

Proposal for MA678 Midterm Project

Yifan Zhang

◆ Personal Statement:

After graduation from the MSSP program, I plan to work as a professional data analyst. Currently, my thought is mastering statistical methods, tests, and models to help leaders make better decisions. At present, data analysis is getting more and more involved in the sports industry, for instance, all players and coaches have access to the percentages of any major statistical category to improve their game plan. Hence, for this midterm project, I decided to look into NBA datasets including box score, players' basic data, and games results to find out how these factors are related to players' salaries and win shares. The reason why I choose NBA is that since the late 2000s cameras were installed in all stadiums to track every game and all the basic data are public, which indicates I can obtain adequate data to do model fitting and hypothesis tests. Utilizing the knowledge I have learned and devoting myself to my favorite industry is always my dream.

◆ Question:

Which factors have the largest impact on games' win share? What kind of player is most favored by clubs now or to say what kind of player earns most in today's NBA league?

◆ Data Source:

https://github.com/MattC137/Open_Data/tree/master/Data/Sports/NBA

◆ Reference:

1. Loke Cohen(2020, February 2), How Data Analytics is Changing the Sports Industry For Good, Driveresearch. <https://www.driveresearch.com/market-research-company-blog/how-data-analytics-is-changing-the-sports-industry-for-good/>

◆ Timeline:

- 1) Data cleaning and EDA: Nov. 8 – Nov. 14
- 2) Modeling and Validation: Nov. 15 – Nov. 23
- 3) Write up: Nov. 24 – Nov. 26