CLUSTER INNOVATION CENTRE (CIC)
(UNIVERSITY OF DELHI)
3rd Floor, Rugby Sevens Building, University Stadium
G. C. Narang Road, University of Delhi, Delhi-110007

- **Project ID/No.:**

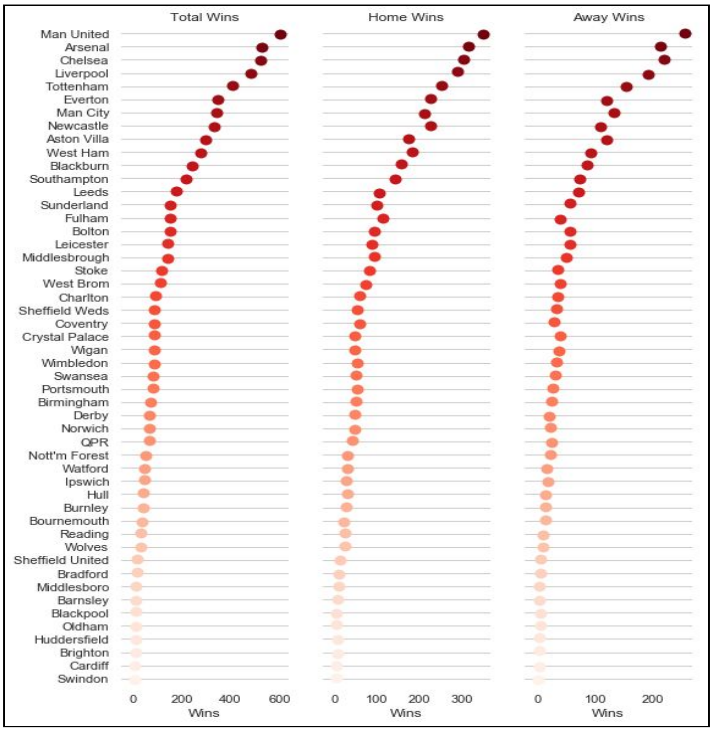
- **Project Name:** English Premier League (EPL)- Data Analysis using Machine Learning

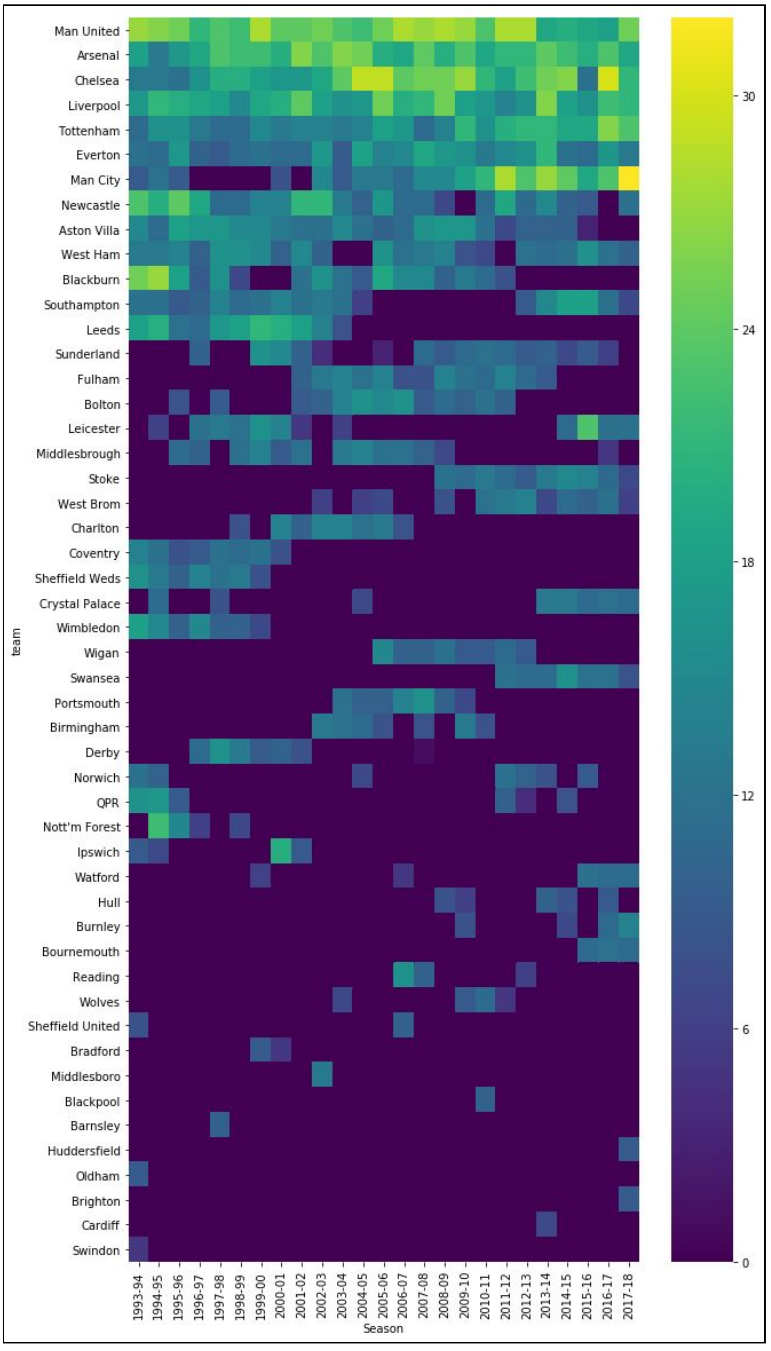
- **Project Members:**
  - Nishant Joywardhan(11824)
  - Satyam Sinha(11834)

- **Abstract:** The following project pursued after completion of Udemy's Machine Learning course (*Machine Learning A-Z: Hands-on Python and R in Data Science)*, is basically intended towards making an analysis of the English Premier League's (EPL) team data available for over the last 25 years (seasons), using Machine Learning Algorithms, and then using that to make match predictions.
  Initially, we started with the basics of Python and Machine Learning with the help of the course.
  After being well versed in the important topics of the domain, we began our project on the EPL Data Analysis:-
  - The first step was to plot the pie chart of winning percentages of the home(local) team, away(visitor) team and the possibility of a draw, on the basis of records taken from the data.
  - Then we have taken many parameters into account such as goals-per season, average goal per game per season, number of games per month, most games being scheduled on weekends, goals per month, goals per matchday, which team wins the most home and away games on an average, etc. These parameters are visualized using box plots and simple graphs.
  - Then we have successfully created a heat map of every team over the course of the different seasons.
  - Finally, dot plots of win per team per home/away game are recorded.

  With the successful completion of the above steps, we have completed the analysis of the EPL data taken.
  On the basis of the above analysis, we can make predictions as to which team has the more likely chances to win a game and even assign a fixed, odds in favor and odds against ratio for each outcome.

- **Project photo:** Final dot plot and the heat map obtained from the analysis:-





- **Conclusion:** This is a very basic data analysis and it has a lot of room for improvement, such as- the introduction of new data points to make better analysis, the introduction of online parameters such as the data available on twitter and its hashtags, etc. We would like to incorporate these changes in the future and make this an even better analysis using Machine Learning.