

Monash University: Assessment Cover Sheet

Student name	Goh	Kai Yuan	
School/Campus		Student's I.D. number	30881919
Unit name	FIT3179 Data visualisation - S2 2022 MUM		
Lecturer's name		Tutor's name	Dr. Grace Chai-Wen
Assignment name	Data Visualisation II Report	Group Assignment:	No
Note, each student must attach a coversheet			
Lab/Tute Class:	4	Lab/Tute Time:	8-10am
Due date:	17-10-2022	Submit Date:	17/10/2022
		Word Count:	956
		Extension granted	<input type="checkbox"/>

If an extension of work is granted, specify date and provide the signature of the lecturer/tutor. Alternatively, attach an email printout or handwritten and signed notice from your lecturer/tutor verifying an extension has been granted.

Extension granted until (date):/...../..... Signature of lecturer/tutor:

Late submissions policy	Days late	Penalty applied
Penalties apply to late submissions and may vary between faculties. Please refer to your faculty's late assessment policy for details.		

Patient/client confidentiality: Where a patient/client case study is undertaken a signed [Consent Form](#) must be obtained.

Intentional plagiarism or collusion amounts to cheating under Part 7 of the Monash University (Council) Regulations

Plagiarism: Plagiarism means to take and use another person's ideas and or manner of expressing them and to pass these off as one's own by failing to give appropriate acknowledgement. This includes material from any source, staff, students or the Internet - published and unpublished works.

Collusion: Collusion means unauthorised collaboration on assessable written, oral or practical work with another person. Where there are reasonable grounds for believing that intentional plagiarism or collusion has occurred, this will be reported to the Associate Dean (Education) or nominee, who may disallow the work concerned by prohibiting assessment or refer the matter to the Faculty Discipline Panel for a hearing.

Student Statement:

- I have read the university's Student Academic Integrity [Policy](#) and [Procedures](#)
- I understand the consequences of engaging in plagiarism and collusion as described in Part 7 of the Monash University (Council) [Regulations](#) (academic misconduct).
- I have taken proper care to safeguard this work and made all reasonable efforts to ensure it could not be copied.
- No part of this assignment has been previously submitted as part of another unit/course.
- I acknowledge and agree that the assessor of this assignment may, for the purposes of assessment, reproduce the assignment and:
 - i. provide it to another member of faculty and any external marker; and/or
 - ii. submit to a text matching/originality checking software; and/or
 - iii. submit it to a text matching/originality checking software which may then retain a copy of the assignment on its database for the purpose of future plagiarism checking.
- I certify that I have not plagiarised the work of others or participated in unauthorised collaboration or otherwise breached the academic integrity requirements in the Student Academic Integrity [Policy](#).

Date: 16/10/2022 Signature: KAIYUAN *

Privacy Statement:

For information about how the University deals with your personal information go to <http://privacy.monash.edu.au/guidelines/collection-personal-information.html#enrol>

Data Science Job Salaries

Assignment 2 Report

Name: Goh Kai Yuan

ID: 30881919

Tutorial Class: 4

URL: <https://kgoh0018.github.io/30881919-Assignment2/>

Number of words: 956

b. A brief description of the domain, Why and Who

The domain that I have chosen is Job. To be more specific, I have chosen Data Science as the theme of this visualization. Data Science is one of the upcoming popular jobs. It is a field of study that combines domain expertise, programming skills, and knowledge of mathematics and statistics to extract meaningful insights from data. These insights can then be used by the company or individual to improve on their sales and their reputation.

As a soon-to-be Data Scientist, I feel that there is a need to make a visualization out of the dataset that I have found on Kaggle. This will provide those Future Data Scientist to have an insight on the salaries that is going to be earned by them.

c. What: A brief description of the data

The link to the dataset is: <https://www.kaggle.com/datasets/ruchi798/data-science-job-salaries>

The author is RUCHI BHATIA (Bhatia, 2022), and it has been upvoted for 1255 times by the time of writing this report. Moreover, it has a usability score of 10 and thus, I know that it is a trusted source of dataset. As per download, I have to extract some columns out to have a cleaner data to plot the choropleth map. Moreover, I also check if there is any NA values or irrelevant values in the dataset. The cleaning data process is being done on RStudio which is a platform to perform R language coding.

Dataset Type: Table

It has items and attributes

Qualitative:

1. employment_type
2. job_title
3. salary_currency
4. employee_residence
5. company_location

Quantitative:

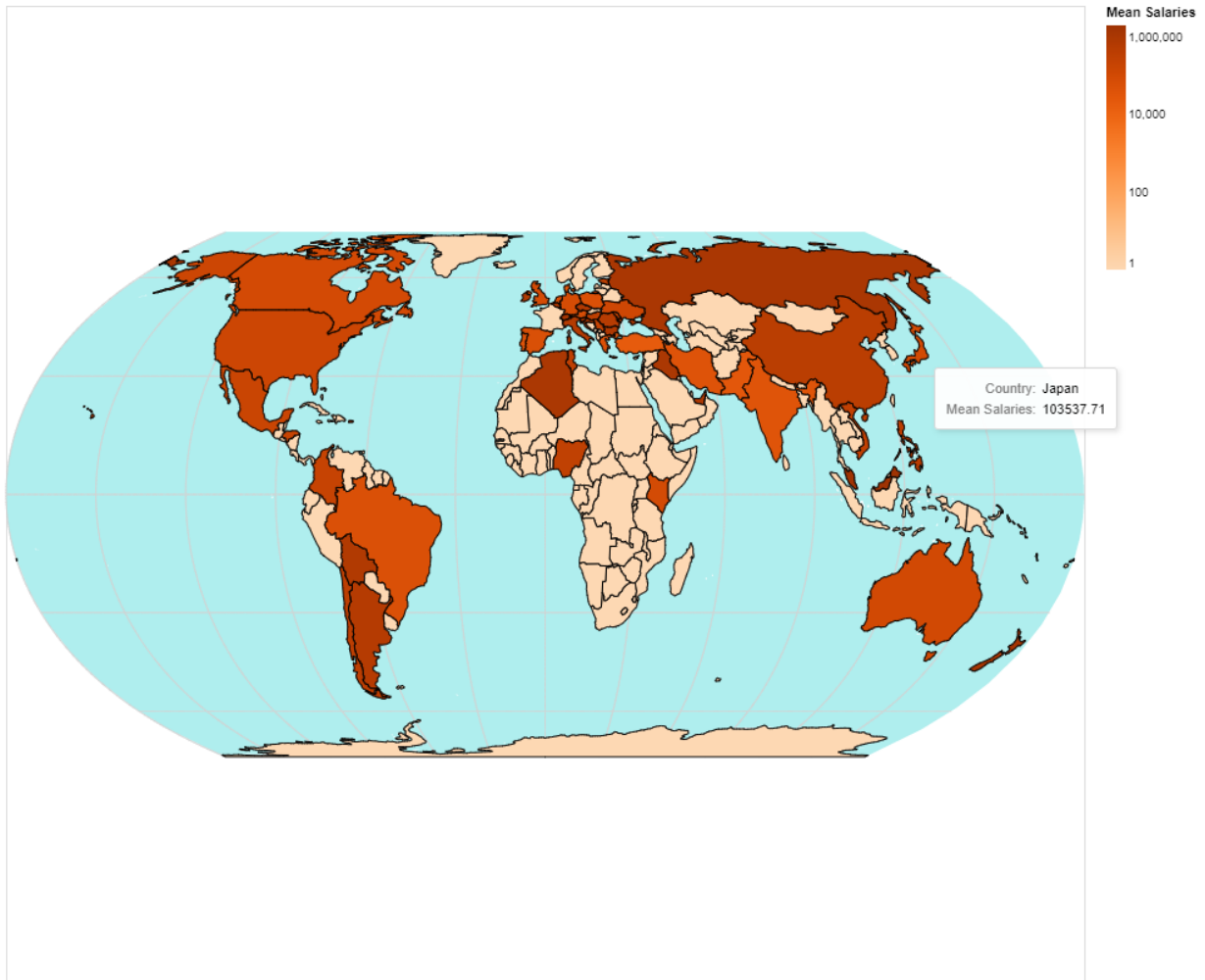
1. salary
2. salary_in_usd
3. remote_ratio

Ordinal:

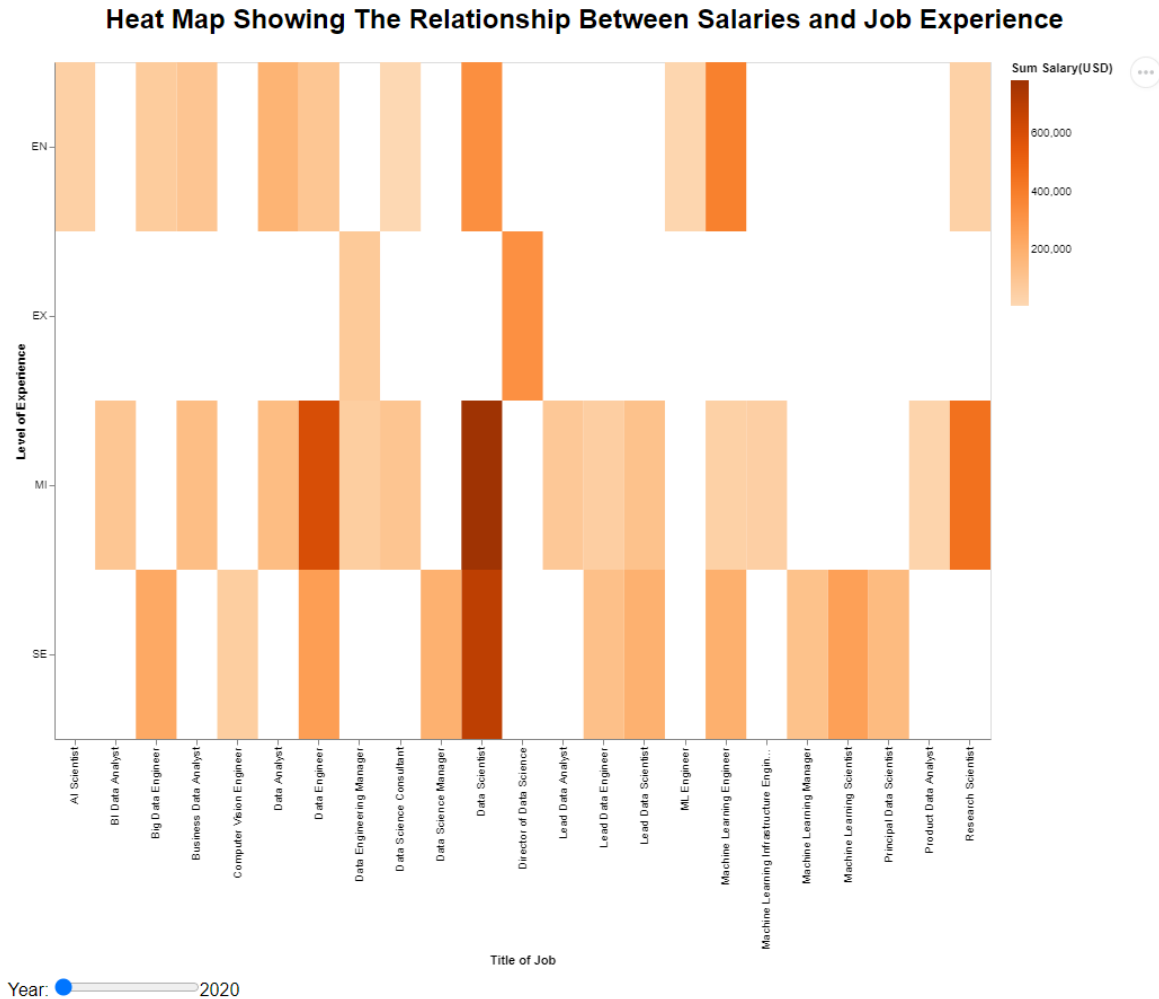
1. work_year
2. experience_level
3. company_size

d. Why and How:

Choropleth Map of Countries Having Data Science Related Job and Salaries

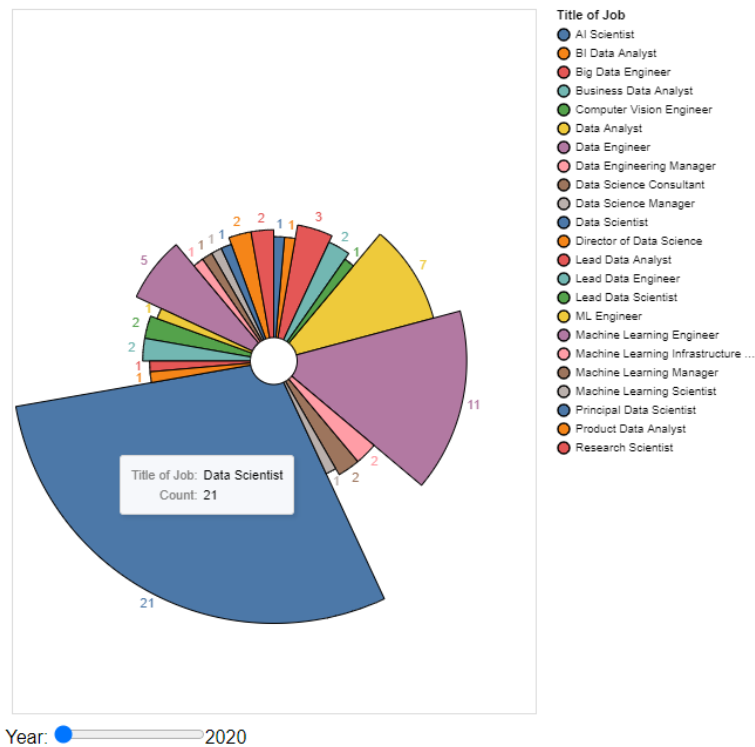


This choropleth map allows me to visualize the mean salaries earned by each country easily. This choropleth map easily tells users where they should look for a data science related job and how much they can earn in a year. However, this depends on the job experience and job title that the user has which I am going to discuss later. The map is sourced from a website called [naturalearthdata.com](https://www.naturalearthdata.com). The map earth easily helps us me to fill up the countries where there is data scientist job according to the dataset.



This heat map easily tells user how much they earn if they have any job title and experience level. This heat map consists of a qualitative data which is the Title of Job, ordinal data which is the Level of Experience and quantitative data which is the Sum Salary (USD). I have learnt to plot this graph by referring to the VegaLite Github page's example (*Table Heatmap*, 2022). Heat map is suitable when I want to compare between different components of attributes. This will help me get an overall relationship between those attributes. For users, they can easily find out which job title is earning the most at which experience level in which year and see a trend from it.

Pie Chart Showing The Count of **Data Science** Related Jobs

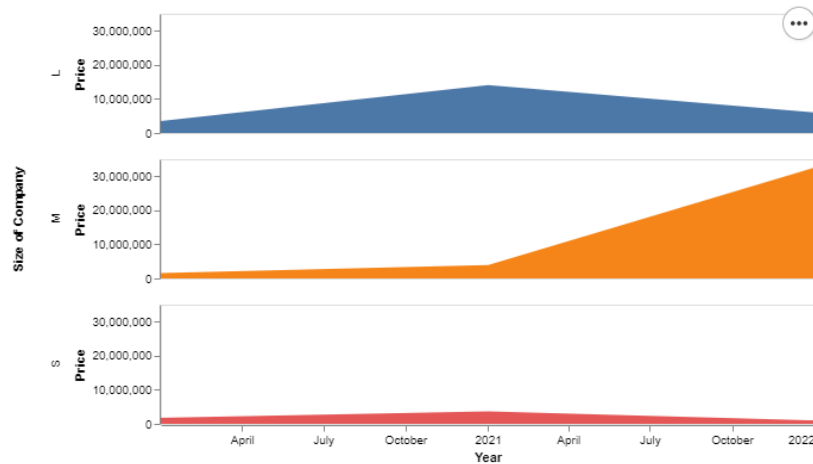


Data Science related job offers various kind of jobs and these jobs are more technical than the others. These jobs are offered much higher pay than the others. As such, various kind of factors such as **experience level** affects the salary that are offered by the company.

As we can see from the pie chart, Data Scientist has the highest number of employment, followed by Data Engineer and Data Analyst. These jobs can be seen that it is more attractive by the people nowadays. The sum of salaries for Data Scientist is the highest which we can see from the heat map above. It is also shown that no matter the experience level, Data Scientist has the majority of the salaries.

I am choosing this specific idiom which is a radial plot to help me visualize the number people that has the job title that is being employed in each year. I have learnt to plot this graph by referring to the VegaLite Github page's example (*Radial Plot*, 2022). The data that is being used here is Title of Job which is the qualitative data and the Count (count of job_title) which is the quantitative data. The reason to create this visualization is to guide user to see which job title has the highest number of people that is being employed. From my heat map visualization, I cannot determine how much people is working under the job title and this radial plot solves that problem.

Relationship Between The Company's Size and Salaries



The salaries also depends on the size of the company as we all know bigger company offers higher salary. However, according to the graph, it seems like medium size companies are offering much higher salary than larger size company for **Data Science** related job.

This may due to the higher number of employees working on the same project, resulting in lesser share of salaries. The other cause might be due to COVID-19 which larger size companies might suffer more lose compare to medium size companies

I have chosen this specific idiom which is a multiple area chart to help me visualize the relationship between the company's size and salaries. I have learnt to plot this graph by referring to the VegaLite Github page's example (*Trellis Area*, 2022). The data that is included in this chart is 2 ordinal attribute which is Year and Size of Company and 1 quantitative attribute which is the Price. Users can use this information to determine the size of the company they want to enter so that they can earn much more than others.

e. Design

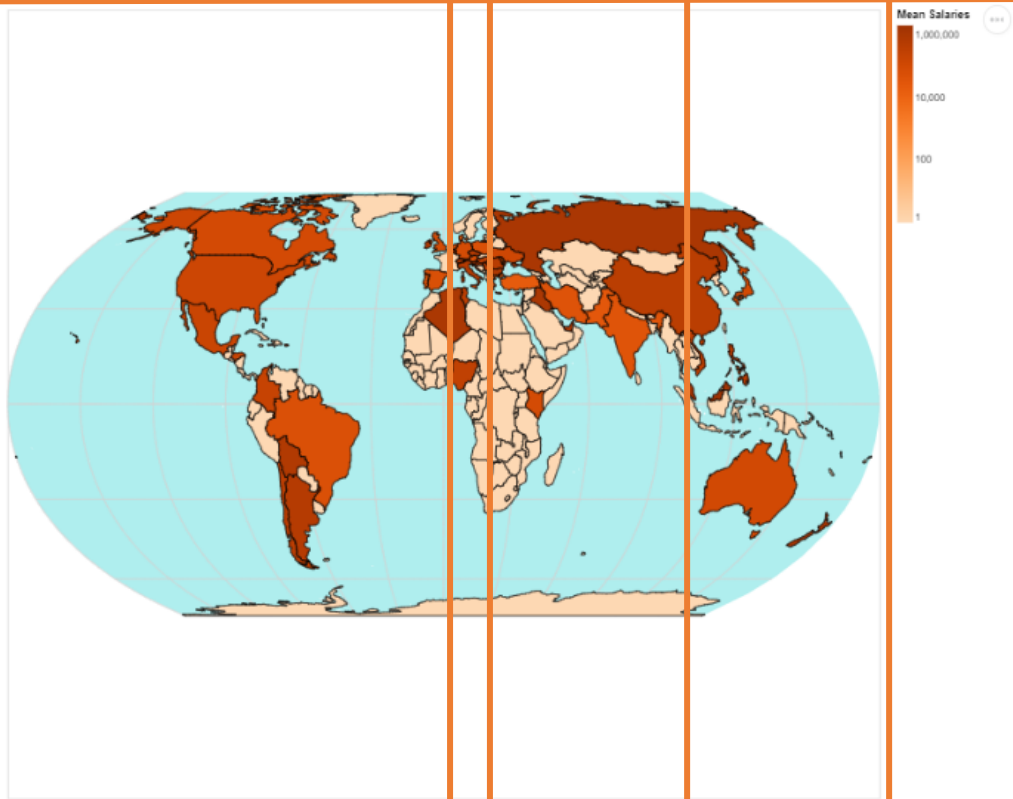
Layout

Data Science Related Job Salaries

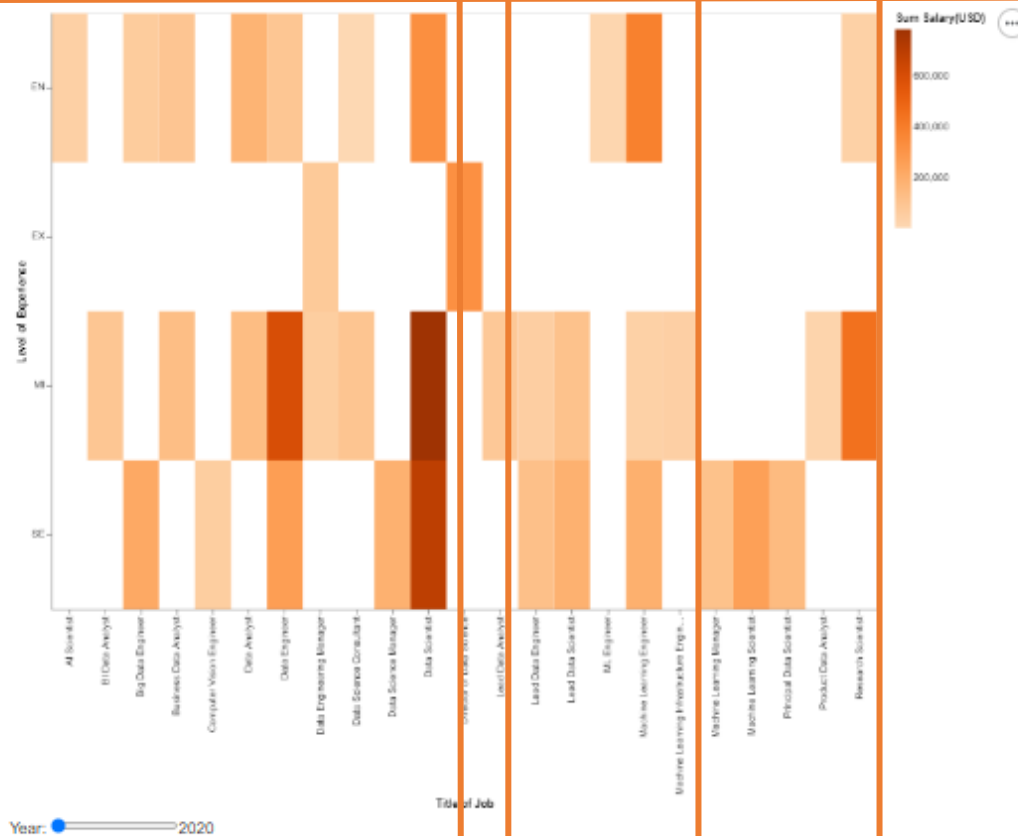
Data Science is one of the upcoming popular jobs. It is a field of study that combines domain expertise, programming skills, and knowledge of mathematics and statistics to extract meaningful insights from data. These insights can then be used by the company or individual to improve on their sales and their reputation.

Data Science related job has become more and more popular and it has more demand compared to the last previous years. At some countries, some companies even offer high salary pay for their employees. Different country receive different salary payment depending on the demand of Data Scientist.

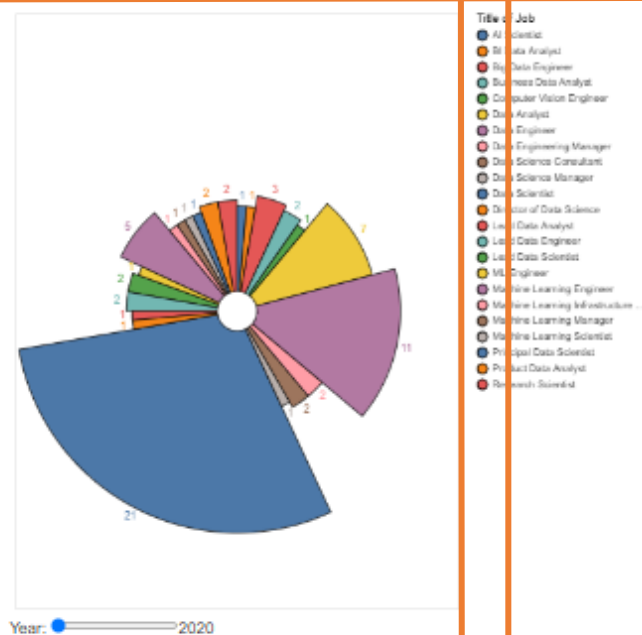
Choropleth Map of Countries Having Data Science Related Job and Salaries



Heat Map Showing The Relationship Between Salaries and Job Experience

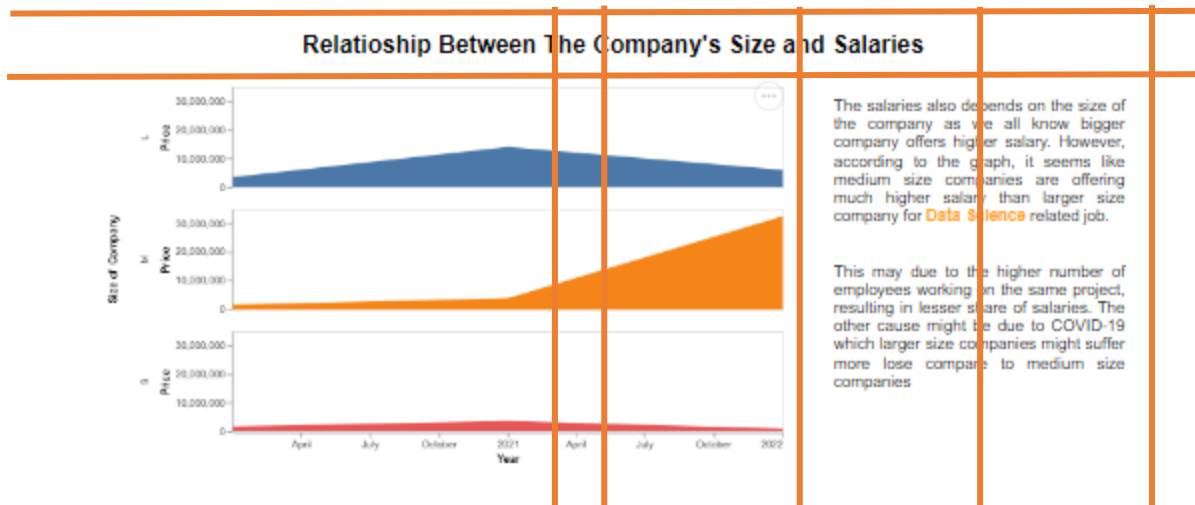


Pie Chart Showing The Count of Data Science Related Jobs



Data Science related job offers various kind of jobs and these jobs are more technical than the others. These jobs are offered much higher pay than the others. As such, various kind of factors such as **experience level** affects the salary that are offered by the company.

As we can see from the pie chart, Data Scientist has the highest number of employment, followed by Data Engineer and Data Analyst. These jobs can be seen that it is more attractive by the people nowadays. The sum of salaries for Data Scientist is the highest which we can see from the heat map above. It is also shown that no matter the experience level, Data Scientist has the majority of the salaries.



My choropleth map is the Visual Center and I have used the most minimum amount of sight lines to build my visualization. The visualization is very balanced arranged. There are no sudden big or small elements that mess up the balance of the whole visualization.

Colour

The colour I have chosen is colours that are available to be seen by a vision-impaired person. I did not use any mix of red and green color together for the charts. The word Data Science is being bold with orange colour to tell the users that Data Science is the topic in this visualization.

Figure-ground

I create a visual hierarchy by having bolder text for the important words such as experience level in my visualization. Title is also being bold, and it is bigger than other texts to show the users that this is the title of this visualization.

Typography

For the typefaces, I have used Open Sans for my entire visualization. As Open Sans is part of Sans Serif, it has better readability on diagrams and maps.

Storytelling

The genre of my visualization is annotated chart. For my visualization, it is mainly about the relationship between the Salary and the Job Title. Choropleth map is mainly visualizing the country

that has Data Science related job. My visualization also shows that in each of the job title, which is the most popular job title and which job earn the most according to the experience level.

Reference

Bhatia, R. (2022). *Data Science Job Salaries*. Kaggle.com.

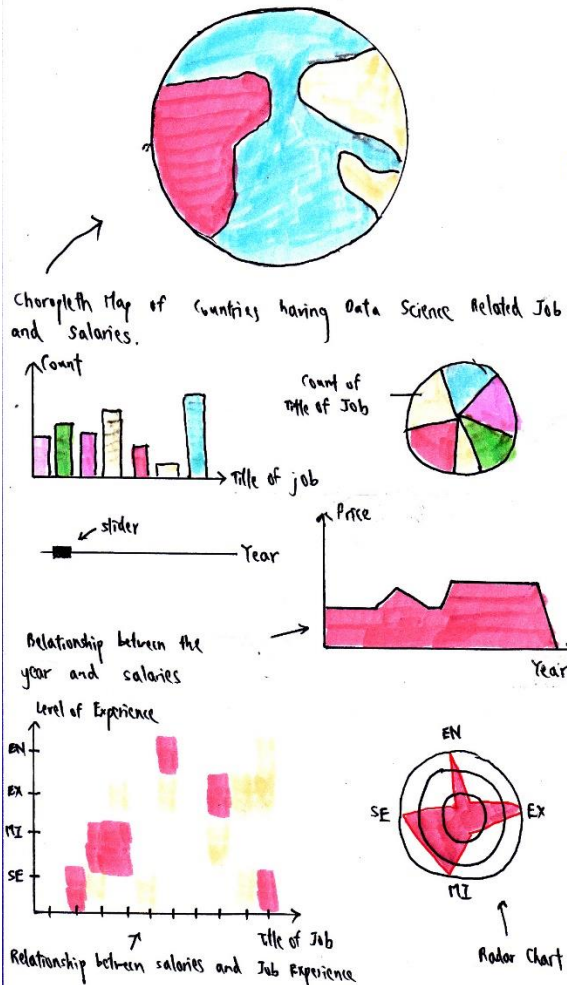
<https://www.kaggle.com/datasets/ruchi798/data-science-job-salaries>

Radial Plot. (2022). Vega-Lite. https://vega.github.io/vega-lite/examples/arc_radial.html

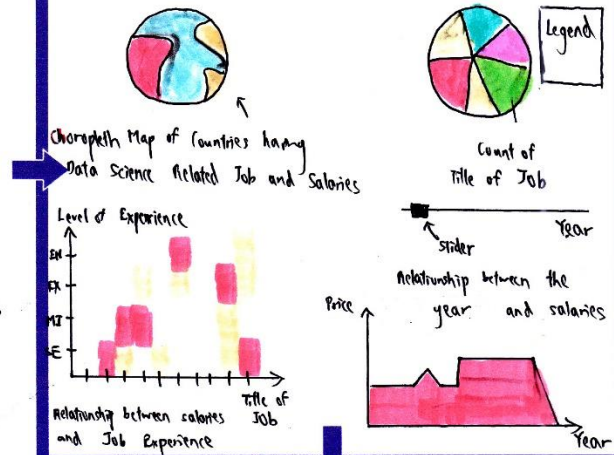
Table Heatmap. (2022). Vega-Lite. https://vega.github.io/vega-lite/examples/rect_heatmap.html

Trellis Area. (2022). Vega-Lite. https://vega.github.io/vega-lite/examples/trellis_area.html

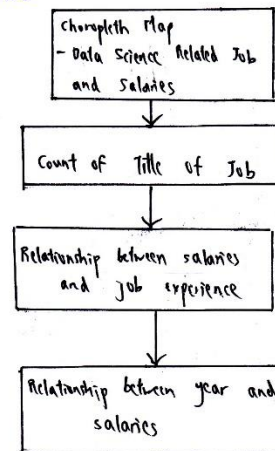
1. Ideas



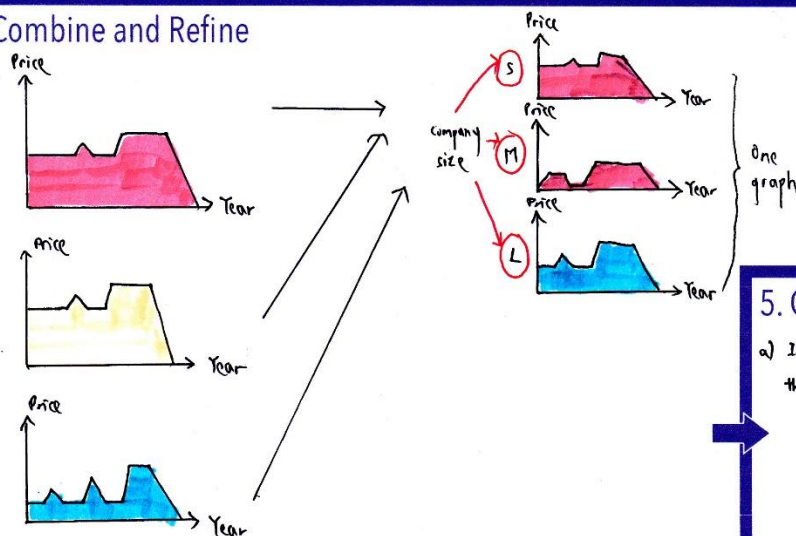
2. Filter



3. Categorize



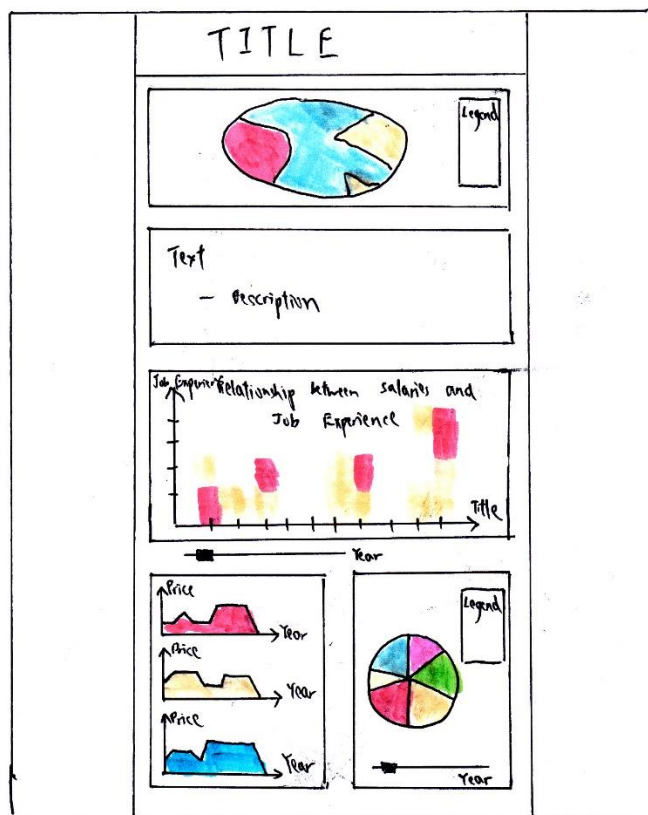
4. Combine and Refine



5. Question

a) Is it available to implement the visualization

Layout



Title: Dashboard view

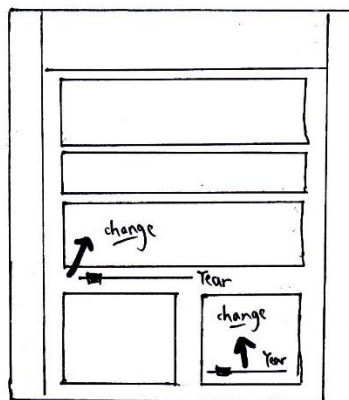
Author: Goh Kai Yuan

Date: 11/10/2022

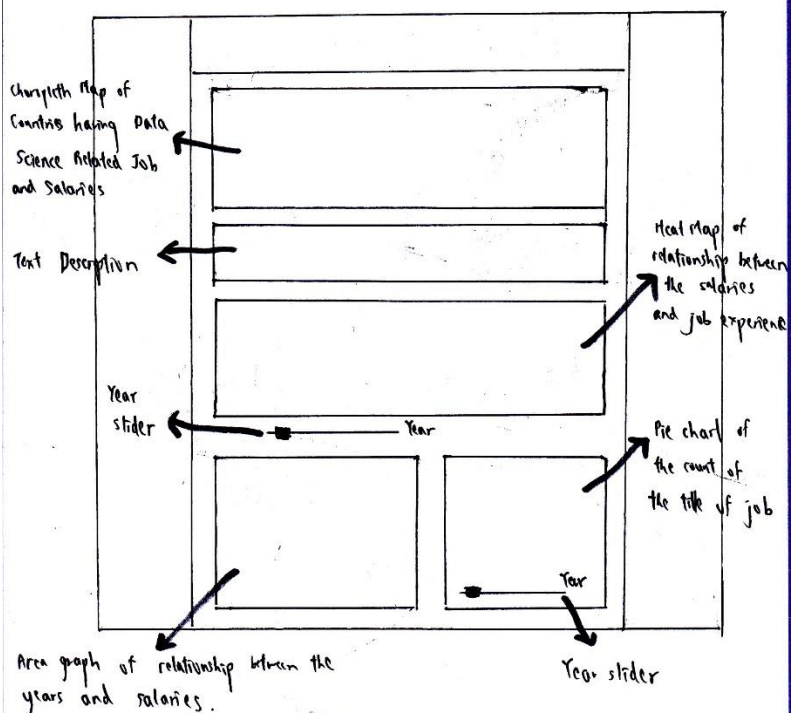
Sheet: 2

Task: Data Science Job Salaries

Operations



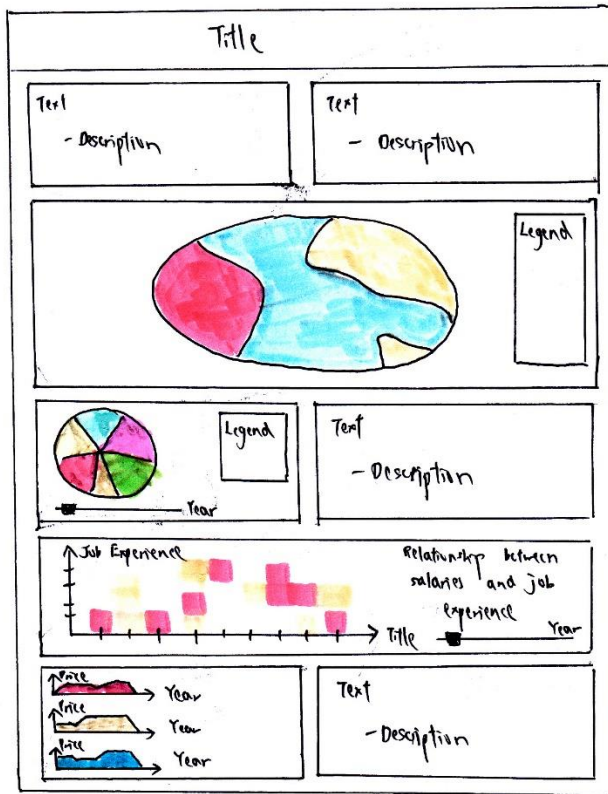
Focus



Discussion

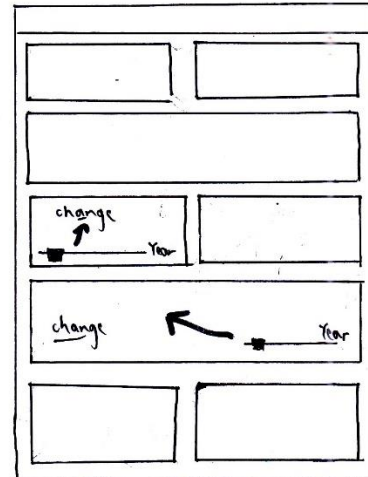
- Symmetry
- No visual center
- Little description about the graph
- Storytelling is abit vague.

Layout

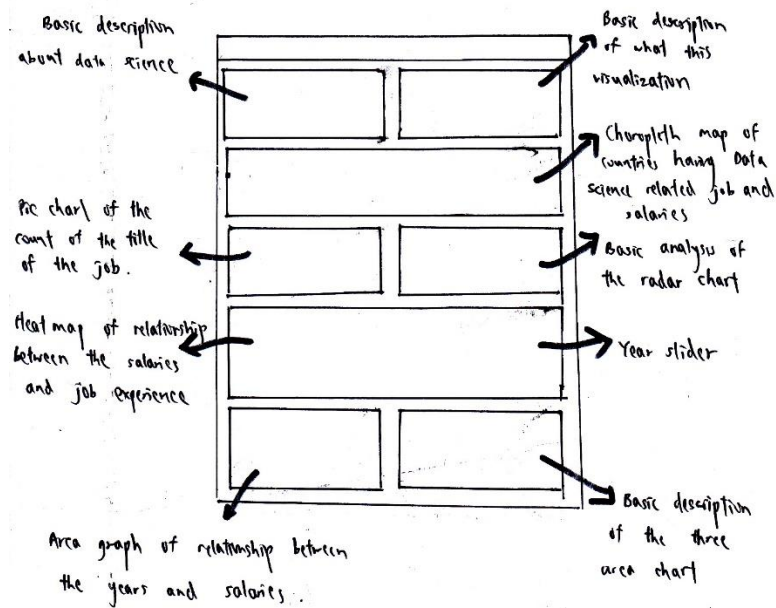


Title: Dashboard view
 Author: Goh Kai Yuan
 Date: 17/10/2021
 Sheet: 3
 Task: Data science Job Salaries

Operations



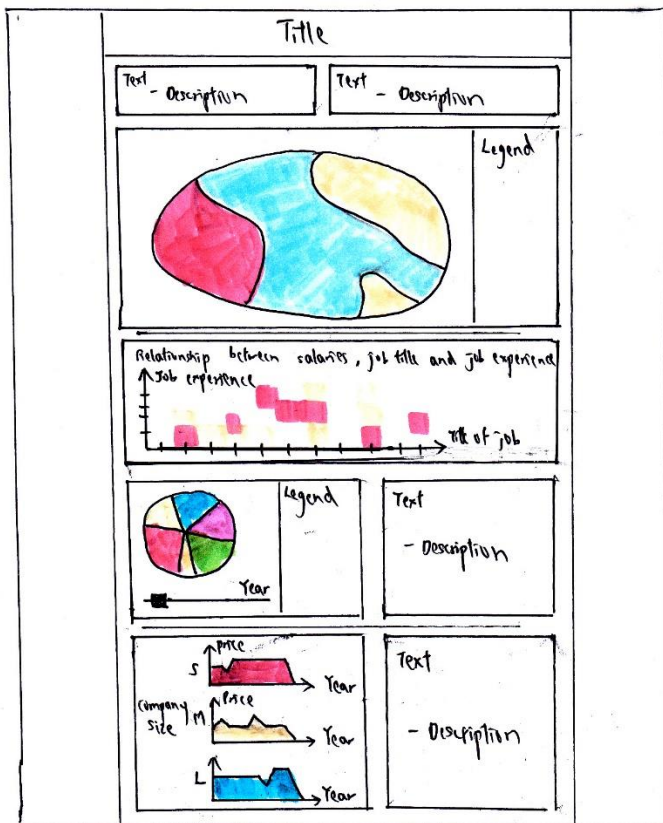
Focus



Discussion

- Symmetry
- Lesser sightlines
- No visual center
- More neat for all the placement and can get a goodflow of the storyline.

Layout



Title: Dashboard view

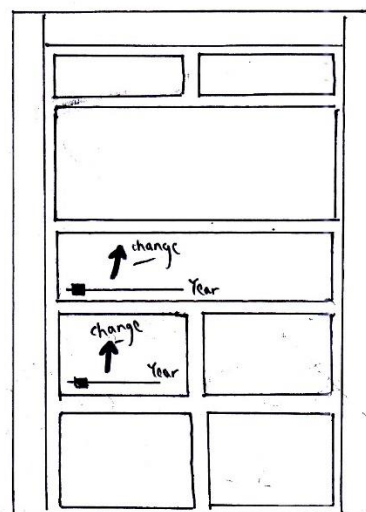
Author: Goh Kai Yuan

Date: 17/10/2022

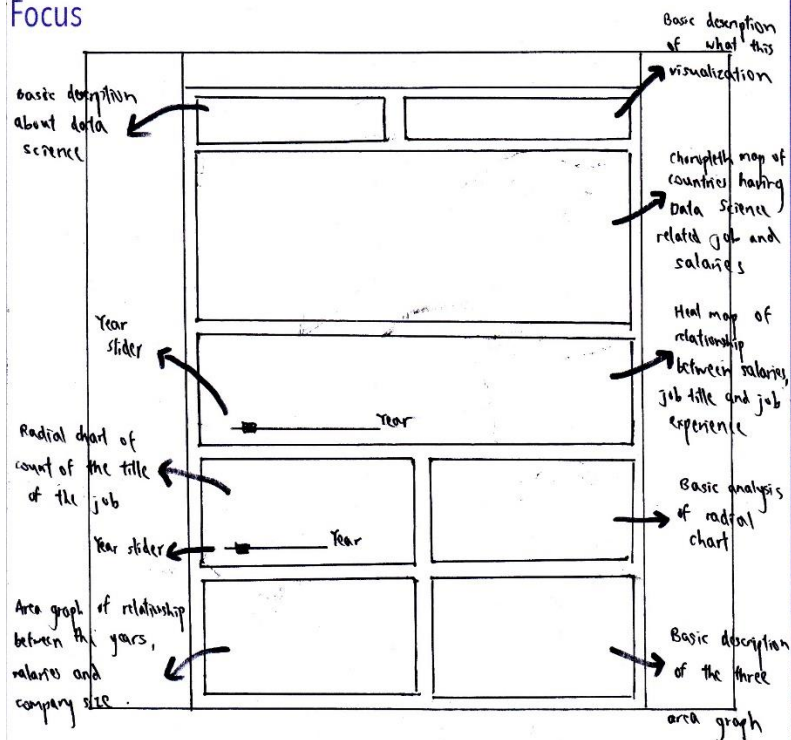
Sheet: 4

Task: Data Science Job Salaries

Operations



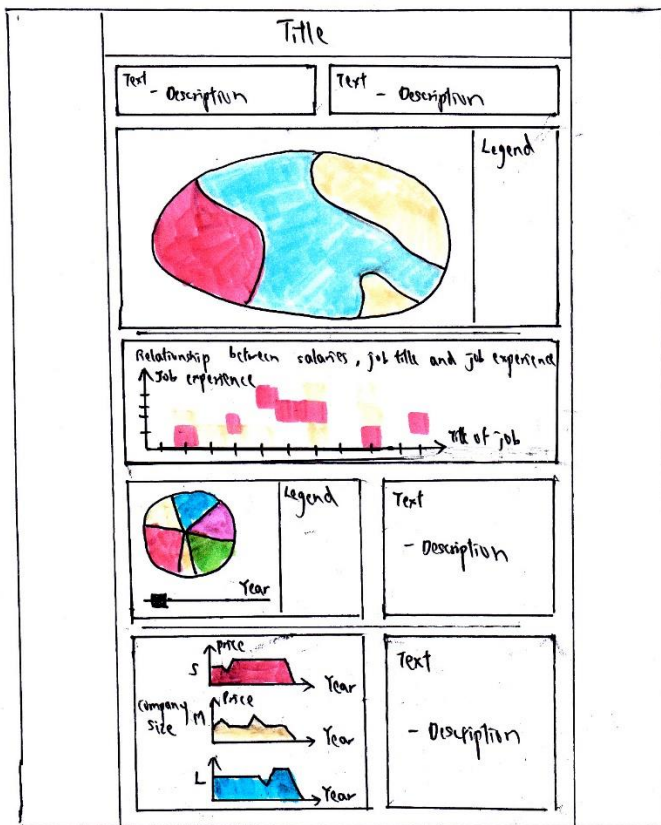
Focus



Discussion

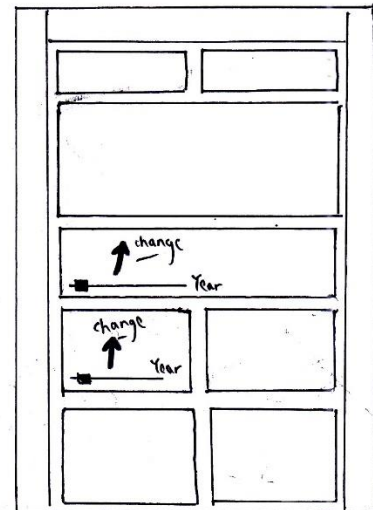
- Symmetry
- Choropleth map as the visual center
- Minimized white space
- Optimized sightlines
- Good flow of storyline.

Layout

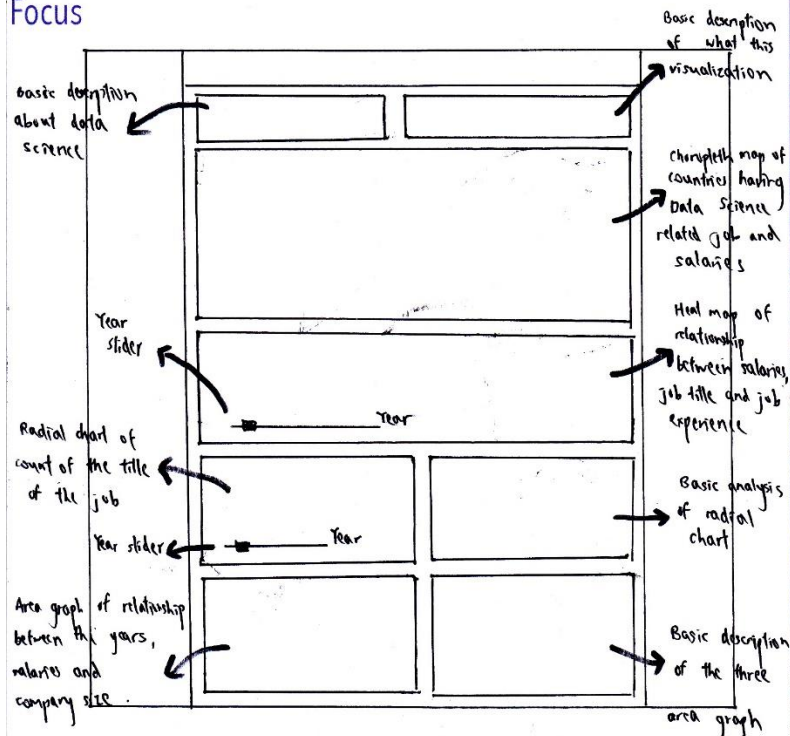


Title: Data Science Job Salaries
 Author: Goh Kai Yuan
 Date: 17/10/2022
 Sheet: 5
 Task: Final View

Operations



Focus



Detail

- Database implemented using csv file
- Clean data
- Time to Build: 1 week