**Case Study**

**Data Scientist**

**Overview**

As a Data Scientist your role encompasses but is not limited to:

* formulating analytical problem statements and forming hypotheses
* designing and building data pipelines
* evaluating models
* presenting your analysis and insights to various stake holders

The purpose of the case study is to assess your ability to:

* interrogate and understand a data
* conduct appropriate analyses and modelling
* extract insights and present potential opportunities to improve user experience
* present your solution concisely with your audience in mind

**Data**

In this example we are looking at a hypothetical sample of data for SEEK AU website and mobile app.  
 **job-ads data:** Each row of the file is a JSON object representing a job ad, the data used to render job ad cards on the search results page as well as detailed job ad view page. The attribute names of the ad are self-explanatory.  
  
**events data:** Auto-generated and anonymised data following some distribution that simulate online user behaviour with clicks and applications in a CSV format. Each row contains one interaction with the information as *interaction time*, *resume id*, *job id*, *event platform* (web, ios and android) and the *kind* (V=view or A=application).

**Task**

Based on the data provided (job ad content and candidate and jobs interaction), your task is to conduct appropriate analysis and extract insights from this data – **goal: analyse, summarize and show the insights from the job market data that can be used to build AI** products (e.g., the data could be useful for auto-suggestion services, recommendations, search, data enrichment, strategy, etc). Any analysis that helps to understand strong and weak points over the data is welcome. Data analysis involving text and having a creative way to present also is well evaluated. To complement your initial analysis, using interaction signals would also be important to understand key point to help when some Data Scientist needs to create machine learning models to recommend jobs to candidate or build a search engine, for example.  
  
The task is open and you have the freedom to go in any direction as soon you identify opportunities and extract insights. Find usage for the data is part of the task. Some suggestions that you can follow to structure your analysis and presentation: data characterization, identify strong and weak points, present user case with examples how data can support your claim.

**Deliverable**

Please do all your coding in Python, R, Java/Scala or SQL (you can use whatever packages you like) and provide clear comments and/or a written summary. One approach might be to create a Jupyter notebook ([http://jupyter.org](http://jupyter.org/) or similar technology) to present the code and the documentation/discussion together. This solution will be evaluated by SEEK Data Scientists.

In addition, put together a **concise PowerPoint presentation that can accompany a half-hour presentation** and discussion of the findings. This presentation and discussion should be targeted at an audience of a data scientists, software engineers and service managers, so tailor your communication as you see fit.

Bring ensure that your presentation is included when you email your solution to the case study.