# Talent Acquisition Data Analysis Findings

## Comments to the Data

The work with the data was performed mostly in Jupyter Notebook with Python/Pandas. Some additional work was also performed in Excel. The choice of Jupyter Notebook and Python over Excel is because it's easier to perform reproducible reports and the analyses can be done both on an aggregated level and on a for example regional level with very little change in the code. The data visualization was in this case performed in Excel.

## Comments to the Dashboard

The goal for this dashboard in to provide a bird's eye view of the efficiency of the Talent Acquisition process on a monthly basis to help guide decision making.

To be able to make this dashboard some assumptions were made:

- Target audience: Head of Recruitment and Head of HR
- Secondary audience: Other executives and HR team
- A recruiter is responsible for all the roles within a position category.
- To guide decision making, we need to focus in on a specific, current month. In this case April.

## Takeaways/Key findings

The data given spans from January 11th to September 2nd and consists of data from Application questions, Applications and additional Source Mapping information.

In April there was 4,658 applications registered and 98 active job postings (based on different job titles that were applied for). The volume of applications for April is quite normal. From February to August the average of application per month was 4,561. August and February had the least applications while May and June had the most. Available positions and the number of applications is highly correlated over time with a slight lag to the following month.

Most applications (1,456) came to positions in the category Home Office, which does not have a specified region. Among the regions Newark received the most applications and Troy the least. This can probably be explained by the size of the regions, since Newark seem to have 14 schools and Troy only 2 schools.

In April in total 62 candidates got hired and 4,151 got rejected. 55 percent of the rejected candidates were rejected because they 'Did not meet screening requirements'.
392 candidates withdrew their applications and the most common reason for the withdrawal was 'Not responsive to recruiter'. In this case they were probably no longer interested in the position.

18 applications that came in in April were still in progress when April ended. 11 of them were in the In-review state and 7 in the Interview state. Most of them where in the 'In-review: On Hold: Pre Screen' state. No applications were in the state Lead, New or Offer. In April, but also in general for the months analyzed, the applications on average spent very short time in the state

Lead and New. This is good, it shows that the applications are attended shortly after they have been submitted. The applications stayed on average most of the time in the state Interview and Offer. This also make sense since for the interviews and the offers you have to take into consideration that interaction with the candidate have to take place which require waiting time for the candidate to get back. In April the average time in the Application process was 36 hours, for February to August the average is 38 hours.

If you break down the applications per position categories there are also some interesting insight to be found. The number of applications must be looked at in context with number of job postings for each category. The application per posting ratio shows that the most popular categories to apply for are College Completion, General Teacher and Operational Leadership. Seen to rejection rate Regional Support Staff and Summer Teaching Fellows have the highest rate. The lowest rejection rate is found with Core Teachers. It's because there are a lot of open positions and not that many applications per posting. This might be a sign that Core Teachers are high in demand, but that they are low in supply. For more information about applications per position category please see the chart in the dashboard.

In April the most applications came via Job Boards, for example Indeed or LinkedIn. Applications from Job Boards takes up the most time in the Applications process in total, but as mentioned earlier it also generates the most applications. So, seen to the time per application for each application channel Job Boards have among the lowest time per application. Applications via Fairs, Internal, Referrals and Smart Recruiters however have the most time per application. For Internal it seems to result in a very low rejection rate. Fairs and Smart Recruiters have a medium high rejection rate, while Referrals tend to generate a slightly higher rejection rate. To make strategic decisions only based on applications and time for different channels in one month is not a wise move. This is something that typically has to be studied over time. If number of applications and rejection rate per channel is studied over a 3-month span (February to April) it appears that some results are pretty much the same from month to month while some tend to fluctuate a little. Ultimately the channels should be studied over an even longer time span and if possible, comparisons should be made from month to month per year to spot seasonal trends.

## Comments about the Data

There seem to be a lot of room for improvement of the data. Below are some of the key issues addressed.

**Data Quality**

- There's a lot of missing data. Can this problem be addressed or explained in some cases? I need more info about the data collection process to be able to suggest changes.
- Some Job Titles have 'Do Not Use' in the title, but they still seem to have been used and there are applications registered to them. I do not have enough information about these job titles to suggest an improvement. Also, these job titles are at the moment included in the statistical findings since they seemed to have been used.
- Fundamentals: Make sure that there are no unintended duplicates (ex. in Source Mapping tab Indeed is duplicated in Source column) and that all columns in the excel sheet have a title that is understandable and structured (ex. three columns in the Applications tab miss titles). Duplicates make the analysis complicated and can also be a problem when

combining data from different sources. When there are missing titles the user has to guess the column content, which in some cases can turn out wrong.

- Different ways to write the same thing. For example, there are 4 ways of writing 'On hold: Pre-Interview' in the Application Status column. If the data is analyzed without finding this problem the result might come out wrong. It's very time consuming to fix this kind of problem. Therefor it would be great if there was only one option for recruiters to use, in other words choose one standardized way to write it. This type of problem is recurring in the data.
- There are applications from early 2019 that still seems to be in the application process, not hired or rejected. This is a quality issue since either it has not been updated correctly or there is a lack of options to put the application in another state that feels right.

**Data Format**

- Multiple answer options seem to be available for some questions in the Applications Questionnaire. For example, the questions about how you heard about the company, the diversity questions and question about license. Within the bounds of legal compliance, it is suggested that the questions are re-written so that the candidate can only choose one alternative, the one that is the overall best match for the candidate.
- Give the Position Category Home Office its own Region label. At this point the Position Category Home Office and the Job Title pool (has no Position Category) has no region. It makes analysis harder when some columns systematically have been left empty.
- A date for when a job was posted and when a person was hired would be nice to have. You can somewhat make a pretty good estimate of this with the existing data, but it takes extra time. When you have a posting date it's interesting to look at when a candidate applied during the timespan the job posting was active.
- In the Application tab the column titles are on the second row. There is a non-working formula floating around in one cell on the first row. This creates a problem when converting the excel file from .xlsx to for example .csv, many programs need a .csv to be able to read the data.

Some changes can be made in how the data is collected, but some problems are easier to address with a line of code that corrects the problem. If a problem is addressed after the data is collected it's important that there is a clear documentation for the process. Also have in mind that it's not good to make changes in the data collection too often. If creating reoccurring reports and dashboard the same code can be reused if they have the same content and uses the same data. In case of changes it also makes it more difficult to make comparisons with earlier results.

To end with I would just like to say that with more time there are topics that would be interesting to explore further. Also, there was neither time nor space to address all of the problems with the data that I experienced during this brief analysis. I would gladly talk to you more about topics for further exploration and other data issues in person.