



PROJECT BY:

- Mirgadir Alakbarov
- Andrew Thornell
- Kirsten Walkup
- Max Atherton
- Umar Farooq



PROJECT DESCRIPTION

It all started when a group of 5 hotshots were hired by Mr. Drake Romaro, a billionaire who made his fortune from betting on sports. While his initial bets were lucky hoops, which made him a fortune, now he thinks he needs to develop a calculated winning strategy.

Mr. Romaro has started his own analytics company to analyze the 2019 – 2020 basketball season data to predict the winning team. He has recently asked fresh graduates from UT Austin Data Camp to help him win big this season.

**WILL THIS NEW TEAM BE ABLE TO
MAKE THE ACCURATE PREDICTION?
LETS FIND OUT!**



RESEARCH

QUESTIONS TO ANSWER

Q.1.

Which are the top 8 teams for each conference in the current season, till the last data point? What are the scores of the top 8 teams? And their win to loss ratio. For each conference.

Q.2.

Does the home team win more or the visitors team win more matches?

Q.3.

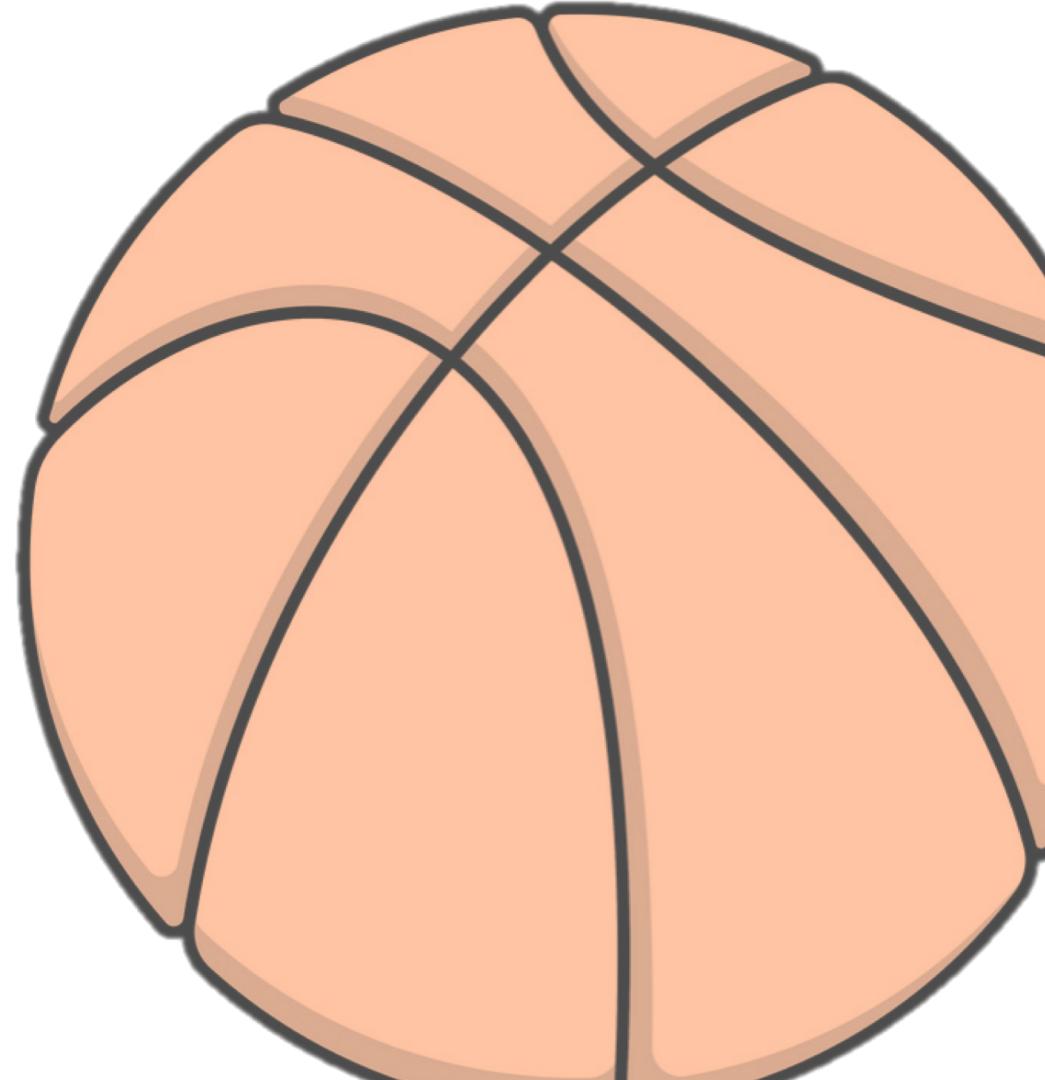
What are the locations of the game? Heat map indicators of teams with more wins.

Q.4.

Calculating correlation between team player stats, team field goal % and other variables against games won by the team.

Q.5.

Defining winning parameters to determine a score for "Win Factor" of each team.



DATA

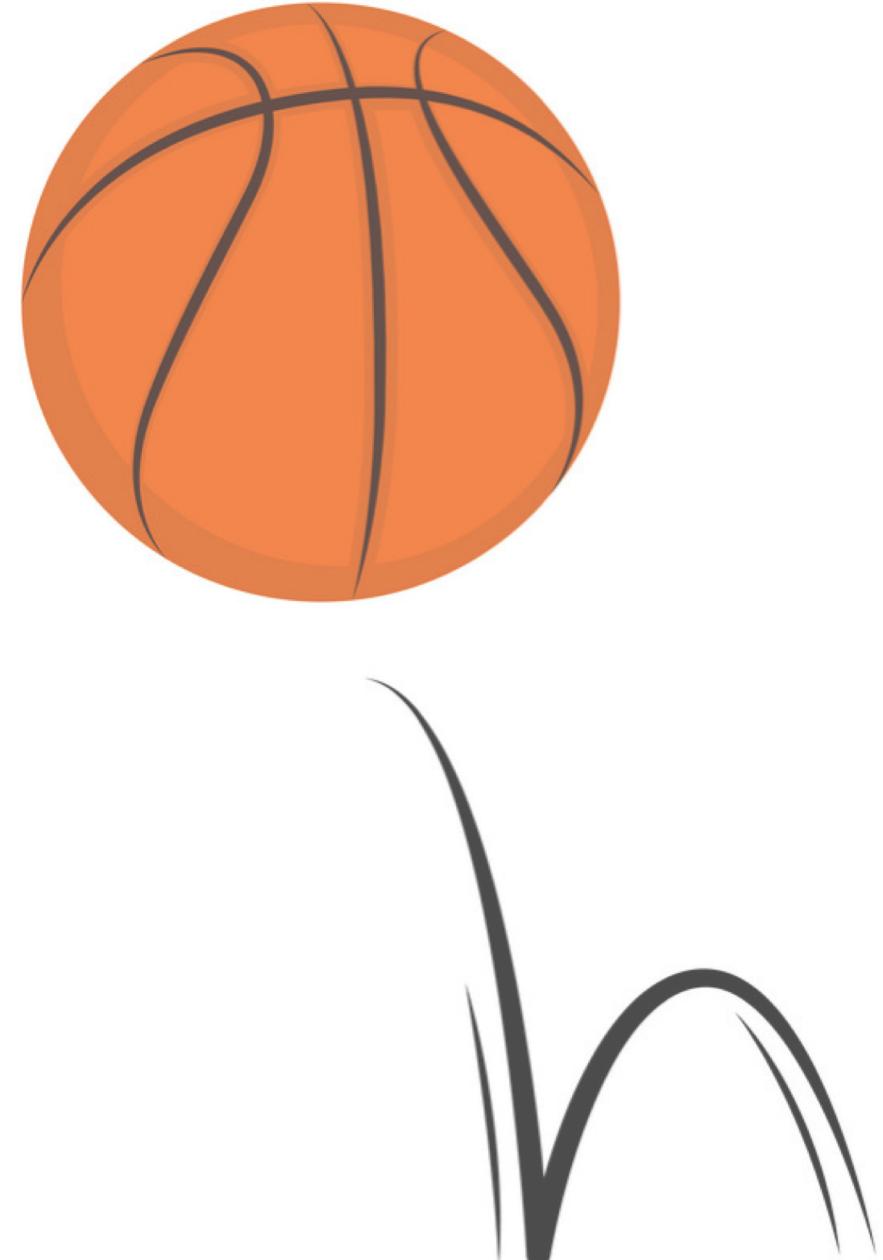
TO BE USED



To access data [click here](#) in slide show mode.

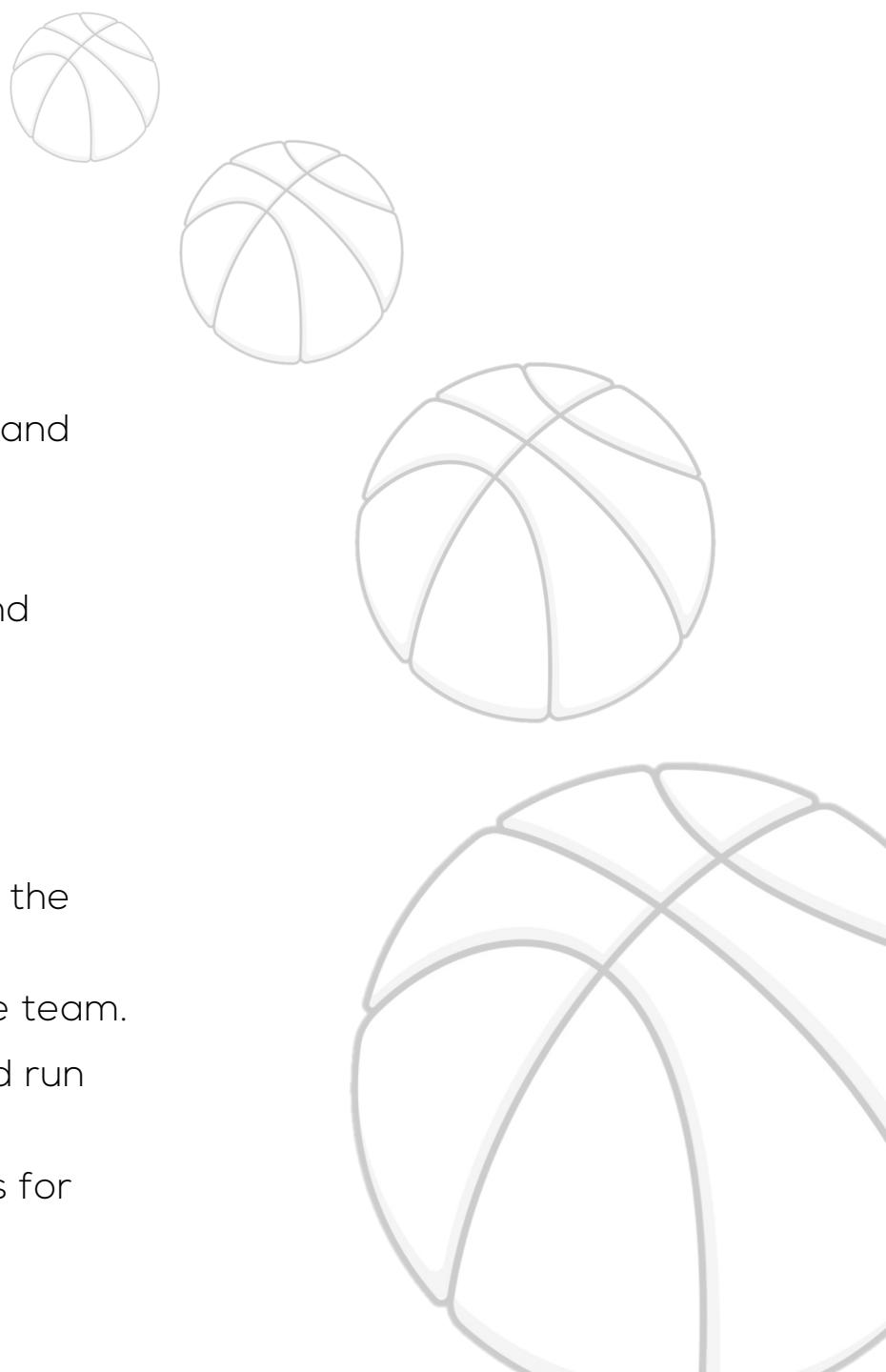


To access data [click here](#) in slide show mode.



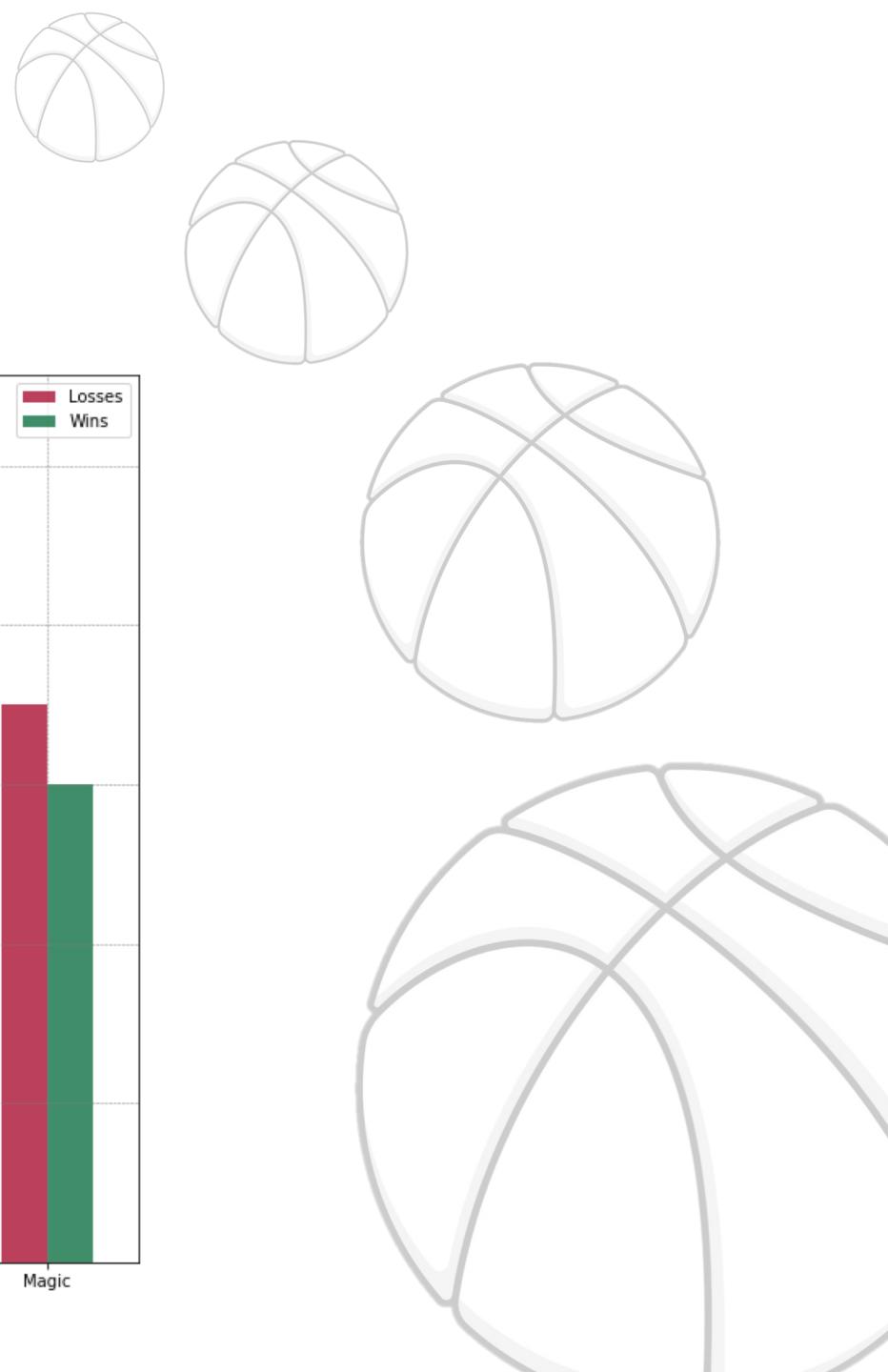
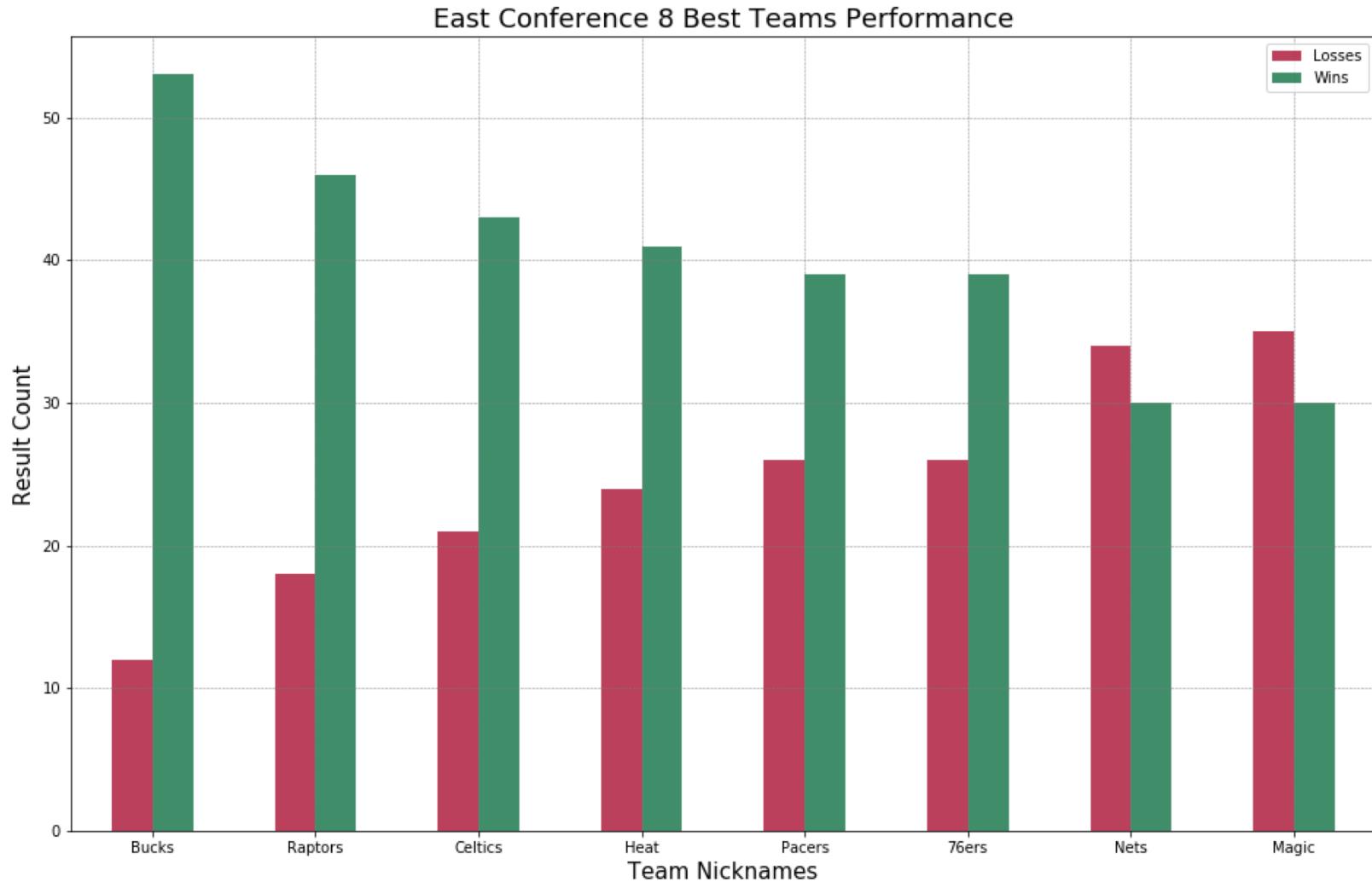
BREAKDOWN OF TASKS

- Importing and merging relevant data files.
- Create data frame of top 8 teams and their respective win/loss scores and ratio.
- Make a bar graph with wins in green and losses in red.
- Calculate the number of win counts by home team and visitors team, and represent them in a pie chart.
- Use Google API to match venue location and extract their latitude and longitude. Map game location on Google Maps.
- Heat map indicators of teams with more wins percentage.
- Extract player points and average their data in new data frame. Sum all the team's player data and match them to their respective team.
- Calculating correlation between team player stats and game won by the team.
- Make scatter plots of Team Win vs. Player Stats and other variables and run regression for each conference to find relationship between variables.
- Make observation as per analysis results and define winning parameters for win prediction.



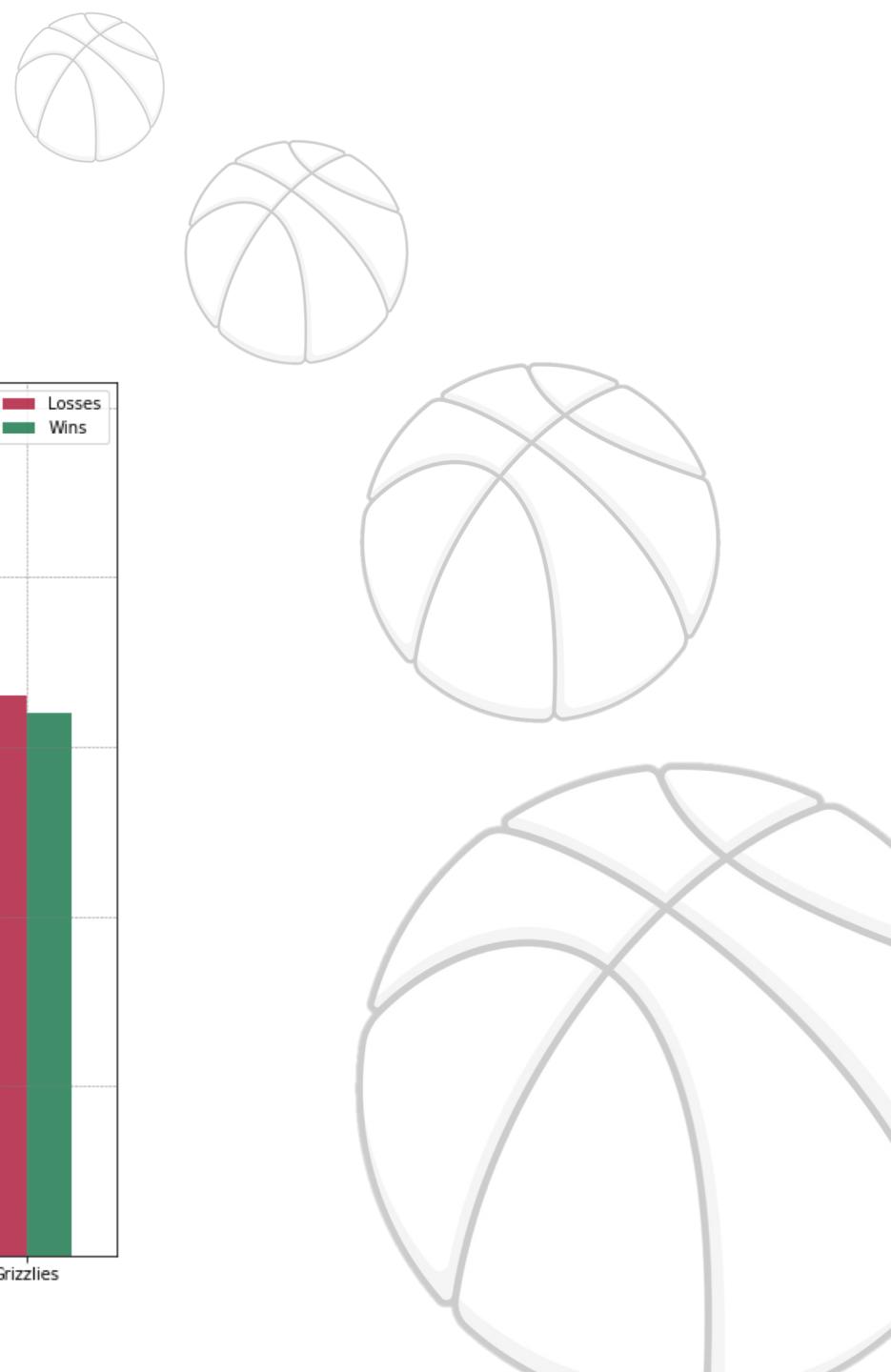
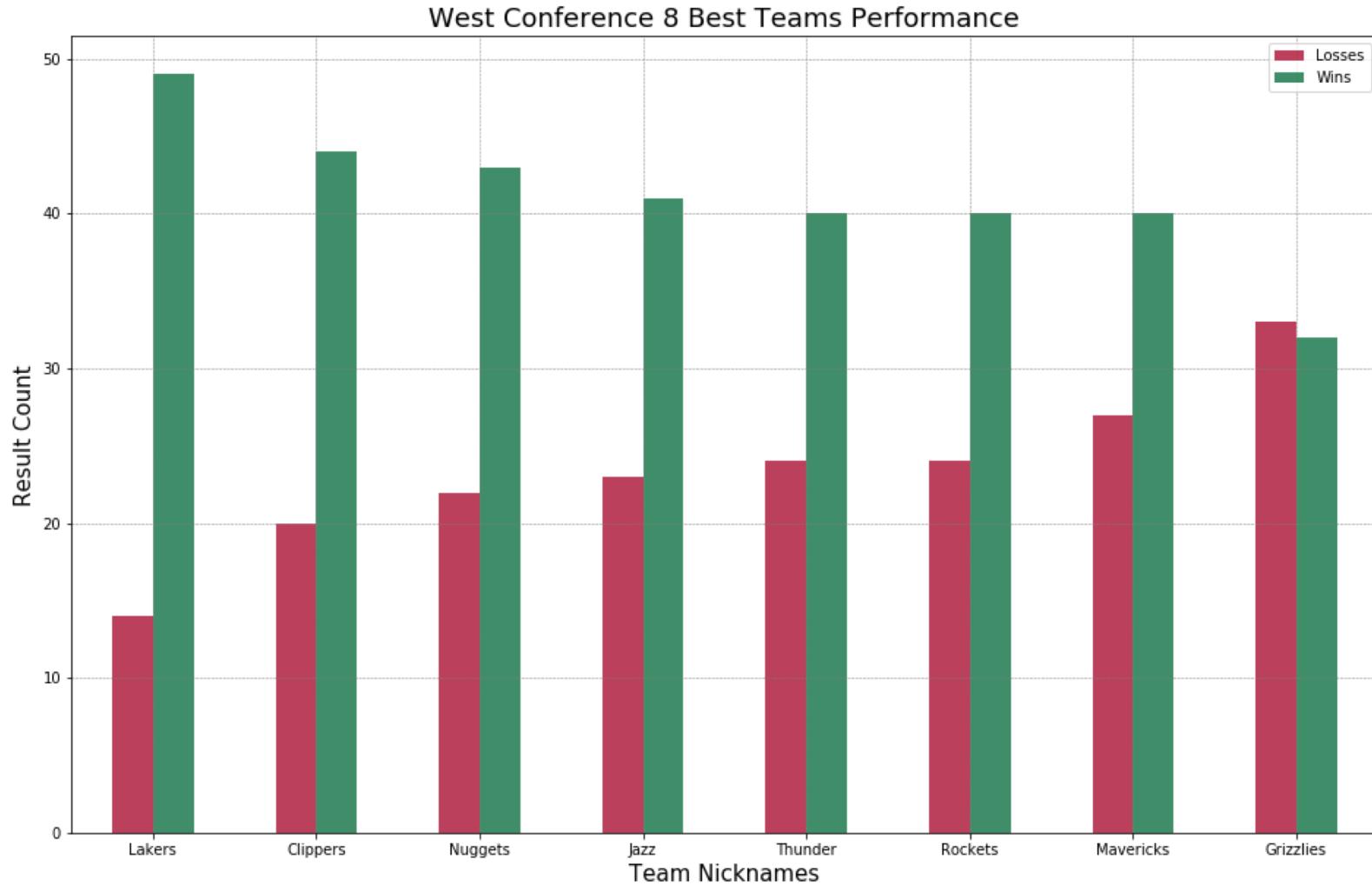
TOP 8 TEAMS

Bar Graph (East)



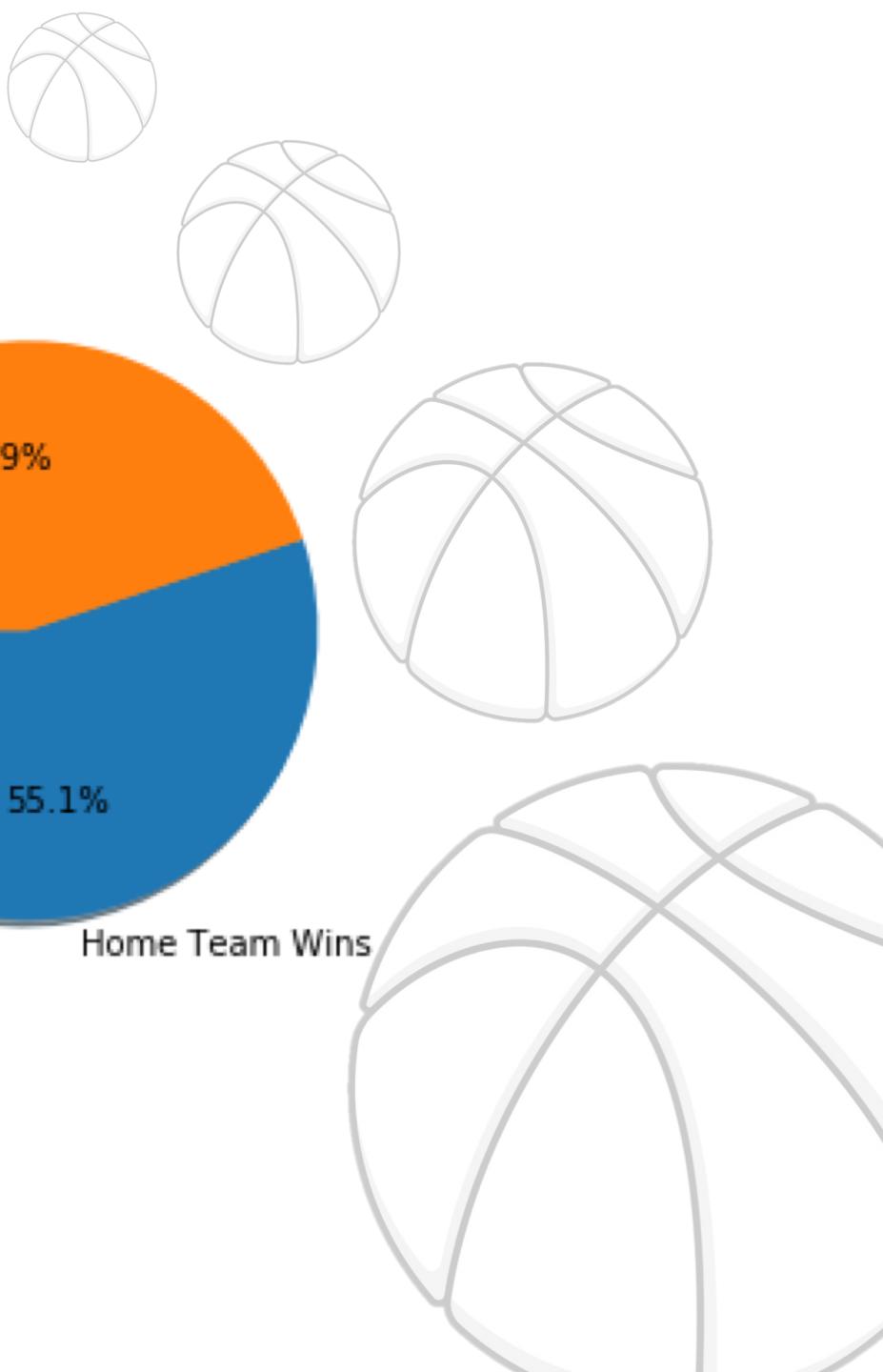
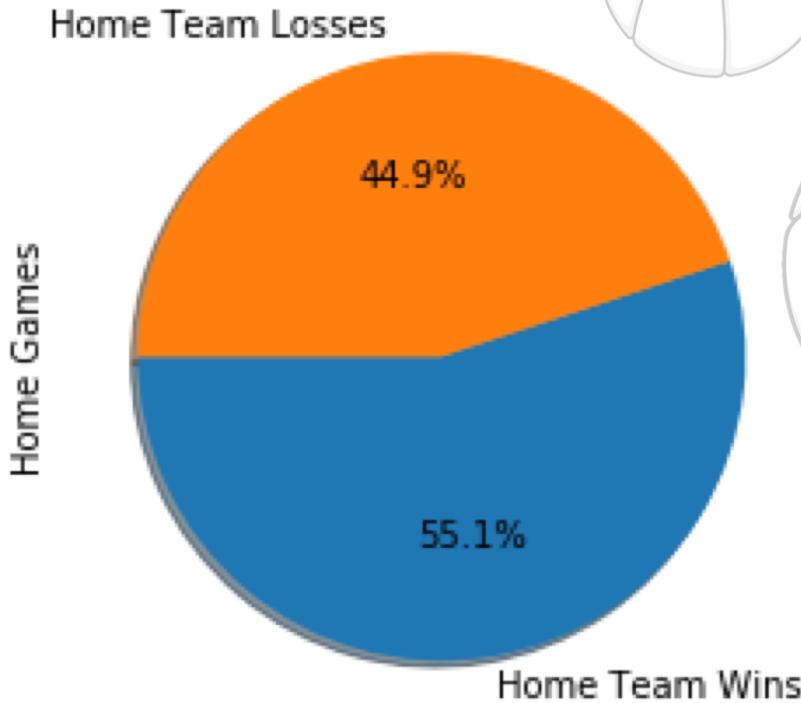
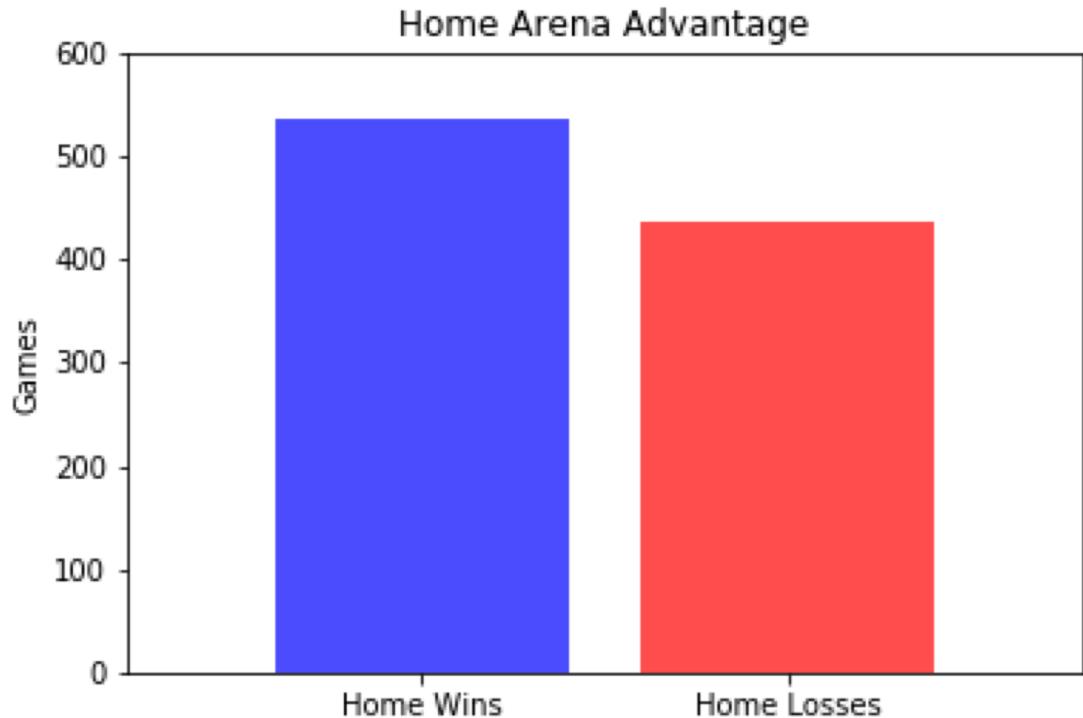
TOP 8 TEAMS

Bar Graph (West)



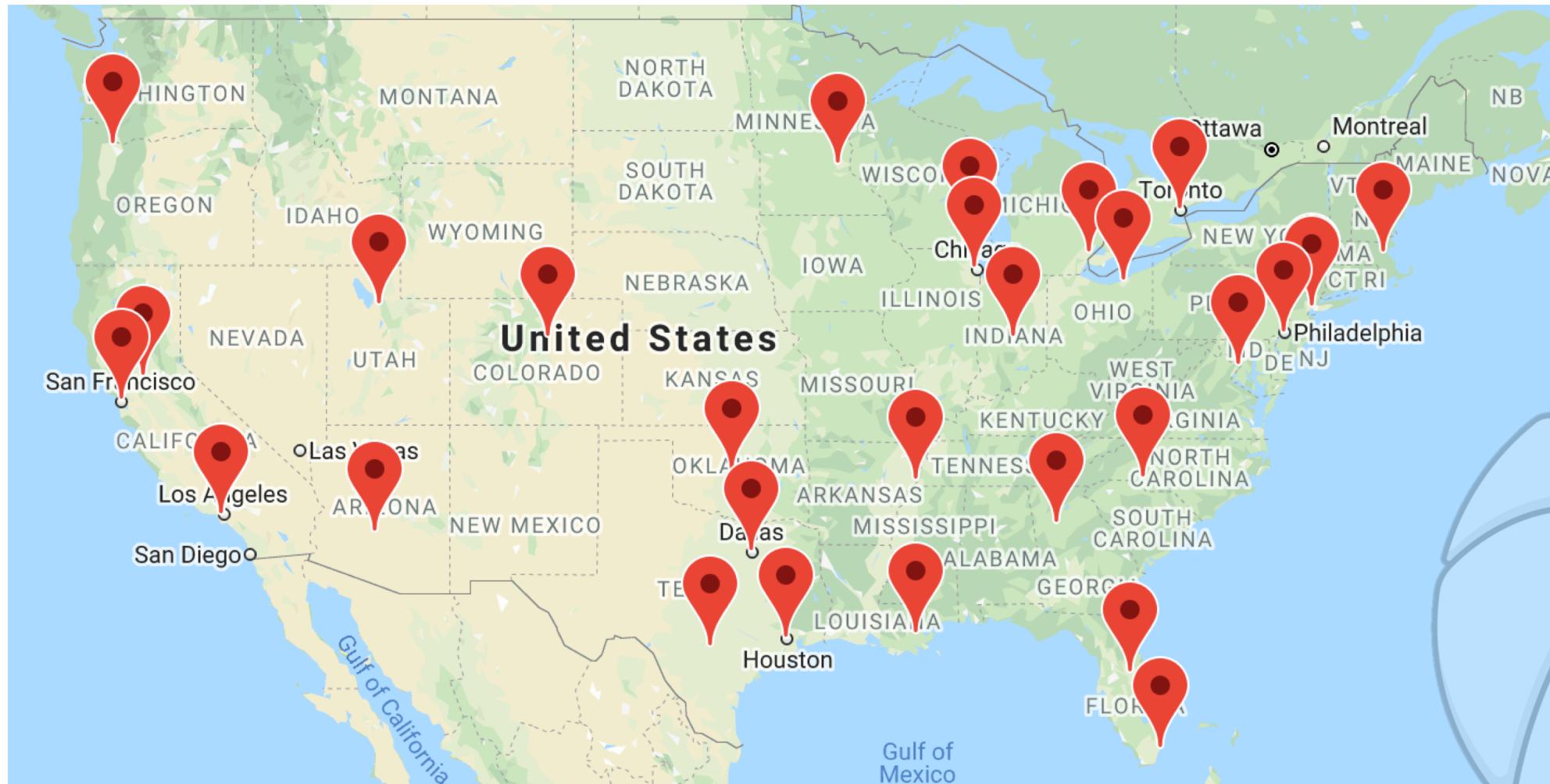
HOME COURT ADVANTAGE

Home Win vs Loss Counts



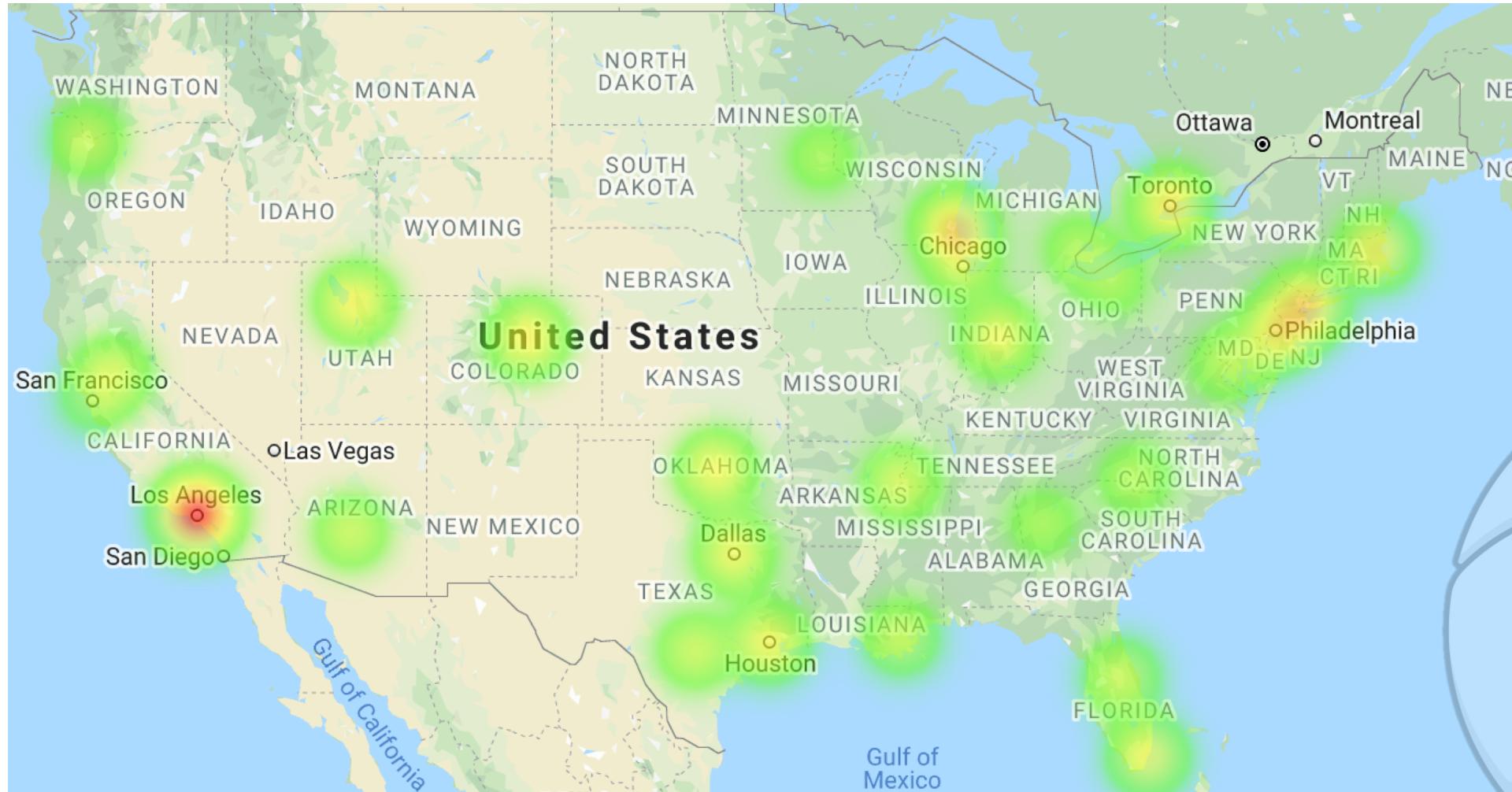
LOCATIONS

Match Venues



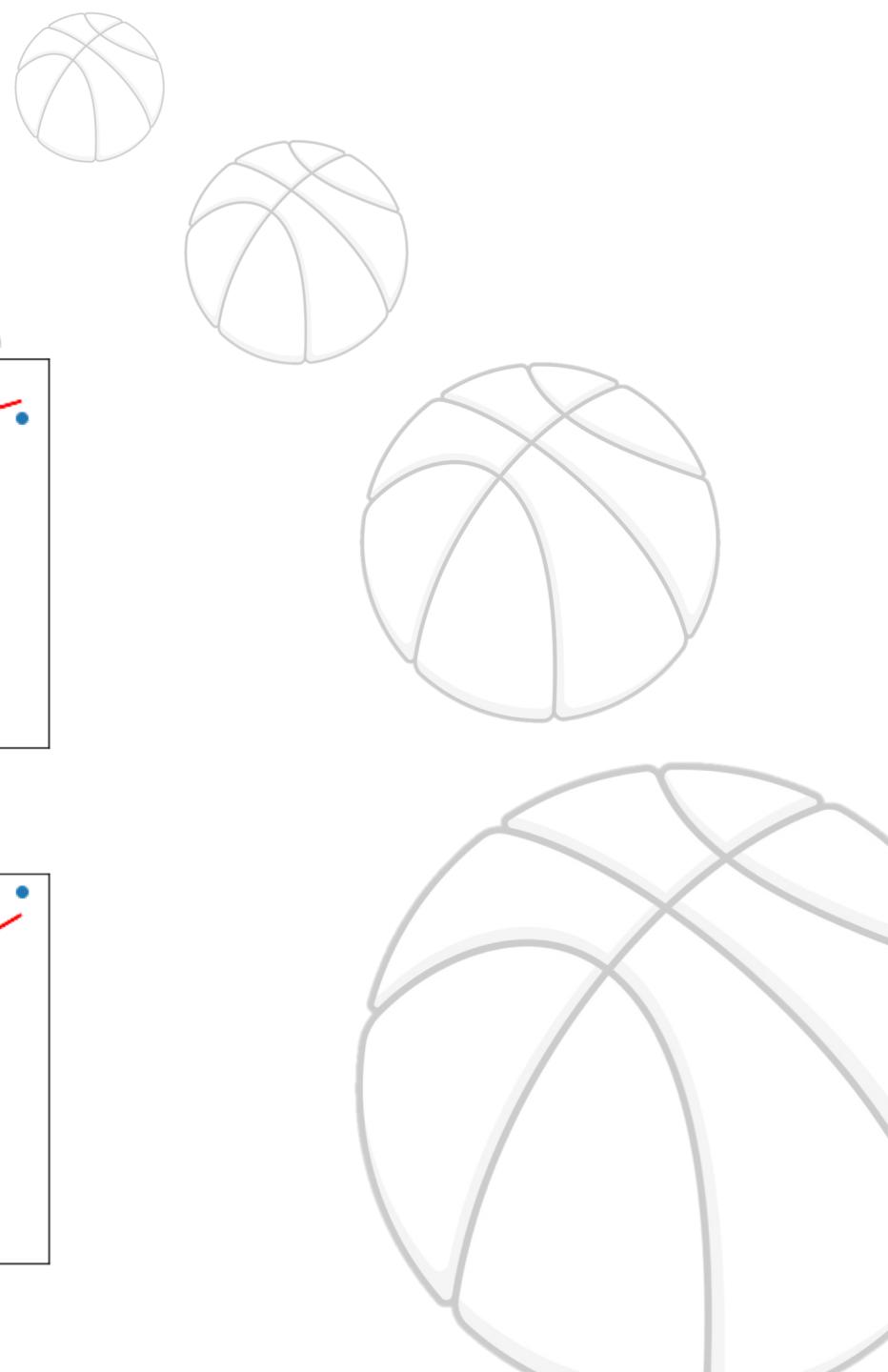
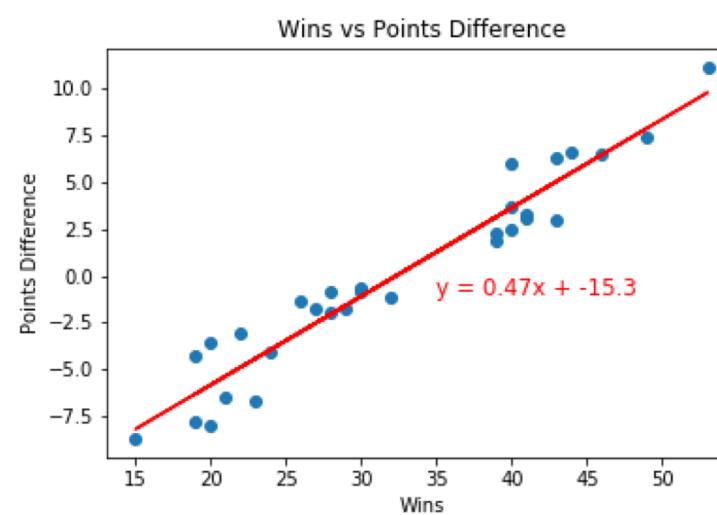
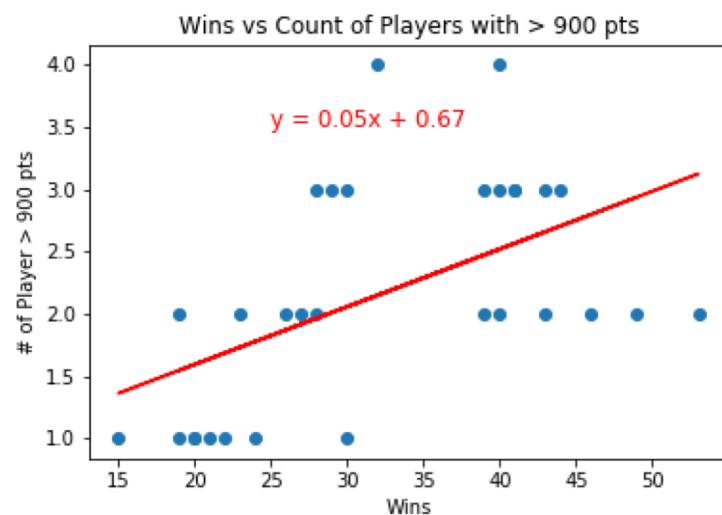
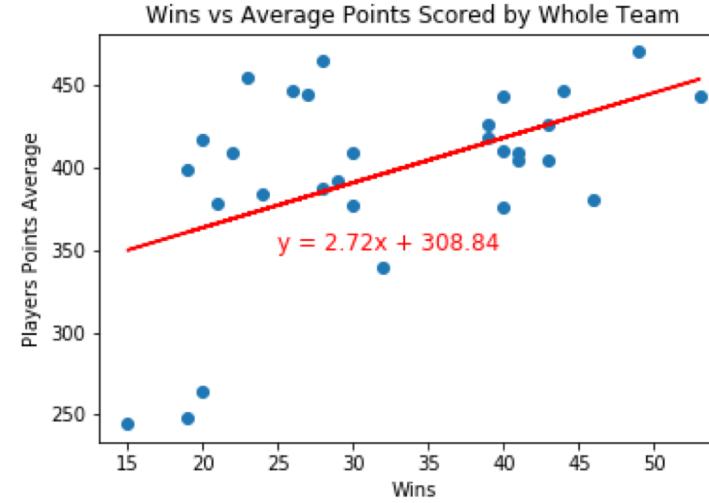
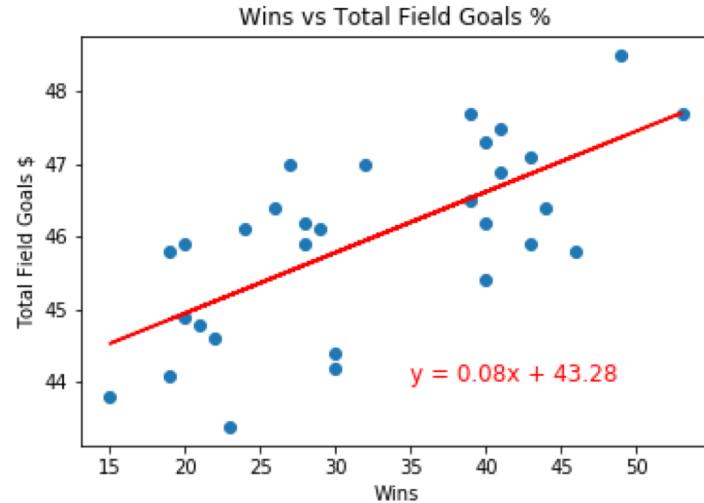
LOCATIONS

Winning Teams Concentration



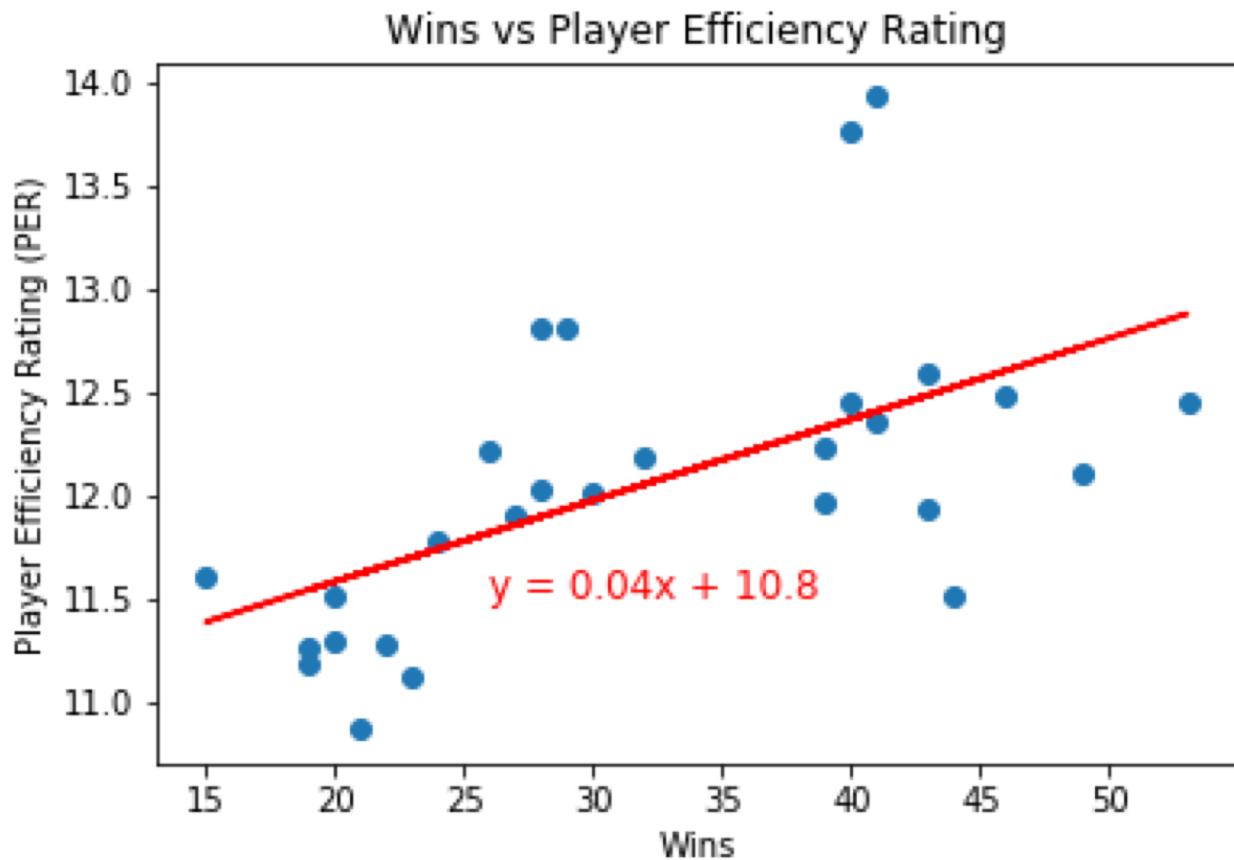
REGRESSION

Win vs Variables



REGRESSION CONTD.

Win vs PER



NBA's Player Efficiency Rating: $(PTS + REB + AST + STL + BLK - ((FGA - FGM) + (FTA - FTM)) + TO)$

REGRESSION CONTD.

Correlation, P-Value & R-Squared

	Correlation	P-Value	R-Squared
Win vs TFG	0.69	2.8821e-05	0.47
Win vs APS	0.49	0.005525	0.24
Win vs GPC	0.53	.00261	0.28
Win vs PD	0.96	8.4217e-18	0.93
Win vs PER	0.58	0.001281	0.33

Win Factor Equation =

$$.08(\text{FGT}) + .47(\text{PD}) + .05(\text{GPC}) + 2.72^*(\text{APS}/1000) + .04(\text{PER})$$

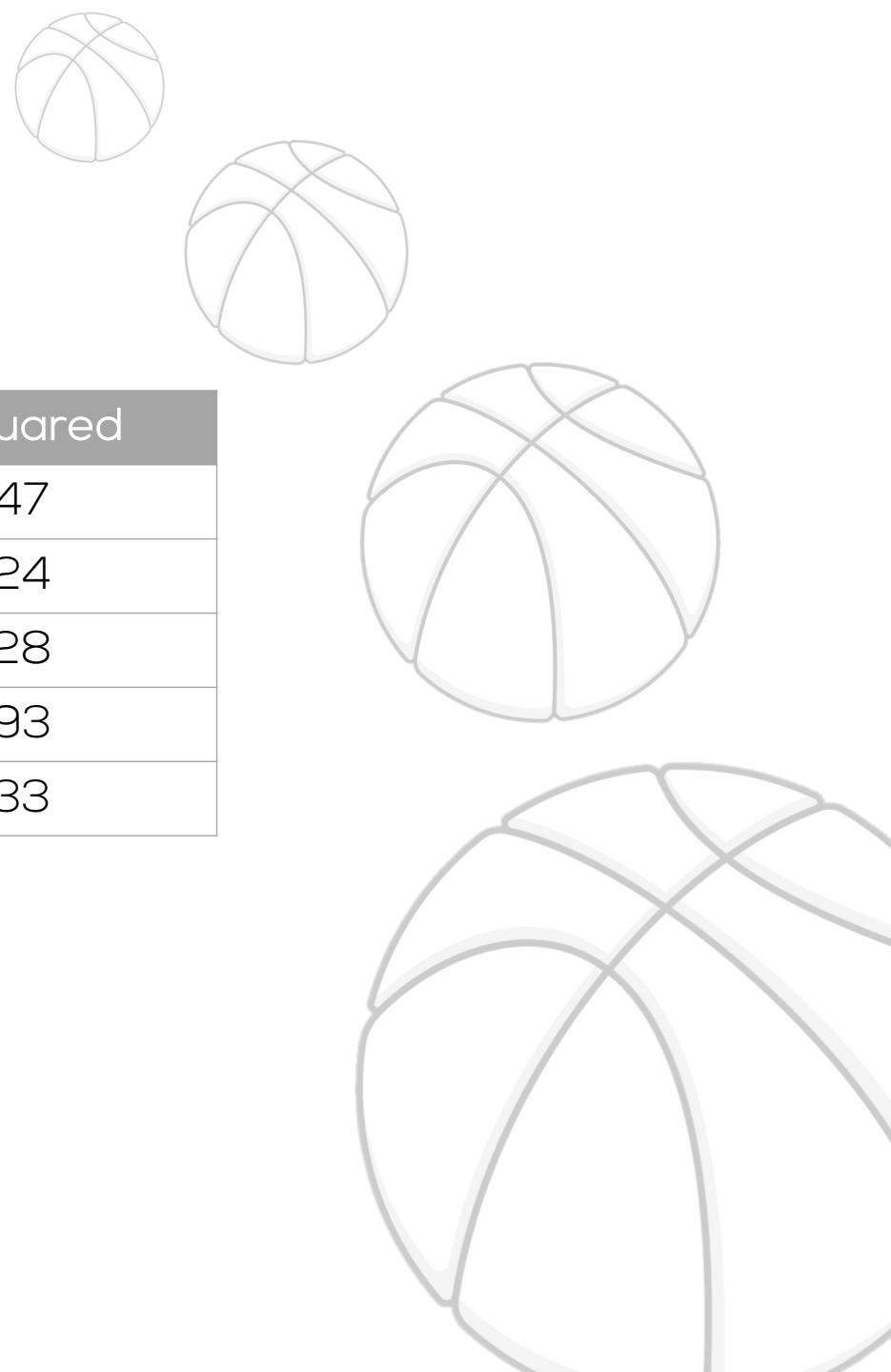
TFG Total Field Goals Percentage

APS Average Player Score of Whole Team

GPC Good Players Count (Pts > 900)

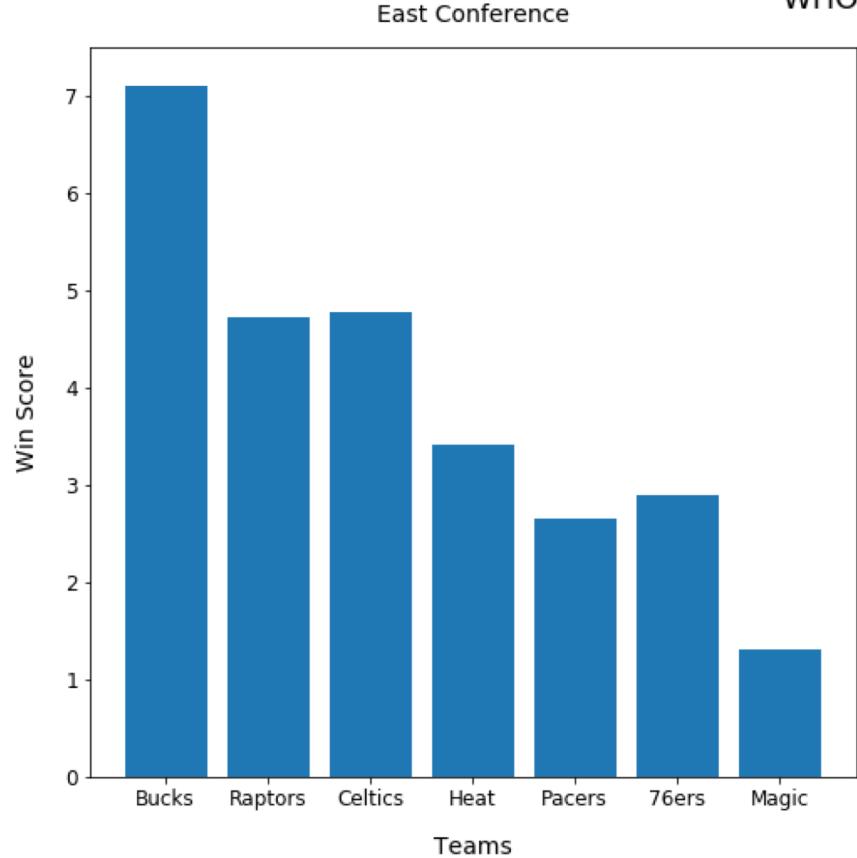
PD Points Differences

PER Player Efficiency Ratio

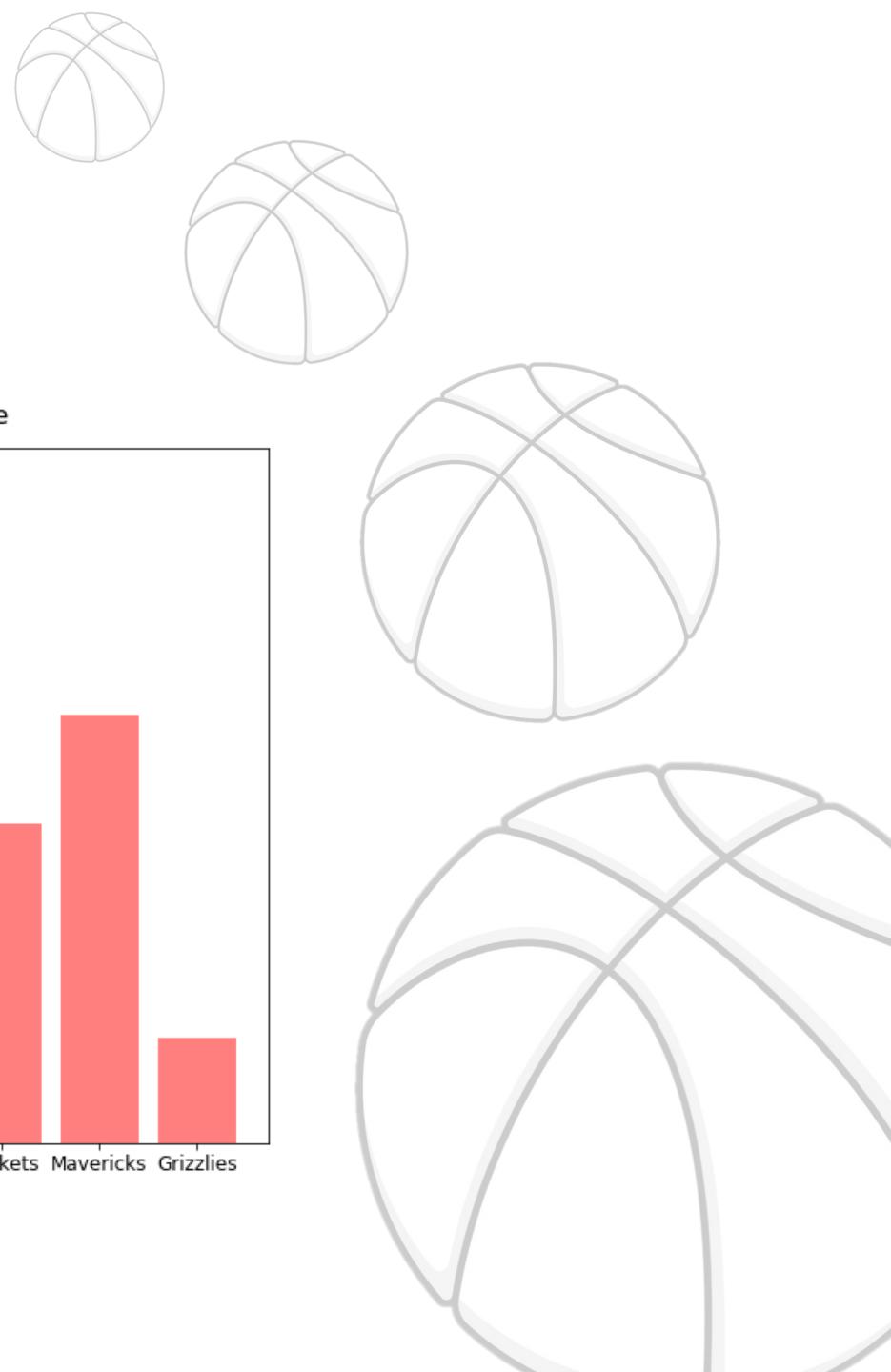
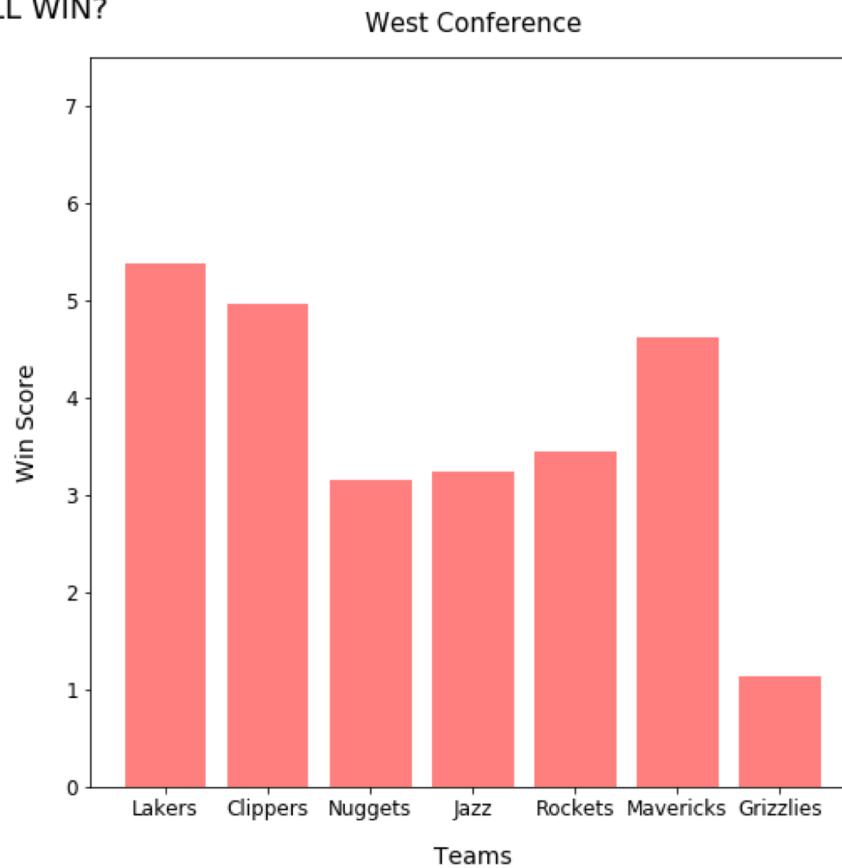


WINNING TEAM

Recommendations



EAST VS WEST!
WHO WILL WIN?



CONCLUSION

Recommendations

Q.1. Who are the top 8 teams for East and West Conference?

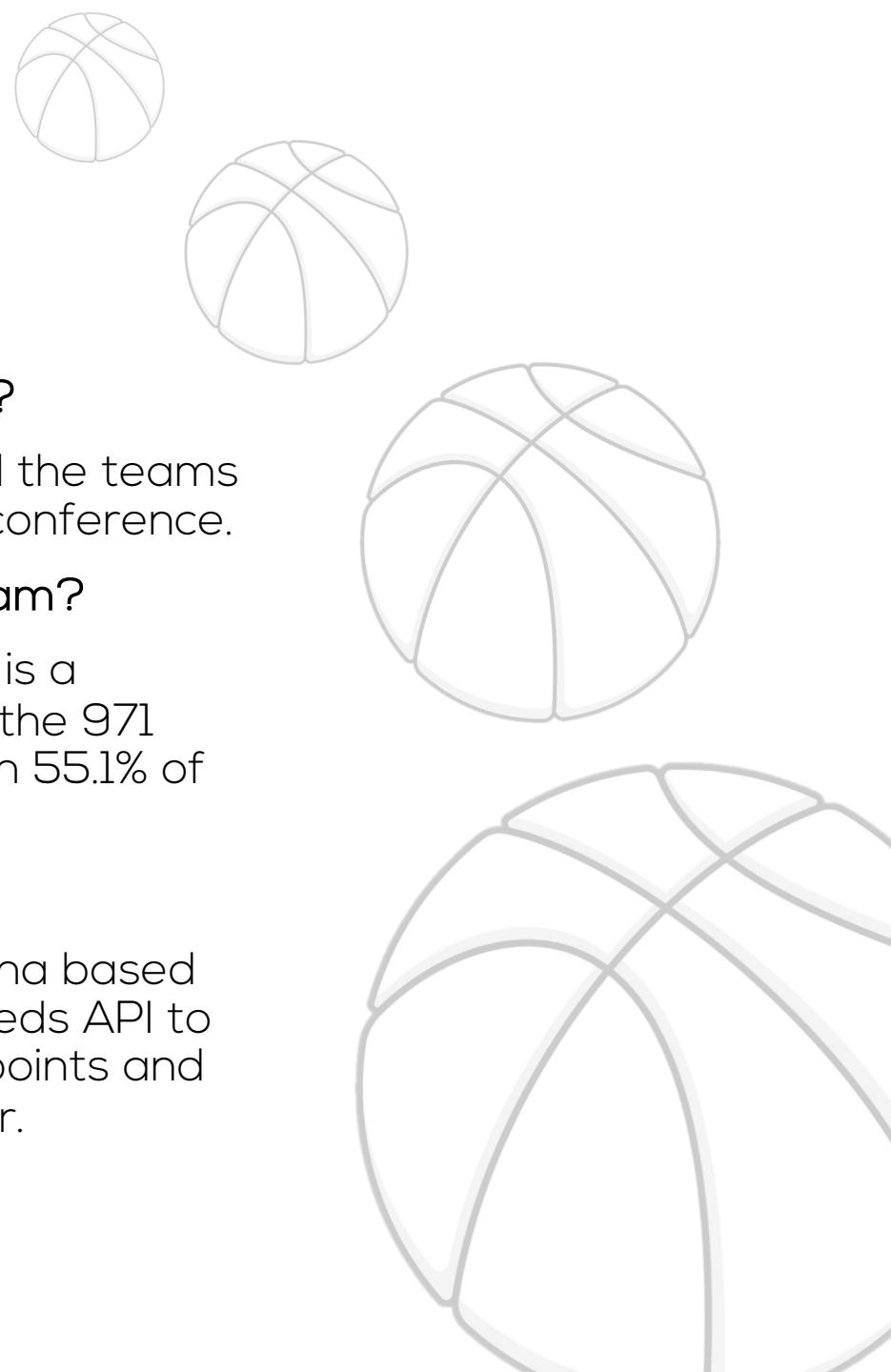
To find the top eight teams, we looked at the win loss ratio of all the teams and found the top 8 teams for both the Eastern and Western conference.

Q.2. Does the home team win more matches or the visitor team?

We found by looking at the 2019-2020 season data, that there is a noticeable advantage to playing on a team's home court. Using the 971 games played for the season for reference, the home team won 55.1% of the time.

Q.3. Where are the locations of the game?

Using google places API, we used a location search for each arena based on the ones listed in the data sets acquired from the Sports Feeds API to then map them across the united states. We then used these points and the teams win percentages as the factor for our heat map layer.



CONCLUSION

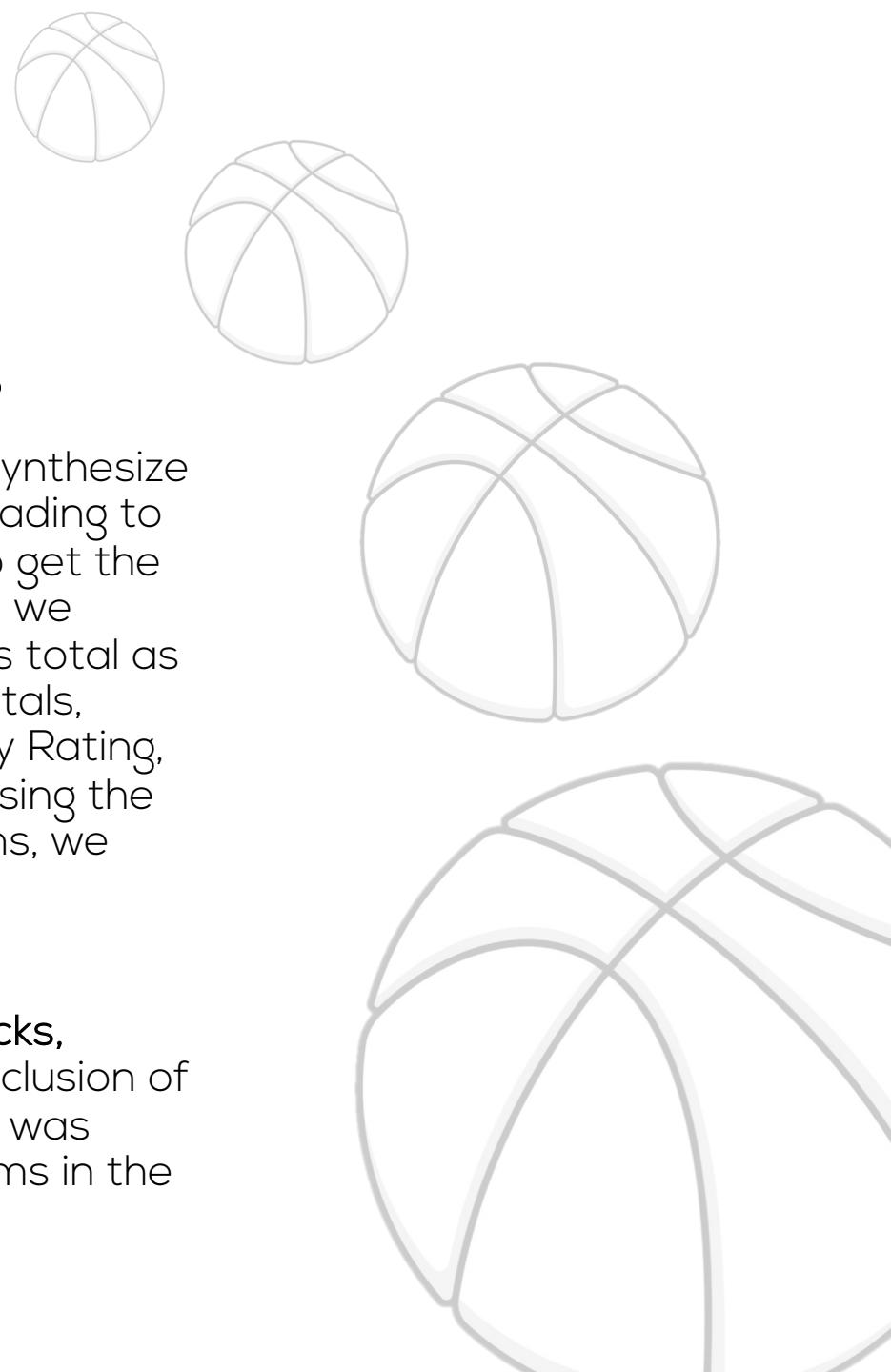
Recommendations

Q.4. How can we determine who has the best chance of winning?

Answering this question took some crunching, but we were able to synthesize a team's win factor by looking at a multitude of factors ultimately leading to a team's success. We found our data using the Sports Feeds API to get the most up to date data. First, we looked at their win percentage, then we looked at how many players on a team's roster scored over 900 pts total as our threshold. Then after running linear regressions on Field Goal Totals, Points Difference, Average Points Scored, and Final Player Efficiency Rating, we were able to devise a formula for getting a teams Win Factor. Using the slope of these individual regressions as a multiplier for the top teams, we came up with the formula to make our final recommendations.

Q.5. Who had the best chance of winning?

From our findings and analysis of the 2019-20 season data, the **Bucks**, **Lakers**, and the **Clippers** had the best chance of winning at the conclusion of the season should it not have been cancelled due to COVID-19. This was based on their Win Factor which can be seen against the other teams in the bar graph titled "East vs West! Who Will Win?"



CONSIDERATIONS

Recommendations

1. Another thing to consider is that the team Win Factor does not factor in the difficulty of schedule, and to factor this in would take some machine learning, and some much deeper analysis. If two teams faced off and it was more of a defensive match, the scores and stats would look low, but you might see more fouls committed, blocked shots, or offensive rebounds, so this is more of a look at the offense of the teams than the defense.
2. Also another thing to consider was that this was a shortened seasons cause of the Covid-19 and full data was not available till the last game.
3. Also the API did not allow us to get data for individual games hence Kaggle data was used to calculate player efficiency ratio. However Kaggle data was missing a few games stats for last 10 days of games.

