BEERVIZ PROJECT REPORT

Objective

To create a fun gateway through which someone who is interested in beer can explore various beer styles/brands based on user reviews and figure out which beer to try.

Audience

Our target audience is beer drinkers (both casual and experts) who would like to explore different types of beer.

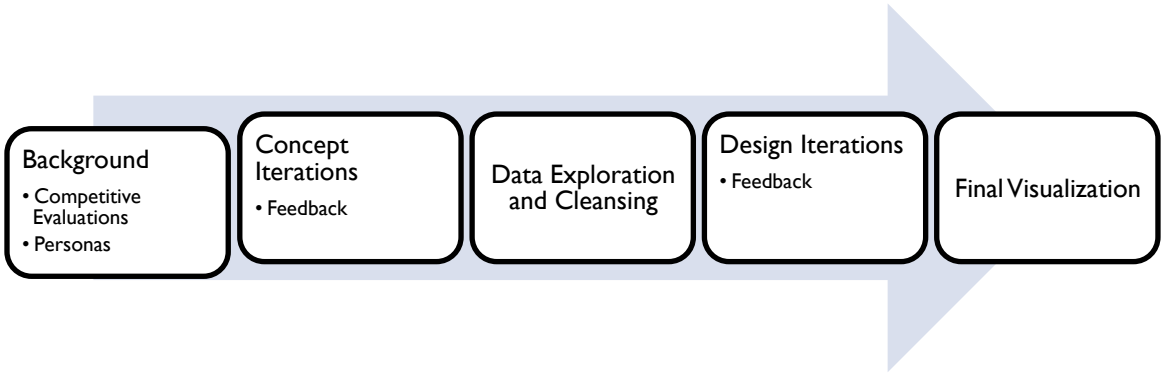
Dataset

We are using a dataset that contains beer reviews, collected over 10 years. (http://snap.stanford.edu/data/web-BeerAdvocate.html).

Variables

Name	Type	What it Represents
Styles	Categorical	Type of beer (e.g. Stout, Ale, Lager, etc.)
Names	Categorical	Name of the beer (e.g. Turbodog, Old Stock Ale, etc.)
SRM	Ordinal	Standard Reference Method (SRM) represents beer color from 1 to 40. (1 = lightest, 40 = darkest)
Appearance	Ordinal	Reviewer ratings on whether the beer is light, medium or dark. Values between 1-5
Taste	Ordinal	Reviewer ratings on beer taste. Values between 1-5
Aroma	Ordinal	Reviewer ratings on beer aroma. Values between 1-5
Palate	Ordinal	Reviewer ratings on beer palate. Values between 1-5
Overall	Ordinal	Reviewer ratings for the beer overall. Values between 1-5

Process



Background

Competitive Analysis

- **One Year of Beer** (http://www.visualizing.org/visualizations/one-year-beer)

  We like the concept of using lines to link the types of beer consumed, and also the use of the timeline. However, since he lists all the beer he drinks for every day of the year, it becomes overwhelming in a static image. It also seems the type of graph is not very well fit, because it does not show any patterns. With no patterns, the use of various colors makes the visualization look busy.

- **Parteispenden** (http://labs.vis4.net/parteispenden/)

  The visualization shows an interactive flow of money from private donors to German political parties. The political parties are ranked in size of donation from largest to smallest, each with its own color. The line flow that represents the donation relationship from private parties to political parties is also shown with the same color. This makes it easy to see and understand the message of the data visualization. We noticed that the parties are listed in alphabetical order from right to left - we would prefer reversing the order to read from left to right since that is the way we are trained to read.

- http://www.visualnews.com/2012/02/09/visualizing-the-american-beer-revival/

  The interesting thing about this visualization is the way it leads the audience toward the information. The graphics are interesting and since it is a video, it establishes a storyline that moves forward seamlessly. However, the format constrains the narrative since it is hard to retain all the numbers and it does not

allow the user to move back and forth. Also, the speed of transition is fast, making it harder to gauge information.

- http://visual.ly/beers-periodic-table

  This visualization explores the wine industry network, and highlights how a few vendors dominate overall sales. There's a similar example for the beer industry, seen at: https://www.msu.edu/~howardp/beer.html. Both these visualizations are good at communicating the dominance of certain key players in the market. From a storytelling perspective however, they are fairly basic.

- http://homepages.cae.wisc.edu/~jamieson/me/BeerMapper.html

  This visualization uses the same dataset as ours, and we looked at it to understand how the data had been interpreted by a different set of designers. This visualization used more of data mining and information retrieval techniques to generate a heat map. While it was interesting, it was lacking the 'fun' element.

Based on our competitive evaluations, we felt there is a need to visualize beers in a way that allows people to explore and discover new brands. Each of us spoke to people to understand how they consume beer, and pick the brands they drink.

Personas

- Dan is a novice drinker and doesn't know anything about beer. He is just getting used to the taste of beer and sticks to the one brand he knows he likes. He would like to try out some others, but is very conservative. He knows he likes light beer, and would prefer to try out different brands so long as they are light.
- Joe, the average beer drinker, is a guy in his mid-20's who has not tried out the famous beer styles. He is a grad student and attends all Thirsty Thursdays (TT). He likes dark beer, but has only tried certain brands (i.e. Guinness Draught). He is hoping to try out various kinds of beer based on information from friends. His only source of information is Louie (a fellow grad student who is a connoisseur of beer).
- Louie has tried many beers, but is always open to exploring new options. He evaluates beers based on how they taste, look and smell. He considers it a matter of pride to have tried various brands of beers and is always on the lookout to try something new.

Concept Iterations

Our first idea was to create a chord-layout that allows users to explore beer styles across a timeline to understand how ratings have changed over time.

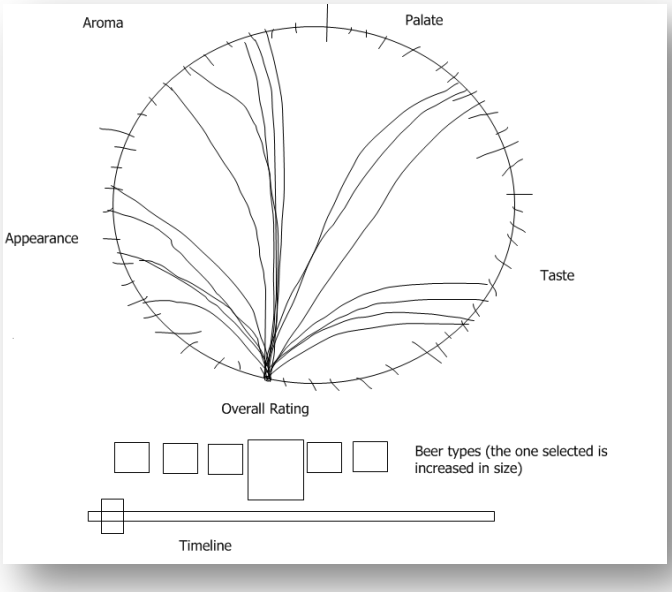


Figure 1: Illustration of initial concept

Feedback: During the in-class concept critique we realized we needed a stronger narrative hook. We had some basic ideas to build this narrative – one idea was to ask people to select a food item and then display beers that best complimented this food choice. This idea received a lot of positive feedback, but getting the data proved to be a challenge.

We worked on a concept where people would pick between 3 broad color ranges of beer, which would then filter the chord to show them ratings for beers in that range. People could explore and pick a new beer, based on similarity in parameter ratings (aroma, taste, appearance, overall). People could also use a color picker tool to match the color of the beer they were drinking to our range of colors.

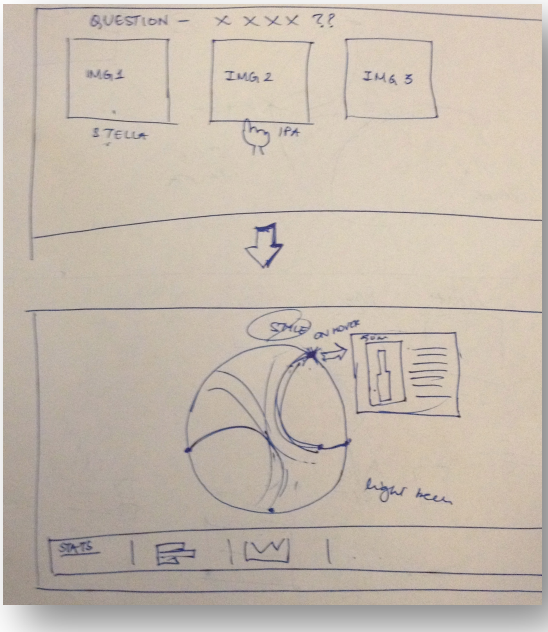


Figure 2: Paper wireframe of concept

Feedback: There was now a narrative, but people weren't able to give feedback on whether the tool was fun since they couldn't understand interactions on a paper drawing.

We also tried an option that involved picking beer from a lineup of bottles, and then showing bar graphs that compared the rating of the selected beer to overall. This was based on feedback received from the initial concept submission.

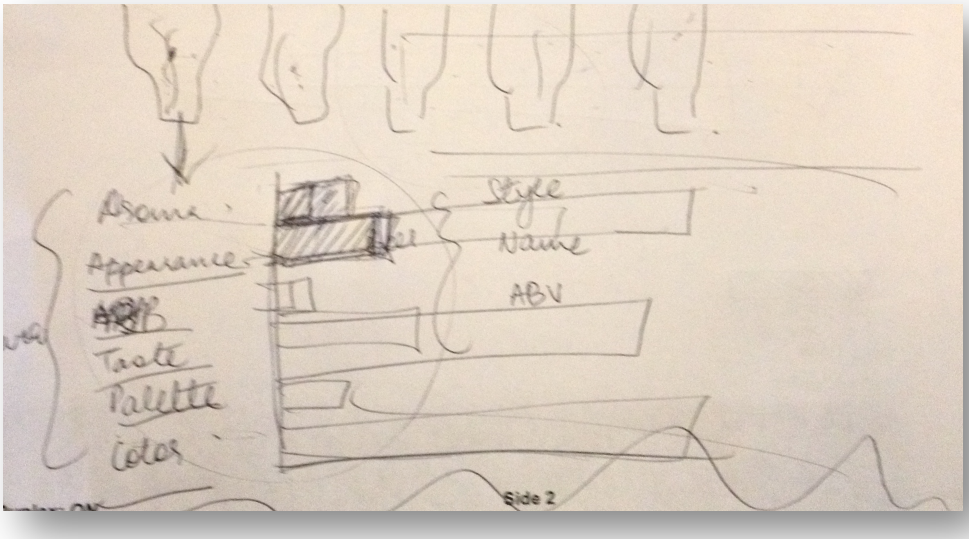


Figure 3: Paper sketch of the Alternate concept

Feedback: People weren't able to explore and compare multiple types of beer. Also, when we looked at the data we realized that the range of reviews was so wide that this comparison wasn't going to allow people to necessarily pick the best alternative.

We tested both layouts with different users and discovered people responded more positively toward the chord. Users said that it looked like a fun way to explore various styles of beer, and so we created a design prototype to get more feedback.

Data Cleansing and Transformation

The raw data was in a text file format and contained over 1.5 million records. We used python for data parsing and Excel for analysis.

Stage 1: Initial parsing

- We read the raw file using python. Removed fields like beerId, brewerId, profileName, and text, which were redundant.
- Converted timestamp into a year value.
- Filtered data between years 2009 to 2012 and dumped it into a csv file.

Stage 2: Analysis

- We used pivot tables to identify trends and created bar charts and line charts for data exploration

Stage 3: Transformation

- For the visualization, it was necessary to convert data into json format
- Read the csv file, transformed special character using Unicode encoding 'utf-8'
- Separate json files were created for aroma, appearance, taste and overall. This helped us modularize our code.
- Separate json was also created for summary charts.
- *In order to run ipython, it is necessary to [install](#) it.*

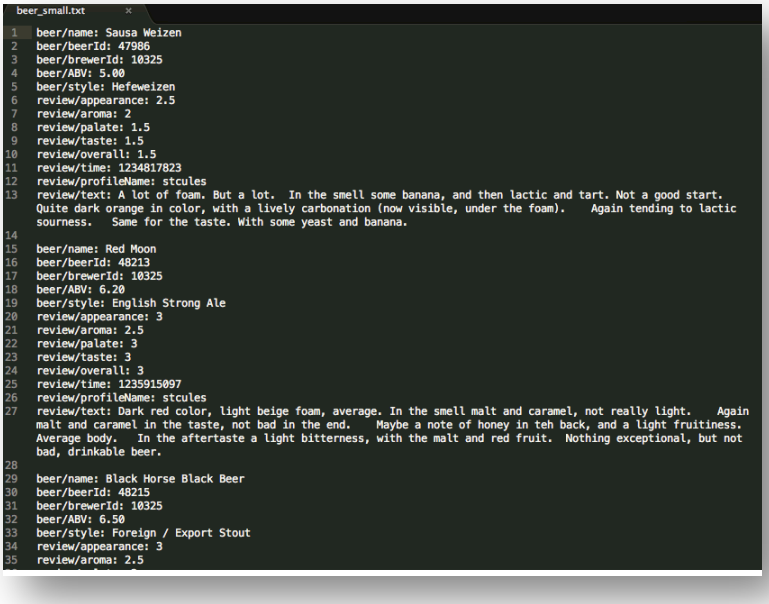


Figure 4: Original Data File

Figure 5: Final JSON file

Design Iterations

We began creating a rough layout by projecting our ideas onto a television, which helped us brainstorm.

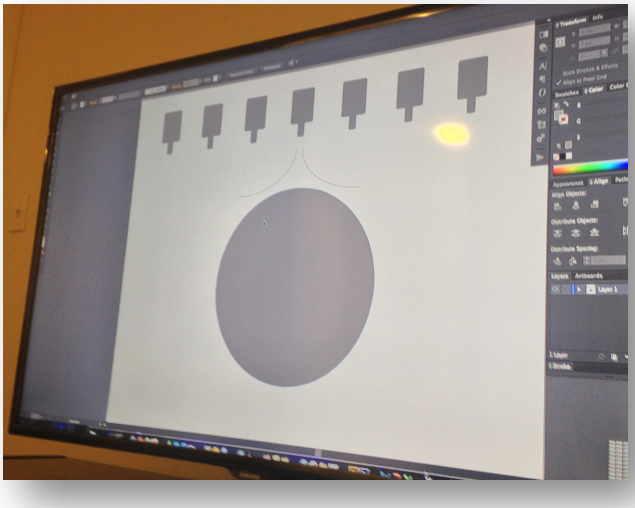


Figure 6: Display of collaborative drawing

We decided to incorporate the idea of allowing users to select a beer color, which would help filter the chord. We created a simple wireframe to outline the elements.

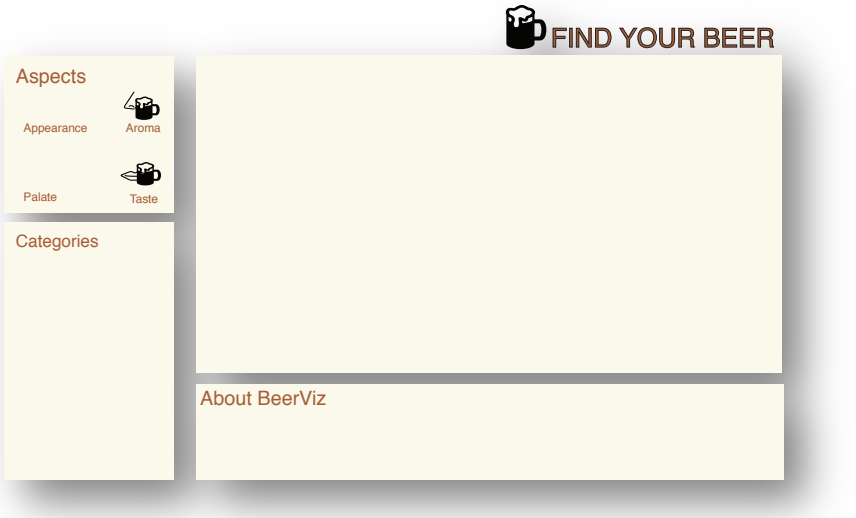


Figure 7: Wireframe of the webpage

The idea was to allow users to filter the chord based on the color of the beer and find beers similar to each other based on parameters. After this, we worked on a design that would incorporate these elements as well as include interactions.

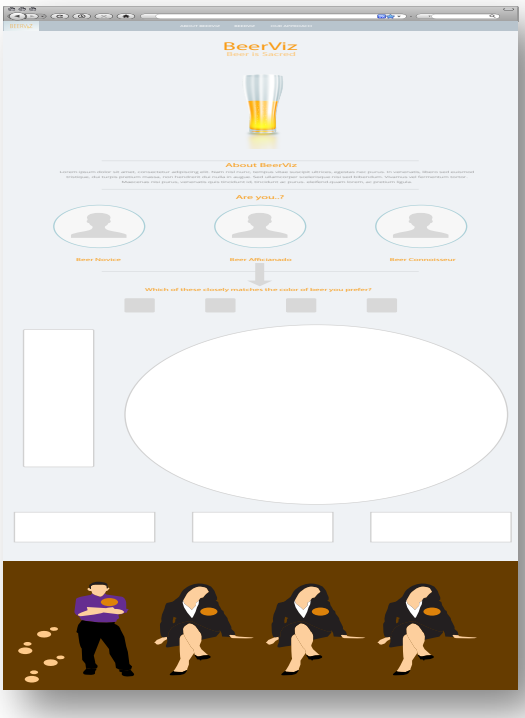


Figure 8: Page design version 1

Feedback: During the in-class design critique, people were unable to state what personality type they were. However, they liked the idea of filtering by color of the beer. We had a chord with minimum data for the critique and people enjoyed exploring it, so we decided to work on improving it.

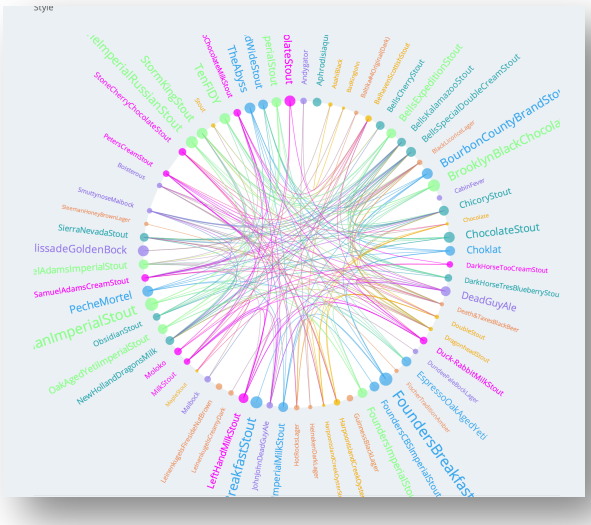


Figure 9: Chord with all beer styles

But once all the data was incorporated, we realized there was a plethora of colors exploding on our chord. In order to simplify this, we worked on clustering. This led to a cleaner layout that had beers grouped by style and then an attribute (e.g. aroma). The resulting visualization for the chord –

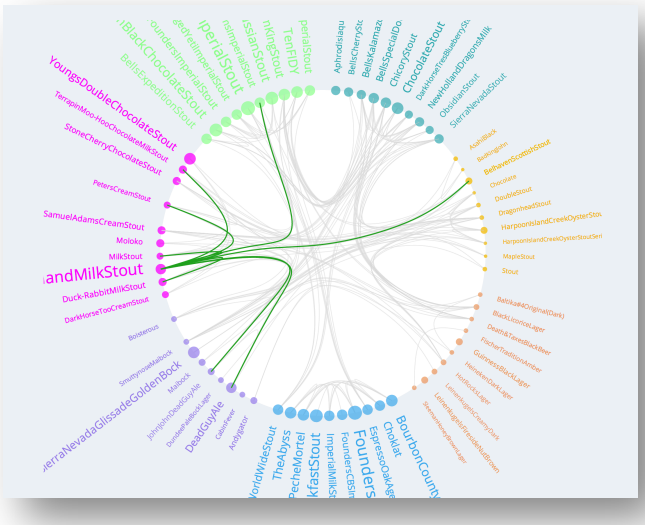


Figure 10: Chord with clustered beer styles

This was our **turning point**. It allowed us to show the relationships (similarity in rating) between different beers, while also showing the clusters of beer styles, and using the bubble sizes to reflect popularity (more the of users who rated the beer, more popular the beer is).

We used 'Agglomerative Clustering', a subset of hierarchical clustering that uses a bottom-up approach where each observation starts in a cluster and pairs of clusters are merged as one moves up in the hierarchy (Wikipedia), The package we used was Michael Bostock's clustering bundle (vendor/packages.js). The design choice of using a circular layout helped us in showcasing the multi-dimensionality of our data.

We then worked on improving the narrative, the interactions and other visual elements in the design, based on feedback. We also worked on incorporating annotations. It took us a couple of iterations to get our page-text right. For example we initially called the parameter select bar "Filter beers by", but that got feedback that users expected the beer sets to change because of the word filter. Once we changed it to 'Find Similar beers by' people were able to understand it. We also redid the colors on the chord to make them more pleasing to the eye. Our final visualization –

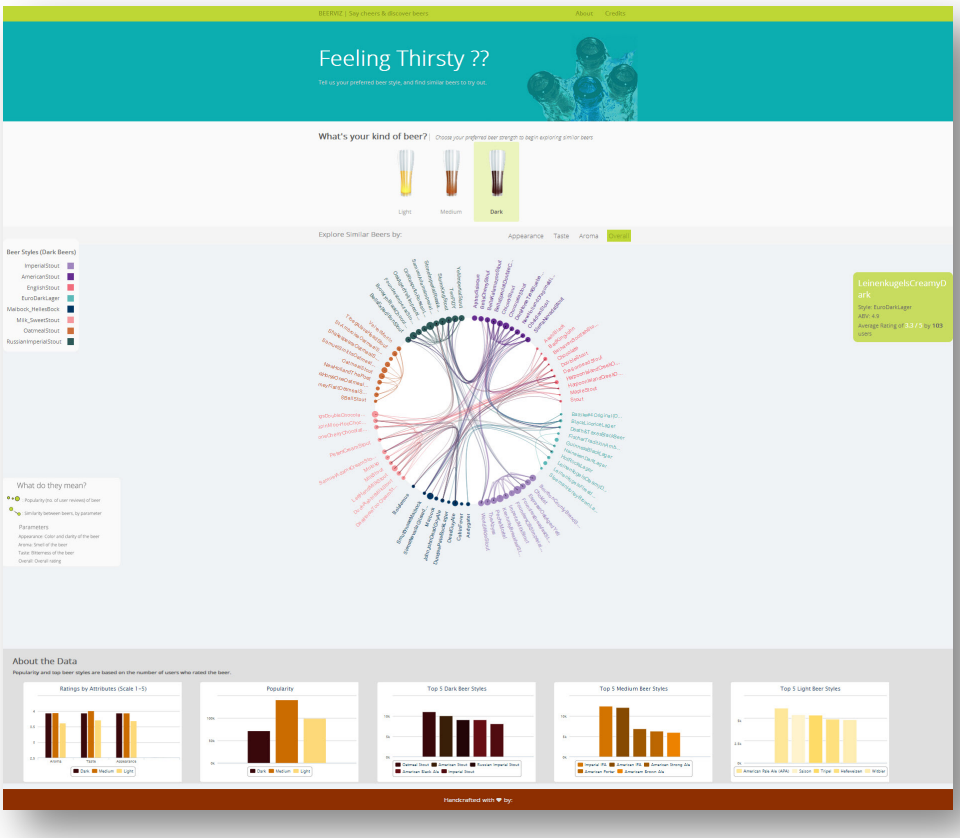


Figure 11: Final Visualization

Feedback: Users found it fun to use, and had a great time exploring various beer styles through the layout. At this point we ended our design iterations.

Color Choices

We chose beer colors of light yellow, brown, and dark brown as the main colors as it is related to the beer colors.

Key Challenges and Design Decision

- **Size of the Dataset:** Our data set had 1.5M records, which was too big for Excel explorations. We used Python to reduce the size of the dataset by removing the field that contained qualitative comments. The reason we chose

this field to eliminate was the necessity to parse them to truly derive meaning from the data. We also decided to use data from 2009 – 2012 to ensure the ratings reflected the most recent user feedback.

- **Data Formats:** Beer names used a lot of special characters that posed a challenge while cleaning. We addressed this through UTF-8 encoding.
- **SVG Limitation:** We were forced to use Camel Case labels so as to allow the hover function to work since SVG text broke linkages when there was a space. This was a technical limitation we couldn't resolve.
- **Beer Name Lengths:** In order to ensure that the chord was displayed within a certain div size we had to truncate beer names to 20 characters. Additionally we added alt-text to help the user.

Task Delegation

We formed a team where each of us had expertise in a certain area so we could all learn from each other and create an effective visualization.

Team Member	Tasks
Divya	User Research, Visual Design, Report, Presentation
Evie	Data Analysis and Key Findings, Blogpost, User Research
Shreyas	Coding, Interaction Design, User Research
Sonali	Data Analysis, Coding, User Research