What keywords should you use?

# Background

Resume keywords are the terms and phrases that describe specific job requirements. They are the abilities, skills, expertise, and values that the recruiter is looking for in a candidate. How important can a bunch of words really be? Your resume will be screened by some form of an Applicant Tracking System (ATS). At one company I’ve worked for, we would often receive several thousand resumes for a single job. Essentially, the ATS software allows a recruiter input a list of keywords into the system, and the system simply scans and counts the number of occurrences of these keywords and phrases, returning a list of potential candidates simply based on term matches and term frequency. If you don’t have the right terms and term frequency, your resume gets thrown in the void and never reaches the recruiter.

This week’s project we are going build a better mouse trap. What terms are used in positions I’m interested in? How does my resume stack up against those terms? To do this we’ll work in combination with our Career Management Class with Ms. Vogler. Your task is to find 20 jobs (or more) that you are interested in and compare the terms (words, phrases, and bigrams) in the job description to your own resume.

So, what kind of resume keywords should you include?

## Setup

* Download the job\_descriptions\_and\_resume.ipynb notebook it will contain the step by step instructions you need to follow.
* Download the **MSBA20 Job Descriptions.xlsx** excel document you will use it as a template as you gather the **20 + jobs** you are interested in.
* Grab **your resume** and save to PDF
  + we need this to be in PDF format to do the analysis

## Task 0 – Gather Jobs

Using the jobs, you’ve gathered and put into excel (use the template) read the excel file into Python. I’ve provided a template file called MSBA20JobDescriptions.xlsx which contains a sample of 20+ jobs. Please **gather your own job** list this is just for convenience.

## Task 1 – Job Description Term & Bigram Frequency Analysis

1. Create a table of term frequencies using Job Descriptions – be sure to **remove the junk.**
   1. Create a word cloud of terms, like this:

A close-up of words

Description automatically generated

* 1. Create a word cloud of terms that start with the letter “a”, like this:

A close-up of words

Description automatically generated

## Task 2. Words after DATA

Data is clearly an important word so what words come after data?

1. Create a bi-gram of Job Description terms starting with the word Data, group by bigram and count them up into a bigram\_term\_frequency table.

|  |  |
| --- | --- |
| word | n |
| data analytics | 8 |
| data analysis | 7 |
| data sources | 7 |
| data sets | 6 |
| data driven | 4 |
| data management | 4 |
| data quality | 4 |
| data analyst | 3 |
| data science | 3 |
| data solutions | 3 |

1. Create a word cloud of data + term combinations, by filter the bigram\_term\_frequency for bigrams starting with “data”, your word cloud should look something like this:

**A close-up of words

Description automatically generated**

1. Create a bar chart of the top 10, data + term combinations, like this:

A bar graph with different colored bars

Description automatically generated

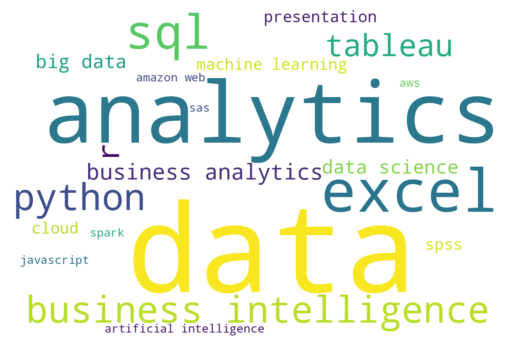
## Task 3. Analyze Technology

1. I’ve provided a sample set of **technology terms** and **bigrams** that you can use as a filter in the notebook. Please add additional as you see fit!
   1. Create a term frequency table
   2. Create a bigram frequency table
2. Filter the term\_frequency table based on the technology\_words provided plus any you’ve added.
3. Filter the bigram\_frequency table based on technology\_bigrams provided plus any you’ve added.
4. Smash the results together into technology\_term\_frequency. you might want to rename things so that words and bigrams match up into a single column. The function pd.concate() will help you append the two results together.
   1. Make as bar chart like this…

A bar graph with different colored bars

Description automatically generated

* 1. Make a word cloud like this…



## Task 4. Your resume

1. Import your resume using pdfminer library and the function extract\_text()
2. Parse your resume, filter out the common words, digits, and any words / terms you want to exclude.
3. Create a word cloud of the remaining words in your resume, for example here’s my resume word cloud, clearly data is a common word in my resume maybe that’s something I want to address?

A close-up of words

Description automatically generated

1. Filter for technology\_words and technology\_bigrams smash the results together and make the following (just like in task 3)
   1. A bar chart of term frequency, like this:

A graph with different colored bars

Description automatically generated

* 1. A wordcloud of term frequency, like this:

A close-up of words

Description automatically generated

## Task 5. Final writeup.

In a **Word Document**, write up a summary based on the analysis of both sets of jobs, what stood out? What are the key skills that you think you’ll need to master. Were there any terms like AWS that you are unfamiliar with? Based on your current resume, can you identify any keyword gaps, for example, SQL something that I’m good at (really good at) but it’s completely missing from my resume. Also based on the word cloud it’s clear to me that my resume is over indexed on SAS & Hadoop skills, which unfortunately for me are dead technologies. Most importantly how could you or what kind of analysis could you think of to improve your analysis? Be sure to touch on the following. Keep it simple, under a page, just easy to read and understand.

* What are the top 10 terms / words that are in common between your resume and the jobs you’ve found?
* What are the top 5 technology words and bigrams that are in common between your resume and the job descriptions?