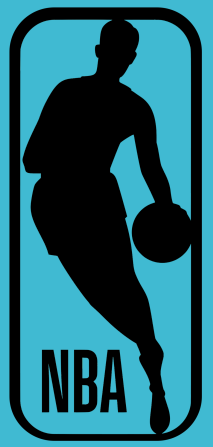
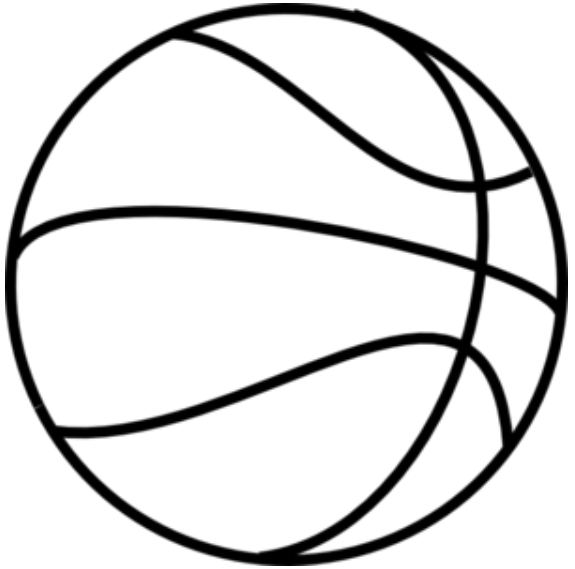


Does playing ***at home*** result in more wins?

DSC 530 Final Project – Bellevue University

Maddie Bauer



Images from: https://www.pngkey.com/maxpic/u2e6i1y3w7r5e6u2/ and http://clipart-library.com/white-basketball-cliparts.html

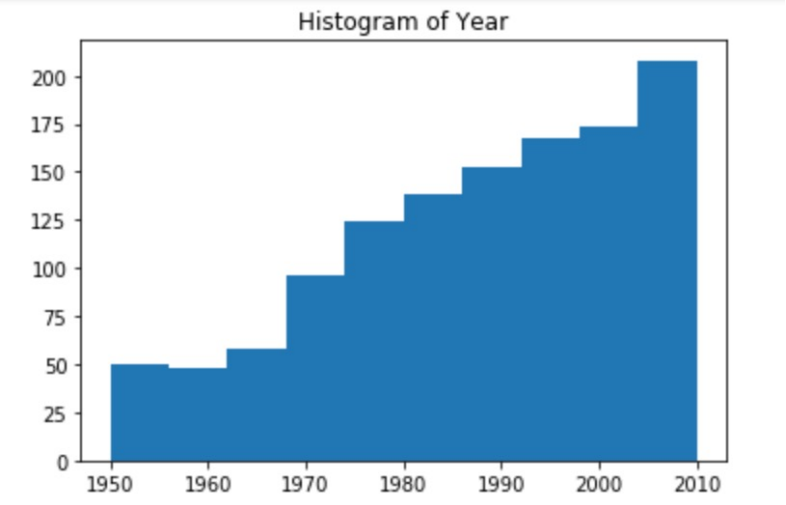
Statistical Question

Have you ever heard of "home court advantage"? For this project, I want to decide if playing an NBA game on a team's home court is truly an advantage. For this project I will look at correlation, hypothesis testing and linear regression. By the end, I will determine if home court advantage is real or just a hoax based on my dataset.

The Data

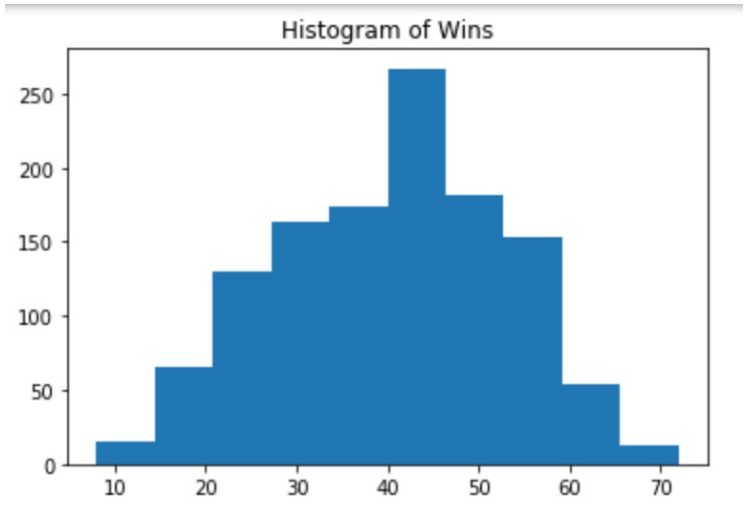
- This dataset has the **home** and **away** performance statistics of NBA teams during the years 1950-2010.
- Single csv file
- 1,216 observations
- 11 variables
- Found on Kaggle (https://www.kaggle.com/drgilermo/home-advantage-in-soccer-and-basketball?select=NBA.csv)

Variables



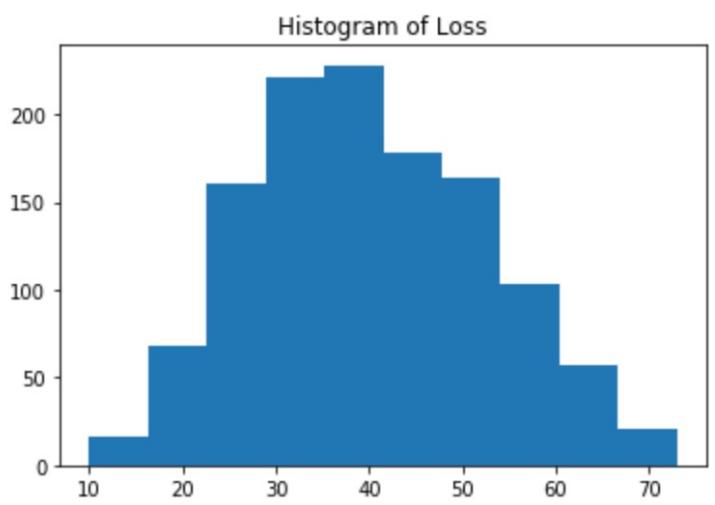
count	1216.00
mean	1987.12
std	15.68
min	1950.00
25%	1976.00
50%	1989.00
75%	2000.00
max	2010.00

Year (int) – The year that the season was played.



count	1216.00
mean	40.18
std	12.45
min	8.00
25%	31.00
50%	41.00
75%	49.00
max	72.00

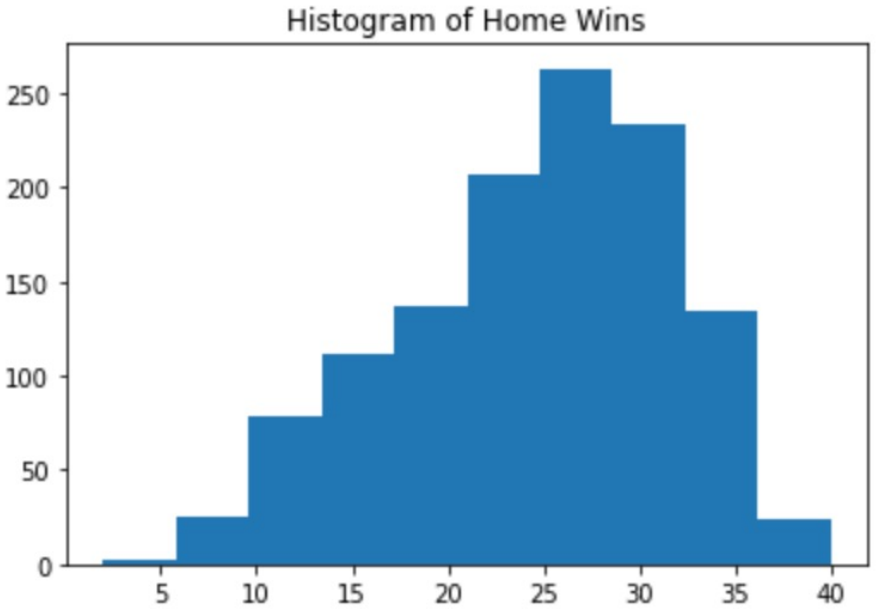
Wins (int)– The number of wins a team had in a season.



count	1216.00
mean	40.18
std	12.43
min	10.00
25%	31.00
50%	40.00
75%	49.00
max	73.00

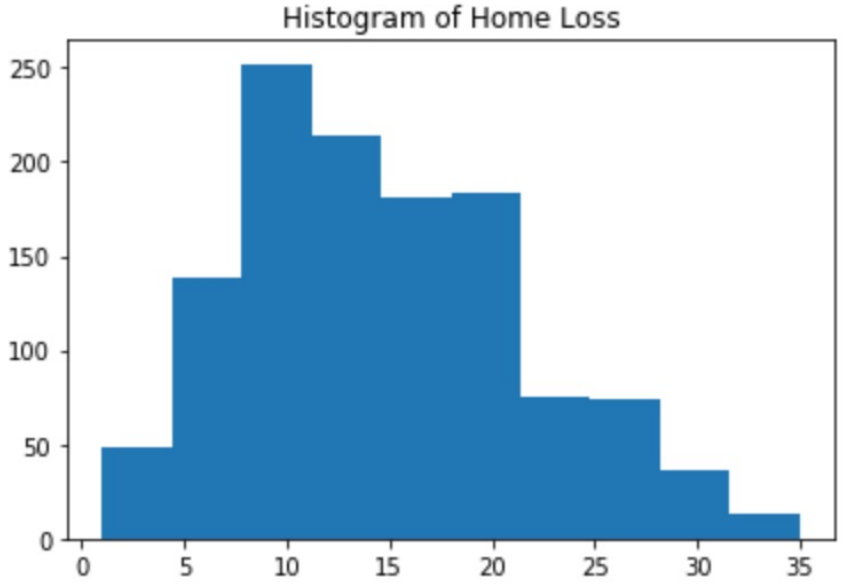
Loss (int) – The number of losses a team had in a season.

Variables Continued



count	1216.00
mean	24.40
std	7.13
min	2.00
25%	19.00
50%	25.00
75%	30.00
max	40.00

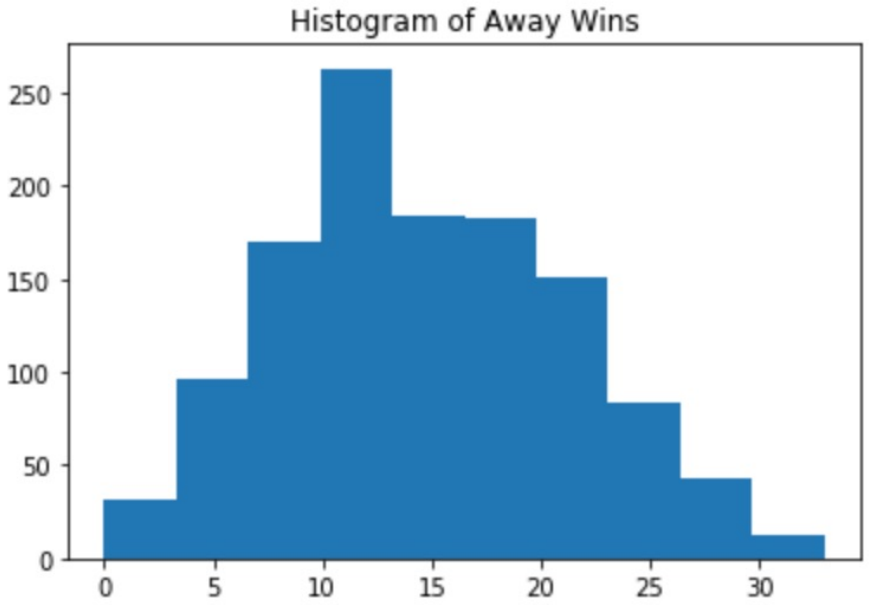
Home Wins (int) – The number of wins for a team at home during the given season.



count	1216.00
mean	14.67
std	6.84
min	1.00
25%	10.00
50%	14.00
75%	19.00
max	35.00

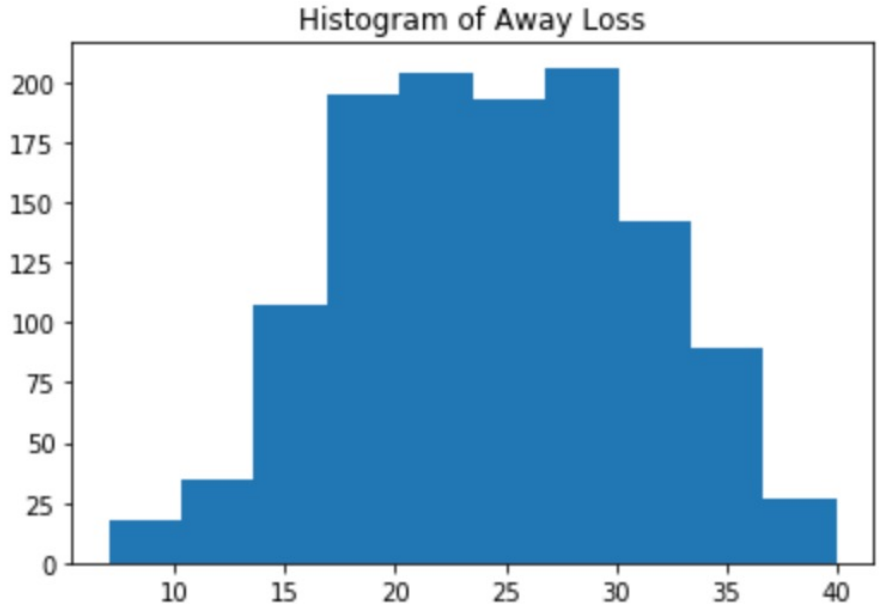
Home Loss (int) – The number of losses for a team at home during the given season.

Variables Continued



count	1216.00
mean	14.67
std	6.56
min	0.00
25%	10.00
50%	14.00
75%	19.00
max	33.00

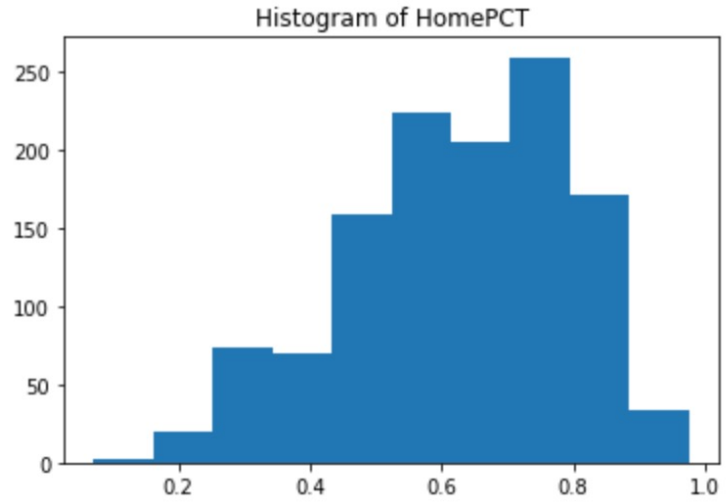
Away Wins (int)- The number of wins for a team on the road during a given season.



count	1216.00
mean	24.41
std	6.60
min	7.00
25%	20.00
50%	24.00
75%	29.00
max	40.00

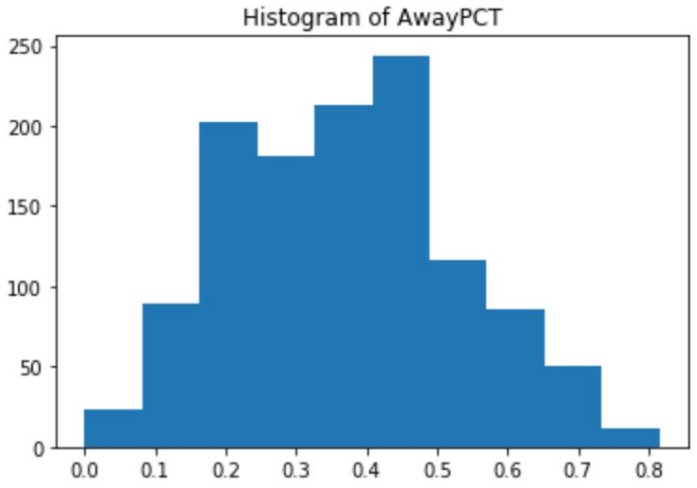
Away Loss (int) – The number of games lost for a team on the road for a given season.

Variables Continued



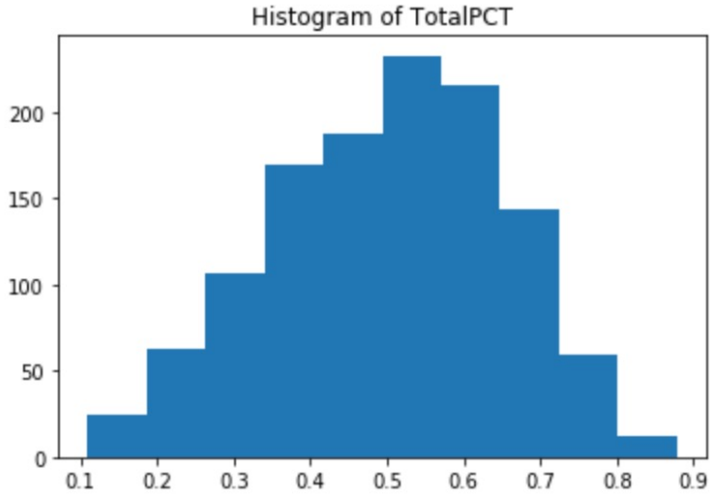
count	1216.00
mean	0.62
std	0.17
min	0.07
25%	0.51
50%	0.63
75%	0.76
max	0.98

HomePCT (float) – The percentage of wins at home for the season.



count	1216.00
mean	0.37
std	0.16
min	0.00
25%	0.24
50%	0.37
75%	0.49
max	0.82

AwayPCT (float) – The percentage of wins on the road for the season.



count	1216.00
mean	0.50
std	0.15
min	0.11
25%	0.39
50%	0.51
75%	0.61
max	0.88

TotalPCT (float) - The percentage of wins at home AND on the road for the season.

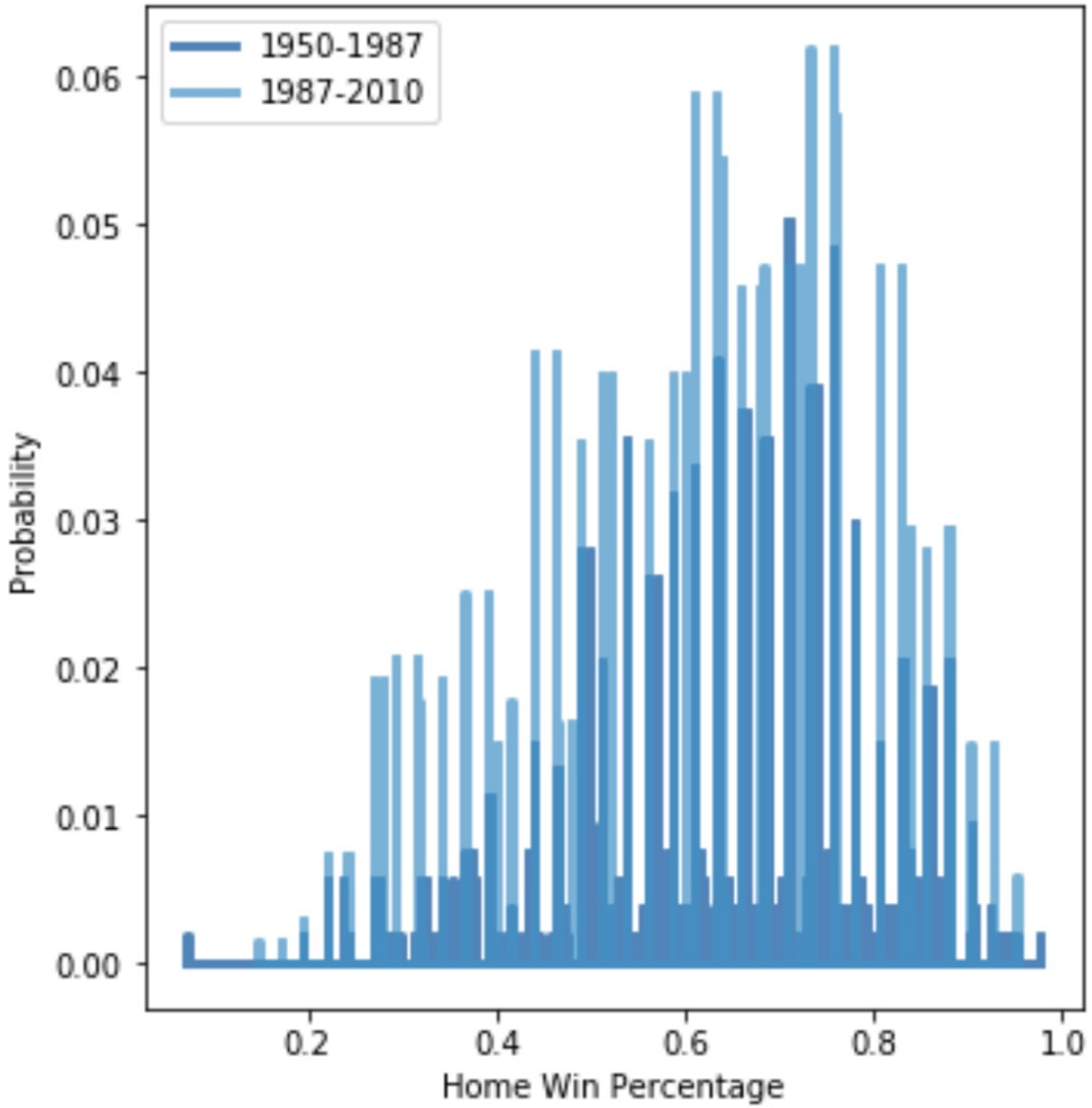
*This histogram is skewed in favor of home court advantage being real.

PMF

The PMF for HomePCT is split by years. The mean year (1987) was the splitting factor.

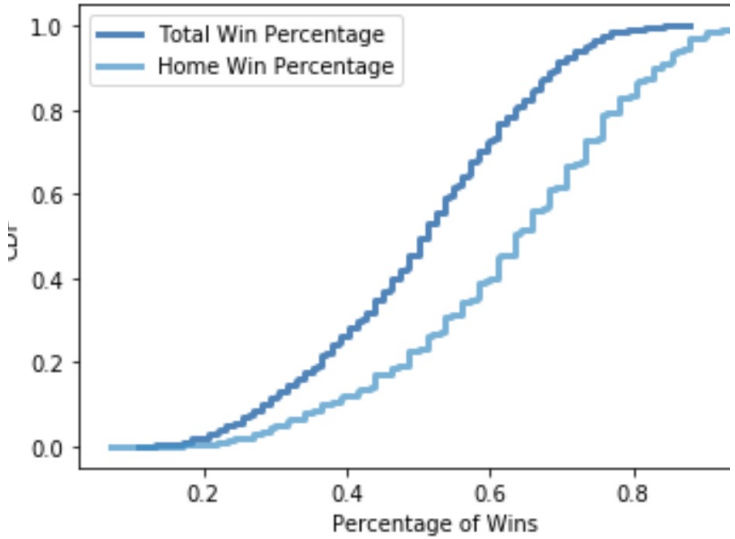
The first group is 1950-1987 and the second group is 1987-2010.

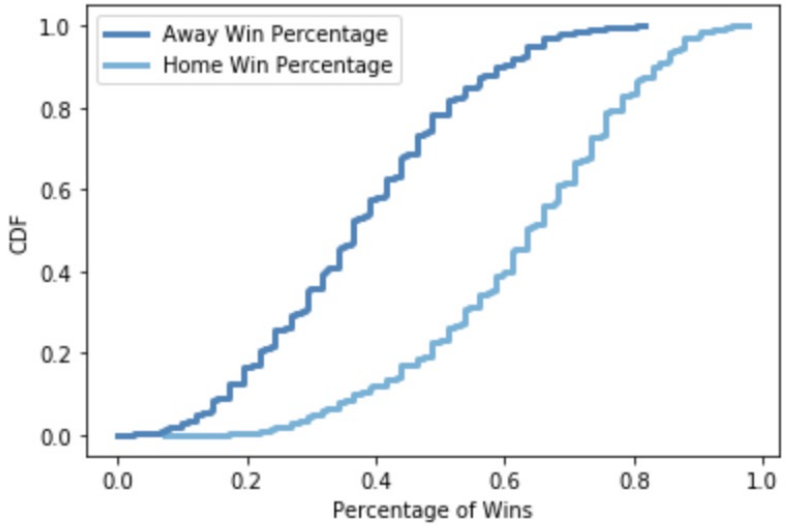
From the graph we can see the proportion of Home Wins for the more current years is much greater. This could be due to many different factors, such as the number of games per season increasing over the years , an increase in off-season training, etc.



CDF

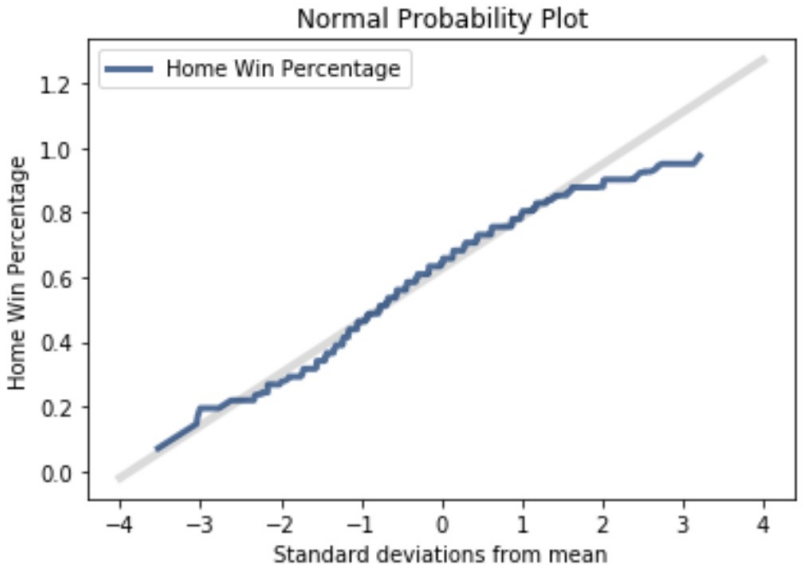
These CDF graphs show us that a team's win percentage is overall better at home when compared to their overall win percentage as well as their away win percentage.

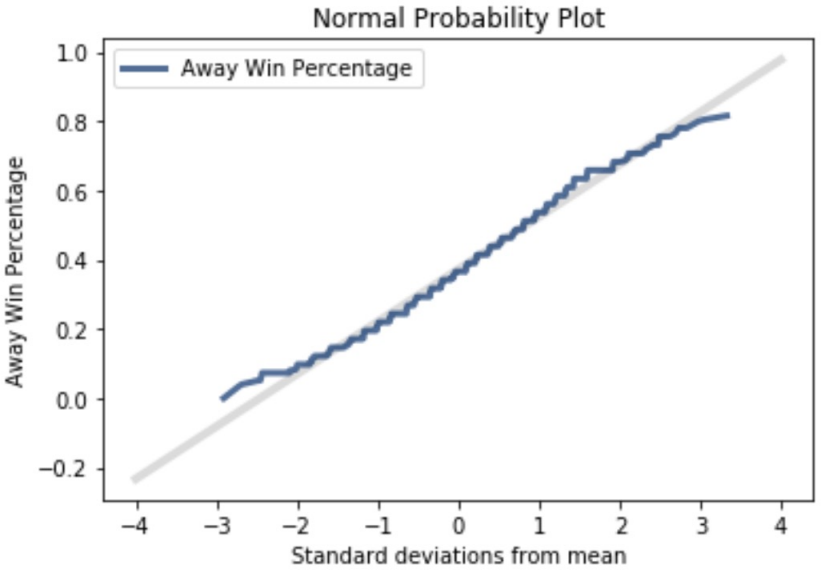


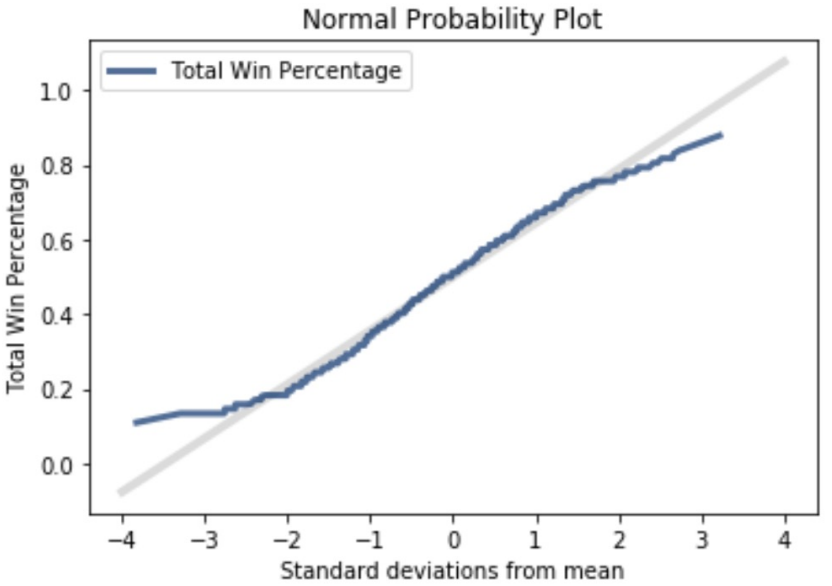


Normal Probability Plots

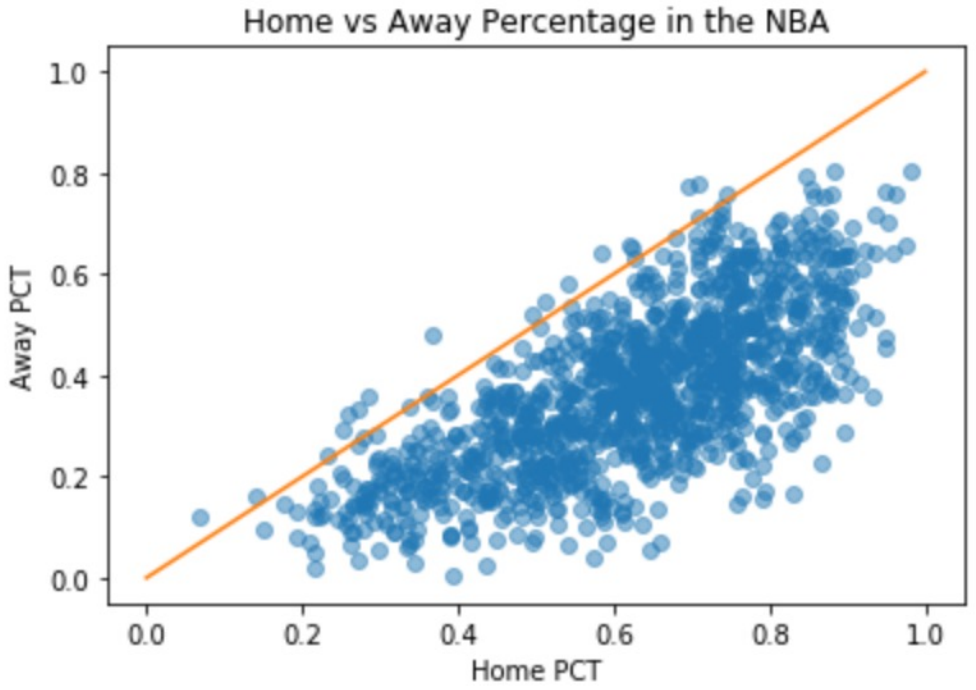
The plots on the right appear to be normal distributions. In each scenario the data lies close to the perfect normal model, which is represented by the grey line on each plot, with the exception of being 2 or more standard deviations away from the mean.





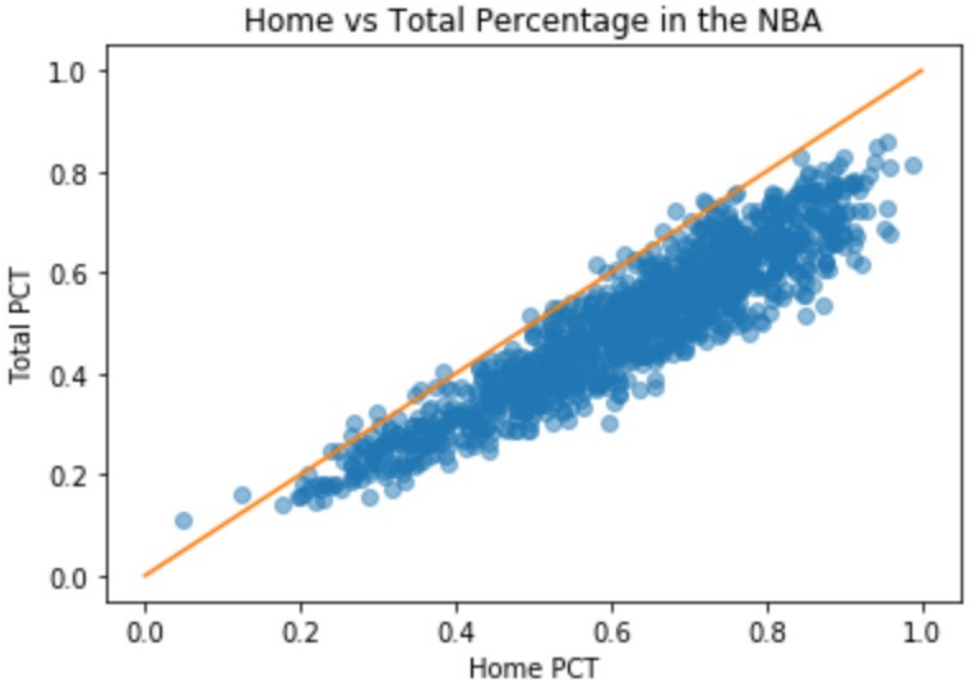


Scatter Plots & Correlation



Correlation: 0.706

There is a strong correlation between home win percentage and away win percentage.



Correlation: 0.927

There is a very strong correlation between home win percentage and total win percentage.

These scatter plots show that roughly 15 teams are represented above the orange line (y = x).
This means that only 15 (out of 1,216) teams have performed better away during their season compared to at home.

Hypothesis Test

Ho: There is *no* difference in winning percentages based on playing at home or away.

Ha: There *is* a difference in winning percentages based on playing at home or away.

Result: I obtained a p-value of 0.0. This tells me we would reject the null hypothesis and accept the alternative hypothesis. There **is** a difference in winning percentages based on playing at home or on the road (away).
Thus, home court advantage is a real thing in the NBA.

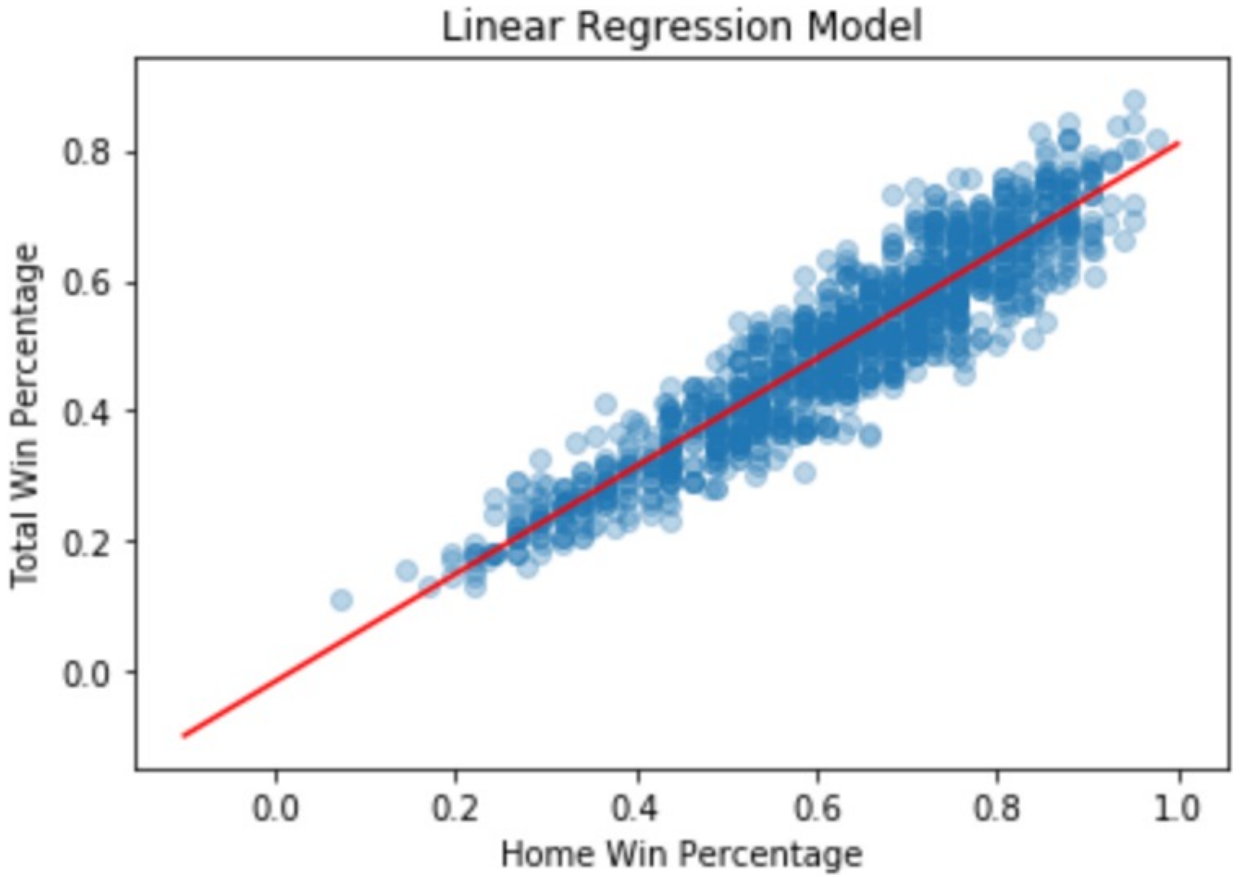
Linear Regression

formula = 'TotalPCT ~ HomePCT'
slope = 0.8254
intercept = -0.0158
R^2= 0.86

This model predicts whether the percentage of home wins affects the overall percentage of wins in a season.

Of course it should/does, so I am not convinced that linear regression was needed in this scenario. However, the results illustrate that 86% of variation in the TotalPCT is related to HomePCT.

My dataset would have needed more variables that could be contributing factors to Home Win Percentage, such as average crowd level, number of fans in attendance, etc. in order to run a more accurate regression model.



References

Downey, A. (2015). Think Stats, 2nd Edition. O'Reilly Media, Inc.

Goldstein, O. (2016). Home Advantage in Soccer and Basketball. Retrieved July 15, 2020 from https://www.kaggle.com/drgilermo/home-advantage-in-soccer-and-basketball