Building a Champion Winning NBA Team

Kyle Lo, Krystal Sung, Jyothi Ranjit & Harry Morton

Table of Contents

- What is the Subject of our Capstone Project?
- Diving into the Source Data
- How is our Data Structured?
- Key Questions for our Team
- Analysis and Relevant Insights
- Next Steps

What is the Subject of our Capstone Project?

What Determines a Championship Winning Team?

Our team wants to **generate a model** that is able to **predict a team's likelihood to win the NBA**.

By training our model on the results of each team in NBA history by season, we can **make a prediction on any given organization** in the future.

We want User Interaction with our Model...

With our model in place, our vision is that users can create a team of NBA players, and receive the likelihood of that team to win!

This estimation will be calculated by understanding the players statistics, injury history, and more!

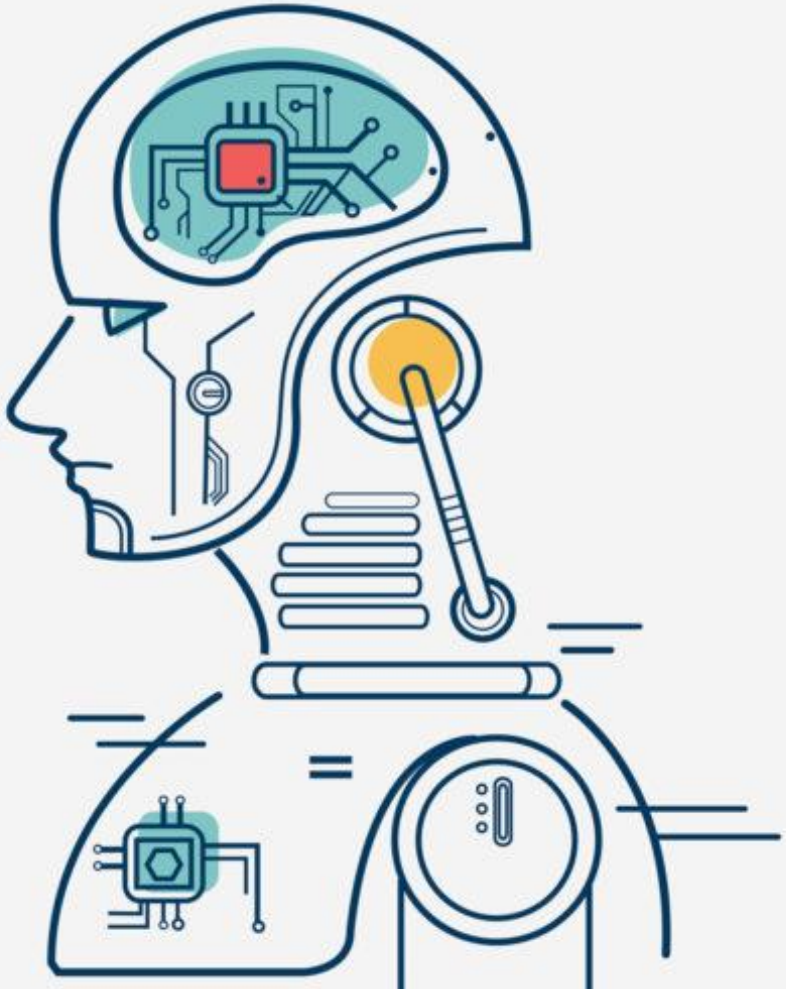
Why Choose the NBA as our Capstone subject?

Applicable Machine Learning Task

Estimating the likelihood of an outcome is a common machine learning task in data science

- This was a bonus as we near the end of the bootcamp!

Thus, the experience of building this model will be valuable as we further our careers



Abundance of Data to be Collected

The NBA has A TON of accurate, historical data that is widely available

The ease of sourcing this data, and the confidence in it's accuracy helps us build the most accurate ML model possible!

Finally... A Fun Way to Apply Bootcamp Skills

What better way to apply skills you've learnt than to a topic that genuinely interests us all?

Tackling our Capstone Project in a familiar topic means we can focus on the concepts learnt in the class, and not on new material!

Diving into the Source Data

We Identified 3 Key Areas for Data Sourcing

1. Player Information

This includes position, historical injuries, aggregated statistics and various salary information

Players are vital in building a championship team!

2. Team Information

Who is the coach? What is the attendance? What is the available salary cap?

Players cannot win the NBA without a sound team as a foundation.

3. NBA Season History

Where did teams placed in the past? What were player stats for each year? Injuries? Other misc. Info

Understanding the rankings gives us training data!

Data Sources: Player Information

Player data includes:

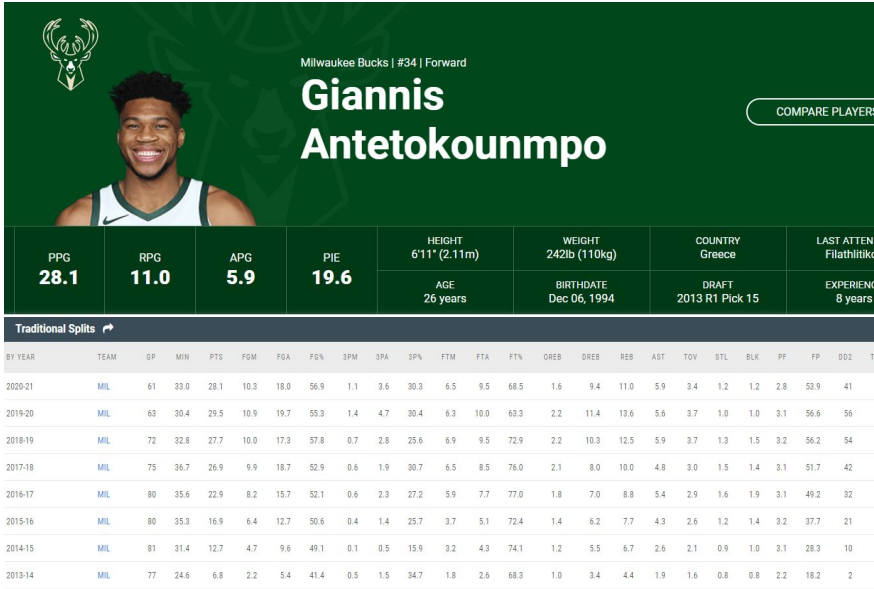
Games played, Points, Field Goals, 3Pts, Free Throws, Steals, Blocks, Turnovers, Rebounds, etc..

Purpose of Data:

Can we determine if a player is on a winning team based on stats?

Can we predict the winning team based on players stats?

How does each feature weigh against each other?



Data Sources: Team Information

Team Info includes:

Team stats, standings, win %, salary,

market cap, attendance, ect...

Purpose of Data:

Does a team stats determine a championship?

Does standing determine which team will win the championship?

Does the money machine go BRRRRR?!

Does popularity of a team attracts better players?

	TEAM	GP	W	L	WIN%	MIN	PTS	FGM	FGA	FG%	3PM	3PA	3P%	FTM	FTA	FT%	OREB
1	Utah Jazz	72	52	20	.722	48.2	116.4	41.3	88.1	46.8	16.7	43.0	38.9	17.2	21.5	79.9	10.6
2	Phoenix Suns	72	51	21	.708	48.6	115.3	43.3	88.3	49.0	13.1	34.6	37.8	15.6	18.7	83.4	8.8
3	Philadelphia 76ers	72	49	23	.681	48.4	113.6	41.4	86.9	47.6	11.3	30.1	37.4	19.6	25.5	76.7	10.0
4	Brooklyn Nets	72	48	24	.667	48.3	118.6	43.1	87.3	49.4	14.2	36.1	39.2	18.1	22.5	80.4	8.9
5	Denver Nuggets							43.3	89.2	48.5	12.9	34.2	37.7	15.7	19.5	80.3	10.5
5	LA Clippers							41.8	86.7	48.2	14.3	34.7	41.1	16.2	19.3	83.9	9.4
7	Milwaukee Bucks	72	46	26	.639	48.1	120.1	44.7	91.8	48.7	14.4	37.1	38.9	16.2	21.4	76.0	10.3

NBA Champions!!!

Data Sources: NBA Season History

History Data includes:

Past championships, past stats, injuries, games played per season, past standings, finals appearances, ect...

Purpose of Data:

Training data to train our model on winning teams.

How does injury affect a team's chance of winning a championship?

Determining probability to win championship

Likelihood to place in the top 8 teams or bottom 8 teams

Explaining the Data Exploration Process

The Data Exploration Process

What are we trying to explore?

Does good stats make a good player/team?

Does a bigger market cap entice better players?

Does a player's salary affect a teams chance of winning?

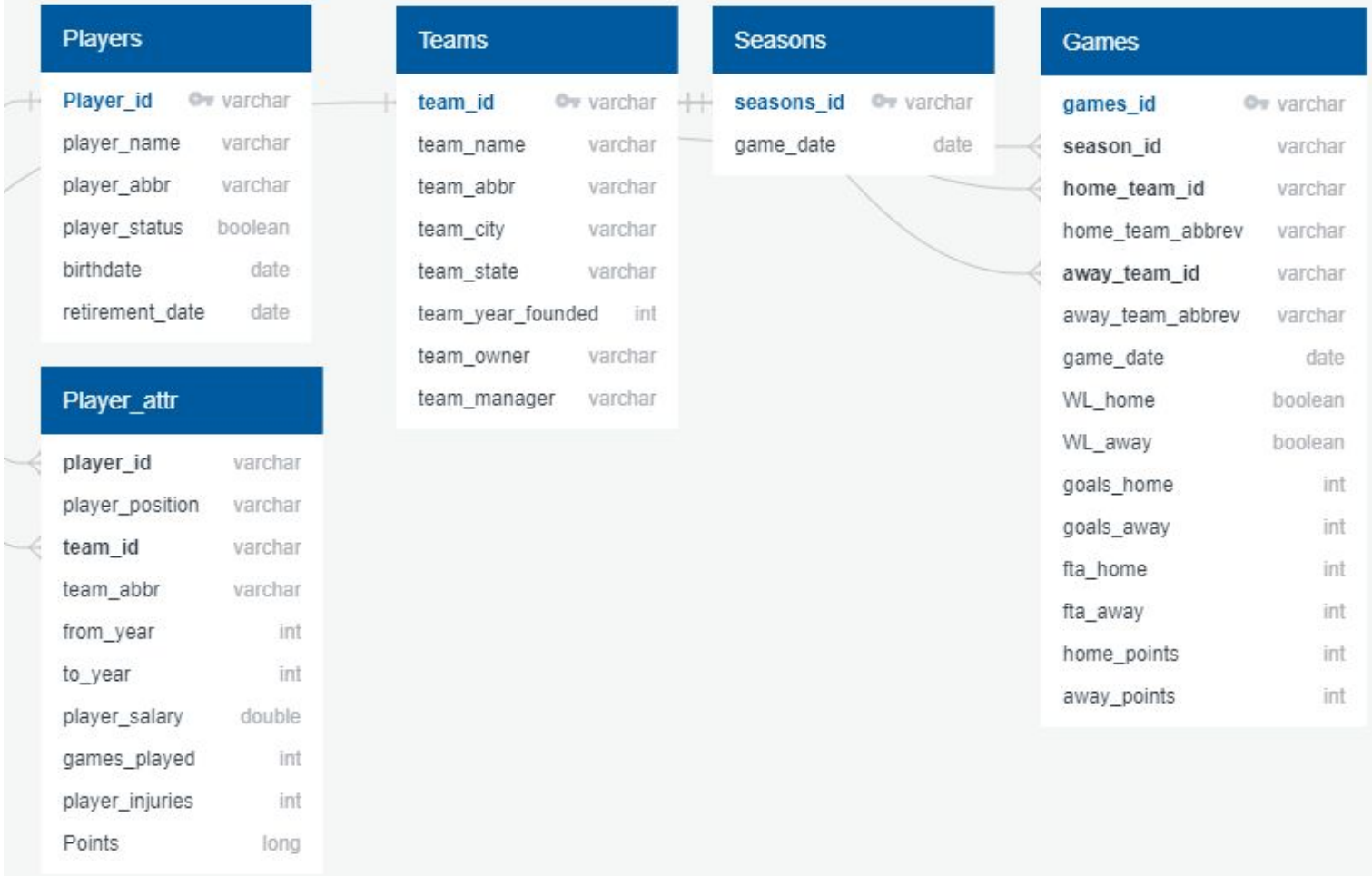
Can we predict this years winner based on past championship game stats?

How is the Data Structured?

Leveraging SQL & Postgres for our Database

- AWS database used to store all of our exploratory data so anyone on the team can access it at anytime.
- We are using Postgres (pgAdmin) to connect to the AWS database.
- Cleaning will be done outside of Postgres
- Database is for clean final exploratory data only

www.quickdatabasediagrams.com



Outlining the Key Questions for Our Team

We Identified 4 Key Areas for Data Sourcing

Kaggle

TGFK: Thank goodness for Kaggle!

Contains massive amount of clean datasets that fits our needs.

Contains previous analysis we can build upon

NBA-API (nba.stats.com)

Used to extract more recent data to build upon existing Kaggle data

Can obtain raw form of data that will need wrangling

fivethirtyeight

Contains RAPTOR data that is a new metric to measure basketball stats.

Contains analysis for dataset like kaggle.

espn.com

Web Scraping needed

Contains salaries

Has a simple layout of leaderboards with stats of players/teams

Analysis of our Data and Relevant Insights

Analysis with The SIX

Who will win the championship? The Machine Learning model will predict based on players stats, who is going to win the championship. To win the game of basketball is simple. Whoever has the most points at the end of the time limit is the winner. We can do a quick analysis and assume that the teams that scores the most points per game are more likely to win.

What makes a championship team? We know that a player with good stats will make a team good, but teams are limited with cap space. So does a players salary determine how good they are? With this knowledge we know that teams are willing to pay individuals a lot more money to have them on their team.

When: Have the attributes that predict a championship winning team changed over time? We can answer this by aggregating various statistics by year for our ideal population

Where did the team place last year in the standings? We know that every season is different, but understanding where teams placed in the standings give us insight in their future investment. In contrast teams usually place depending on where star players move.

Why does a team win? Does offense or defense win a game? Does injury play a bigger role than having a team with good stats? From this analysis alone we can determine that if a star player is injured then it is unlikely for your team to win. Since your star player takes a majority of the salary cap, a team will have worser players on the bench. Looking at the standings will only tell a person how a team performed during the regular season. When playoffs come around many players are usually injured.

How does playoff game stats differ from the regular game stats? When analyzing the different game stats, it is best to differentiate the regular season stats with playoff stats. In every game there is an unknown factor called “clutch.” How clutch is a player? A team or a player can have amazing regular season stats but chokes in the playoffs when it matters.

Next Steps

Bringing it all together

Start answering our questions for analysis

Build out dashboard to visualize analysis

Build machine learning model to determine what makes a championship team

Start putting together the right data for our story

Thank You.