An Analysis of ODI matches in Cricket

Sneha KK

Introduction

• With over 2.5 Billion followers, cricket is one of the most liked sport in the world . Its highly uncertain nature keeps us rooted to our seats but also makes it difficult to predict the matches with accuracy.
• In this project we will analyze data gathered from cricket websites on ODI matches to find patterns.
• We will also try to use machine learning algorithms and data on past winners to predict ODI match winners . This prediction can be purely for entertainment purposes or can be used for planning and betting.

Dataset Description

We will be using data from two files: ODI-data-1971-2017.csv(3932 rows x 7 columns) and ODI-data-2017-2021.csv (495 x 8 columns) .

The columns in the dataset are:

- Team 1,Team 2: Teams that compete in an ODI match
- Scorecard: ODI match no
- Winner : ODI match winning team
- Margin : Won by runs/wickets
- Ground : Cricket ground where the matches were held
- Match Date: Date of the match

The data was collected from the website : https://stats.espncricinfo.com/ci/content/records/307851.html

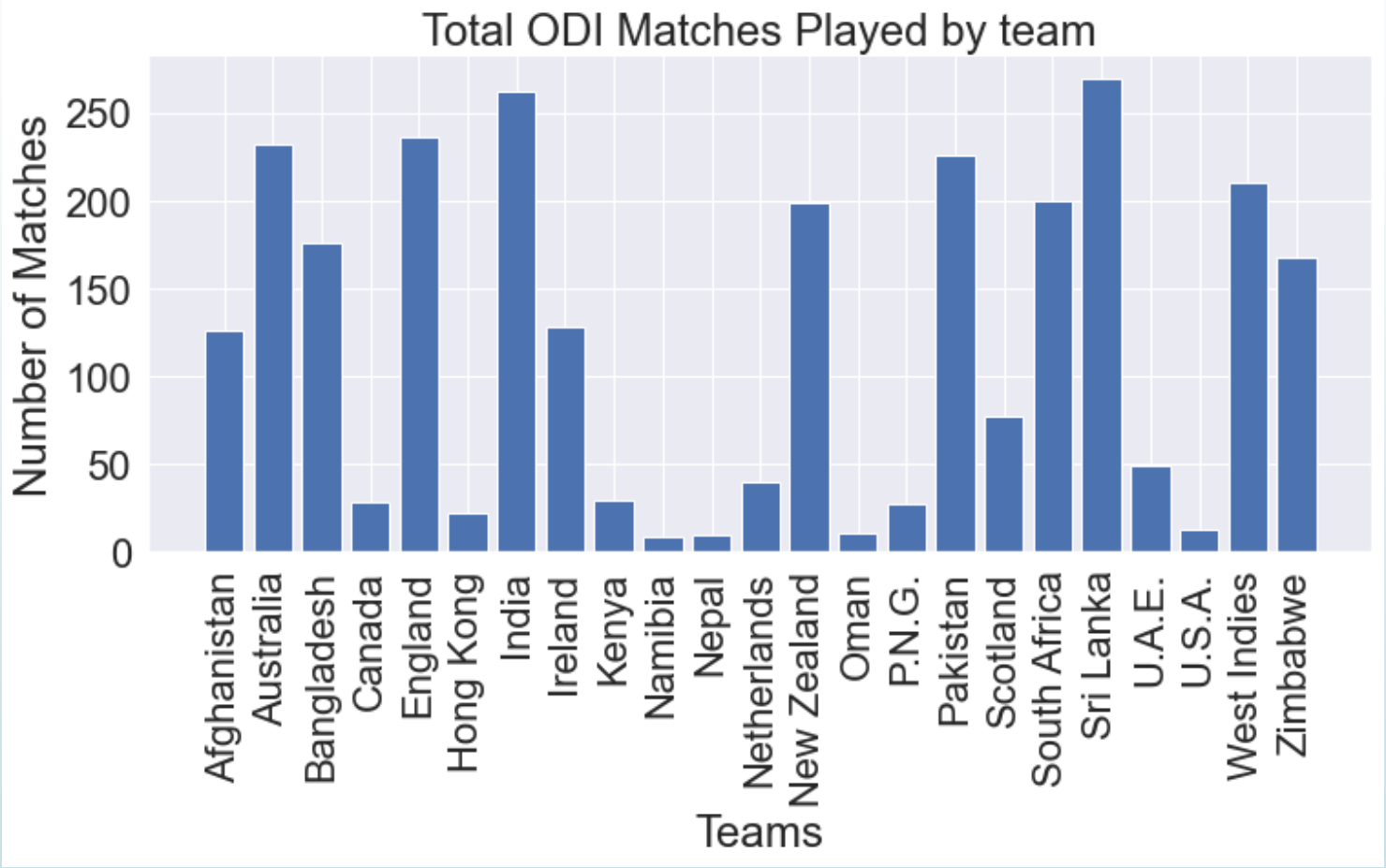
Approach

Many factors influencing and ODI match has changed over the years (like players, teams, methodologies and even rules), so it would not be accurate to use very old data to predict current scenario . Hence we will be using data starting from 2010 only.
After data cleaning and visualization, we will use a dendogram to test the relationship between variables.
We will proceed with feature engineering and feature selection to get an idea about the important influencing features.
We will use different ML algorithms and choose the best one out of it based on the predictions , and analyze the metrics of our models

Data Analysis

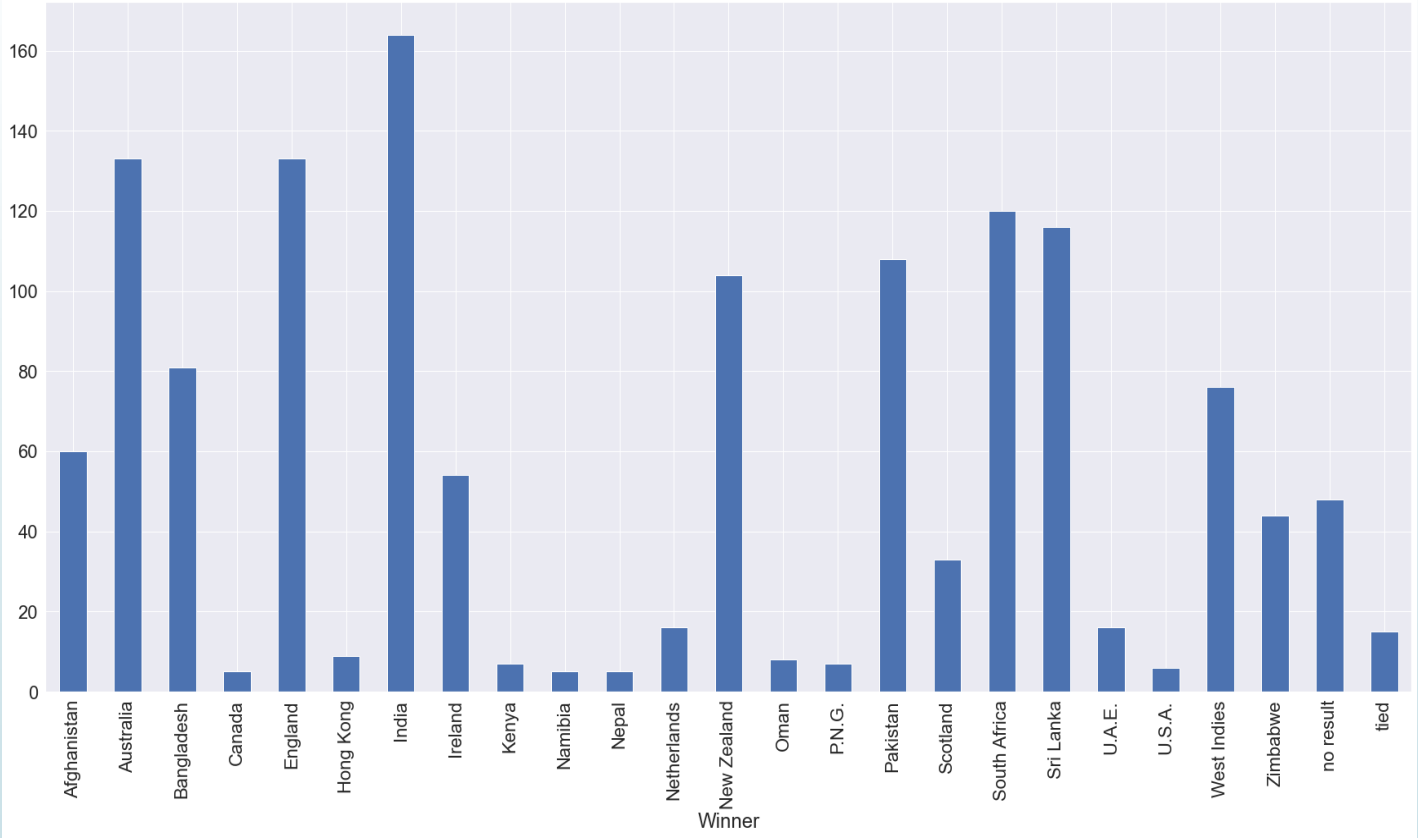
Total ODI matches played by teams

- Sri Lanka and India have played the most number of matches.
- Australia, England, Pakistan, New Zealand and West Indies have also played a lot of matches.
- USA, Oman, Nepal and Namibia have played very few matches.



Winners

- India has won the highest number of ODI matches, followed by Australia, England, South Africa and Sri Lanka.
- Canada, Kenya, USA, Nepal, Netherlands have won very few matches.
- There are many matches with no result and few which were tied.

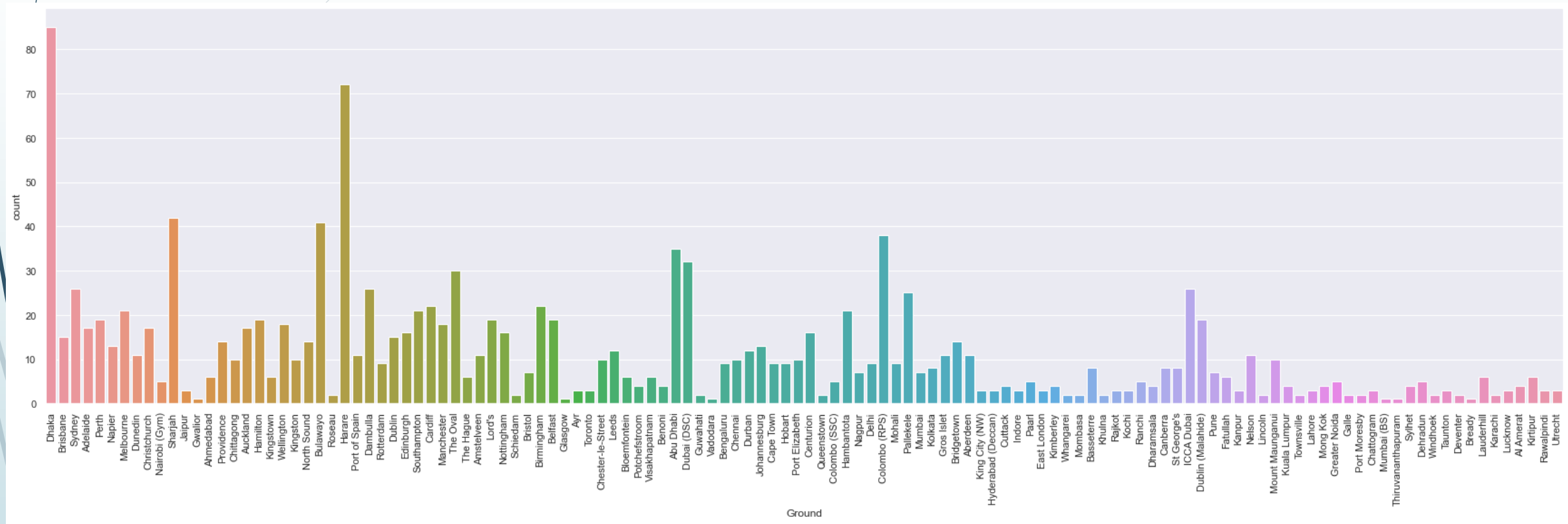


Ratio of matches won

- Although our result says Oman and Namibia have won most of the matches they have played, we can neglect them as they have played very little matches.
- India has the highest won/played ratio followed by South Africa, Australia, England and New Zealand.
- Zimbabwe. UAE and West Indies have the least ratio (excluding Canada, Kenya, PNG)

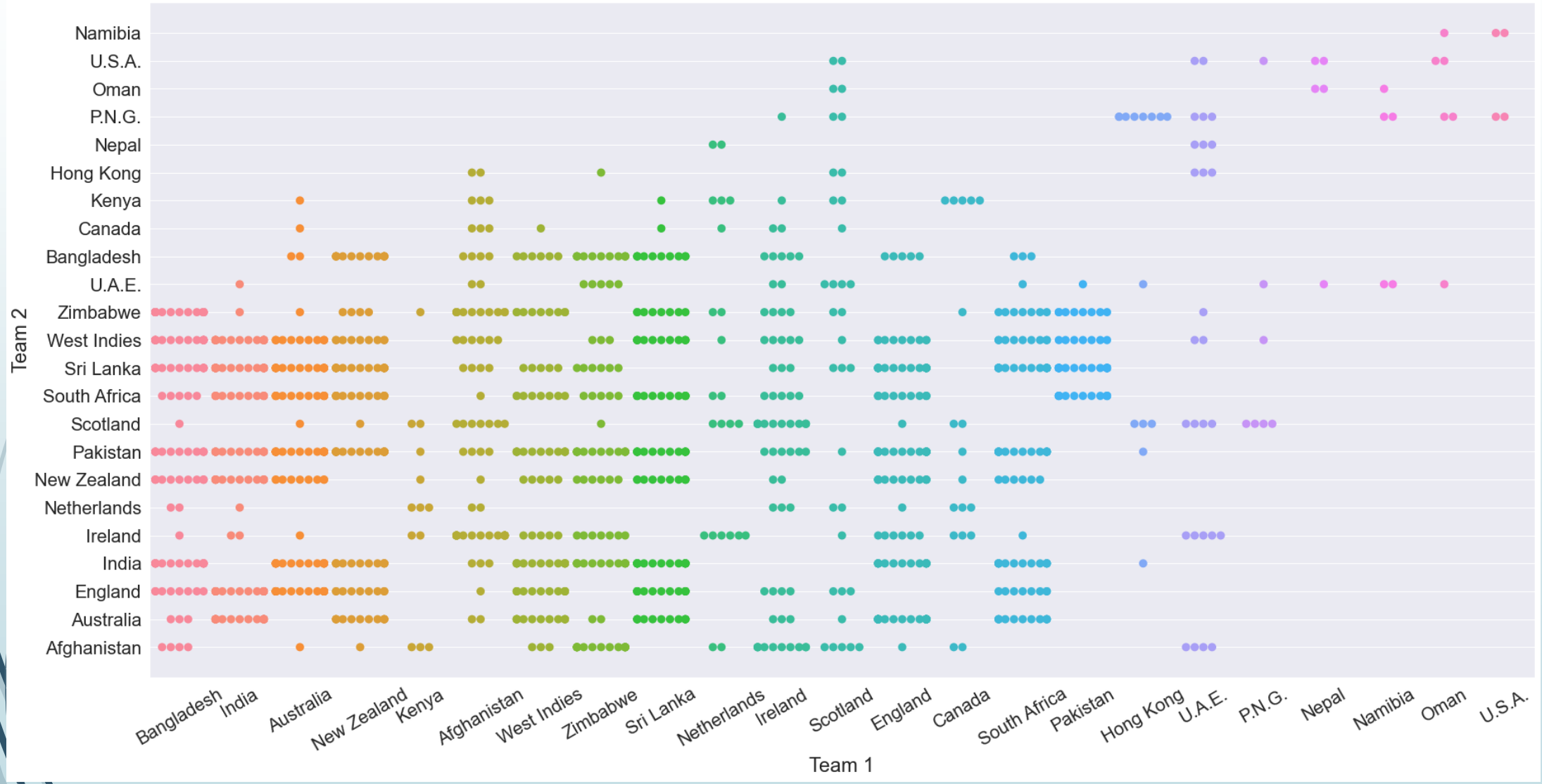
	Team 1	Team 2	total	won	ratio
Oman	6	5	11	8	0.727273
India	133	129	262	164	0.625954
Namibia	5	3	8	5	0.625000
South Africa	94	106	200	120	0.600000
Australia	125	107	232	133	0.573276
England	141	95	236	133	0.563559
New Zealand	119	80	199	104	0.522613
Nepal	5	5	10	5	0.500000
Pakistan	51	175	226	108	0.477876
Afghanistan	75	51	126	60	0.476190
U.S.A.	4	9	13	6	0.461538
Bangladesh	117	59	176	81	0.460227
Sri Lanka	112	157	269	116	0.431227
Scotland	34	43	77	33	0.428571
Ireland	68	60	128	54	0.421875
Hong Kong	14	8	22	9	0.409091
Netherlands	23	17	40	16	0.400000
West Indies	93	117	210	76	0.361905
U.A.E.	27	22	49	16	0.326531
Zimbabwe	89	79	168	44	0.261905
P.N.G.	7	20	27	7	0.259259
Kenya	13	16	29	7	0.241379
Canada	18	10	28	5	0.178571

Grounds

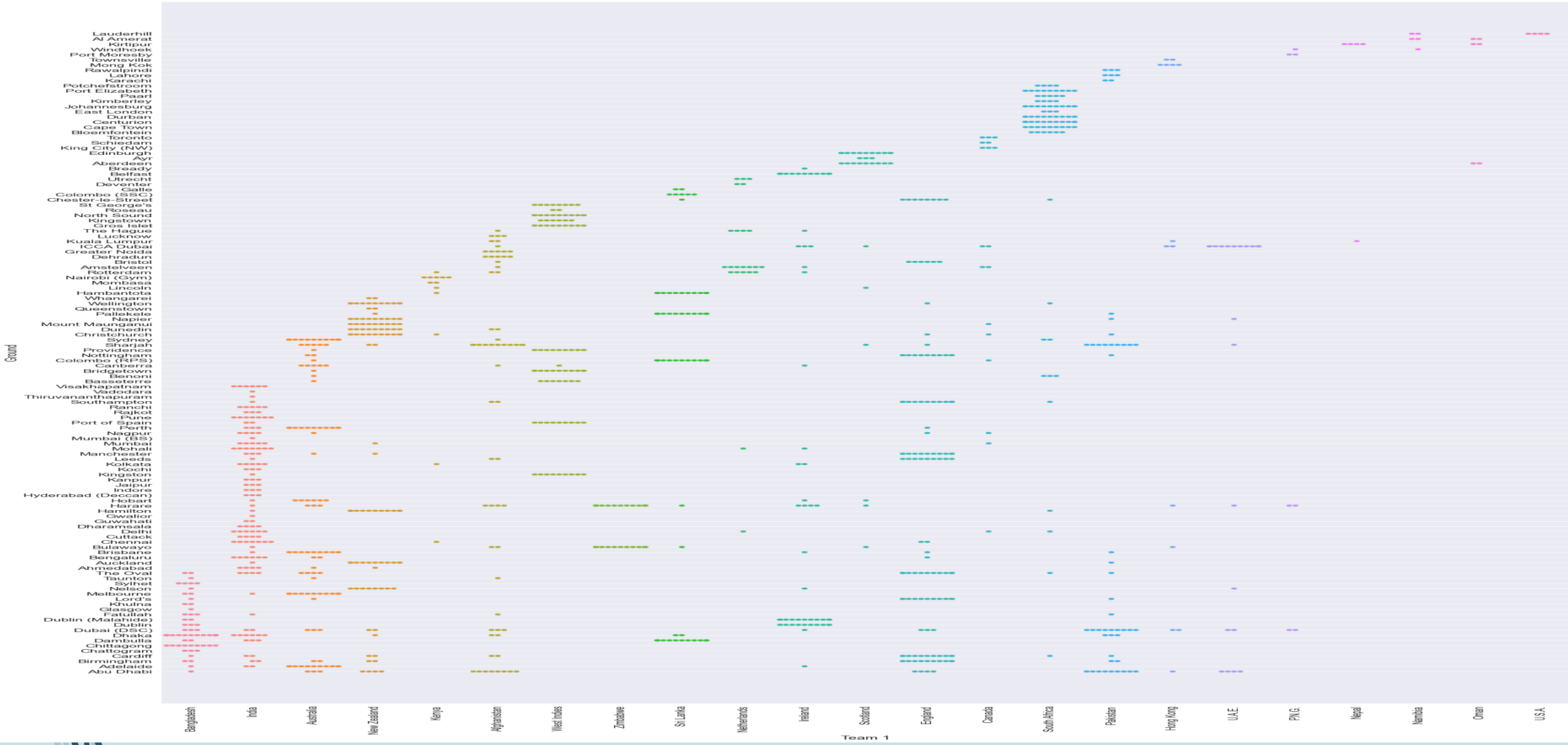


Many matches were conducted in Dhaka, Harare, Abu dhabi, Bulawayo, Colombo (RPS) and Sharjah

An overview of the matches played

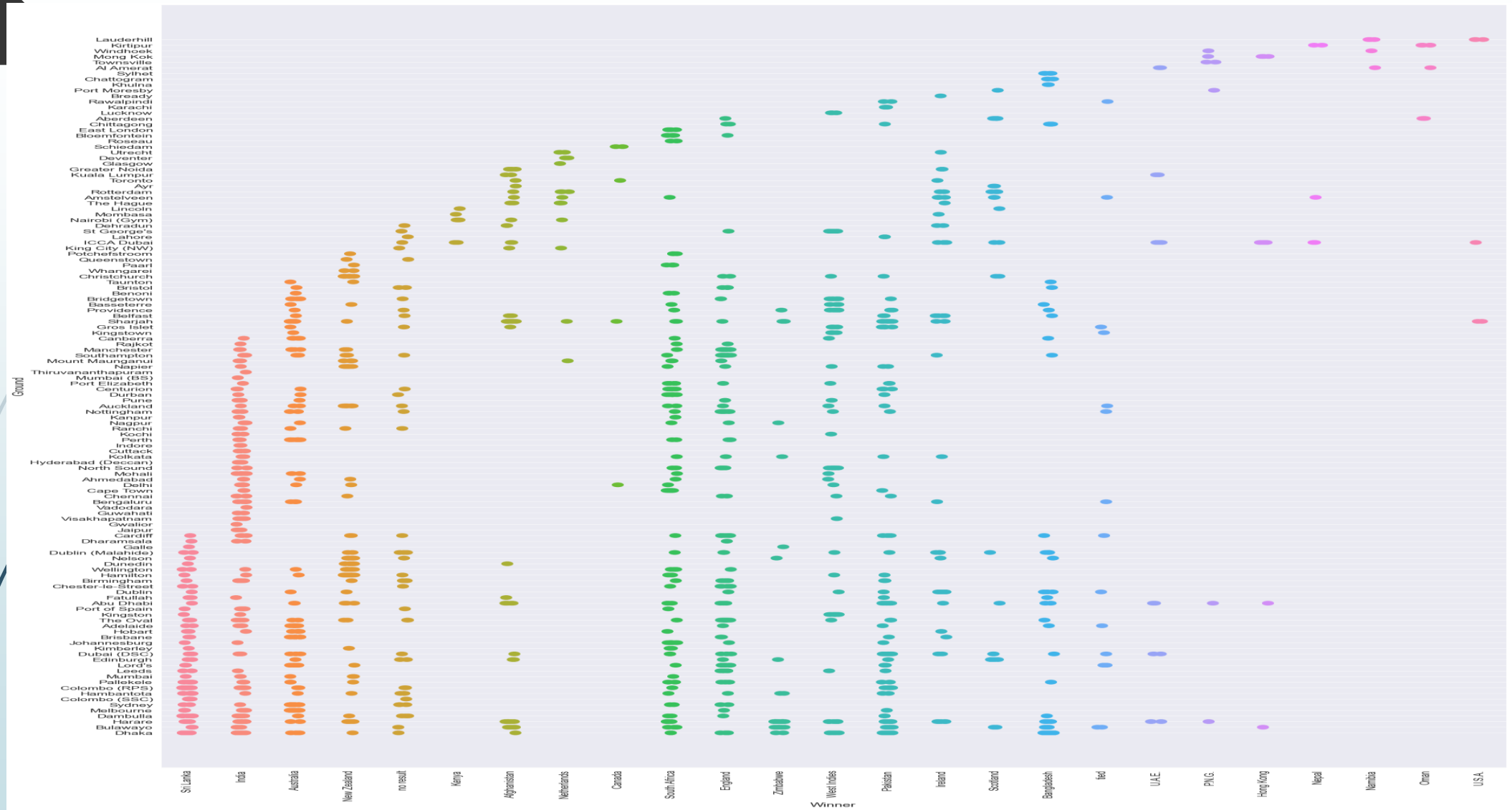


Ground vs teams

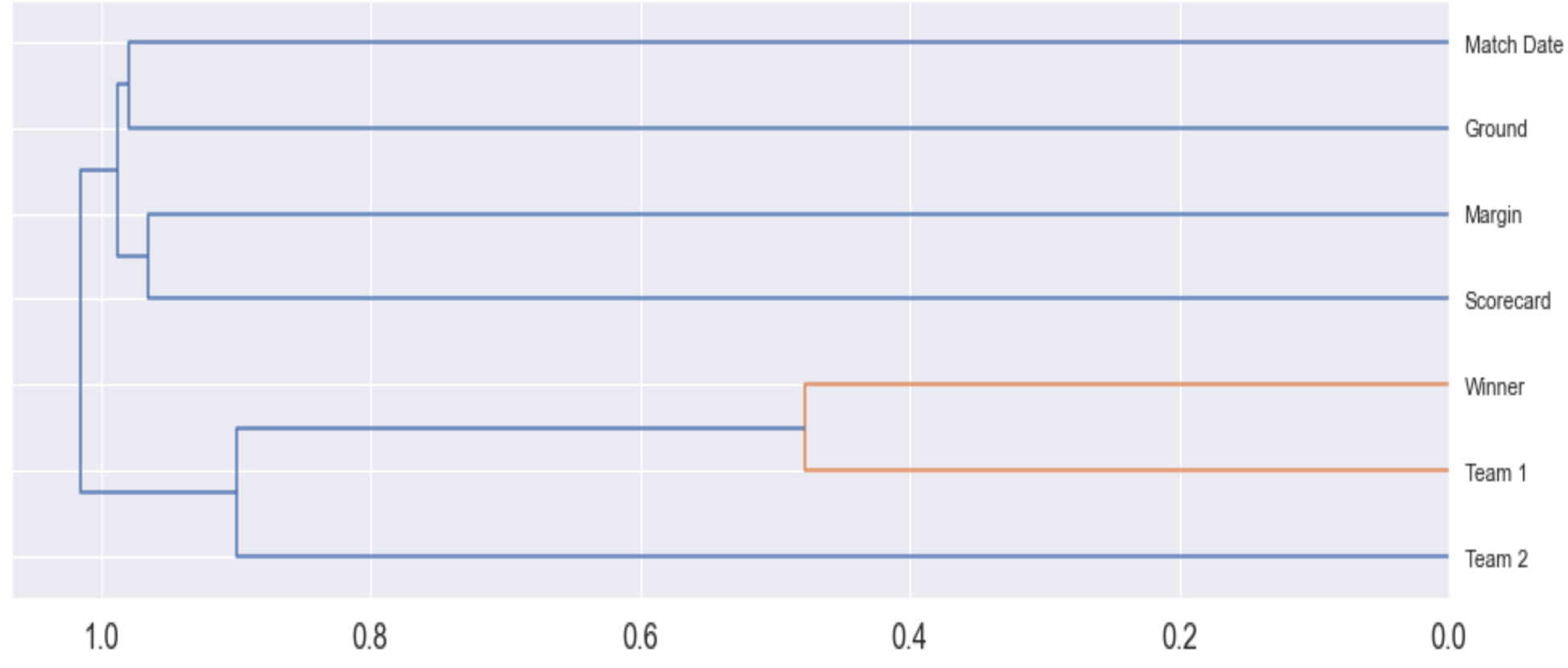


Some teams play at a certain ground repeatedly, especially in home grounds (eg Bengaluru, Chennai, Mohali, Pune for India and Adelaide, Melbourne ,Brisbane, Perth and Sydney for Australia)

Matches won in different grounds



Here, we can easily see which team has won more matches in which ground, for example, India has won most of its matches in Bengaluru, Mohali, Delhi, Dhaka (mostly home grounds)



Team 1 and winner are strongly correlated

Modelling

Model	Accuracy
Logistic Regression	64%
Random Forest Classifier	66%
KNN	57%
Naïve Bayes	44%
SVM	76%
Bagging	75%

Results and Conclusions

- Sri Lanka and India have played the most number of matches (between 2010 and 2021), and India has won the highest number of matches and also has the highest win ratio followed by South Africa, Australia and England
- Many matches were conducted in Dhaka, Harare, Abu Dhabi, Bulawayo, Colombo (RPS) and Sharjah
- There is a strong correlation between matches won by teams and the grounds (mostly home grounds)
- We were able to build a model with 75% accuracy. The seemingly low level of accuracy in most of the algorithms is due to the uncertainty of the game, and hence boosting is necessary. A lot of physical and psychological factors could affect the outcome of a particular match. Support Vector Machine outperformed all the other models and can hence be used to predict the winners of ODI matches.

Thank you