PROJECT PROPOSAL

DATA SCIENCE – GRAD SCHOOL FINDER

Pranavnath Dommata
u1068846
u1068846@utah.edu

Neha Pathapati
u1069211
u1069211@utah.edu

Project Repository: https://github.com/u1068846/VisProject.git

BACKGROUND AND MOTIVATION:

The United States of America is a home to a large number of schools that are well acclaimed for their Data Science programs. Hence, for an individual intending to purse a graduate degree in Data Science in the US, it may be overwhelming to zero down upon a list to schools that best suit his requirements, which he can apply to. Hence, we propose to design an interactive visualization tool that assists an individual in easing up this process.

PROJECT OBJECTIVES:

With the gigantic increase in the volume of data, there seems to be a huge demand in the industry for professionals who are trained to create value out of this data. Generally, such professionals are designated as – Data Scientists, Data Analysts, Data Engineers, and so on.

To cater to the rising need in the industry for data professionals who can mine enormous datasets to extract insights, several universities have begun to introduce Data Science programs into their curricula at the graduate level. Ultimately, they intend to produce graduates who can contribute to stemming the shortage of data scientists in the industry.

For an individual wanting to pursue a Data Science degree in the United States, there are undoubtedly a broad spectrum of options. The "Data Science – Grad School Finder" is a tool that allows one to shortlist a set of schools offering Data Science degrees based on his/her set preferences for parameters such as – type of the program (Masters', Doctoral, or Certificate), delivery mode of the program, world rank of the program, pre-requisites needed for the program, and the state of the United States where the school is located. Additionally, the user can explore the shortlisted schools in more detail, draw up two schools to compare them closely, and sort the shortlisted schools based on his/her preferred parameter.

To reiterate, our objective is to build a comprehensive and an interactive visualization tool that assists potential data-science grads to shortlist, explore, and compare and contrast schools offering Data Science degrees in the US of A.

PROBLEM STATEMENT:

The "Data Science – Grad School Finder" tool finds an optimized set of schools in the US offering Data Science graduate degrees based on the user's preferences set for – the level of the program (Masters'/Doctoral), the delivery mode, the state of US where the school is located, and

pre-requisites for the program. From the list of schools, the user can explore details like – the location of the school (state, city), the several Data Science programs offered by the school, the departments within the school offering these programs, the world rank of the program, a link to the program's website, and so on.

Additionally, from the obtained list of shortlisted schools, a user can choose any two schools to compare them more closely. Lastly, the tool also enables the user to sort the shortlisted schools based on his preferred parameter, for instance, the world rank, faculty-student ratio, research citation score, and so on.

DATASET:

An existing data set of 951 entries will be employed for this project. The source of the dataset is stated in [1].

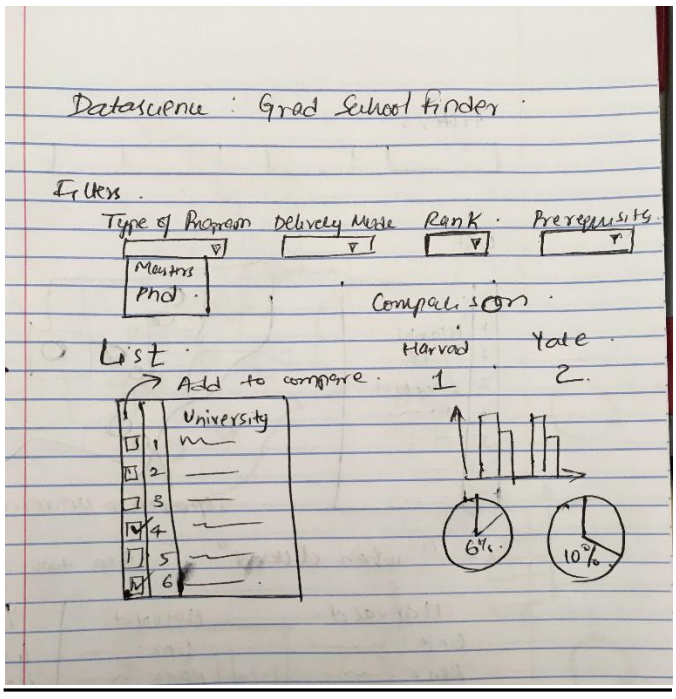
The dataset comprises of a comprehensive set of attributes that are sufficient for all components of our project. The attributes of the dataset include – School, State, City, Program, Type, Department, Delivery, Duration, Prerequisites, Link to the program website, Longitude and Latitude of the school, World Rank, Teaching Score, International Score, Research Score, Total Score for the school, Student Population Size, Student Faculty Ratio, International Student Population, and Male to Female Ratio.

DATA PROCESSING:

The dataset obtained from [1] was in CSV format. Hence, we are using our data in the raw form and did not find the need to subject it to any form of processing.

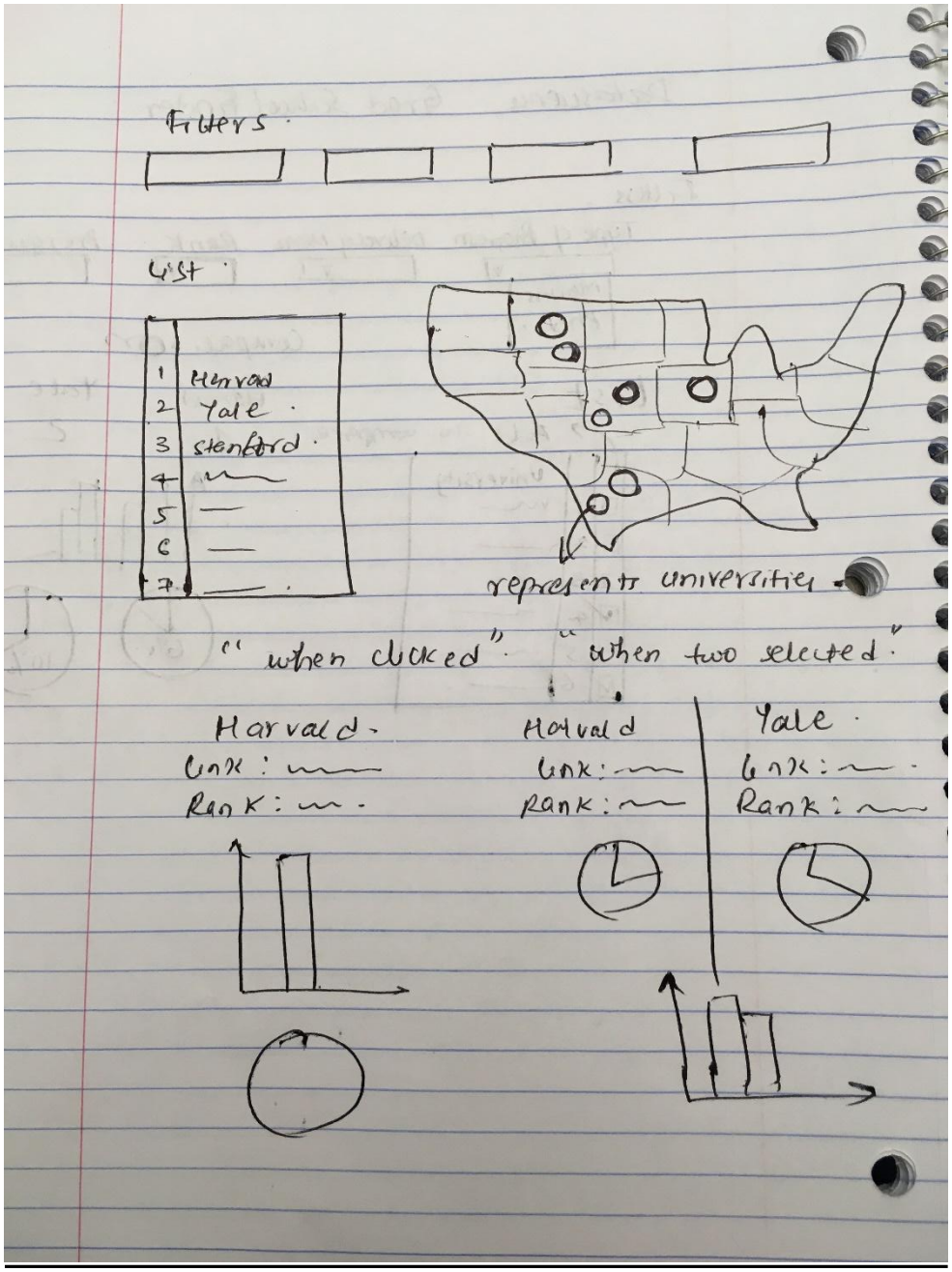
VISUALIZATION DESIGN:

DESIGN #1:



This is our first design based on our original ideas. Here, our visualization had three components – First, filters to allow the user to set his preferences for the type of Data Science program he wishes to pursue, the list of schools derived from the dataset based on the filter values, and the comparison of two schools selected by the user in the list using bar charts, donut charts, tiles, scatterplots, and so on.

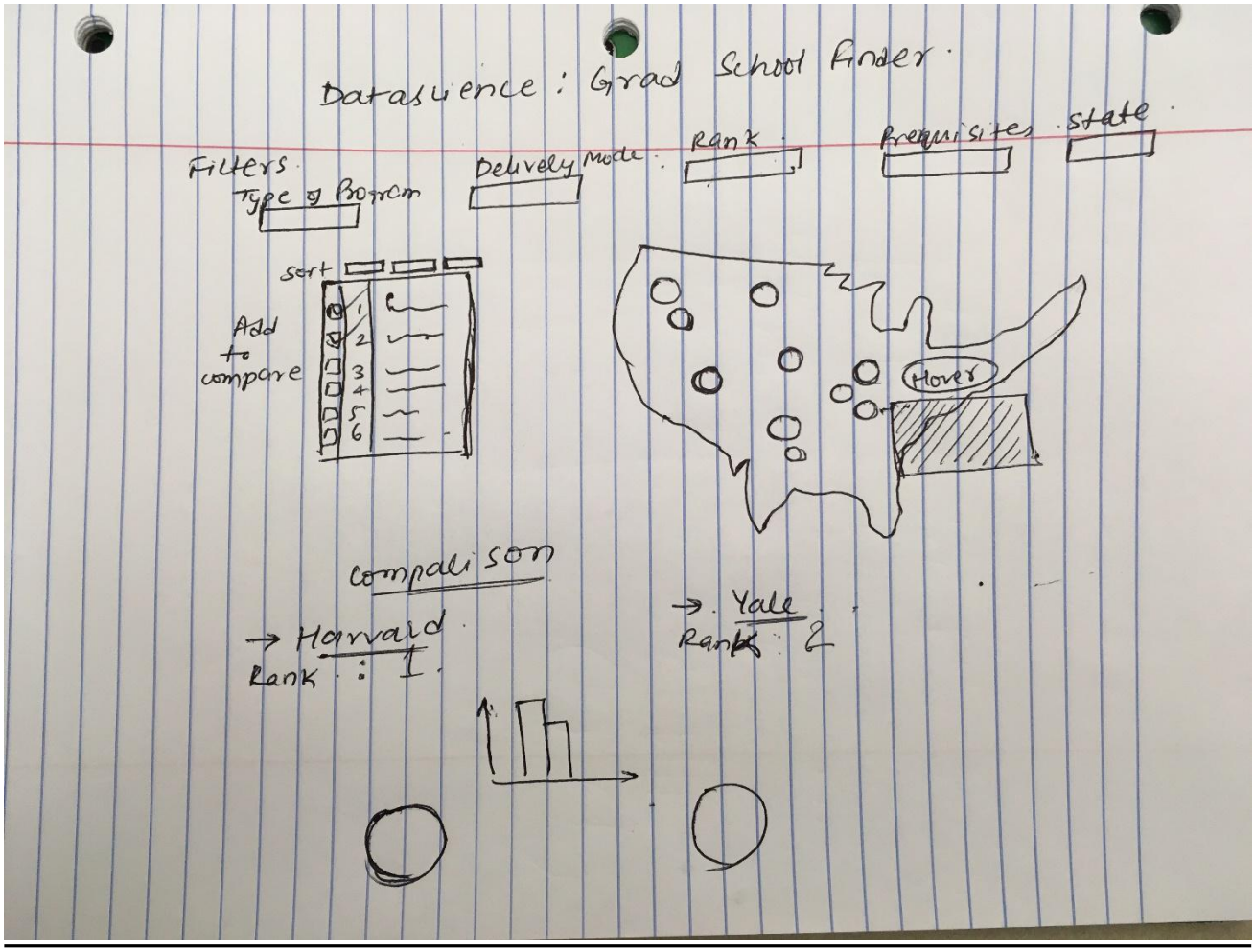
DESIGN #2:



In our second design, we made the following amendments: First, we decided to add a map of the US and highlight the location of the schools shortlisted by the tool to give the user an idea of the location of the school. This decision was primarily influenced by the presence of the latitude and longitude attributes for the schools in our dataset. Second, when the user clicks upon a highlighted school on the map, additional details of the school such as – the location, the degree names, the departments within the school offering these degrees, the world rank, a link to the program's website, etc. are

displayed below the map. Third, when the user selects two schools on the map, a comparison of the two schools is provided using bar charts, tiles, scatterplots, donut charts, and so on.

DESIGN #3:



This is our best design till date. Here, we added functionalities such as – allowing user to hover over a school on the map and get details of it in a tooltip, and sort the schools shortlisted by the tool based on his preferred parameter like for instance, the world rank, the student-faculty ratio, and so on,

MUST HAVE FEATURES:

1. Filters provided for the user to define more precise dataset he/she wants to look into.
2. Display a map of the US with interactive features like hovering over the university (represented with circles) to explore more about the school and clicking adds the university to compare with the university the user clicks next.
3. Visualizing the parameters using different designs like using grouped bar charts, pie-charts for representing percentages, donut charts, etc. to encode the data in a visibly effective manner.
4. The user is given an option to sort the list of universities displayed based on the parameter he chooses.

OPTIONAL FEATURES:

1. Enabling the comparison for more than two university selections.
2. Being able to add to compare from the list displayed instead of selecting from the map.
3. Zoom feature in the map which makes it more fun to interact with.
4. Selection of the criteria upon which the user wants to compare the two selected universities (multiple selections possible).

PROJECT SCHEDULE:

Week	Date	Deadline	
		Neha	**Pranav**
1	Nov 4	Filters	Comparison
2	Nov 11	Map	University List
3	Nov 18	Map Interaction	Sort
4	Nov 28	Zoom	Comparison designs

OVERVIEW OF PROJECT DELIVERABLES:

An interactive visualization that has the following functionalities:

1. Allows the user to set his preferences for – the graduate degree level (Master's/Doctoral), mode of delivery, the location of the school, and the prerequisites for the program.
2. Obtains a set of schools based on the user's preferences and highlights the schools location on the map of the US of A.
3. Upon hovering over a school on the map, allows the user to explore the school to a greater detail by displaying - the location of the school (state, city), the world rank of each of the programs, the student population size, the student faculty ratio, the research citation score, and so on.
4. Select two schools on the map and compare them closely.
5. Sort the schools shortlisted by the tool based on the user's preferred parameter like – the world rank, research citation score, and so on.

REFERENCE:

[1] Srihari Rao. [Online]. Available: https://www.kaggle.com/sriharirao/datascience-universities-across-us