Project Proposal

Project Title : Visualization of Indian Cricket Premier League

Team Members: Zeeshan Hakim (u1143824) , Abishek Penujuri Nataraj (u1080870)

Link to Project Repository: https://github.com/zeeshanhakim/dataviscourse-pr-2017

Our project idea is based on **"Visualizing Indian Premier League"** dataset.

Background and Motivation:

Almost every kid in India and Pakistan grew up playing cricket and idolizes its players. It is one of the most loved games in the world. USA has football, Europe has Soccer and we have Cricket.

After introducing T20 format, the craze for it increased. One of the most successful premier league based on this format is IPL (Indian Premier League). IPL brought together players from all the countries to play in different teams. The Dream of seeing players from different countries playing for the same team came true.

So, while deciding on the Project, our main interest was to find something interesting related to Cricket. We found a dataset related to IPL and after going through its features and asking opinion from TA's, we decided to go with this Project.

Our goal here is to visualize various trends covering 9 seasons of the IPL. The data has lot of features which can be used to generate good visualizations, the preliminary designs are shown in the upcoming sections. we also show interesting insights such as how participants from global countries change with respect to years especially in the context of India and Pakistan.

Project Objectives:

The main objective of our project is to visualize the trends and season, match, and player specifics of IPL.

Things we are going to answer with our visualization are:

1. One is visualizing the number of matches played throughout the country, showing on the map (maybe by using google API). We can learn where most number of matches are being played.

We can also find the change is trends from each season by observing where the most games have been played for the season

2. From the map, using hover feature we can see other important details regarding the matches.

3. We are also going to visualize various statistics of the IPL, for example player is awarded Orange Cap and Purple Cap based on their performances in the season, we would like to visualize these for every season and other statistics like Man of series, number of sixes and number of fours etc.

4. Also, we like to find out about player skills, like which all players are Right-Handed batsman or Left-Handed and their bowling skills, whether they were captains or not, whether they umpired or not, etc.

5. One of the important things which we try to understand is the trend in international players in the IPL, like how many players are there for each country in the given season.

There may be many other things, which can be answered using our visualizations, on forward we hope to convey many interesting trends and visualizations and answer many questions in this project

Data:

we collected our data set from Kaggle

https://www.kaggle.com/harsha547/indian-premier-league-csv-dataset

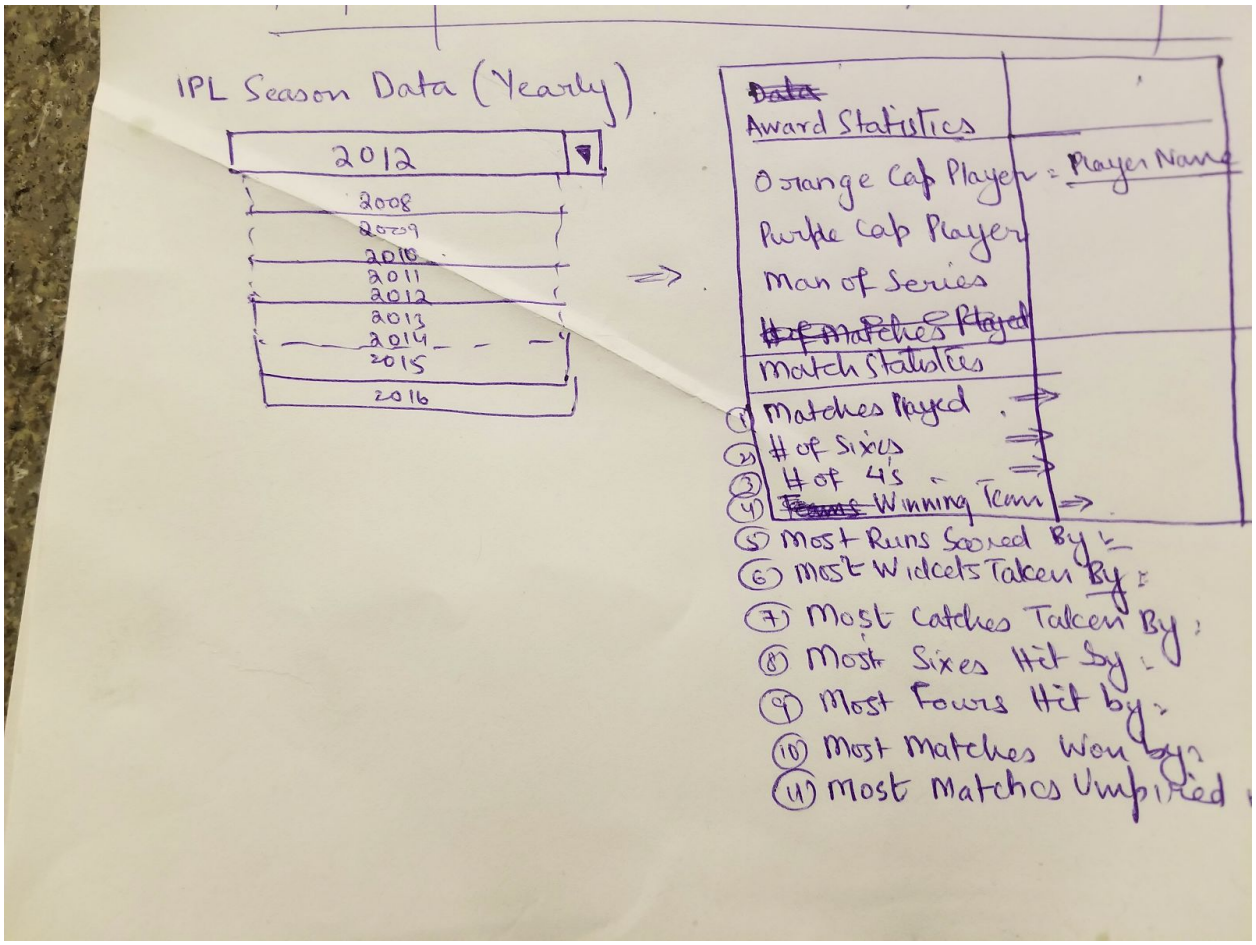
Data Processing:

Yes, we may need to do some data cleaning to remove unwanted elements.

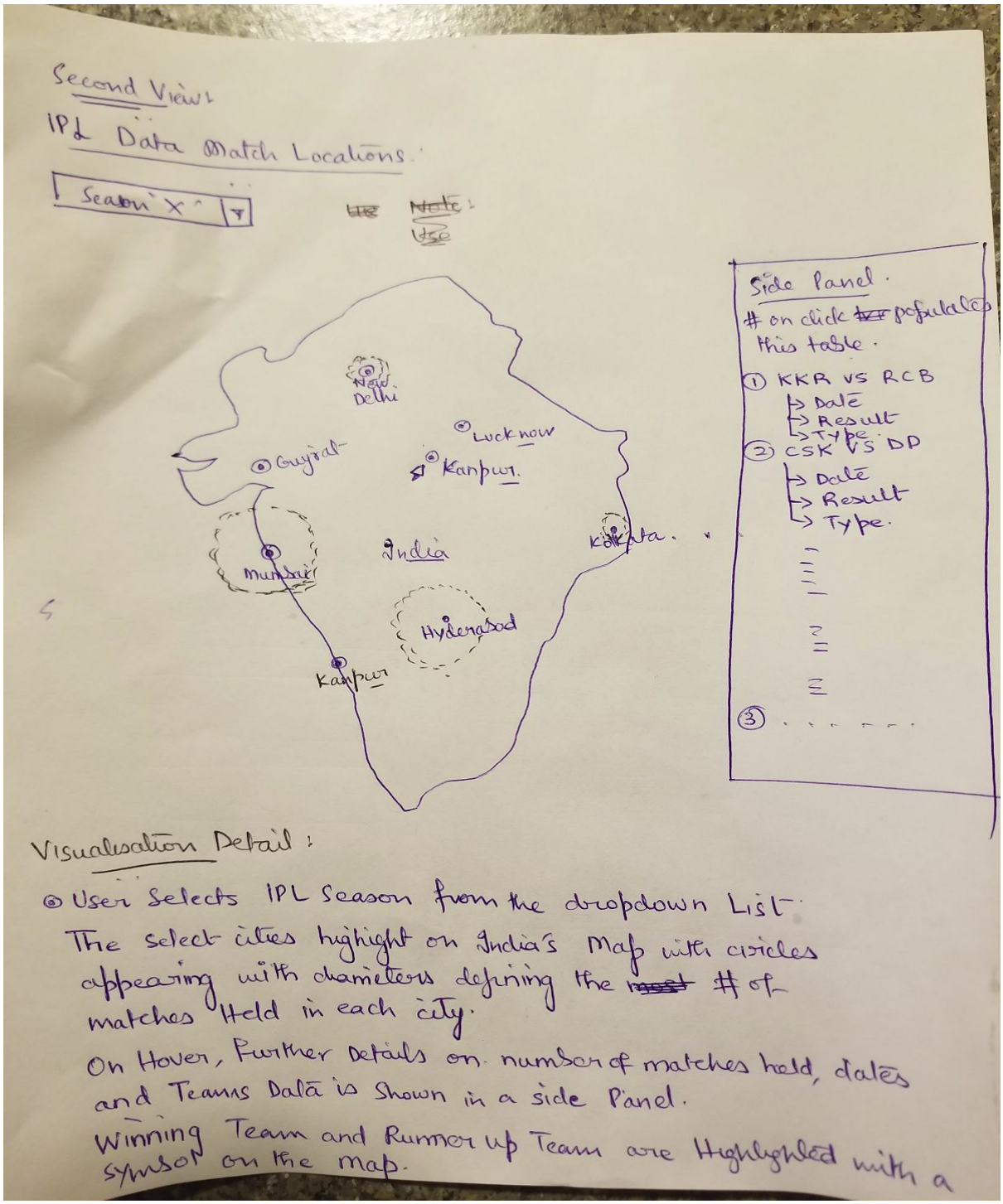
We have the data in the CSV files, we may need to data processing to convert some of the data into JSON Files. In total we have 6 data files covering all the data related to 9 seasons. There is player data, match data, season data and team data. All these data files are interlinked with common id’s

Visualization Design:

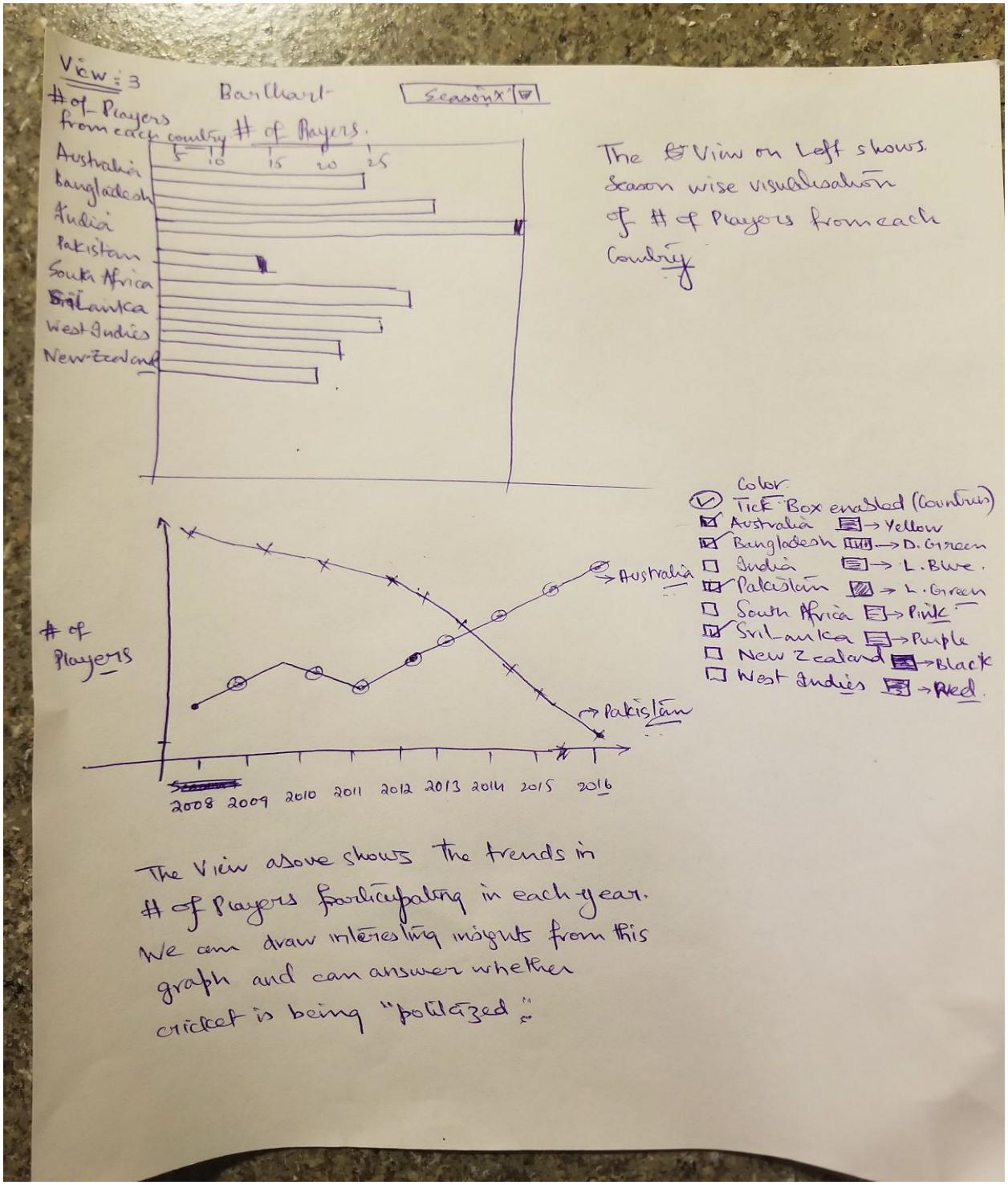
VIEW 1 : Showing the IPL Seasons Statistics



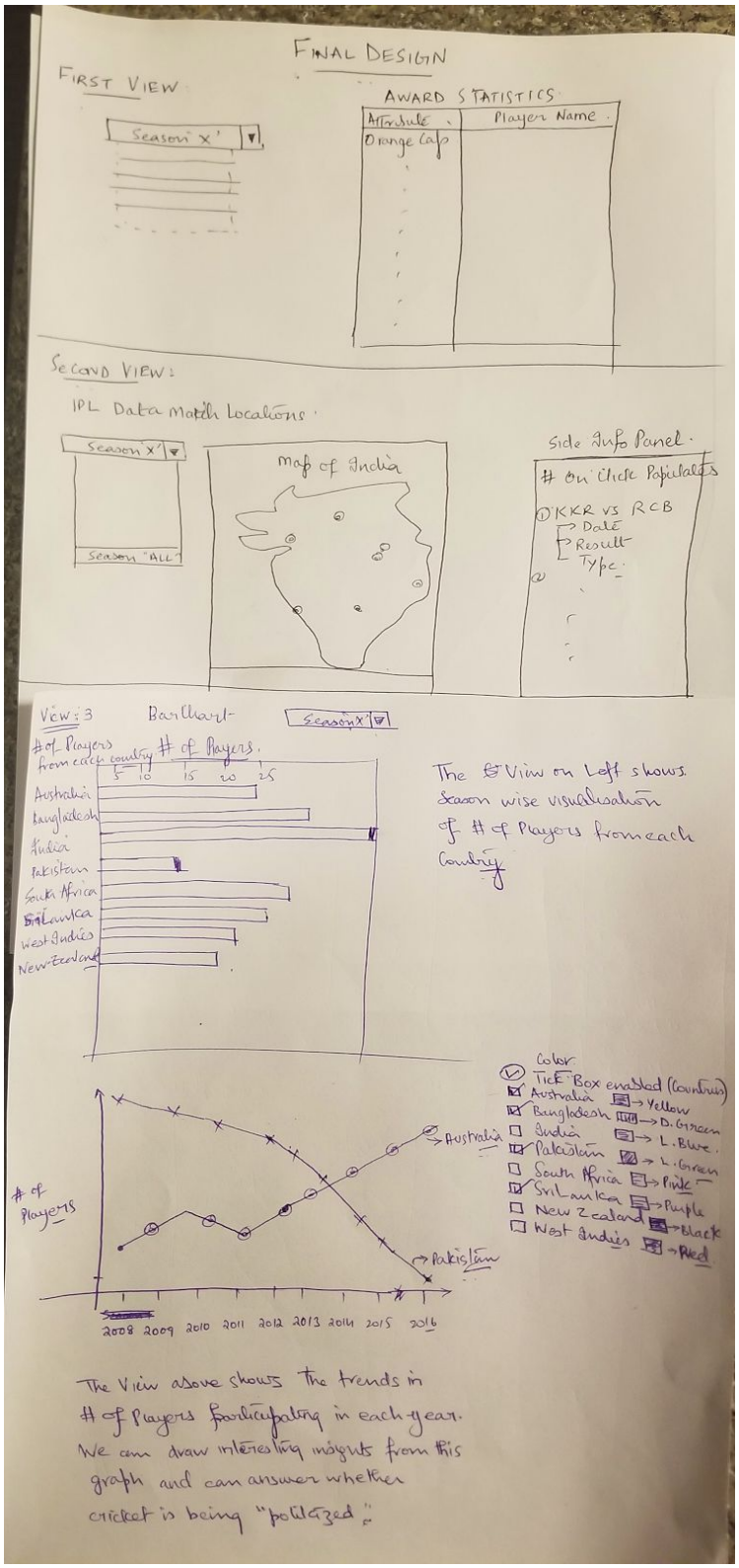
VIEW 2 : Showing the Season Matches Venues



VIEW 3 : Global Countries Participation In IPL Seasons



Final Design:



MUST-HAVE / NICE TO HAVE FEATURES:

- The View 1 showing season's statistics should have basic information about each season.
  A Nice to have feature in this visualization would be displaying more interesting options for users to choose from. From Example, a similar drop down list of showing each team would display team specific data.
- The View 2 would have a map of India showing each venue and the circles' diameter representing the # of matches played in each venue for the Season.
  A Nice to have feature would have a side panel in the side which shows venue specific match details for the season i.e. What matches played? Between which teams ? and Winning Team. This side panel can contain more interesting information for the visualization.
- The View 3 is the Star of the project. This visualization shows the global countries participation in the events and how the trend changes with respect to years/season.
  A nice to have feature would be drawing further trends showing the performance of players from each country like Avg runs scored by players from each country, Which Country's players scored most 6's etc.

Project Schedule:

We are planning to commit around 10 hours per week.

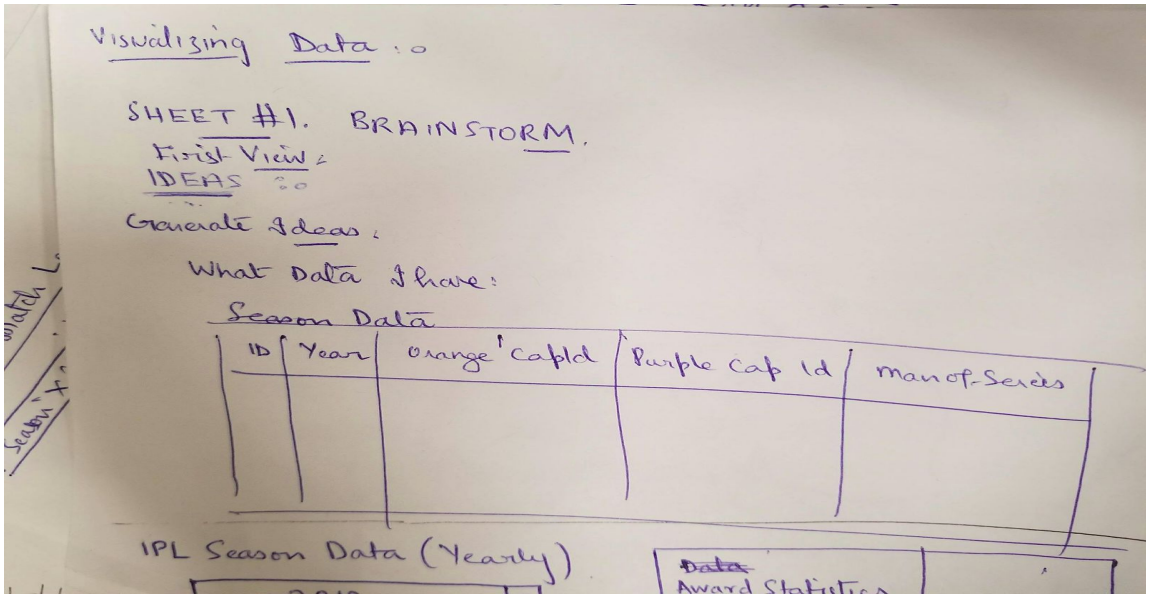
Week1: Data Cleaning, Data Processing, Initial Setup

Week2: View 1 and View 2

Week3: View 3 and Process book

Week4: Final design and Process book

Brainstorming on another view angle:



Developing Views and Data Processing:

The main hurdle so far for us was to perform the data processing. Having raw .csv files in our dataset which are heavily dependent on each other as well as to draw various insights from them has been a huge preliminary task in order to allow us to carry our visualizations implementation.

Having successfully done data processing and finding ways to employ the raw data in usable form, we have put a rough outline of how our visualizations would look like. For now, we have constructed disconnected but useful visualizations very close to what we proposed in our proposal.

For View 2, as we suggested in our proposal we need to visualize a map of India with IPL hosting cities. For this (for the sake of simplicity) we have included a Googles Maps API visualization but in some days, we plan to incorporate **Google Geo Charts** that will have a wide variety of visualization options including the one we proposed in our proposal i.e. Accoding to the diameter of a circle hovering on a city on map is the number of matches hosted per city in a season. We have currently tried to

visualize our ‘host city for season 1 using a bar chart’ **as an evidence of our successful data processing.**

We have also tried to implement a side panel that was proposed in our proposal aimed to contain information regarding matches that are were held at a venue.

Moreover, another interesting insight we draw in our prototype visualization is the Runs by which a team loses the match. This is a bonus visualization which we think can be important to view Season wise trends.

We have started working on the basic design of the view 1. We have visualized a basic table structure reading our data and have implemented a functional dropdown feature to select the season which updates the table. The data in the table is static and must be updates with actual names instead of ID’s which we have shown, we are consequently working on it to update the table by selecting the season in the dropdown.

Plan to do next: for now, the drop-down feature is only referring to our view 1 chart, further we want to make the same drop down call all the views and update our charts based on the season number. Want to make our table dynamic, we are also thinking adding some more features to the table which we want to calculate using the data and update the table.

If possible, we want to add another table to show for the complete seasons, just populating the important stats of the season.

Again, I think we were successful in achieving our big milestone of data processing and now we are all set to create some solid visualizations without much effort. Due to the complex nature of our data comprising of various dimensions and the depth of story we are trying to visualize, data processing was specifically the task we needed to invest much time in.

Peer Feedback

We got feedback from Hannah Swan, she gave us the feedback based on the questions given in the Feedback form, have mentioned all the feedback below based on her opinions

Are the objectives interesting to the target audience?

It was interesting to people who like sports, it gives a basic understanding for those who are not familiar with cricket.

Is the scope of the project appropriate? If not, suggest improvements.

The scope of project is good and is doable. If all the must have features are implemented, it would have a good story to tell

Is the split between optional and must-have features appropriate? Why?

Yes, the split between the optional and must-have features is appropriate as must have features are required for proper visualization and optional are the one which will tweak the visualizations

Is the visualization innovative? Creative? Why?

The view 1 and view 2 are standard visualizations purely inspired from designs in class and homework. The view 3 is creative and has interesting analytics

Does the visualization scale to the used dataset? Could it handle larger but similar datasets?

Yes, the visualization scale to the used dataset and the techniques can be implemented easily for other sports related data

Is the project plan detailed enough? Is a path to the final project clear?

The project plan is detailed and is also flexible to make changes and work on it.

Is an interesting story told?

Hannah doesn't love sports however she felt we could tell an interesting story using the visualizations

Does the visualization follow the principles used in class?

Many things learnt in class like marks, channels, geo-spatial, bar graph, line graph and scales are clear in their plan for implementation

What is the primary visual encoding? Does it match to the most important aspect of the data?

The primary visual encoding is set of charts, as each view will show different features and all of them are derived from important aspects of the data

What other visual variables are used? Are they effective?

Marks, channels are clearly stated, and color is based on the team jerseys and all these are used effectively

Is color sensibly used? If not, suggest improvements.

Yes, the color is used sensibly

Is the interaction meaningful? If not, suggest improvements.

Planned interactions are meaningful, and all the charts would be better interconnected

If multiple views, are they coordinated? If not, would it be meaningful?

Multiple views are coordinated, but having a same dropdown for all the views would be sensible

Is there any animation planned? Is it clear? Is it intuitive?

Animation is not planned for now, when the implementation of interconnectivity between views are done, having good animations help in conveying a story or change.