Data Scientist Job Change Analysis

Chuqiao Xiao

Objectives

The project aims to execute the exploratory data analysis based on a dataset regarding data scientists' job changing. The requirement is to use as many complex ways of data visualisation as possible while limiting the number of questions to 5.
The exploratory data analysis including factors relevant to the job changing decisions of data scientists were investigated.
The visualisations seek to create a detailed map of insights about how data scientists' behaviours vary in terms of job selection according to some factors. The key assumption is that this report is constructed from the perspective of an HR business analyst who is working in a start-up and is trying his/her best to explore the talent market of data scientists from an incomplete HR dataset.

Data Description

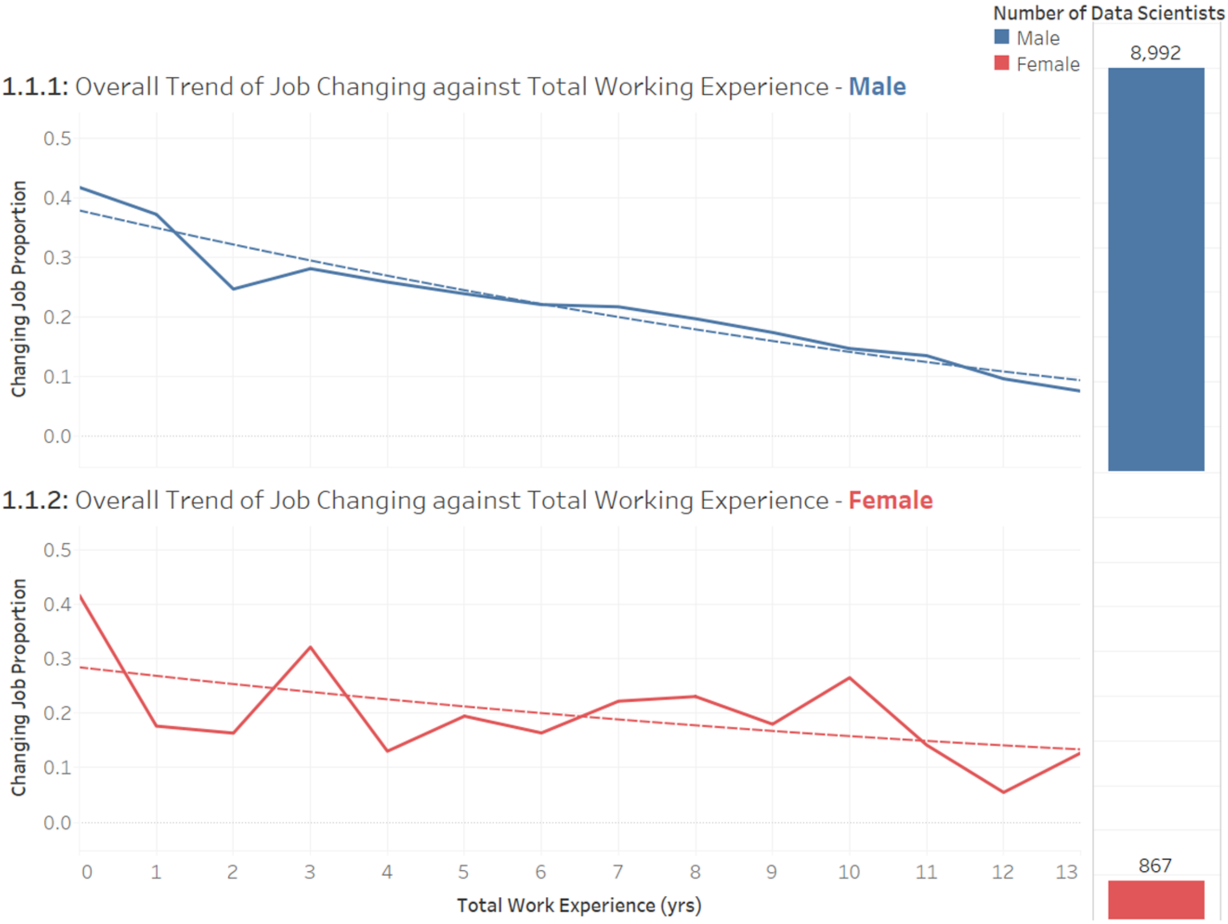
Data can be retrieved from Kaggle:
https://www.kaggle.com/datasets/arashnic/hr-analytics-job-change-of-data-scientists
Among which 19,158 records with 14 columns were used as the analysis data.

Data lexicon

enrollee_id : Unique ID for candidate
city: City code
city_ development _index : Developement index of the city (scaled)
gender: Gender of candidate
relevent_experience: Relevant experience of candidate
enrolled_university: Type of University course enrolled if any
education_level: Education level of candidate
major_discipline :Education major discipline of candidate
experience: Candidate total experience in years
company_size: No of employees in current employer's company
company_type : Type of current employer
last_new_job: Difference in years between previous job and current job
training_hours: training hours completed
target: 0 – Not looking for job change, 1 – Looking for a job change

Visualisation #1 (Section 1)

S1.Q1 – How does Data Scientists' Job Changing Intention Vary between Genders and Total Working Years?



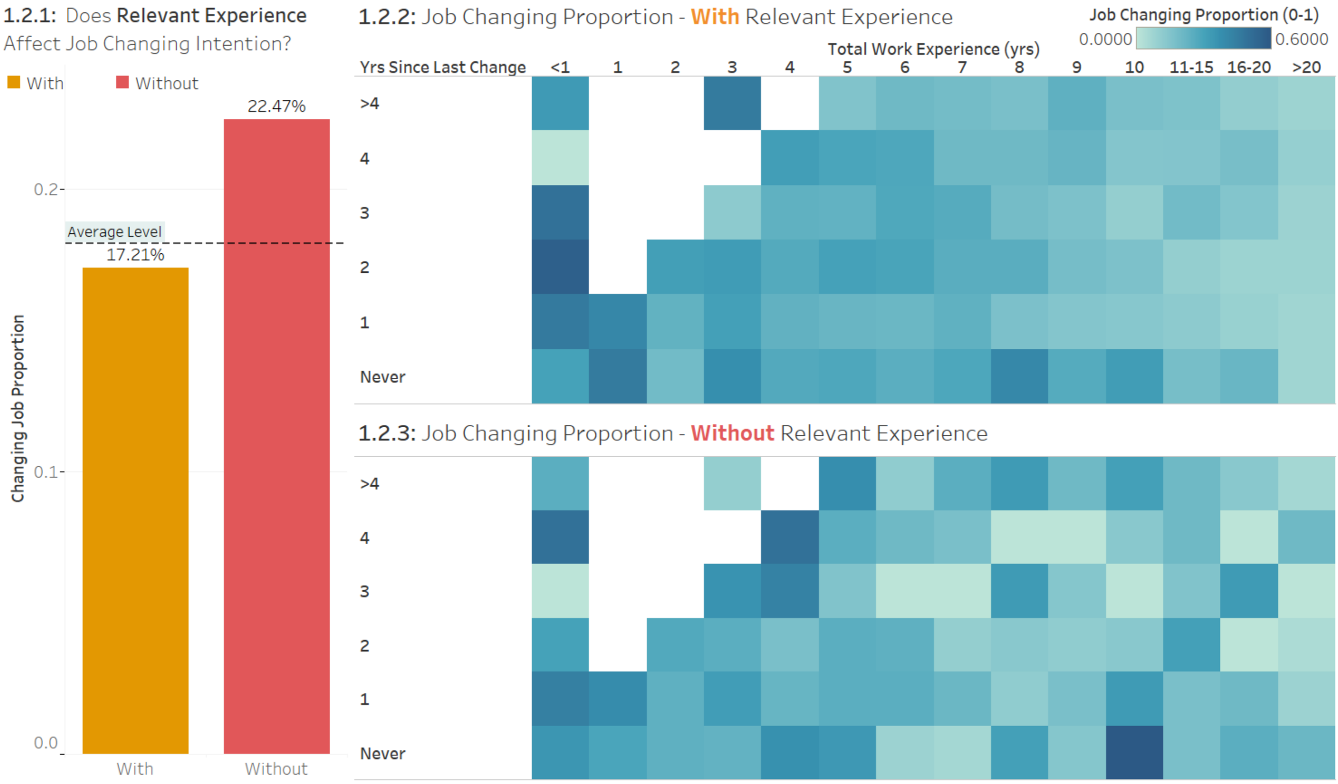
Companion description

In section 1, the job changing probability is investigated. Question 1 in S1 seeks to reveal the relationship between the ***job changing rate*** and employees' ***gender*** and ***total working experience***. The visualisation shown above are divided into upper and lower panels by ***gender***. Each panel shows both the ***number of employees*** on the right and the trend line of ***job changing rate*** against ***total working years*** on the left.

Firstly, according to the bar chart on the right of the diagram, there are much more ***males*** (8992) than ***females*** (867) in the market. Both line charts indicate a slow decreasing trend of ***job changing rate*** as employees' ***total years of working*** increase, this may indicate a lower intention to switch job in employees' later career. The line chart of ***females*** seems to be more volatile than the one of ***males***, which indicates that women may have a less stable career journey. Alternatively, it can be explained by a relatively limited ***female*** sample size which results in relatively more disperse samples.

Visualisation #2 (Section 1)

S1.Q2 – How Data Scientists' Working Experience Influences the Job Changing Intention from Different Dimensions?



Companion description

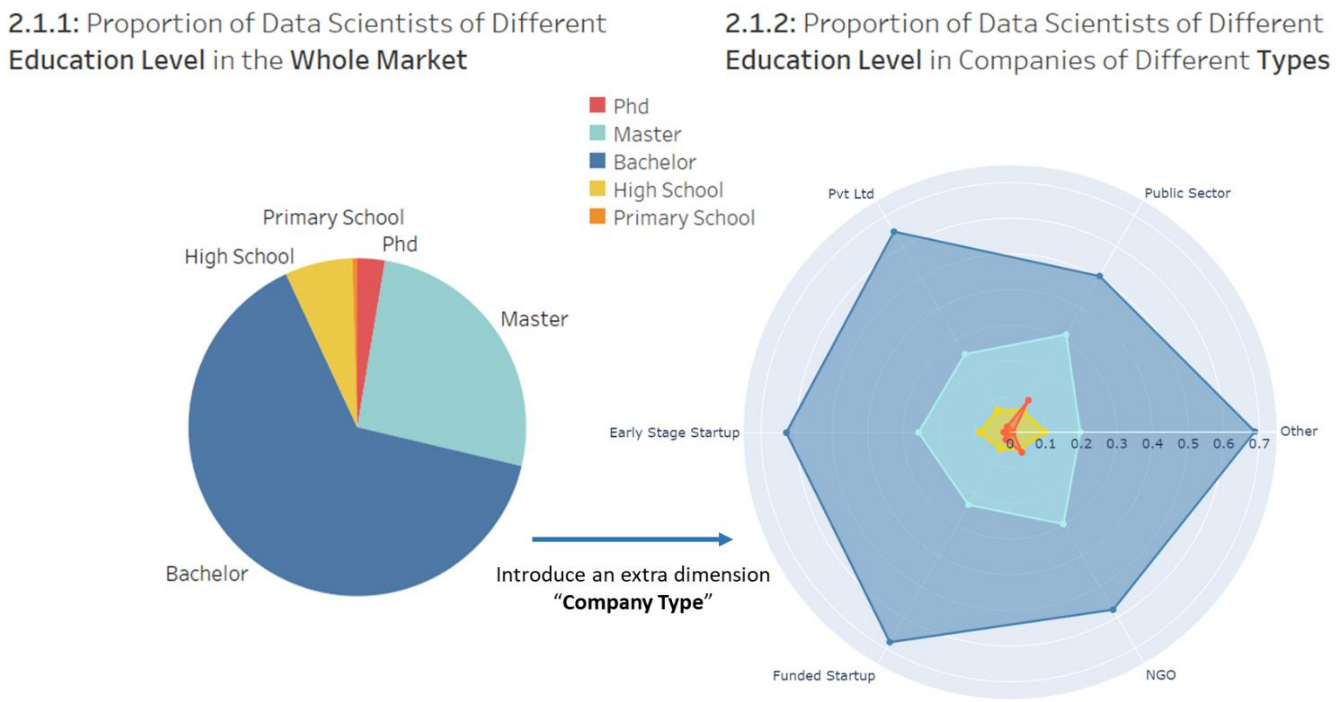
Question 2 in Section 1 seeks to further investigate how job changing rate varies among data scientists. Except the ***total years of working***, the ***relevant experience*** indicates whether an employee has relevant working experiences previously, the ***time intervals since last job changing*** indicates how many years are there since last time employees change jobs.

The bar chart on the left compares the variation in ***job changing rate*** of employees with and without ***relevant experiences***. The red bar of employees ***without*** *relevant experience* shows a significant higher ***rate of job changing*** (22.47%) than both ***average rate*** and rate of employees ***with*** *relevant experiences*. This indicates that employees without relevant experiences are more likely to switch jobs.

Heat maps on the right show how ***job changing rate*** vary within different combinations of ***total working experiences*** and ***years since last change***. The blank area above the diagonal line indicates on the top left implies that very few employees were unemployed, except candidates with working experiences less than 1 year. Overall, the lower graph of employees ***without relevant experiences*** shows a chaotic distribution of ***job changing rate*** than the upper graph. The colour fades out along the horizontal axis **total years of working** within both graphs. It can be inferred that behaviours of employees ***without relevant experiences*** are less predictable yet more likely to change jobs. This further consolidates the previous conclusion, which argues that experienced employees are more stable than unexperienced employees. No significant pattern regarding ***time intervals since last change*** is observed.

Visualisation #3 (Section 2)

S2.Q1 - How do Data Scientists Distribute in the Market?



Companion description

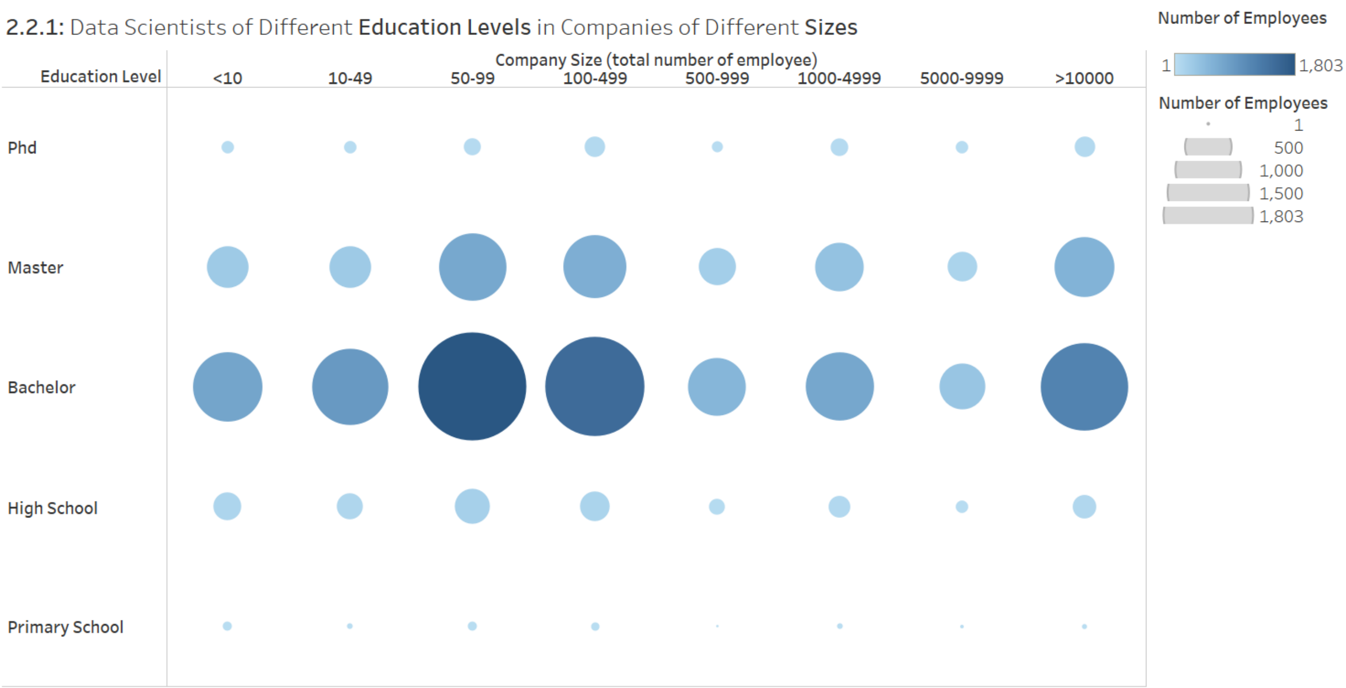
In the Section 2, behaviours of data scientists of different education backgrounds are investigated. Employees of higher education levels (Master, Phd) are main targets in this report as the fictitious company seeks to hire data scientists of high education backgrounds who can provide valuable insights. Question 1 in S2 aims to get an overview of distribution of employees of ***different backgrounds*** in the whole market and different ***types of company***.

Pie chart on the left indicates most of data scientists in the market have ***bachelor***'s degrees, followed by those with ***master***'s degrees. There are very few employees with high school and lower degree.

By introducing an additional dimension ***company type*** into the chart, a radar chart is displayed on the right. ***Proportions of employees*** of different ***education levels*** in different ***company types*** are shown above. In particular, the distribution of ***Phd*** employees show significant trend of working in ***public sector*** and ***NGO*** (non-profit organisation), as ***Master*** also shows a slightly similar pattern. This might because employees with ***higher educational degrees*** tend to pursue more meaningful achievements in their visions and leverage their knowledge to a larger cause. ***Bachelor*** is less likely to take roles in an NGO or a public sector.

Visualisation #4 (Section 2)

S2.Q2 – How Do Data Scientists Distribute in Companies of Different Scales?



Companion description

Question 2 in the S2 seeks to uncover the distribution of employees from different education backgrounds in different size of companies.

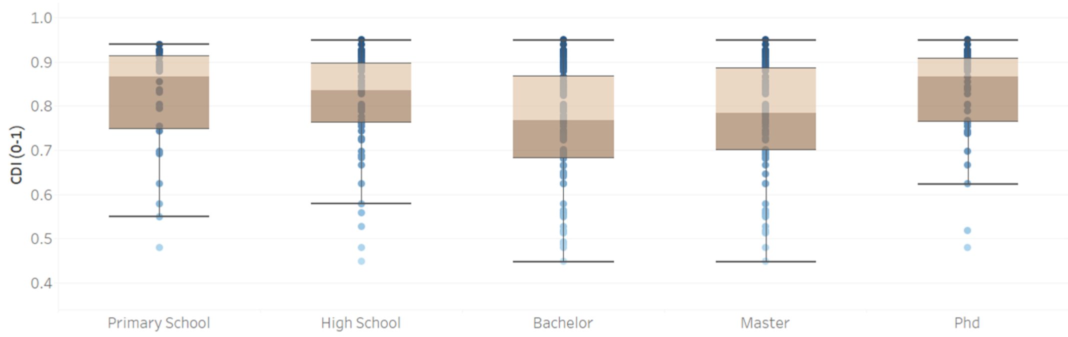
A bubble plot is used to show how ***number of data scientists*** varies through different combinations of ***education level*** and ***company size***. The darker and larger a circle is, the higher ***number of data scientists*** the type of company has.

Initially, it's anticipated that a ***larger company*** would have ***higher number of data scientists***. On the contrary, the visualisation indicates that companies of ***50 to 499 employees*** have the highest ***number of data scientists***, while larger companies have relatively lower proportions of data scientists. An explanation to this trend may be that the ***number of companies*** of different sizes in the original dataset is unequally distributed, thus resulting in an unequal distribution of number of data scientists in companies of different sizes. Besides, due to the incomplete information, it can't neither be identified if there are employees who work in the same company. According to the **limited information**, we may interpret the outcome as maybe a larger company have more outsourcing analysts compared with the smaller companies, as the scale requires them to operate in a more flexible way.

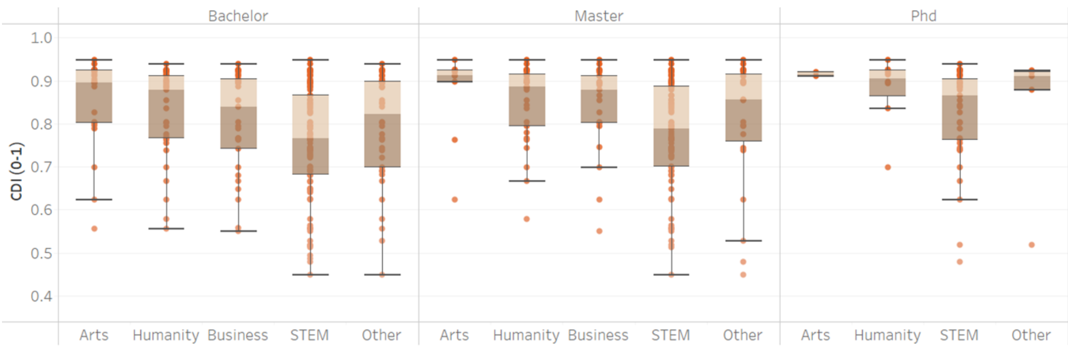
Visualisation #5 (Section 2)

S2.Q3 – Do Data Scientists Value the Development Level of the Working Locations (cities)?

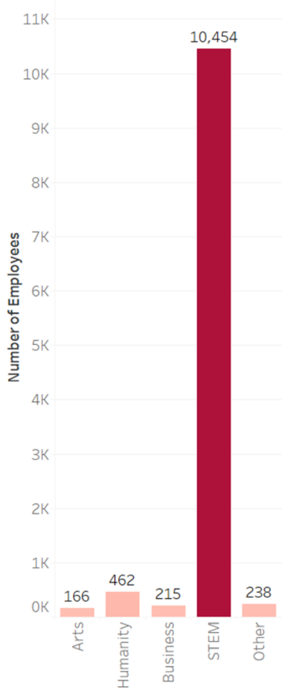
3.1: City Development Index (CDI) - **Education Level**



3.2: City Development Index (CDI) - **Professional Field**



3.3: Employee Numbers - **Professional Field**



Companion description

Question 3 aims to reveal relationship between employees' ***education backgrounds*** and their ***cities' development levels***.

Box plots of ***city development index (CDI)*** are categorised by ***education levels*** and ***professional fields*** on the left half of the panel, while the ***numbers of employees*** of different ***professional fields*** are shown by the bar chart on the right. ***City development index*** is an indicator for measuring the city development's level (Böhringer, 2007).
The bar chart on the right indicates an outstanding higher number (10,454) of ***STEM*** backgrounds employees than other backgrounds.
It's shown in the upper-left that employees of ***Bachelor*** and ***Master*** have a relatively lower ***average CDI*** than others. However, the causal relationship between these two variables is vague. This is because ***Masters*** and ***Bachelors*** are supposed to have stronger competencies than employees with lower education backgrounds, probably leading to a better living environment. An explanation may be that cities which offer more ***data scientist vacancies*** may have lower ***CDI***. The lower-left plot further uncovers their relationship by specifying the box plots into ***professional fields*** of ***Bachelor***, ***Master*** and ***Phd***. In these plots, majors other than ***STEM*** show a higher ***average level of CDI***. Although employees from STEM are supposed to have higher levels of skills compared to other majors, they still live in cities that have relatively lower CDI, which indicates that the different distribution in CDI is not totally related to professional competence and can be explained by other factors like personal preference. The variations may also be influenced by the limited size of samples of other majors.

References:

Böhringer, C. and Jochem, P.E.P. (2007) ‘Measuring the immeasurable — A survey of sustainability indices’, Ecological Economics, 63(1), pp. 1–8. doi:10.1016/j.ecolecon.2007.03.008.