various models to understand the feature importance and salary prediction

* Skills,
* Job Type
* Location

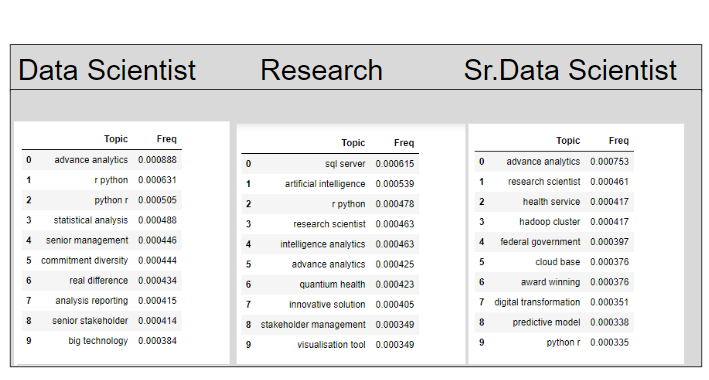
Determine the salary.

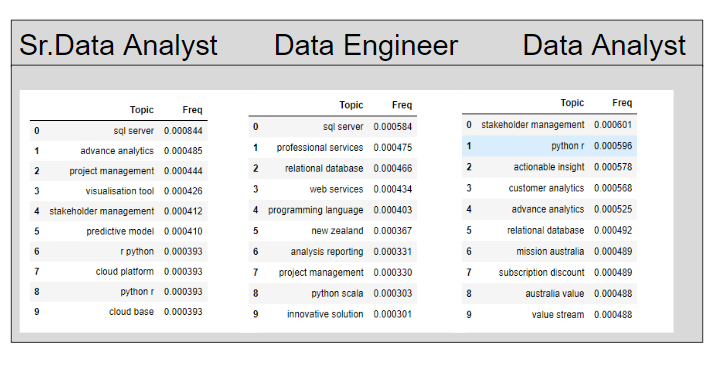
****

****

**Problem 2:**

Job titles prediction from Job Description(Unsupervised Learning)

****

****

Conclusion:

Although it is possible to predict titles with some clarity via Unsupervised Learning, there is a lot of overlap as the descriptions are similar for most data related jobs.