**ADSEE Project**

**Job Market Signalling**

**Exploring Policy and Research Signals in Advertisements**

**Approach Analysis Whitepaper**

*ADSEE - Job Market Signalling*

Exploring policy and research Signals in Advertisements

Dr. Alan Berg March, 2021



*A simple data pipeline*

Table of Contents

[Overview 3](#_Toc82170214)

[Who we are 3](#_Toc82170215)

[Background 3](#_Toc82170216)

[Case Study 4](#_Toc82170217)

[Workflow 6](#_Toc82170218)

[Breaking your work up into small pieces 6](#_Toc82170219)

[Understanding Job market Signals 6](#_Toc82170220)

[References 7](#_Toc82170221)

#

# Overview

This course describes how you deliver actionable information from Big sets of online Job Advertisements. You will be introduced to a basic data science work flow and text mining techniques. The aim of the course is to support Economists, policy makers and researchers who wish to quickly explore their questions.

**This case study is fictitious**

We place you in the role of an adviser to a Government Committee which has to quickly react and advise you about changes in the job market due to a new pandemic. Your role is that of a ***data story teller***. You are not responsible for drilling down into complex details and economic models. You want to give the committee an idea of where problems are emerging.

We will take a pragmatic approach to teaching based on concrete examples. For the curious, there will be many links to supplementary material. The courses examples are based on small pieces of R code. R is a programming language widely used by data scientists.

## Who we are

The course content was generated as part of the [ADSEE](https://adsee.eu/) project. The main objective is to deliver a useful educational and training program in data science through the development of educational modules, and the adaption of contents and methods according to the envisaged needs of the groups.

## Background

(Provost and Fawcett 2013) *At a high level, data science is a set of fundamental principles that support and guide the principled extraction of information and knowledge from data.*

Data Science has its own work flows. With modern programming Languages such as Python1 and R2 , these work flows are relatively easy to actualize. Labour Market Intelligence (LMI) is actionable processed Labour Market related information. Job Market Signalling (JMS) is Labour Market Intelligence that can trigger actions such as policy or legal changes. JMS relies on Data Science techniques and data pipelines to process data such as a Big3 collection of Job advertisements. When the right research questions are asked, then your workflows will turn the data into actionable insight.

Online Job advertisements are an opportunistic Big authentic data source that if collected across the whole of a country or continent, provides broad coverage of employee demand, skill shortages

1. <https://www.python.org/>
2. R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. URL [https://www.R-project.org/.](https://www.r-project.org/)
3. Big as in Large volumes of data that are inconvenient to store on a laptop computer and has many dimensions to it such as name of company, educational level, salary, location, job title, etc.

and practices. There are biases in the collected data: for example, one job advert is not the same as the one position to be filled. However, we can consider and approximate.

Three advantages of online advertisements are:

1. The advertisements represent authentic demand from the Job market
2. You can collect a comprehensive set within a reasonable time
3. Even with simple data science /text mining methods, you can visually explore and explain features

The data is available and waiting on you to explore. A commercial company that collects Job Advertisements is Burning Glass4, which trawls over 40,000 online sites. An example of a European Commission related organization that does the same is CEDEFOP5 which has the facilities to collect Job advertisements across the whole of European. Both CEDEFOP and Burning Glass have gathered hundreds of millions of Job advertisements.

**Exercise 1 Discussion**

Which labour market themes are mentioned on the CEDEFOP and BurningGlass websites? How are they relevant within your context?

# Case Study

It is your first day on the job. You have been handed a red folder with an urgent assignment and indirectly have access to a Big set of Job advertisements. The minister has heard that a number of countries are investing heavily in supporting their IT sector. The head of the statistics bureau is thinking of sending out an urgent survey to large companies. However, they are not sure which companies are most affected by the pandemic and want you to share some ideas on this theme.

As it is your first day, you are provided with older data that is open to the public. 20,000 Job advertisements from the UK from 2013 were donated by MonsterBoard6. You will use the example data to create and practice your own work flows. You will then ask the IT department to run your queries and process the data. After a while they will trust you and you will gain access to the data and compute directly. To explain what you are doing, you will need to create a reproducible data pipeline. That is, you start off with raw data, clean, process, visualize and look for patterns, and finally report. Later, you will go through the cycle again but with more refined questions. As you want a reproducible pipeline, you will write a little R code and share.

1. <https://www.burning-glass.com/>
2. <https://www.cedefop.europa.eu/>
3. [https://monster.com](https://monster.com/)

Luckily the example data was already pretty clean so the first thing you practice is looking at the volumes of Job Advertisements across occupations. You pass your code to the data owners and, for practice, they then process against 60 million job advertisements7 from the US job market. From the initial plots, it’s clear that demand is increasing in the health sector and decreasing in the IT sector. Looking further, you can see that demand has stabilized at new levels.



**Figure 1,2:** *Percentage of job advertised for Health and IT skill clusters*

From these simple statistics, you can anticipate gaps in supply and demand. With well-known techniques combined with understanding the context in which jobs are advertised, you can efficiently gain insights.

From past experience in a Marketing department of a soft drinks company, you know that tweets about your company products contain emotional words. From simple counts of those emotional words, you can estimate the sentiment around specific products. This was why the red-hot chili drink pilot was taken off the market. The tweet sentiment was immediately very negative.

Taking a standard dictionary of positive words, you count the average number of positive words per job description. Not surprisingly you see an apparent trend.

1. As an example, the graphs are from BurningGlass data comprising roughly 2 years’ worth of data which is around 3,000,000,000 words within the job descriptions.



**Figure 3:** Positive sentiment in the US job market.

However, when looking at the job descriptions some of the words are being used out of context. To make what you measure more reliable, you decide to make your own dictionary and have checked by another expert. And so, you enter into the specialization of Job Market Signalling.

# Workflow

The minister is enthusiastic about your work and now wants to expand your exploration into new themes. She wishes you now to look at huddles in the job market for different audiences such as gender or those entering an occupation for the first time.

To avoid criticism, you understand that you will need to create a reproducible case. You will need an idea of which signals exist in the job market. You need to know the biases in your data, have an idea of how to explore your data, clean up your data and report it. You decide to use the R language because it is well documented, has a large community of support, and can be extended by many packages. An added bonus graphics with code is easy to find on the Internet. And you do not have to be afraid of the code because if you need to learn a little coding, then there are a lot of good free books available.

**Exercise 2**

**Review** the books available on the bookdown website8. List the ones that are of help to you with a brief description of why. If you are in a group, share the list and select the group favourites.

**Discuss** if the books you have found are helpful for your needs. If not, what is missing?

# Breaking your work up into small pieces

The minister has taken an interest in your work and you do not want to disappoint them. To be able to start advising, you will need knowledge about the strength and the weaknesses of your data.

## Understanding Job market Signals

So, what types of signals have already been found in the data that you are using?

8 <https://bookdown.org/home/tags/>

There are many types of signals in the job market that are actionable; for example:

* Job Volumes
* Huddle words that place barriers in front of candidates such as the need for driving licenses or working at certain times of the day
* Gendered wording
* Personality traits also described by the use of specific wording
* Evidence of criminal activity

Bigrams are two words that are next to each other. The following is a Table of the top 10 most frequent Bigrams for 2019-01-04 in the US job market as captured by Burning Glass. Which of the bigrams suggest huddles?

|  |  |  |
| --- | --- | --- |
| **word** | **n** | **rank** |
| job description | 84219 | 1 |
| **education level** | 57514 | 2 |
| **level experience** | 56369 | 3 |
| company information | 56201 | 4 |
| week salary | 56097 | 5 |
| **shift shift** | 55994 | 6 |
| **experience license** | 55992 | 7 |
| company phone | 55977 | 8 |
| information company | 55966 | 9 |
| company direct | 55960 | 10 |

**Table 1:** *Bigram Frequency* for the day 2019-01-04

Not only are there many potential signals depending for what you are looking at but the field is also changing rapidly, so you will need to keep track.

**Exercise 3**

The minister has asked you to find out which signals are possible to track. Visit Google scholar and look for the term “Job Market Intelligence.” Review only for papers written in the last year. List the title, a brief summary of the abstracts and write a note if the paper is useful for finding Signals.

**Discussion:** Which signals do you consider important and why?

# References

Provost, Foster, and Tom Fawcett. 2013. “Data Science and Its Relationship to Big Data and Data- Driven Decision Making.” *Big Data* 1 (1): 51–59. <https://doi.org/10.1089/big.2013.1508>.

12 <https://bookdown.org/yihui/rmarkdown/>