NBA Analytics

Will the home team win?

By: Jason R. Tracey

Image Source

Problem:

- Is my favorite team going to win the game at home??

Stakeholders:

- NBA Teams
- Betting Agencies
- Fantasy Basketball Aficionados
- Sports Enthusiasts

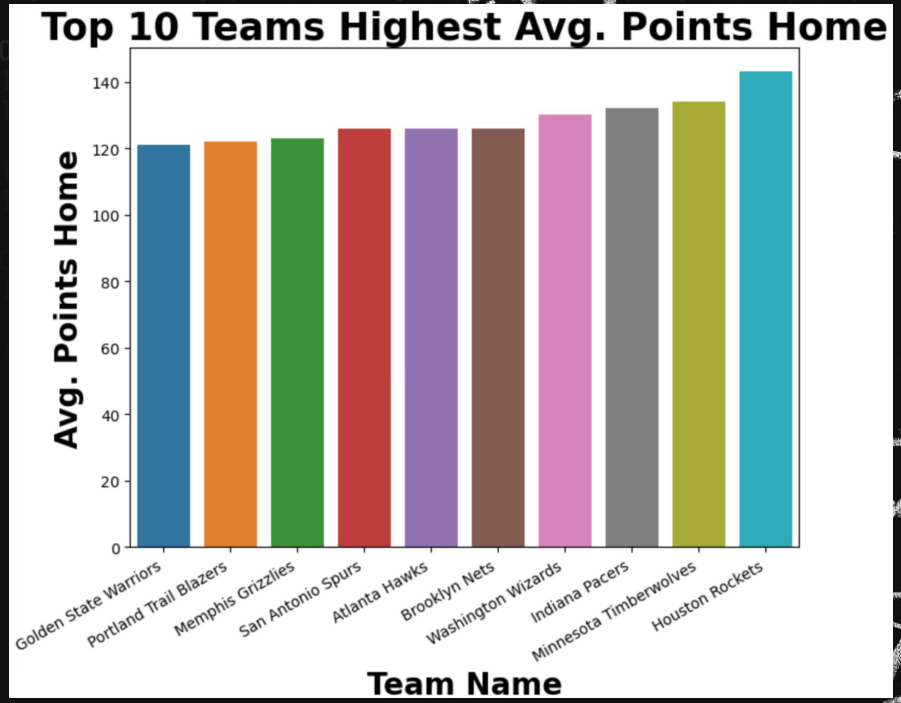
Image Source

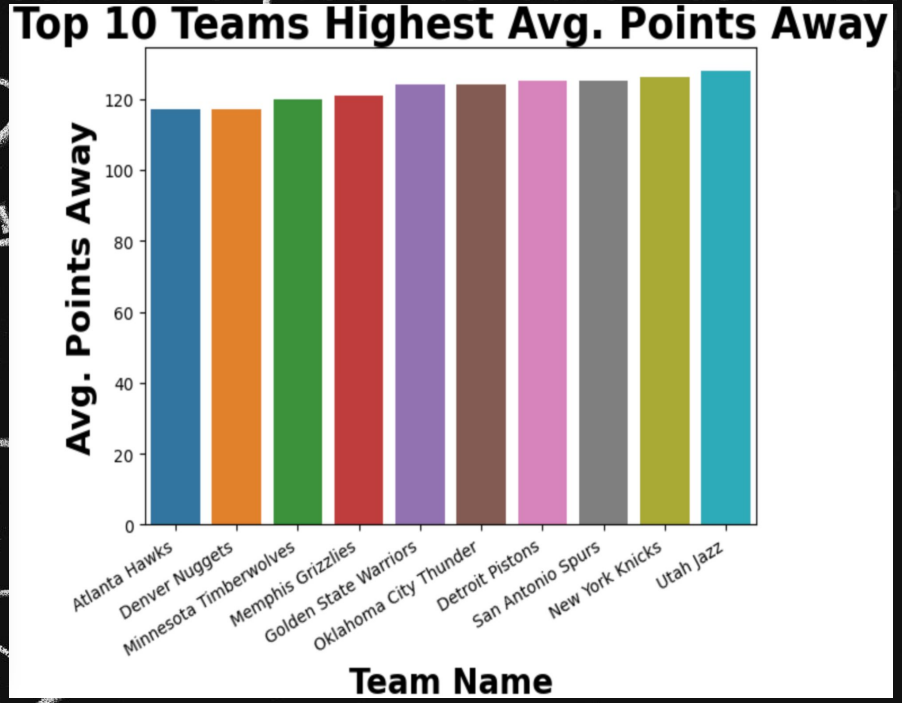
About the Data:

- **Kaggle Data Source:**
- **2 different files:**
  - **games.csv: High Level Game Statistics from 2004 to mid 2022**
  - **teams.csv : Team Info for All Teams**
- **21 columns:**
  - **Home Team: FG%, FT%, 3FG%, Assist, Rebounds, Points Scored**
  - **Away Team: FG%, FT%, 3FG%, Assist, Rebounds, Points Scored**
  - **General Game Info: Game ID, Team ID's(4), Game Status, Season, Game Date**
  - **Target: Home Team Wins**

Image Source

A Look At The Visuals:

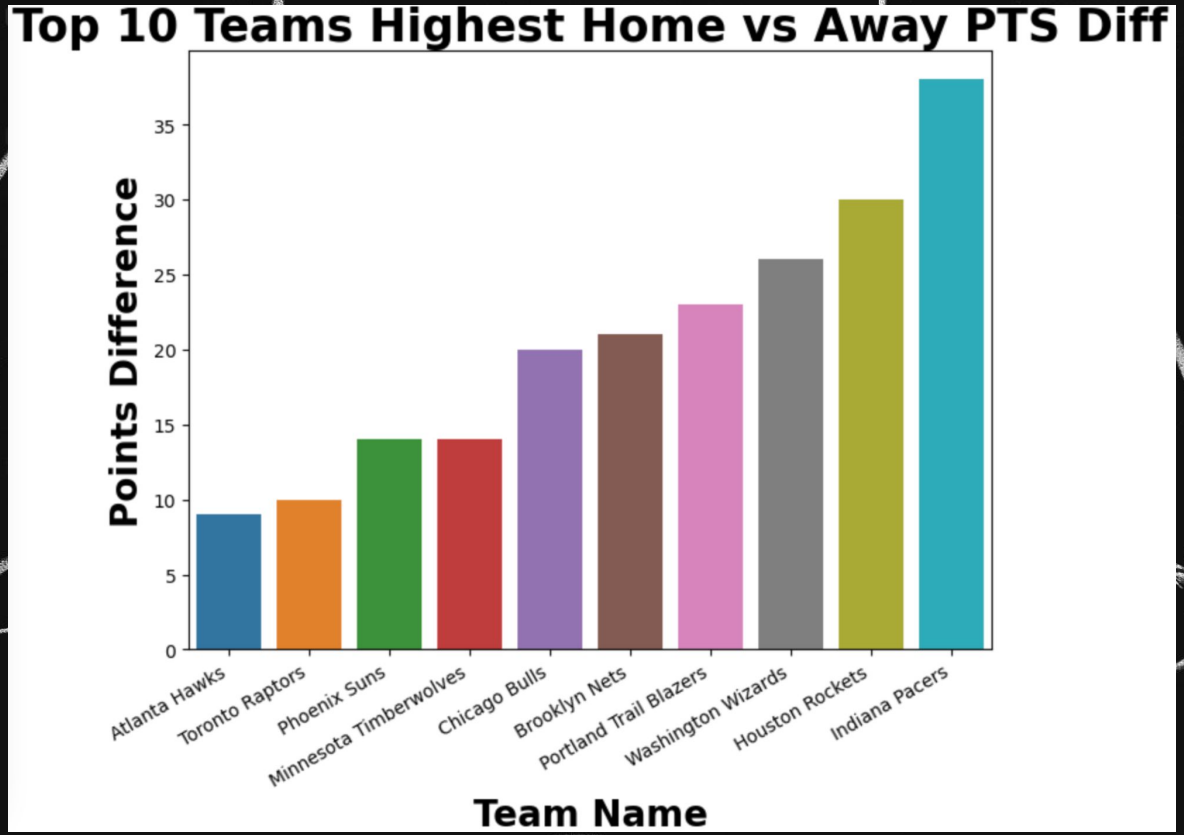




These bar plots show the Top 10 teams that average the most points at home and away respectively.

Image Source

Another Look At The Visuals:



This bar plot shows the Top 10 Home Scoring vs Away Scoring Differential

Image Source

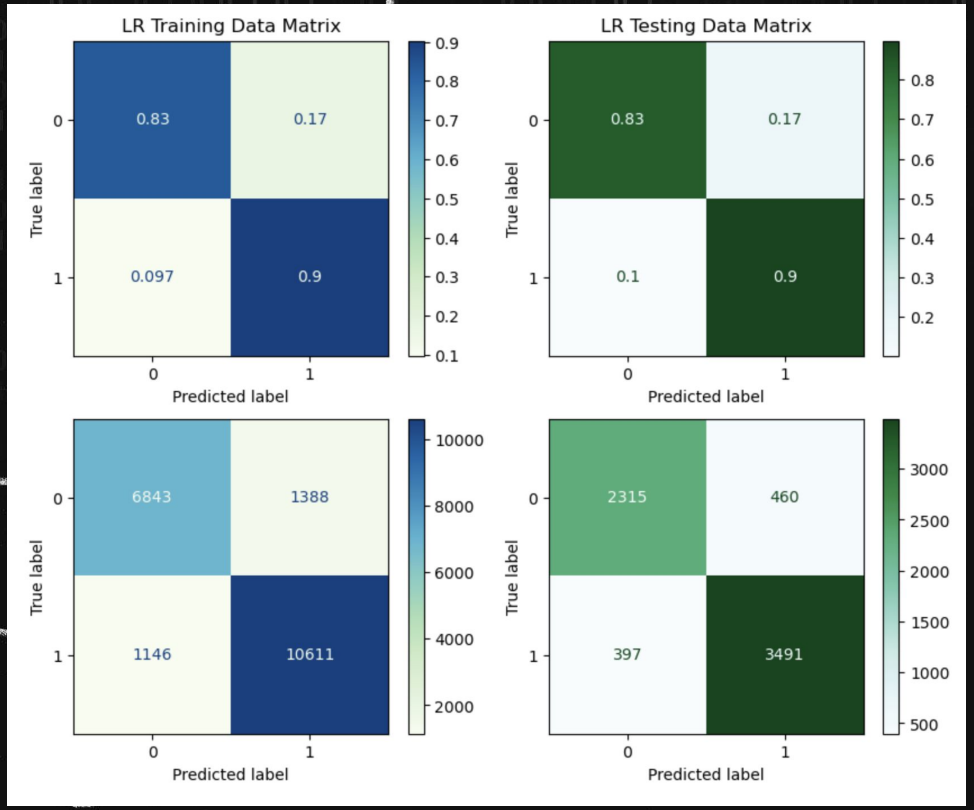
A Look At The “Best” Model:

Logistic Regression Training Data Metrics:

	precision	recall	f1-score	support
0	0.86	0.83	0.84	8231
1	0.88	0.90	0.89	11757
accuracy			0.87	19988
macro avg	0.87	0.87	0.87	19988
weighted avg	0.87	0.87	0.87	19988

Logistic Regression Testing Data Metrics:

	precision	recall	f1-score	support
0	0.85	0.83	0.84	2775
1	0.88	0.90	0.89	3888
accuracy			0.87	6663
macro avg	0.87	0.87	0.87	6663
weighted avg	0.87	0.87	0.87	6663



Default Linear Regression Model performs the best

Image Source

A Further Dive Into the Model:

Strengths:

- 87% accurate
- 86% precise at picking losses
- 88% precise at picking wins
- Good statistical data

Limitations:

- 6,663 data points
  - Need more data
- 397 false negatives
  - Underdog scenario
- 460 false positives
  - Overconfidence
- Data is gathered late

Image Source