

Detail Report of the Design of the Dashboard

A) Process Component

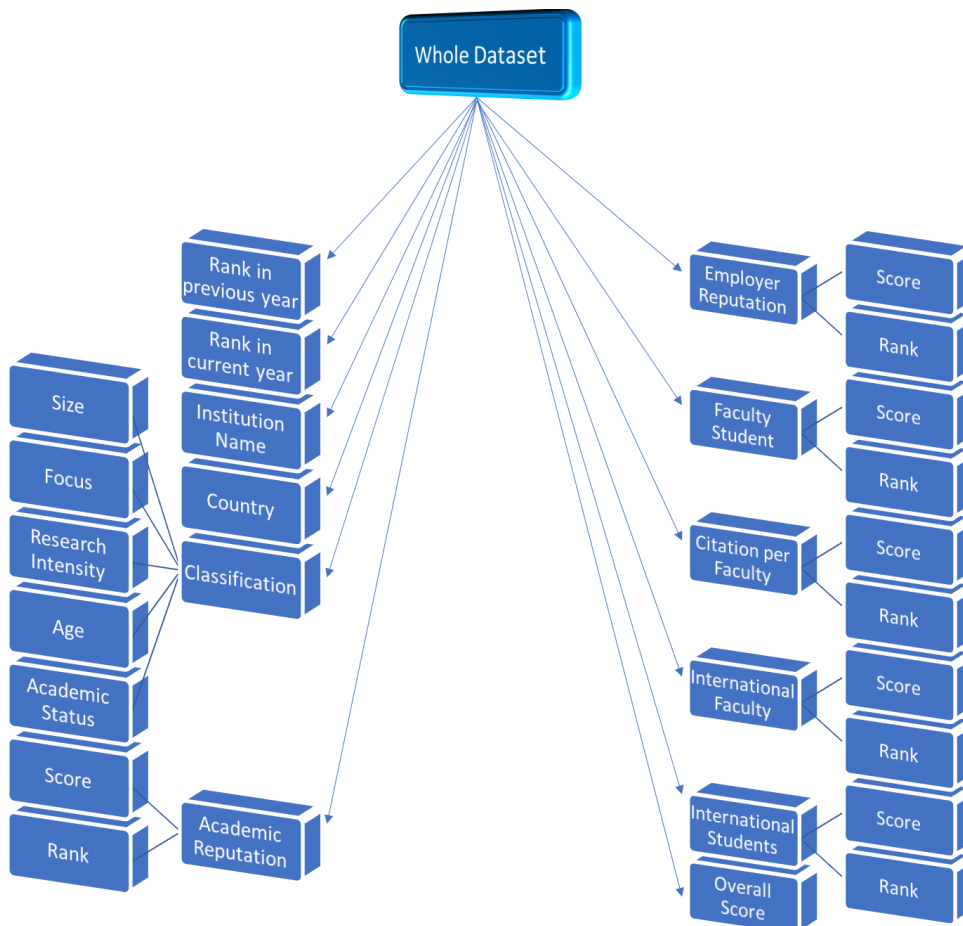
1) Describe the audience for your Data Visualisation (to be written under AUDIENCE page)

As the main objective of this project, the outcome of the data visualization is intended to support UniSelect whose business routine is to advice students in their study decisions. More specifically, UniSelect elicits prospective students' requirements including courses and university attributes and finally assists them to locate the best university institution matching to their requirements. In this sense, UniSelect staff must be able to answer questions regarding to institution reputation, the courses they cover, how intensive the study is, geographical distance from student's ground, etc. All of these functionalities will be provided within the dashboard to ensure the above objectives of UniSelect will not missed.

2) Identify Data Elements (to be written under DATA ELEMENTS page)

The dataset contains two sections: 2019 and 2020. Each of them holds exactly the same elements.

All elements are listed below:



Data description:

- [Rank in previous year] is the global rank of the institution in the previous year.
- [Rank in the current year] is the global rank of the institution in the current year.
- [Institution Name] is name of the institution.
- [Country] is the country where each institution locates.
- [Classification] describes the major characteristics of the institution which are included below:
 - [Size] is the capacity of the institution to accommodate students.
 - [Focus] is the number of faculty area that institution operates.
 - [Research Intensity] is the intensive level of research of the institution.
 - [Age] is the time period that the institution has been operating.
 - [Academic Status] is type of business proprietorship includes (A) public, (B) none profit and (C) private.
- [Academic Reputation] is academic fame of the institution in the global scale.
- [Employer Reputation] is fame of the institution learned by employers.
- [Faculty Student] is number of faculty members to student ratio.
- [Citation per Faculty] is production level of citation per faculty.
- [International Faculty] is the number of international faculty members the institution has.
- [International Students] is the number of international students the institution has.

The last six elements above, all contains subset elements of [Score] and [Rank]:

[Score] is the measurement of decimal number, scaled from 1 to 100.

[Rank] is the measurement of discrete number, scaled from 1 to the total number of available institutions.
- [Overall Score] is the summary of the overall performance of each institution in numerical scale.

Data type classification:

- [Rank in previous year] is a categorical and ordinal element.
- [Rank in the current year] is a categorical and ordinal element.
- [Institution Name] is a categorical and nominal element.
- [Country] is a categorical and nominal element, representing location.
- [Classification] contains its subsets of:
 - [Size] is a categorical and ordinal element.
 - [Focus] is a categorical and ordinal element.
 - [Research Intensity] is a categorical and ordinal element.
 - [Age] is a categorical and ordinal element.
 - [Academic Status] is a categorical and nominal element.
- [Academic Reputation] contains its subsets of:
 - [Score] is a numerical and ratio element.
 - [Rank] is categorical and ordinal element.
- [Employer Reputation] contains its subsets of:
 - [Score] is a numerical and ratio element.
 - [Rank] is categorical and ordinal element.
- [Faculty Student] contains its subsets of:
 - [Score] is a numerical and ratio element.
 - [Rank] is categorical and ordinal element.

- [Citation per Faculty] contains its subsets of:
 - [Score] is a numerical and ratio element.
 - [Rank] is categorical and ordinal element.
- [International Faculty] contains its subsets of:
 - [Score] is a numerical and ratio element.
 - [Rank] is categorical and ordinal element.
- [International Students] contains its subsets of:
 - [Score] is a numerical and ratio element.
 - [Rank] is categorical and ordinal element.
- [Overall Score] is a numerical and ratio element.

3) Describe at a high level what types of charts you might use to display the data. (to be written under THE RIGHT FIT page)

UniSelect collects the preferences from students regarding the course and university features. After analyzing the information, the ultimate objective of UniSelect is to pinpoint a few universities which best match to students' requirements. In this sense, the dashboard must provide filtering options containing essential features such as university characteristics and geographical location which can help UniSelect to trim off the unnecessary data and easily locates the targeted universities. Most of the data features are categorical which are very suitable for table visualization. Table can be used either to present options for filtering or shows detail of the data. Map can also be used for categorizing geographical area for a support in filtering data. Further UniSelect needs to know if an institution is better than another. So, we can use the numerical data, such as scores of the institutions, to form the visual ranks to help highlight the potential institutions. In this case, bar chart is one of the best candidates of visualization because it is easily distinguishable even the values are very close.

4) Finally, describe any data ethics considerations with the data you intend to use.
(to be written under ETHICS page)

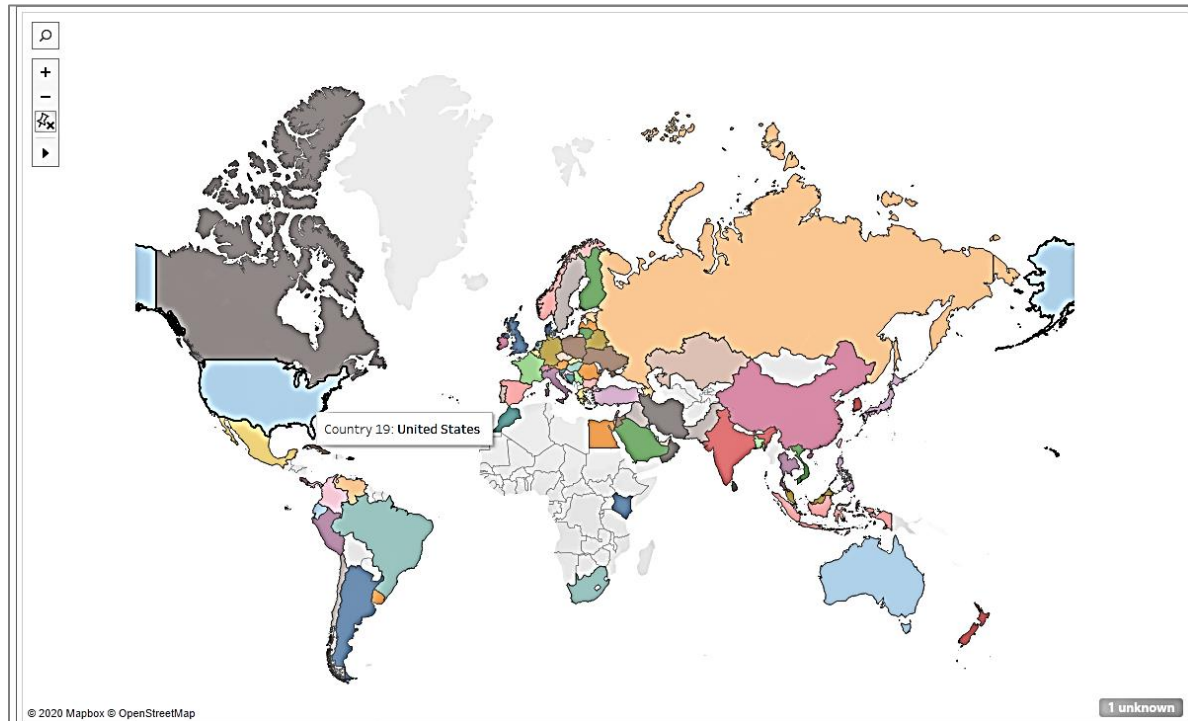
Describe any considerations you think you should keep in mind while developing and sharing your visualisation.

Data privacy is always a concerning issue when processing or presenting data. My personal plan is to balance between the information-rich optimization for a greater insight of data and the possible mitigation of data privacy. Looking back at the data features, categorical features are less brutal than numerical features when discussing about the privacy. For instant, sub-features of [Classification] contains the condensed versions of much wider and more accurate measurements which make it less sensitive for public. The feature [Rank] is more identifiable to a certain extent because it tells how good an institution is in the global competition. However, the total number of candidate institutions changes annually, and thus it is vague to say the exact ratio. Further, some rank values are in a ranged measurement which has already deidentified the data uniqueness. The feature [Score] is the absolute sensitive feature since it can tell the exact performance of an institution within 0 to 100 range. So, aggregated form of [Score] and [Overall Score] is important to take into consideration to conceal its privacy in the dashboard.

B) Solution Component

Story:

I personally select a number of visual charts for this dashboard project. First of all, I prioritize the visual map which allows user to see the classification of institution locations across the globe. Beside the classification, it provides the view of geographical distance that broadens the comparison options beyond the classical parameters. Each of country boundary is colored distinctively for a precise and convenient selection on the map. In the designed dashboard, the map is used to present the country locations to where each institution belongs. This map assist user to filter the unwanted countries and narrow down the amount of institution to only one preferred country.



Next, I choose table + circle shape visualization to depict various classification options of the dashboard. Since the data features for institution classification are categorical, table is the most suitable candidate to visualize this classification. Circle shape is selected to present the values in the table because it helps explain its functionality as if it is an option button used for filtering data. All sub-classification features of institution ([Focus], [Research Intensity], [Size], [Age] and [Academic Status]) are presented into table format with their values shaped as circles. All circles, except [Academic Status], are colored using solid blue gradient. The darkness of the color expresses the higher values. [Academic Status] has solid blue for all of its values since they are nominal. The reason I select this color is that it symbolizes the sophistication and correctness. It also gives great contrast against white background making it easy to spot.

[Classification] chart offers great options for user to trim off the unmatched institutions and jump right into a much less crowded data which contains the set of institutions the user is looking for. UniSelect asks for the preferences from students in order to suggest the institutions they want. This chart will simplify and speed up the searching experience of UniSelect based on those preferences and bring an enormous step closer to final selection in no time.

Classification:

Focus

FC

CO

FO

SP

Research Intensity

VH

HI

MD

LO

Size

XL

L

M

S

Age

5

4

3

2

1

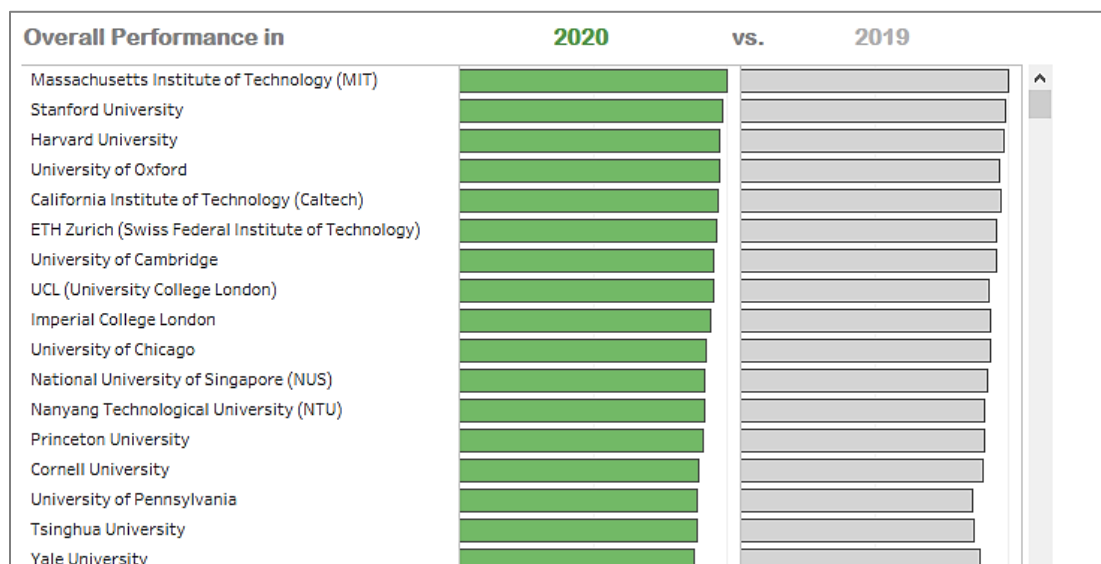
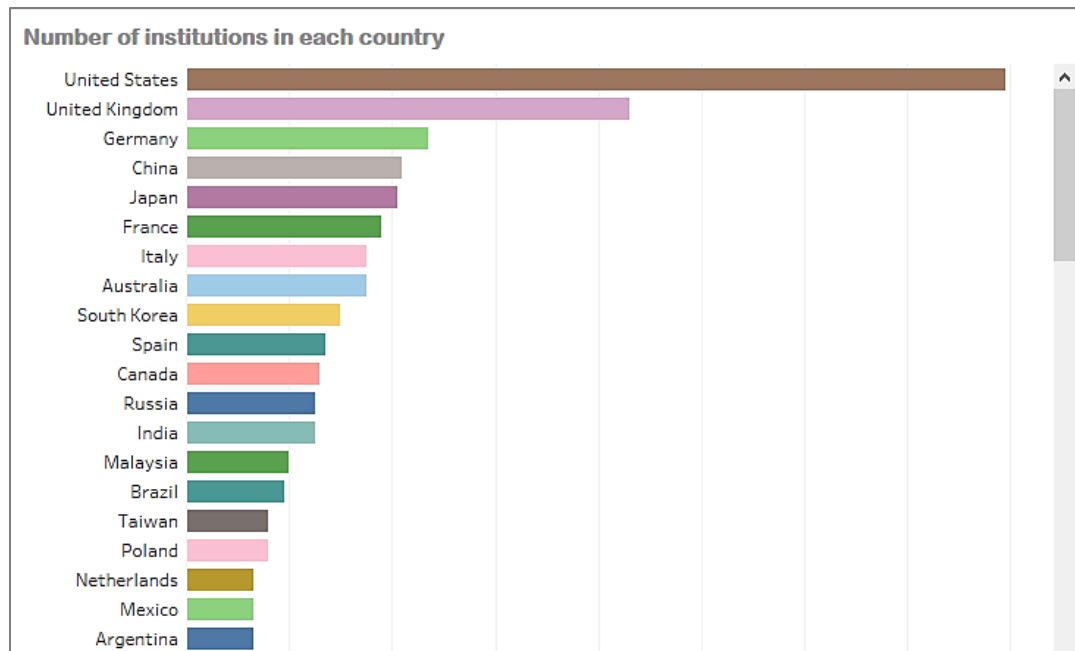
Public

N-Profit

Private

The third type of chart I use is horizontal bar chart. It is a perfect of type visualization when we want to compare between data categories. The shape of the chart is friendly to recognize, and it can capture the precise distinction between data, even at a close match. In the designed dashboard, the bar chart is used to display the overall performance of each institution and the total number of institutions in each country. [Number of institutions in each country] graph comes with mixture of color variation. Those colors are synchronized colors that match with individual country map. In [overall performance] graph, color green and gray are used to mark the overall performances of each institution in 2020 and 2019 respectively. The higher value of overall performance brings the positive indication of the institution, therefore green color matches well to the positivity. Gray shows the past overall performance in 2019 and it suggests user to pay more attention to the green bars for 2020.

[Number of institutions in each country] graph gives the idea of how many available institutions in each country and allows user to jump into more detail of each individual country. [overall performance] acts as a quick guidance for user to spot the top institutions and enables user to further filter and drill more detail of each institution. These interactions meet the need of UniSelect who wishes to query for specific country and quickly identify the top famous institutions.



The last type of visualization that I use is table + data bar. The purpose is to show a mixture of text and graphical information. This visualization contains different type of data values including detail and aggregated one (sensitive data), therefore table is suitable type of chart that can hold both types of data. Soft blue is used for marking the data bars. This color gives a sense of uniformity with [classification] chart and they both have the same meaning. In the designed dashboard, this table is used to display the detail which can support user's conclusion. After all of the filtering operations, the final list of matching institutions is displayed in this table. In the real application, UniSelect can view this table, sort by any specific feature. For instance, UniSelect may be interested to find the best institutions which are well known in the international spectrum by sorting [International Faculty] in descending order. By scanning through this table, UniSelect can finally suggest the most potential institutions for students.

Institution names and their current scores of different areas									
Institution Names	Rank 2020	Rank 2019	Rank 2018	Academic	Citations/Fact.	Employer Repu.	Faculty Stud.	Int. Faculty	Int. Student
Aalborg University	324=	343=	379						
Aalto University	134	140	137=						
Aarhus University	145	141=	119						
Aberystwyth University	484=	432=	481-490						
Abo Akademi University	541-550	531-540	551-600						
Abu Dhabi University	701-750	701-750	751-800						
Adam Mickiewicz University	801-1000	801-1000	801-1000						
AGH University of Science and Technolo..	801-1000	801-1000	801-1000						
Ain Shams University	801-1000	701-750	701-750						
Airlangga University	651-700	751-800	701-750						
Ajman University	751-800	801-1000	Null						
Ajou University	601-650	651-700	651-700						
Al-Farabi Kazakh National University	207=	220	236=						
Alexandria University	801-1000	801-1000	751-800						
Aligarh Muslim University (AMU), Aliga..	801-1000	801-1000	801-1000						
American University	601-650	541-550	471-480						
American University in Dubai	601-650	561-570	601-650						
American University of Beirut (AUB)	244	237=	235						
American University of Sharjah	371	376=	411-420						
Amirkabir University of Technology	489=	498=	501-550						

Finally, I would like to share some comments on the dataset. I found some problems in the data. Firstly, it contains a misplaced column [Academic Reputation] in 2019 sheet. More, some data values of [Overall Score] come with some ranged values when they are supposed to be numerical values. Lastly, the null (missing) values are originally marked with dash '-' symbol which turns all numerical features [score] into categorical features in Tableau. Before I could use the data, I need to rearrange the misplaced column, change all [Score] into numerical features, and unify the ranged values in [Overall Score] into single values.

HomePage:

How Beautiful Data can help you?

Design and Development Service

We handle interactive visualization project which brings you the vision of data that you never expect. With our support, you never feel complicated and confused with data anymore. Instead you will love the data as much as we do.

Based on Tableau, our team help you build the interactive dashboard that streamline and democratize your critical data. With our great experienced professionals, we ensure the dashboard optimizes the need of your business routine activities while minimizes the risk of data privacy and security. You will never be able to imagine how useful and beneficial the final product will be for your company which ultimately takes you to the realistic edge of a smart businessman.

More than an ordinary service, Beautiful Data brings the best support at all phases of development. We care not only during the development, but our great care begins at the first time we meet and throughout your business life.

Our Experience and History:

Beautiful Data has been on business for more than 10 years. We've experienced all kind of data variations and problems. The difficulty has strengthened our team and our endeavor to sharpen the skills and professions has never stopped. Our years of success have been symbolized by the smiles of our business clients and partners. Their feedbacks have kept us on track and reflected the highest quality of products we made. UniSelect is one among many of our happy clients whose dashboard is shown as a model in our site.

ABOUT US Page:

I am a master student of Data Science and intern at Beautiful Data Inc. I am very keen working with data visualization. It is a great opportunity for me to be part of Beautiful Data and carry out the competing project of creating interactive dashboard for one of Beautiful Data clients, UniSelect.

Although is it my first hands-on experience in data visualization, I can sense the relationship between me and the data. I like the job of being creative and having freedom of idea that could bring any unexpected result. On top of that, I am the kind of person who prioritizes learning and sharing, and thus UniSelect project is really meaningful for me.

C) Personal development Reflection (600 words)

I have been always aware of my roadmap and objectives that I set for myself toward the future career. As a data science student, my future success does not rely on any single skill but a set of skills, experience and even routines which I need to build. So far, I have absorbed a lot data analytical expertise from this unit which takes me to a new level of understanding of data. Every progress I have made through the unit, I encountered difficulties that not only sharpened my technical skills but also leveraged my routine of self and time management.

There were times when the challenges were unbreakable without support. Being a second language learner, I had the moments that my language limitation halted my assignment progress. Thanks to Language and Learning Advisers who took their hands seriously in providing useful guidelines in writing as well as organizing the work. With their support, I was able to plan and control my assignment progress and use the right expressions concisely to minimize any ambiguity in my assignment report.

Every assignment has added new boundary to my existing knowledge. There was no difference when it came to the arrival of the third assignment. However, this assignment lied over the edge of the former assignments. Instead of calculating and predicting the data, this time, I had to work on a different perspective of how to look at the data in visual forms. This assignment not only tested the visualization skill but also the comprehension of data nature, design creativity, business application and anticipation of potential risk of data. For instance, I needed to analyze the requirements of the business client to understand their problems and needs. Secondly, I must explore and observe the data nature and match

it to the needs of the business client. Next, I had to put all the skills of data visualization that I learned from classes as well as self-study into the development of the dashboard. Design creativity must be in place to ensure the friendliness, navigability and practicability have been counted in one nutshell. Lastly, I had to balance between the juice of the information and the risks of data privacy which, to a certain extent, held back the ultimate details producible by the dashboard. All of these aspects of the assignment have assembled various parts of my learning into one final productivity. This achievement interpreted the volume of learning that I have absorbed from this unit so far.

It has been a new milestone of achievement toward data science profession to be able to complete the third assignment and had a thorough experience of data visualization. I feel that I am geared with potential competency that promisingly brings me to another big step forward along the way through my personal roadmap. Although I find the experience fruitful, there is still room for improvement. It is often the first experience when a student takes a journey through a unit, and for that reason, there always things that we could do better. In my personal journey, time management was not efficient enough to allow a proper timeframe for self-discovery of Tableau features and techniques. I underestimated other project which resulted in the delay to begin this assignment. The problem had a great impact on my ability to vary the visual projections within the dashboard. If I could make it, this variation would help user to easily distinguish the components and functionalities of dashboard elements.

Having learned from the unexpected experience, I believe that it is essential to flatten all of the detail of any future assignments. This way, I can estimate the precise amount of effort and time for completing a given assignment. I know that controlling the time is not easy as said and done. Therefore, I hope that this meaningful experience helps push me to another step closer to the perfection.