# *CS6630: Project Proposal*

## **Title**: U.S. Dream - Permanent Visa Applications Visualisation

## **Team**:

Sanket Deshmukh - u1076955@utah.edu
Ashish Gupta - 1078564@utah.edu
Anmol Ahuja - u1056311@utah.edu

**GitHub:** https://github.com/sanketd11/dataviscourse-pr-US-VISA-Applications-Visualization

### *Background and Motivation:*

In this Project we will be visualizing dataset of US Applications over the period of time. Our motivation behind choosing this project was due to the fact these days US Applications is one of the most debatable topics these days. There are many reasons for choosing this project such as:

- We want to know from which country maximum applications are being filed.
- Which sector is filing the most visas applications.
- In which category of Visa there has been an increase dramatically over the years.
- What's the success rate of each type of visa getting approved?
- What's the salary range being offered under different type of visas?

There are personal interest also behind choosing this project as we would also want to be hired on H1B visas in future. So, using this project we can analyse the overall trend and we can get fair amount of information which can be really beneficial for us.

### *Project Objectives:*

The objectives of the project are:

- To visualize the data in a much better way so that it is more presentable.
- To form a story or to conclude a pattern of different type of visas over the years.
- Different type of design will be used which will co-relate with each other.
- We will cover every aspect of the data to depict pertinent things.

### *Data:*

We searched for our dataset on the internet and found various datasets on different websites. We then compared all of the datasets with each other and chose the best dataset

which had most of the attributes and more importantly which had more suitable attributes to cater to the needs of the project. Finally, we chose the dataset from www.kaggle.com.

The complete link to our dataset is provided below:

https://www.kaggle.com/ambarish/eda-us-permanent-visas-with-feature-analysis/data

Our Dataset has appropriate data from 2011 to 2016. It also includes information on employer, position, wage offered, job posting history, employee education and past visa history, associated lawyers, and final decision.

## *Data Processing:*

We have done substantial data clean-up on the dataset we got from www.kaggle.com.

There were many redundant and duplicate columns present in the original dataset. We took only distinct columns in our final dataset which we plan to use towards our project.

Also, there were many rows which had Null values for some of the attributes/columns. We did not include those records in our final dataset as well.

Moreover, there were many columns in the original dataset which had no data. We did not include those columns as well.

We did many data validation checks also on our dataset and chose only those records which had correct values for Dates, numbers in Salary and text in Employer_Name, Country_Of_Citizenship etc.

The quantities/attributes which we derived from the original dataset out of 130 columns/ quantities/ attributes include :-

Case_Status, Class_Of_Admission, Country_Of_Citzenship, Decision_Date, Employer_Name, Employer_State, Job_Title, US_Economic_Sector, Wage_Offer, Wage_Offer_Unit.

## *Must-Have Features:*

Some of the Must-have Features which we think are most salient for successful implementation of our project “U.S. Dream” are as follows :-

- Case_Status- It tells what the result of visa application filed is. Whether application was “Accepted”, “Denied”, and “Withdrawn”.
- Class_Of_Admission- It tells what type of Visa application is applied by a candidate. For ex- “H1-B”, “F-1”, “J-1”, “B-1”.
- Country_Of_Citzenship- It tells from which country the Visa application has been applied from by the candidate. We can showcase this feature on a world map which can tell the countries with highest number of applications.

- Decision_Date- This feature will be used for telling the time-frame on visualization. We have the data from the year 2011 to 2016.
- Employer_State- This feature tells the state within U.S. where the candidate would be working (if visa is approved). It is an important feature to show on a U.S. map, giving pertinent information about where are most of the Employers located within U.S. who are more willing to file Visa applications for their employees.
- US_Economic_Sector- It is basically a categorization of all the Employers who have filed for Visa applications during the time-frame of 2011 to 2016. This is more general attribute which we can use to categorize the applications rather than the Job Title.
- Wage_Offer-This feature tells about the salary which the visa application candidate would be getting (if visa is approved). We can visualize this feature to show a trend/pattern giving information about which Employers and which States give more salary to their employees compared to others.

## *Optional Features:*

Some of the Optional Features which we think are nice to have in our project but are not critical for successful implementation of our project are as follows :-

- Employer_Name- It tells the name of the Company/Employer which has filed the Visa application for the candidate. It's a nice to have feature in our visualization but not an important feature.
- Job_Title- It tells the Position Title of the candidate in the company which files the Visa application. We can categorize visa applications based on this feature but since there are a lot of different titles in different companies, the visualization would look much more cluttered when using this feature.
- Wage_Offer_Unit- This feature tells the Unit of Pay. Whether the wage being offered is "Hourly", "Weekly", Bi-Weekly", "Monthly" or "Yearly".

## *Project Schedule: (tentative)*

***Week 1:***

In this we will be finalizing the design after getting the Peer Feedback, we will incorporate all those changes in our design. Similarly, we will meet TA to get feedback which is most important.

***Week 2:***

In this week we will be coding the individual components of the design coherently.

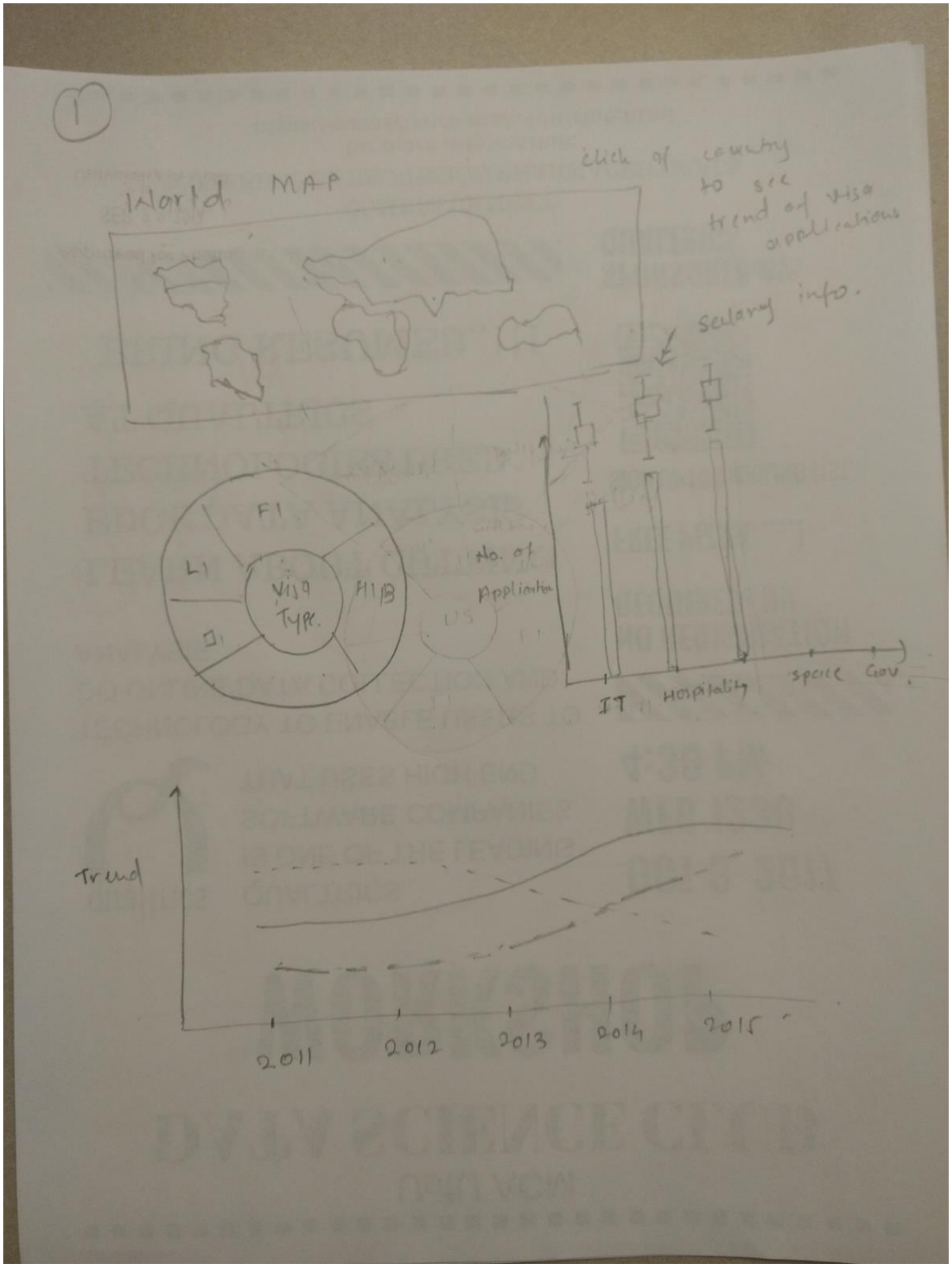
***Week 3:***

After the coding part is done we will then link all the individual designs and will make sure interaction is done.

***Week 4:***

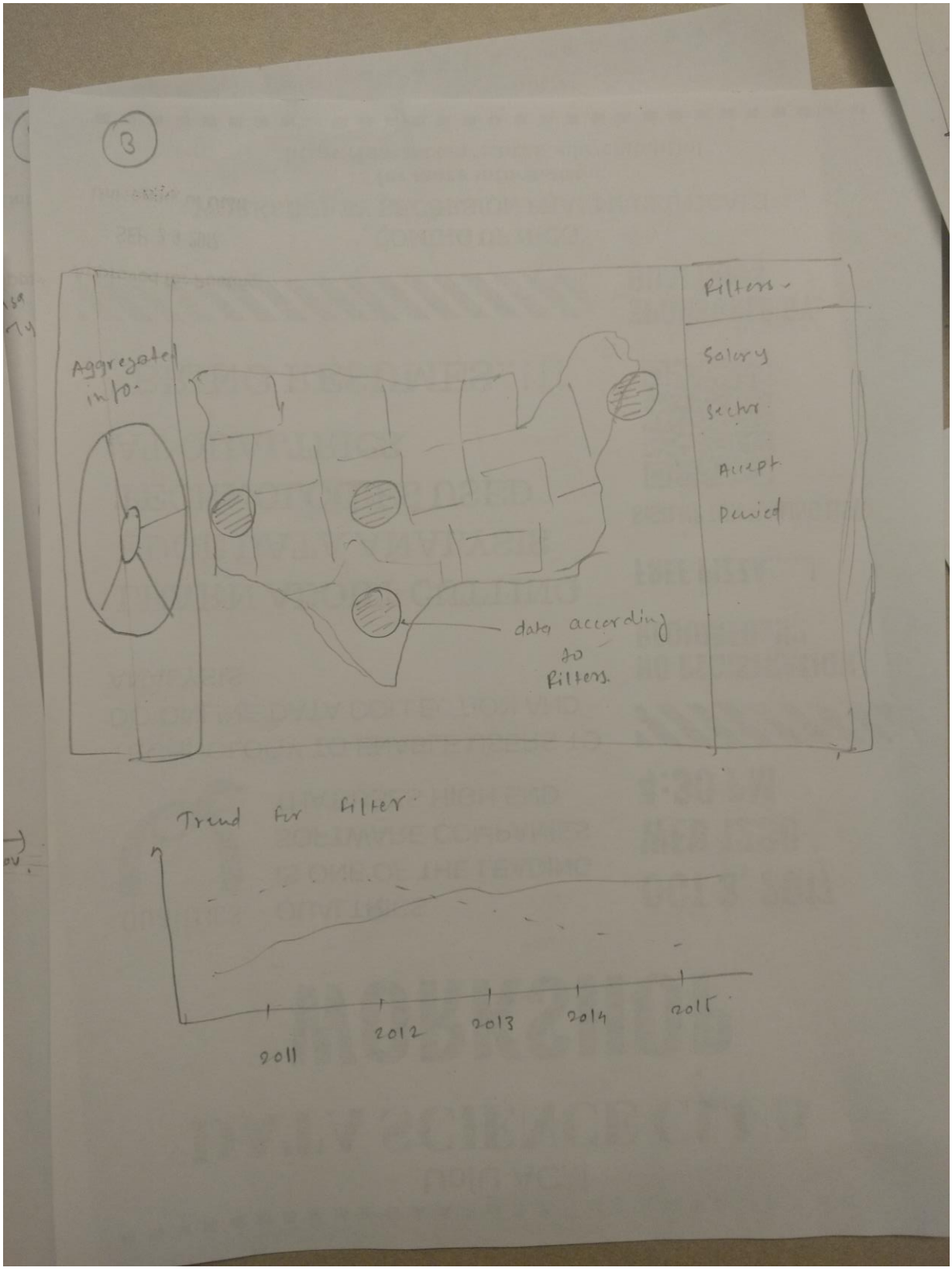
Lastly, we will see everything is working and will make any change if required and will show it to TA.

## Design: 1



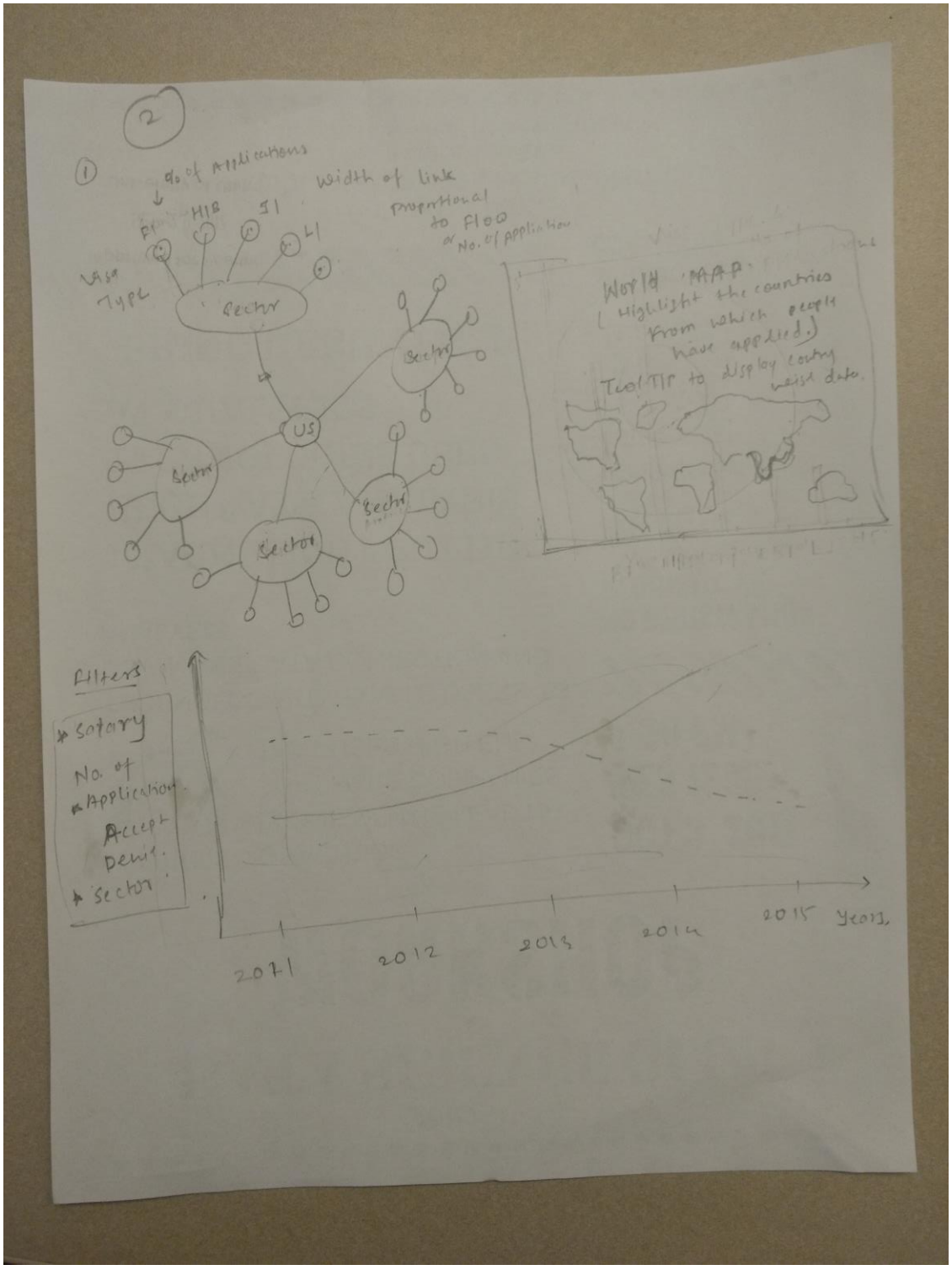

After the first load of the page, aggregated data including Visa types, their distribution over the industry sectors, the box plot for salary values, etc  will be displayed for all the countries. Clicking on the country will filter out the same information for the specific country. The Time Series will display trends in the selected quantities using brush charts.

## Design: 2




In the above design the size of the circle will display magnitude of the data that is selecected from the filter. The Time Series Graph will show the trend of the selected data. The Donut chart will show the Visa distribution for the selected data category from the filter

## Design: 3




Combination of design 1 and 2 will be the best way to represent the given dataset.