Job Market Analysis Report

Students:

s... - <Name>

s... - <Name>

Part 1 – Data Preparation and Preprocessing

The data set consists of 318477 job postings from seek.com from the 1st of October 2018 - 13th of March 2019. There are 13 attributes which can be found as follow:

Attribute	No. of Variations
ID	318477
Title	168065
Company	40628
Date	163
Location	65
Area	19
Classification	30
Sub-Classification	338
Requirements	234287
Full Description	250901
Highest Salary	11
Lowest Salary	11
Job Type	4

- The number of variataions for job title, requirements and full description is quite large as well, which indicates that there does not seem to be a uniform standard of choosing job titles.

- There are 163 different dates, which cover the date range of the datasets job posts from the 1st of October 2018 - 13th of October 2019

- Classification and Location are categorical variables, which have variation sizes of 30 and 65, seem to provide interesting information to analyse the spatial distribution and popular job sectors within Australia. SubClassification is also a categorical variable, which further divides each Classification into more specific sections.

- Area is a categorical variable that divides a location into smaller areas, and has 19 variations. At a first glance it seems like only the main cities such as Brisbane, Sydney and Melbourne have a value for Area, however, we will investigate the attribute further during analysis.

- Lowest and Highest Salary are numerical variables with 11 different ranges, which will be used for analysis with respect to Location and Classification, in order to see where employees are paid better and what job sectors have the best potential to earn high salaries.

- There are 4 job types, and the attribute can be used to determine whether there is a relationship between salaries and job type