Data Analytics Bootcamp – Project One

Hypothesis

Is there a correlation between investment in an NBA team and their performance in their league?

Questions

What contribution do high paying star players make towards the success of the team?

How is success of a sport team measured?

What is the relationship between investment in a team and the number of wins by year?

Is there a correlation between the investment and the region in which the teams are in (Eastern Conference vs Western Conference)?

Possible Datasets

https://data.world/datadavis/nba-salaries

https://www.kaggle.com/nathanlauga/nba-games

https://rapidapi.com/api-sports/api/api-nba

Team Recap:

We asked more questions about the story we are telling to narrow down what data points we needed. We looked at the datasets we had previously had to see what additional data we might need. We narrowed our analysis down to needing data for the following:

Years of data from 2015-2019 inclusive

Players making 25mil or more correlated to wins for the team

Salaries based on team wins

Measuring Success = regular season wins

Team Investment = salaries of all players or Team Investment = coaches, players etc.

Star player effect on valuations

Game attendance vs team wins (home games and total games)

Data we need for 2015-19:

Team wins per year

Player salaries

Teams Names and 3 letter codes

Teams by conference (regional analysis)

Game attendance for the regular season

Team valuations

Between now and Monday Class:

Michelle-Collecting/cleaning game attendance

Jeremy-cleaning salary.csv for salary info

Paul-Collecting/cleaning team wins data

Jarrod-Collecting/cleaning Team valuations data

Project work summary for Monday:

Found missing salary data and finished cleaning

Created a reference table for name codes and team names

Reworked our questions to keep the scope manageable

Reworked clean data frames to keep consistent format

Discussed plot types and data to use

We worked on getting all our branch commits up to date with the master branch

Planned a meeting for Tuesday from 7-8

## Possible Plots

\* Bar chart of total attendance for each of the 5 years. Compare total home vs road attendance for each of the years.

\* Line graph for player salaries over 5 years

\* Line graph of total attendance vs team valuations

\* Scatter plot of wins vs vip salaries. Add linear regression

\* Wins vs attendance what kind of chart?

\* Bar chart of total wins in Eastern Conference vs Western conference for 5 years.