**INTRODUCTION**

Cricket is a game which has two batters and a bowler at the forefront of action. Test Cricket is a format of the game of cricket which is played over 5 days and is generally considered the most challenging format, both physically and mentally. Popular cricket statistics revolve around matchups between batters and bowlers and bowling teams try leveraging these statistics. A common counter strategy employed by batting teams is to shield the weaker batter in a particular bowler matchup that is tilted towards the bowler. An ideal counter for a bowling team to such a batting strategy would be identifying the bowlers who are good against a pair of batters.

**METHODS**

Building a Bipartite Graph:
A bipartite graph is a graph in which the nodes can be divided into two disjoint and independent parts. Since the game of cricket is essentially a batters vs bowler game, I have considered the pair of batters as the first set and the bowlers as the second set in the bipartite graph. By considering all the batter vs bowler interactions that ever took place in the past, I have determined that the number of runs scored per out is 31.84. I have used this as the cost of an out. To produce edge weights for a directed bipartite graph from bowlers to batters, I have devised the following Edge_Weight metric.

\[
\text{Edge}_\text{Weight} = \frac{\text{Sum of all runs scored}}{\text{Sum of all Out Counts}} = 31.84
\]

Building a Modified Maximal Bipartite Graph:
The Bipartite graph generated by the above method is then used to build a modified maximal bipartite graph for a match scenario. That is, given a set of 11 batters of a team and a set of bowlers in the opposing team, the maximal bipartite graph draws a single edge between a pair of batters and the most effective bowler.

**Dataset:**
A total of 694 csv files comprising of ball-by-ball data of all Men’s International Test Cricket games from the year 2000 to 2016 have been used for creating the Bipartite graph as shown in Fig. 1. The dataset is obtained from the public cricket data repository (https://cricsheet.org).

**Case Study:**
A game that was played between Australia and New Zealand’s men’s cricket teams from 11th February to 14th February 2016 has been considered as a case study to generate a Maximal
Graph and test the model. 11 batters from New Zealand team were pitted against 5 designated bowlers from the Australian team. Since there are 11 batters on the New Zealand team, a total of 55 pair combinations are possible. By searching the relevant batting pairs from the bipartite graph and extracting the bowler with the highest edge weight corresponding to the batter pair, I was able to generate a Modified Maximal Bipartite graph for the Australia bowlers. Each batter combination has an ideal bowler that the bowling team can leverage. Fig. 2 shows the best Australian bowler to be bowled to a pair of New Zealand batters.

Fig. 1: A sample of the Bipartite Graph
DISCUSSION & CONCLUSION
Although there is a clear indication from the Maximal Bipartite graph that certain bowlers have a clear advantage over certain batting pairs, bowling teams have not been using that knowledge to their advantage. For example, [JR Hazlewood] who had the best edge weights to the batting pair [MJ Guptill, TWM Latham] from the past data only bowled 11% of all the balls
bowled to the pair in the game. In fact, only 18.75% of the balls are bowled overall with the optimal bowler. An increased adoption of this model can yield better results for the bowling teams. Also, Fig. 3 highlights the highest performing bowlers against the New Zealand batting pairs. A Count chart (Fig. 3(a)) can reveal how many batting pairs is the bowler the best against and the Sum of Edge Weights Chart (Fig. 3(b)) can reveal the overall impact a player might have on the game. These statistics can help the coaching staff pick a team selectively against an opposition. However, there are other things to look at before pitting a bowler against a batter combination. Cricket is a game that is heavily reliant on the surface on which the game is played. Different surfaces react differently to bowlers and their natural efficiencies might be diminished or enlarged by the surfaces. Also, the game situation might also sometimes affect the bowlers’ efficiencies. Player fatigue is another issue to look at. Since this is a 5-day game, players need to be fresh enough to perform well over that period. Overuse of any bowler could prove detrimental. This Maximal Bipartite matching model has not been tested in real time. None of the teams in the games gone by have consciously adopted such a strategy. It would be interesting to see how well a team performs if they were to use this model.

AUTHORS
Jitendra Sai Kota, Dr. Andrew Wilson, Presented at Analytics Day 2022.