# Contents

# 1 Problem description

What do I want to achieve with this tool ?

## 1.1 Definitions

### 1.1.1 What is the Target Data ?

I want to have access to the key data that make a job or a carreer path interesting. Here is a non-exhaustive list.

1. is it easy to access ?
   (a) is it new ?
   (b) what formation is required ?
   (c) what is the entrance cost ?
   (d) what success stories ?
2. demand > supply
3. loan
4. re-usability
   (a) tremplin
   (b) multiple use case
5. connection with your parcours

### 1.1.2 What problem sould this tool solve ?

what is it you want to solve / answer ?

1. can there be a tool to help me find a job / seek for opportunities ?
   (a) can this be quicker the reading / regular digging ? What this tool could/should bring :
      i. have an overview of field to dig further in
      ii. needs to be quick to browse : efficient summary
      iii. needs to be reliable : data quality / second check

## 1.2 Objectives

### 1.2.1 Needs

1. need to find a job quickly
2. need to be able to have a market overview
3. need to be able to propose a data product
    (a) for my portfolio
    (b) as an open source webapp

### 1.2.2 Will

1. want to have alerts for job opportunities
2. want to reveal career paths / market trends
3. want to keep track of my interests to have suggestions of paths

### 1.2.3 Use

1. target opportunities
    (a) sheets of wanted words
    (b) query matching algorithms
2. data exploration
3. cluster
    (a) nlp
4. find jobs I didn't know about
5. get warned if new opportunities
6. use it as a model for finding my perfect match in the world / exploring the economy
7. make it open source and useable by anyone

#### 1.2.4 Features

1. Update
2. Clustering
3. Visualization

## 2 Approches

There are two approches to keep in mind.

### 2.1 stathead / qualitative

use available data, stats and probability

### 2.2 scout / quantitative

Get humanly involved in the question to avoid using data science where it is not needed and to know where it can be of cruitial help. Helps keeping DS work pertinent.

## 3 Project management

How do I overcome :

### 3.1 Time constraints

#### 3.1.1 run test / small steps / prototypes

### 3.2 Limits

what is shaky about this approach ?

#### 3.2.1 quantitative unknown

1. former predictions / insights : find and evaluate
2. trending/ubiquitous ideas/concepts

#### 3.2.2 is it useful ?

1. what would be the user case ?
   (a) for individuals : orientation
   (b) for firms : market analysis

#### 3.2.3 data acquisition is costly

1. if the approach is not efficient : pb
2. **TODO** can only be tested if large amount of data
3. need to gain faith / enthusiasm
   (a) take small step achievements
      i. small jobs market
      ii. data science / computing
   (b) select exciting data sources
   (c) readings
      i. signal and the noise
      ii. use cases
      iii. data culture / world
   (d) methods / milstones
      i. doubt is good
      ii. **TODO** testing is good

### 3.3 Crossovers

#### 3.3.1 test features / ideas

to gain confidence / go in the right direction

1. UI / UX
   (a) possible on small amounts of data
   (b) test on real users (friends, colleages)
2. broader analysis
   (a) only possible on larger amount
3. application to other fields ?
4. sponsoring ?

#### 3.3.2 build prototype project to show

# 4 Plan

## 4.1 Blogging

### 4.1.1 Org babel

### 4.1.2 Website

## 4.2 Courses

### 4.2.1 Databases

### 4.2.2 Visualization

### 4.2.3 Machine learning

### 4.2.4 NLP

### 4.2.5 Hash tables / numpy computation

### 4.2.6 Proba / stats

## 4.3 Jobs seeker

## 4.4 Implementation

### 4.4.1 Start a clean project

1. **TODO** git
   (a) a branch per functionality
2. **TODO** projectile
3. file system
   (a) /
      i. org
      ii. scraper
      iii. database
      iv. explorer
4. database
   (a) sql ?

(b) csv ?

5. org babel file / emacs env
   (a) snippets C-c & ... Tables C-c C-t is snippet mode for test
   (b) **TODO** track time
   (c) track habits
   (d) decide what goes public and what does not at expension