Final Project

PREDICTING NBA RESULTS

FEB 16, 2023

Tetiana Fesenko

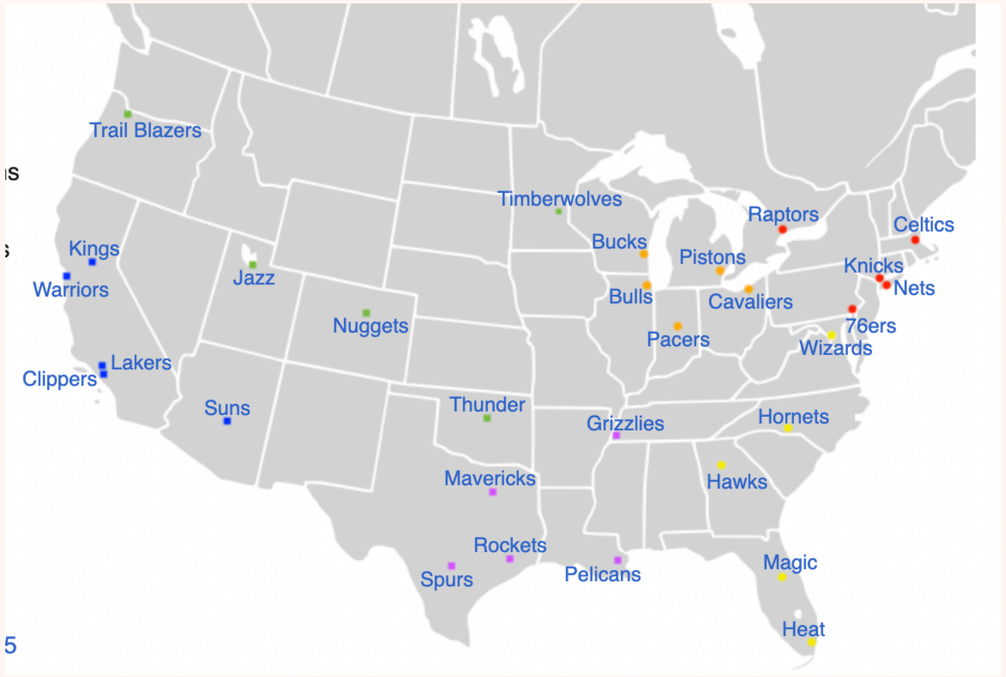
TORONTO, ON

GOALS

- Use BALLDONTLIE API data to build a machine learning model that will predict the outcome of NBA games.
- Complete Data Cleaning, Eda and Feature Engineering to select best prediction model
- 2 best ways to predict basketball game results:
  - Traditional team statistics
  - Advanced team statistics

NBA GAMES - FACTS

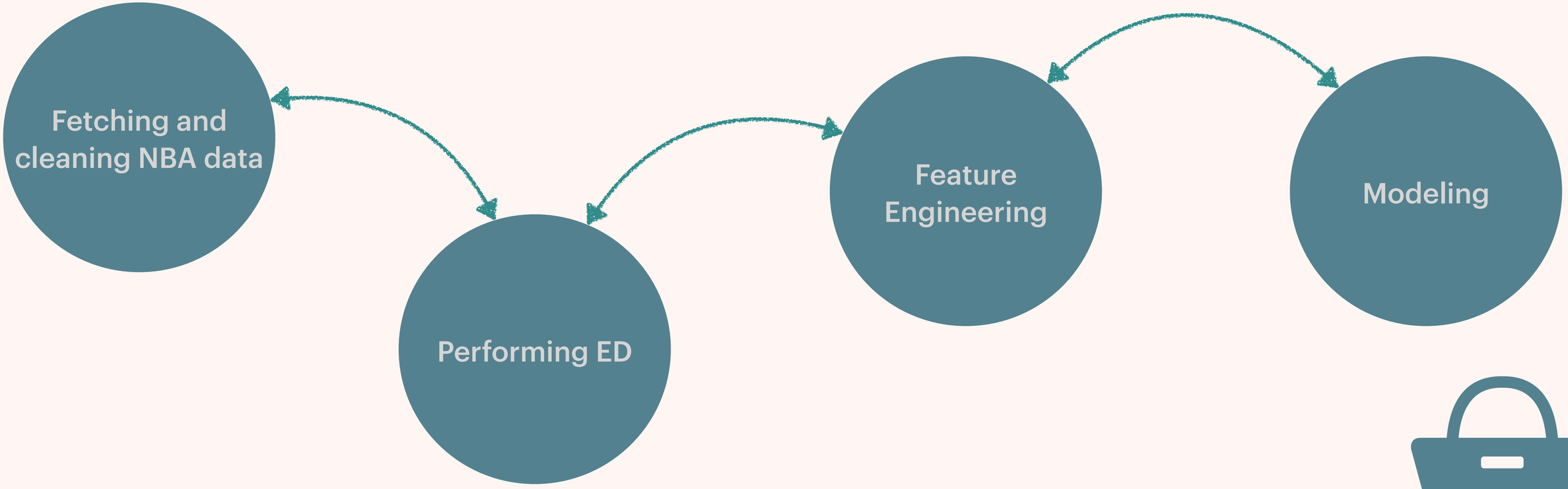
- **NBA - National Basketball Association: professional basketball league in North America. The league is composed of 30 teams (29 in the USA and 1 in Canada).**
- **The regular season runs from Oct to Apr with each team playing 82 games.**
- **Another area where NBA data became important is betting industry.**
- **The average annual spend per capita on NBA betting is $111.49.**



HYPOTESIS

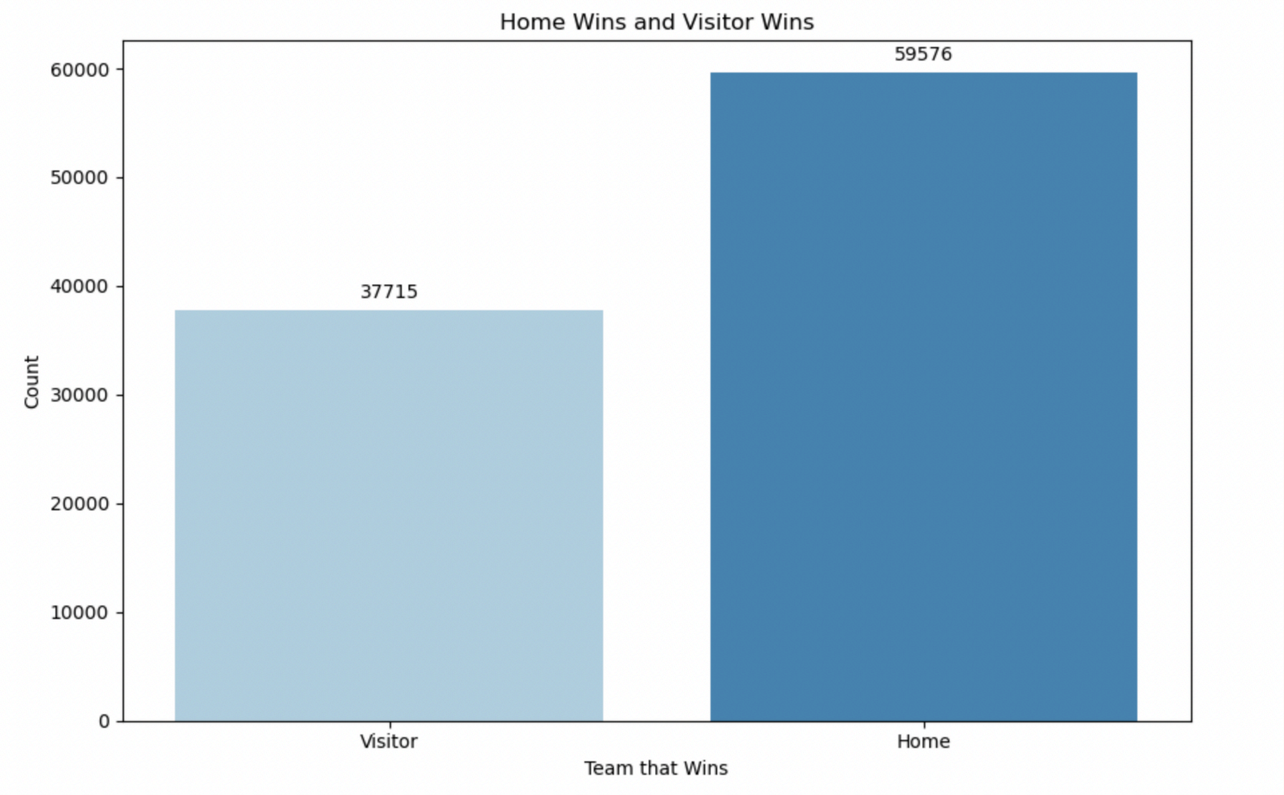
- **Possible outcomes before working with the data:**
  - **Teams that play at home are more to win the match**
  - **Teams that usually have higher scores will more likely to win**
  - **Advanced teams statistics will provide higher accuracy**

WORKFLOW


Fetching and cleaning NBA data
Performing ED
Feature Engineering
Modeling

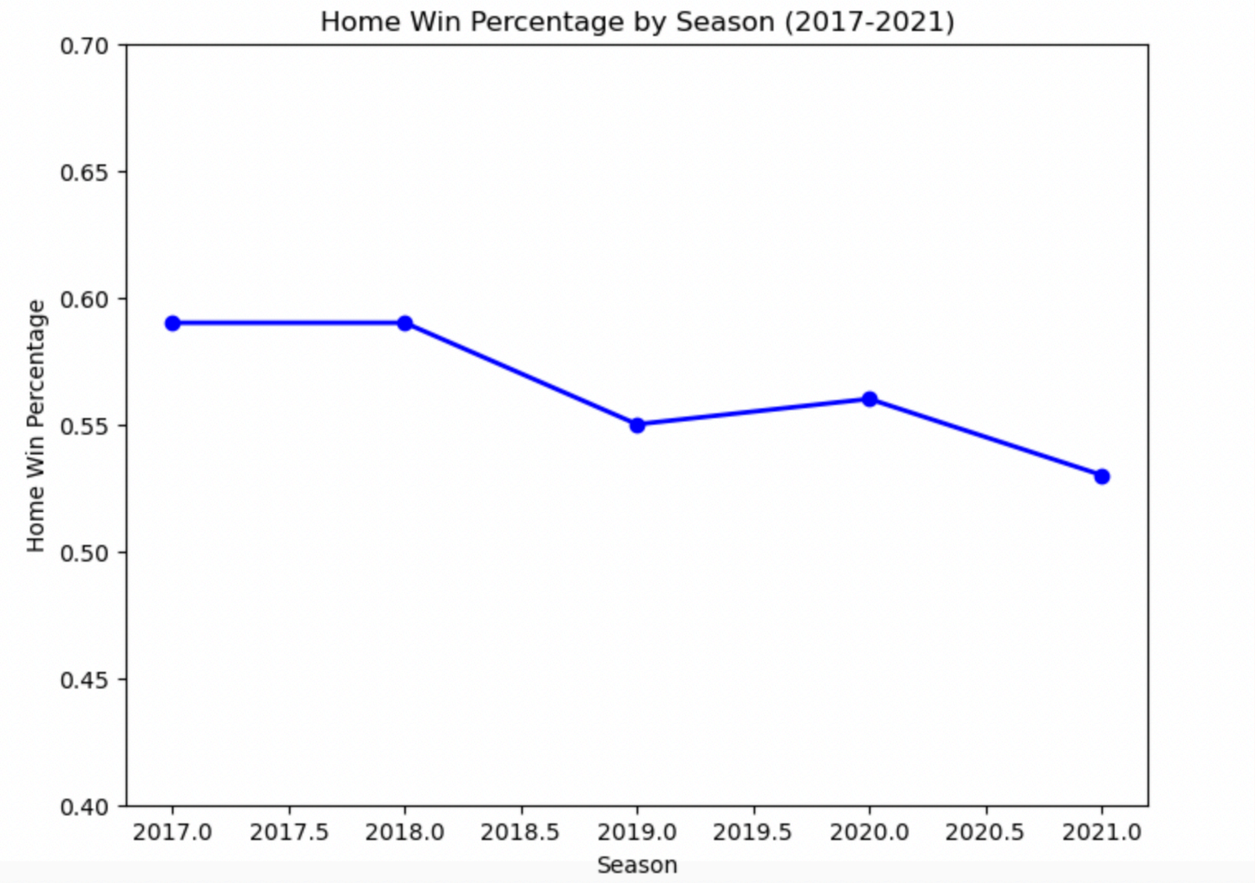
EXPLORATORY DATA ANALYSIS

- **Key findings:**
  - **Teams are more likely to win at home**
  - **Visitor have less changes to win**



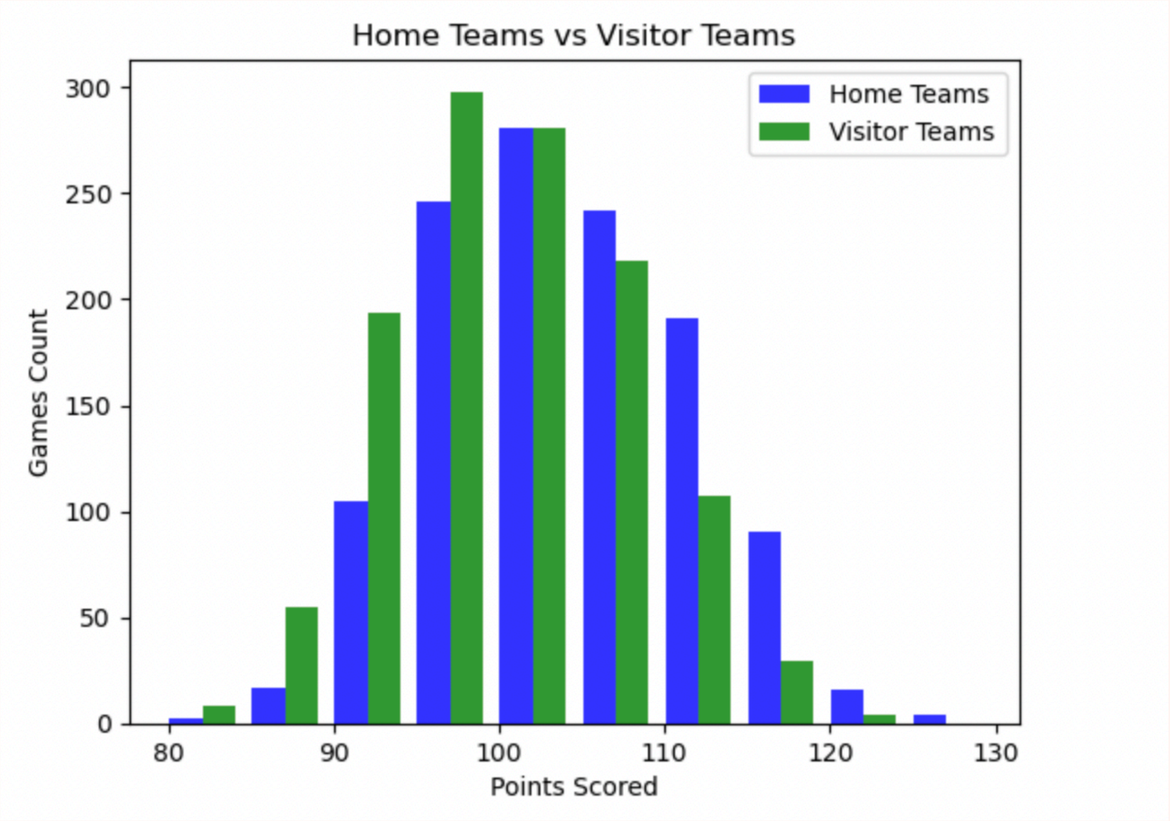
EXPLORATORY DATA ANALYSIS

- **Home Win % by season:**
  - **2017: 59%**
  - **2018: 59%**
  - **2019: 55%**
  - **2020: 56%**
  - **2021: 53%**
  - **I also compared the data to previous years and looks like 20-30 years ago this % was higher: 1988: 67%, 1998: 62%**



POINTS SCORED

- **Home teams tend to score more points per game**
- **Visitor Teams score less when playing away per game**



FEATURE ENGINEERING

- **For Feature Engineering we used over 20 different feature per team based on the traditional basketball statistics**
- **Examples are: Steals, Assists, Free Throws, Total Points scored and more**
- **Average from the previous 25 games to predict how the team will play in the future game**

Name	Description
Rk	Rank
G	Games
MP	Minutes Played
FG	Field Goals
FGA	Field Goal Attempts
FG%	Field Goal Percentage
3P	3-Point Field Goals
3PA	3-Point Field Goal Attempts
3P%	3-Point Field Goal Percentage
2P	2-Point Field Goals
2PA	2-point Field Goal Attempts
2P%	2-Point Field Goal Percentage

FT	Free Throws
FTA	Free Throw Attempts
FT%	Free Throw Percentage
ORB	Offensive Rebounds
DRB	Defensive Rebounds
TRB	Total Rebounds
AST	Assists
STL	Steals
BLK	Blocks
TOV	Turnovers
PF	Personal Fouls
PTS	Points

MODELING

- **Traditional team statistics:**
  - **Logistic Regression - 65%**
  - **Naive Bayes - 63%**
  - **XGBoost - 59%**
- **Advanced team statistics:**
  - **Logistic Regression - 67%**
  - **Naive Bayes - 64%**
  - **XGBoost - 61%**

- **Advanced team statistics - the scores are 2-3% higher than when we use traditional statistics**

CONCLUSION

- **Consideration for future:**
  - **Analyze ELO feature: measure relative strength of a player compared to other players**
  - **Test and analyze betting data for NBA games**

THANK YOU