Evolution of Cricket

By Abhiram Eswaran

Details:

- Hosted Url: https://goo.gl/RGIy4V
- Hosted on: Edlab server.
- Direct access url: http://elnux2.cs.umass.edu:13798/

Overview

Ever since the advent of T20 cricket the gentleman's hasn't been gentle on the bowlers. Thicker bats, shorter boundaries, multiple power-plays, free hits, switch hits and all in all the audacity to chase down impossible targets has undeniably changed the game of cricket. The visualizations presented here tells you the evolution of the Gentleman's game to Power hitters' game.

We look at the following in particular,

- We look at time-wise Average scores and observe the number of 300 plus scores made over the years. We also observe the team wise run rates over the years. The percentage of matches going over 300 with a run rate greater than 6 is alarmingly high.
- The batsmen have been proportionally scoring at a higher striker rate. We also see the rise of power hitters who have very good strike rates but lesser averages like David Miller and Kedhar Jadav
- The economy of bowlers have increased. The changes have been cruel on the bowlers and that reflects on the average economy rates.

Interest and Questions answered

As an avid cricket fan, this project gave me an opportunity to explore cricket data and how the game has evolved over the years.
Questions answered with this project:

Among team data,

1. How has the game changed over the years, this will be shown by the average runs scored by teams by the year?

I present a horizontal bar chart to answer this.

2. How do teams fare in home and away games? This will answer whether teams perform better at home than away games.

Instead of analysing this I present a holistic view of wins and losses of teams over the years. This is done with a bar chart and a map chart.

Among batsmen record,

1. We look at the records of batsmen, specially the best performing batsmen.
   This is done with a line chart which shows a comparison of players.
2. We explore the hypothesis that the best perform are good performers either home or away.
   This is done with a scatter plot which shows the best performing players for each team.
3. We pick few top batsmen and compare their histories, centuries and individual stats.
   This is done with a drop down selection of individual batsmen and their performances presented in a bar chart.

Among bowlers,

1. We explore the records of the best performing bowlers.
   This is done with a line chart which shows a comparison of players.
2. We pick the top bowlers and see how their game evolved over the years.
   This is done with a drop down selection of individual bowlers and their performances presented in a bar chart.

Data Analysis

- Initially the entire dataset was built from scratch by scraping data from Cricinfo, a website which hosts cricket data.
- From the data obtained the following were computed mainly because these stats were critical to the story being told and were not readily available from the scraped data.
  - Average statistics by teams
  - Average statistics by the years
  - Average statistics of players
  - Combined statistics of batsmen and bowlers.

  This was done using Pandas[1] and Numpy[2] in python.
- Filtering: After obtaining the required data, individual filters were added for each visualization which would present the requirements of the visualization. Like individual players statistics etc. This was done using CrossFilter[3]

The Story

The story presented is the evolution of the game of cricket. It is largely believed the game of cricket has moved in the favour of batsmen with larger scores being scored and

chased down. With this project I support this belief with facts. By analysing the data I found the following trends.

1. There has been increased in the average strike rate of batsmen over the years.
2. We see a new breed of players known as power hitters rise who average less but have very good strike rates.
3. The bowlers economy rates have steadily increased.
4. Few teams have lost their shine over time like West Indies for example.
5. Teams are consistently crossing 300+ in one day internationals.

The entire story can be read at the project url, https://goo.gl/RGIy4V

The Implementation

The implementation was done in 3 parts,

1st part:

Data collection and analysis: This was done in Python. The data was scraped with BeautifulSoup, Loaded and formatted with Pandas and analysed with Numpy and Python and saved as CSV.

2nd Part:

Database and the backend app: The CSVs were then loaded into MongoDB to enable easy querying. The backend app was developed with Python using Flask. The app queries mongodb for data and forward the data to the front end.

3rd Part:

Front-end and the visualizations: This was developed using D3.js, Dc.js, CrossFilter, Bootstrap, CSS and Html. The data is fetched from the flask app apis into the html templates developed and loaded on the go.

Interaction guide: Each visualization is accompanied by a tip for the supported visualizations.

DataSet and Url

The final scraped dataset is presented with the code. The dataset was scraped from here,
http://stats.espncricinfo.com/ci/engine/stats/index.html?class=2;filter=advanced;orderby=team_score;size=200;spanmin1=05+Jan+1994;spanval1=span;template=results;type=team;view=innings

The url for the visualizations is mentioned at the beginning of the report.

References

1. Pandas http://pandas.pydata.org/
2. Numpy http://www.numpy.org/
3. CrossFilter http://square.github.io/crossfilter/
4. Inspirations for layout: Stack overflow developer survey. https://insights.stackoverflow.com/survey/2017
5. Development references for DC.js: https://dc-js.github.io/dc.js/
6. Stackoverflow for general references
7. Data obtained from Espn Cricinfo, http://www.espncricinfo.com
8. BeautifulSoup for scraping data, https://www.crummy.com/software/BeautifulSoup/
9. CSS help from: Bootstrap, https://github.com/twbs/bootstrap

Course Suggestions

There was a disconnect between visualizations and analytics in the course, it would have been great if they both went together, Like assignments involving a mix of analytics and visualization would have been better. But, all in all a great course, learnt the concepts, learnt a new language and set up an end to end system. Lot of things to take away. Thank you.