School of Engineering and Computer Science
Te Kura Mātai Pūkaha, Pūrorohiko

PO Box 600
Wellington
New Zealand

Tel: +64 4 463 5341
Internet: office@ecs.vuw.ac.nz

Cricket Data Visualisation App

Kevin Jose

Supervisor(s): Stuart Marshall

Submitted in partial fulfilment of the requirements for:
Bachelor of Engineering, majoring in Software Engineering

Abstract

Cricket is a complex sport which provides a great wealth of online data for potential analysis and discoveries. The problem is that this data is not very accessible or digestible because it is presented in a table format and in large amounts. The purpose of this project is to implement a data visualisation application for cricket data. The following are the resources I will need to arrange: $200 worth of koha for the user-testing participants, a GitLab repository and a lab room. The expected output of this project will be a web-based application which allows users to visualise cricket data. I plan to test and evaluate my solution with a user-study.

1. Introduction

Cricket is the second most popular sport in the world. According to [2], it is estimated that globally, around 2.5 billion people follow cricket. There are three different formats in which cricket is played: test cricket, one day international cricket (ODI) and twenty20 (T20) cricket. Each format is unique and provides different datasets. In this project, I will be focusing on the batting data within the one-day international (ODI) format.

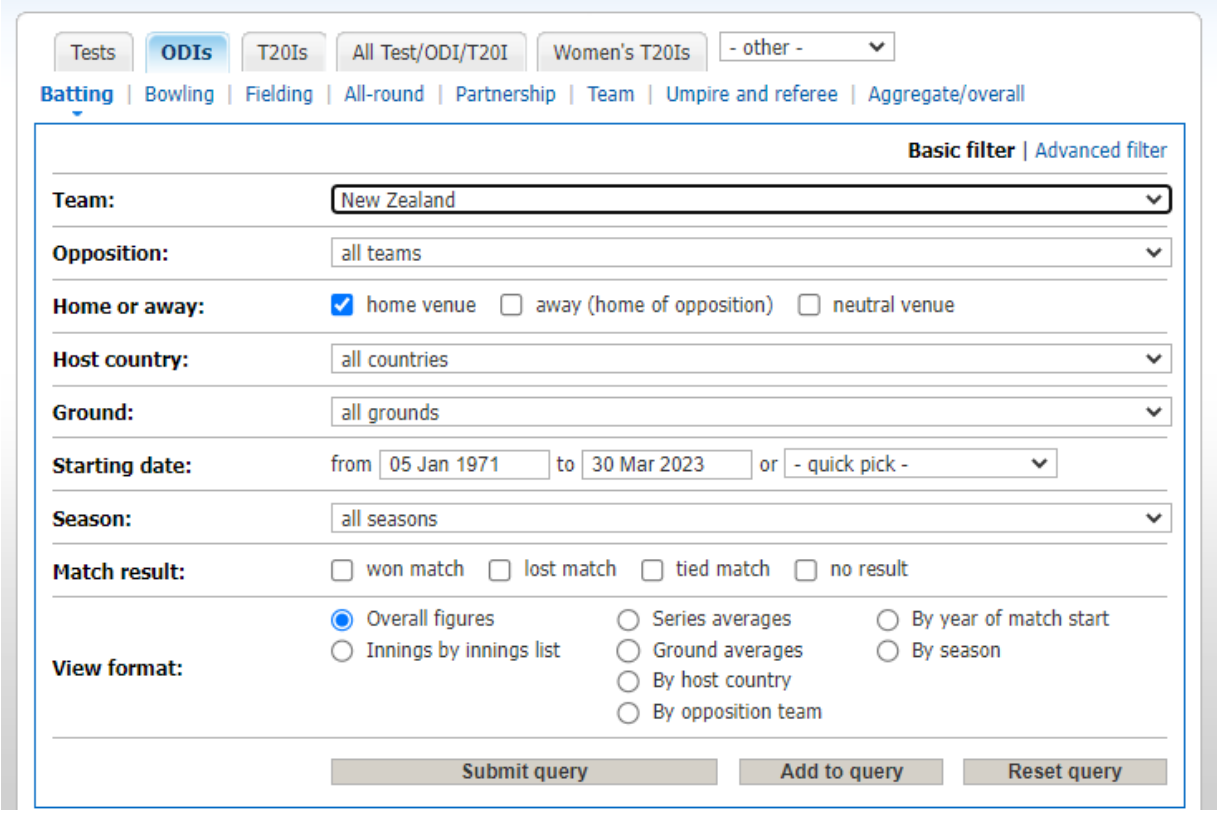
There is a great amount of cricket data available online (e.g. ESPN cricinfo) but this data is not very accessible or digestible because it is presented in a table format and in large amounts. If this data was presented in a visualised format, users would be able to explore and analyse the data in order to find patterns and trends and make predictions. The purpose of this project is to implement a web application that enables users to visualise cricket data.

2. The Problem

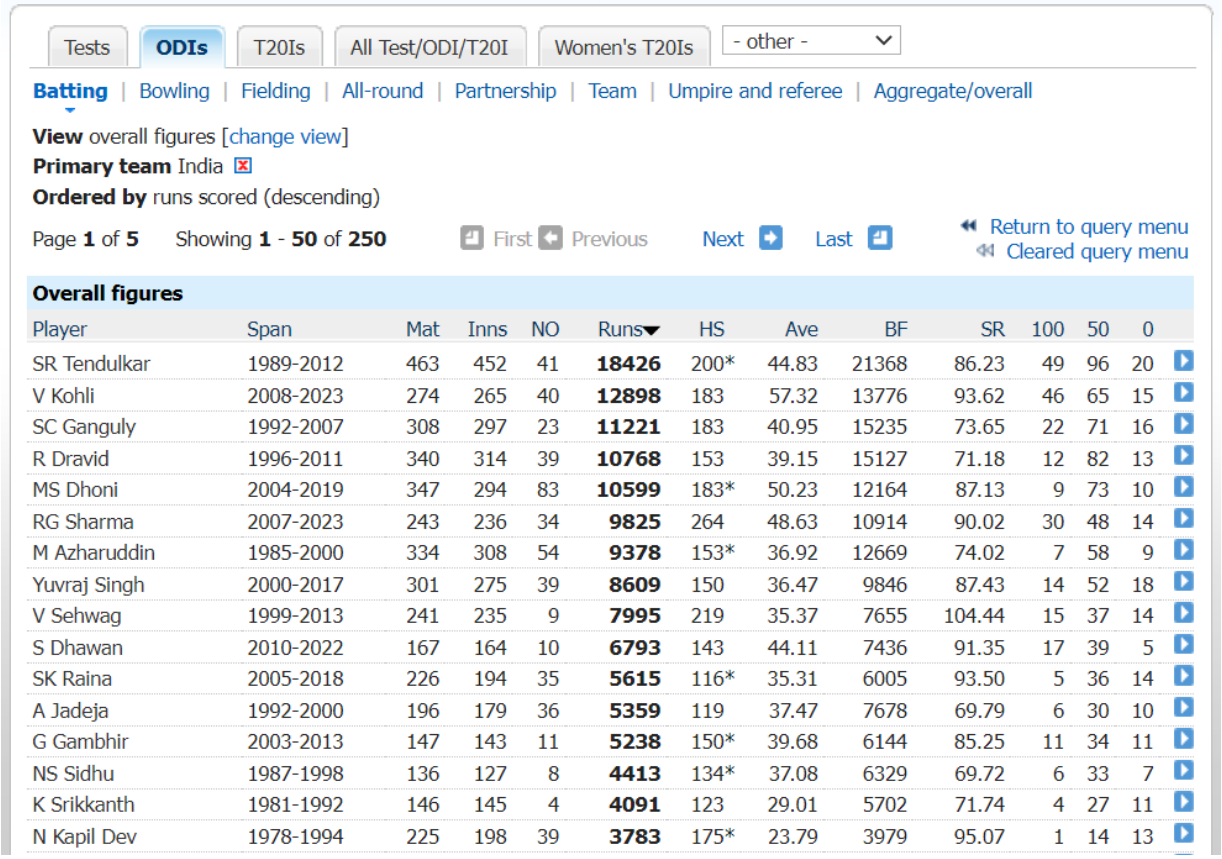
Cricket is a complex sport which provides a great wealth of data for potential analysis and discoveries. Statsguru [3] (by ESPN cricinfo) is an online tool which allows a user to query data from its cricket statistics database. After the user submits a query, Statsguru returns a table of data which contains all the resulting rows (as shown in screenshots below). Statsguru contains the largest online database of officially recorded data in ODI cricket. Its data ranges from the beginning of ODI cricket (1971) to the present. This tool is powerful in the sense that it allows users to choose from a range of filters and it contains all of the data in the history of ODI cricket.

Statsguru

Querying (choose from the filter options)



Results (in table format)



Tests | **ODIs** | T20Is | All Test/ODI/T20I | Women's T20Is | - other -

Batting | Bowling | Fielding | All-round | Partnership | Team | Umpire and referee | Aggregate/overall

View overall figures [change view]
Primary team India
Ordered by runs scored (descending)

Page **1** of **5** Showing **1** - **50** of **250** First Previous Next Last Return to query menu Cleared query menu

Overall figures

Player	Span	Mat	Inns	NO	Runs	HS	Ave	BF	SR	100	50	0
SR Tendulkar	1989-2012	463	452	41	**18426**	200*	44.83	21368	86.23	49	96	20
V Kohli	2008-2023	274	265	40	**12898**	183	57.32	13776	93.62	46	65	15
SC Ganguly	1992-2007	308	297	23	**11221**	183	40.95	15235	73.65	22	71	16
R Dravid	1996-2011	340	314	39	**10768**	153	39.15	15127	71.18	12	82	13
MS Dhoni	2004-2019	347	294	83	**10599**	183*	50.23	12164	87.13	9	73	10
RG Sharma	2007-2023	243	236	34	**9825**	264	48.63	10914	90.02	30	48	14
M Azharuddin	1985-2000	334	308	54	**9378**	153*	36.92	12669	74.02	7	58	9
Yuvraj Singh	2000-2017	301	275	39	**8609**	150	36.47	9846	87.43	14	52	18
V Sehwag	1999-2013	241	235	9	**7995**	219	35.37	7655	104.44	15	37	14
S Dhawan	2010-2022	167	164	10	**6793**	143	44.11	7436	91.35	17	39	5
SK Raina	2005-2018	226	194	35	**5615**	116*	35.31	6005	93.50	5	36	14
A Jadeja	1992-2000	196	179	36	**5359**	119	37.47	7678	69.79	6	30	10
G Gambhir	2003-2013	147	143	11	**5238**	150*	39.68	6144	85.25	11	34	11
NS Sidhu	1987-1998	136	127	8	**4413**	134*	37.08	6329	69.72	6	33	7
K Srikkanth	1981-1992	146	145	4	**4091**	123	29.01	5702	71.74	4	27	11
N Kapil Dev	1978-1994	225	198	39	**3783**	175*	23.79	3979	95.07	1	14	13

Although Statsguru is a good data querying tool, its limitation is that it does not provide any features for data visualisation and therefore limits the data's accessibility and analysis possibilities. As previously mentioned, Statsguru returns a table of raw data and this can be difficult to analyse and comprehend because of the complexity and factors involved in the sport. Due to the lack of data visualisation (such as graphs) in this tool, users are missing out on a lot of learning and findings from this data. In the last 52 years of ODI cricket, there have been many factors that have changed and evolved the sport [4]. Some of these factors include: rule changes, improvements in equipment, improved sensor technology. In order to discover these factors and their effects on player and team batting performance, users must be able to visualise the data as charts or graphs.

It is evident that cricket fans and other inquisitive users need a new tool which provides analysis of cricket data in the form of data visualisations. Data visualisation is an effective way of presenting information to end-users because it makes it much easier for the human brain to understand and pull insights from [1]. Statsguru's dataset contains over 2500 cricket players and each player contains 13 attributes. With such a large and complex dataset, we need a program which condenses this data and makes it more presentable so that the user can explore, interact and visualise the cricket statistics.

3. Proposed Solution

In response to the above problem, the aim of this project is to implement a web application that visualises ODI cricket data. The app must be interactive and intuitive and must allow users to query cricket data (relating to ODI batting) and then generate visual charts for further analysis. The impacts of this data visualisation app are that users can learn more about cricket and its rich history by

discovering patterns and trends in the data. This will give fans a greater appreciation for the sport, in addition to educating those who are new to cricket and wish to learn more.

The duration of this project will be around 7 months (spanning across trimesters 1 and 2). It is currently week 5 of trimester 1. There are 9 weeks remaining in this trimester, including the breaks. During the 3-week period of 3rd - 24th April, I plan to do the following: project research, learning the D3 JavaScript library and beginning development of the web application. During the following 3-week period (25th April – 15th May) I will continue development of the web application and will begin writing the preliminary report.

4. Evaluating your Solution

In the later stages of the project, I will evaluate the web application by performing user-testing. I will conduct a user-study with 10 people in order to identify any usability issues in the system and to assess the overall user experience. By doing so, I can make the system more intuitive and user-friendly.

5. Resourcing and Ethics

In this section, I will detail any resource requirements such as hardware, software or access.

5.1. Hardware

This is not of concern to my project.

5.2. Software, Datasets and Models

There is no specialised software involved in this project. Although, I will be using a dataset which has been downloaded from Kaggle.com as a CSV file. In addition, I will be accessing some data from Statsguru [3].

5.3. Space, Virtual Resources and Access

In order to run user-studies, I have access to an HCI lab (CO 140). In addition, this system is a web-based application that can be run on my laptop therefore I have the option to test in alternative lab rooms as well.

Version control is an essential aspect of software engineering. In this project, I will be using a GitLab repository to store the codebase, to manage issue tracking and to document the project.

5.4. Budget

I will need $200 worth of vouchers as koha for the user-testing participants.

5.5. Ethics

I will be applying for user-testing approval from the Human Ethics committee at the end of Trimester 1. The user-study will occur in the middle of Trimester 2.

5.6. Safety

I will ensure that I am following healthy work practices:

- Working on a desk
- Sitting with proper posture
- Taking regular breaks to avoid eye-strain

5.7. Intellectual Property

I will be using open-source libraries and publicly available data. There is not anything immediately marketable about the system.

References (IEEE)

[1] Brush, K. and Burns, E. (2022) *What is data visualization and why is it important?*, *Business Analytics*. TechTarget. Available at: https://www.techtarget.com/searchbusinessanalytics/definition/data-visualization#:~:text=Data%20visualization%20is%20the%20practice,outliers%20in%20large%20data%20sets (Accessed: March 31, 2023).

[2] Smith, T., Sakr, M. and Worthington, J. (2022) *How popular is cricket around the world? [statistics]*, *SQaF*. Available at: https://sqaf.club/how-popular-is-cricket/#:~:text=Is%20Cricket%20a%20Popular%20Sport,Pakistan%2C%20Sri%20Lanka%20and%20Bangladesh (Accessed: March 31, 2023).

[3] *ESPN CricInfo: Statsguru* (no date) *Cricinfo*. Available at: https://stats.espncricinfo.com/ci/engine/stats/index.html (Accessed: March 31, 2023).

[4] *The evolution of ODI cricket | Andy Bull* (2015) *The Guardian*. Guardian News and Media. Available at: https://www.theguardian.com/sport/2015/feb/10/the-evolution-of-odi-cricket (Accessed: March 31, 2023).