Logo, company name

Description automatically generated

**Individual Coursework Submission Form**

Specialist Masters Programme

|  |  |
| --- | --- |
| **Surname: Xiao** | **First Name: Chuqiao** |
| **MSc in: Business Analytics** | **Student ID number:** |
| **Module Code: SMM635** | |
| **Module Title: Data Visualisation (PRD1 A 2022/23)** | |
| **Lecturer: Simone Santoni** | **Submission Date: 12/07/2022** |
| **Declaration:**  By submitting this work, I declare that this work is entirely my own except those parts duly identified and referenced in my submission. It complies with any specified word limits and the requirements and regulations detailed in the coursework instructions and any other relevant programme and module documentation. In submitting this work, I acknowledge that I have read and understood the regulations and code regarding academic misconduct, including that relating to plagiarism, as specified in the Programme Handbook. I also acknowledge that this work will be subject to a variety of checks for academic misconduct.  We acknowledge that work submitted late without a granted extension will be subject to penalties, as outlined in the Programme Handbook. Penalties will be applied for a maximum of five days lateness, after which a mark of zero will be awarded. | |
| **Marker’s Comments (if not being marked on-line):** | |

**%**

**Deduction for Late Submission: Final Mark:**

# Visualisation #1

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

## 

## Companion description (max 200 words)

The following visualisations aim to create a detailed map of insights about how data scientists’ behaviours vary in terms of job selection according to some of their features. Assuming the 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.

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, it should be noticed that 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. The line chart of ***females*** seems to be more volatile than the one of ***males***, while it can alternatively be explained by a relatively limited ***female*** sample size.

# Visualisation #2

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

## Q1.2

## Companion description (max 200 words)

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%) compared with both ***average rate*** and rate of employees ***with*** *relevant experiences*.

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 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. No significant pattern regarding ***time intervals since last change*** is observed. It can be inferred that behaviours of employees ***without relevant experiences*** are less predictable yet more likely to change jobs.

# Visualisation #3

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

## 

## Companion description (max 200 words)

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.

By introducing an additional dimension ***company type*** into the chart, a radar chart is generated 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*** has the highest proportion working in funded start-ups and early-stage start-ups, which indicates that start-ups may tend to hire undergraduates.

# Visualisation #4

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

## Q2.2

## Companion description (max 200 words)

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

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

## 

## Companion description (max 200 words)

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).

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. 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.