7030ICT Introduction to Big Data Analytics

Trimester 2 2020

Assessment 2 (Part 1): Written Assignment

Group 27: Ceaira Cottle (s2161226), Hanh Nguyen, Susana Pentland (s5241528)

Due Date: Monday 3rd August 2020 at 12:00 midday

Marks: 20% of overall grade

Word limit: 1000 words

The objective of this assignment was to apply data analytics tools to study and analyse the SEEK job market dataset provided. We used data preparation and pre-processing techniques to understand the characteristics of the dataset. We also performed various data analysis techniques to gain a better understanding of the dataset. Our findings are detailed below.

Part 1 – Data Preparation and Preprocessing

The raw SEEK job market dataset contains 149,999 records representing jobs posted over 44 days (from 1st October 2018 to 13th November 2018). Over 26, 000 different companies advertised positions within this time frame on SEEK.

A total of 13 attributes were collected including: 'Id', 'Title', 'Company', 'Date', 'Location', 'Area', 'Classification', 'SubClassification', 'Requirement', 'FullDescription', 'LowestSalary', 'HighestSalary' and 'JobType' (Table 1).

Table 1: Description of attributes in dataset with column type and number of unique values

|  |  |  |
| --- | --- | --- |
| **Attribute** | **Type** | **Number of variations** |
| Id | int64 | 149999 |
| Title | object | 82115 |
| Company | object | 26318 |
| Date | datetime64[ns] | 44 |
| Location | object | 65 |
| Area | object | 19 |
| Classification | object | 30 |
| SubClassification | object | 338 |
| Requirement | object | 121324 |
| FullDescription | object | 127676 |
| LowestSalary | int64 | 11 |
| HighestSalary | int64 | 11 |
| JobType | object | 4 |

Job postings were from 65 different locations with most of the jobs located in large cities, i.e. Sydney (46,357) and Melbourne (35,315). Jobs were also categorized by area with the majority of jobs being offered in *CBD & Inner Suburb* areas (24, 302) and *CDB, Inner West and Eastern Suburbs* (22,079).

Jobs were posted across 30 different sectors, with the majority of jobs in the Information and Communication Technology sector (Table 2).

Table 2: Number of job postings per sector

|  |  |
| --- | --- |
| **Sector** | **Number of job postings** |
| Information & Communication Technology | 16661 |
| Trades & Services | 14125 |
| Healthcare & Medical | 12515 |
| Hospitality & Tourism | 11818 |
| Manufacturing, Transport & Logistics | 9608 |
| Administration & Office Support | 7636 |
| Accounting | 7075 |
| Education & Training | 7033 |
| Retail & Consumer Products | 6496 |
| Sales | 6281 |
| Construction | 6254 |
| Government & Defence | 5926 |
| Engineering | 4812 |
| Mining, Resources & Energy | 4679 |
| Community Services & Development | 3528 |
| Banking & Financial Services | 3481 |
| Human Resources & Recruitment | 3233 |
| Call Centre & Customer Service | 3127 |
| Legal | 3071 |
| Marketing & Communications | 3052 |
| Real Estate & Property | 2729 |
| Design & Architecture | 1379 |
| Insurance & Superannuation | 1144 |
| Consulting & Strategy | 905 |
| Sport & Recreation | 750 |
| Science & Technology | 696 |
| Farming, Animals & Conservation | 683 |
| Advertising, Arts & Media | 644 |
| CEO & General Management | 585 |
| Self Employment | 73 |