PROJECT: Discovering Insights from NBA Free Throws Data

Eduardo Gil González-Madroño

Date: 09th November, 2020

Contents

- The Key Features of our DataSet
- Data Quality Level Score
- Relevant Insights from EDA
    - Data Cleansing
    - ML Model Development
- Further Analysis / source of data

The Key Features of the DataSet

- **Data Context:** The data set includes information of free throws performed in NBA in seasons spanning from 2006–2007 towards 2015–2016.

- **Data Content:** The data was scrapped from ESPN website (crawler not available) and the main figures are:
  - File: CSV of ~74MB size
  - Dimensions: 618019 observations (free throws performed) – 11 features

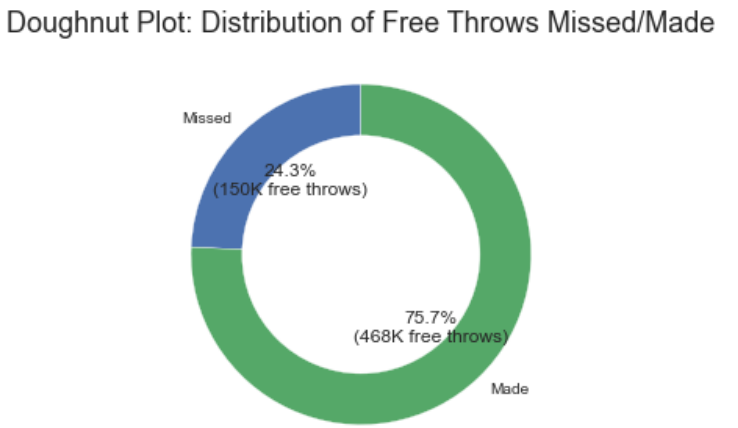
- Original Features of dataset:

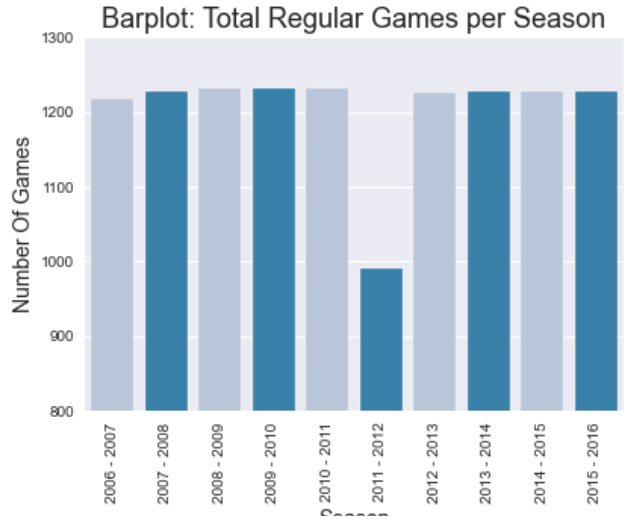
· end_result
· game
· game_id
· period
· play
· player
· playoffs
· score
· season
· shot_made
· time

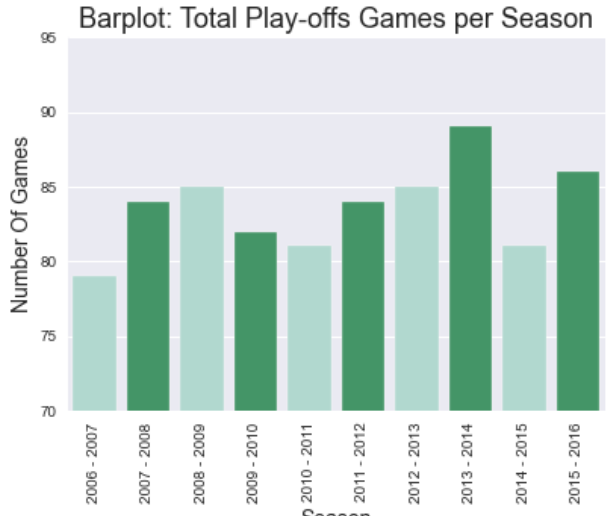
Data Quality Level Score

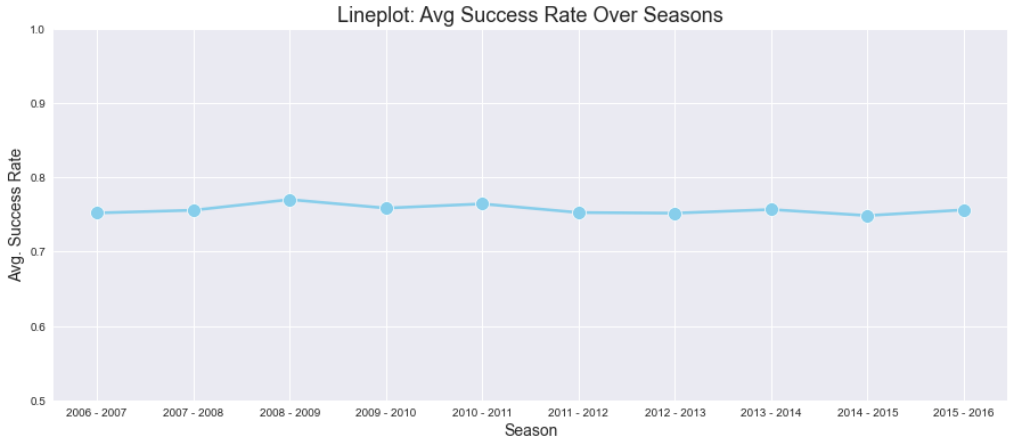
- **Data Missing:** No records were missing across the entire dataset
- **Data Duplicates:** Only 1 observation was duplicated across the entire dataset
- **Data Mispelled:** Only few typos present on feature "play" of the dataset. The rest of features, even being strings of numbers were consistent and features related/obtained from other columns were consistent (i.e. shot_made feature was extracted from the information given in "play")
- **Features Engineered:** Several features were created to simplify the EDA analysis:
  · home_team · player_team · time_lapsed
- **DATA QUALITY LEVEL:** Great. The dataset was greatly consistent and with valuable info.

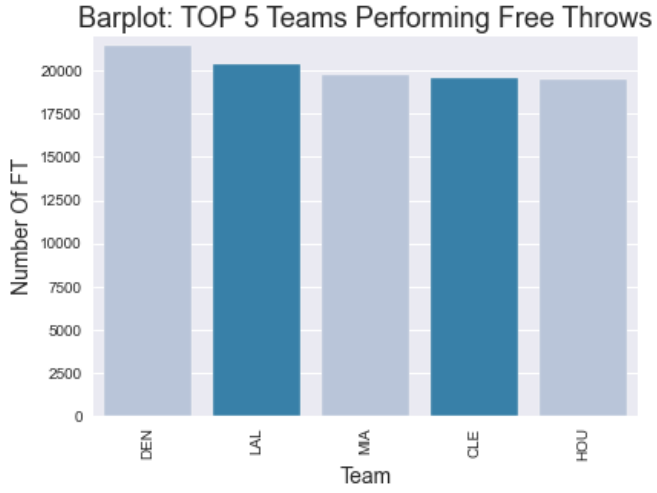
Exploratory Data Analysis: Data Balancing

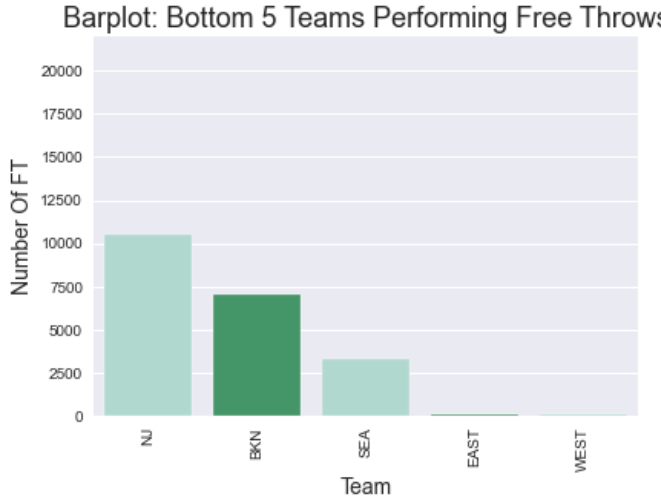




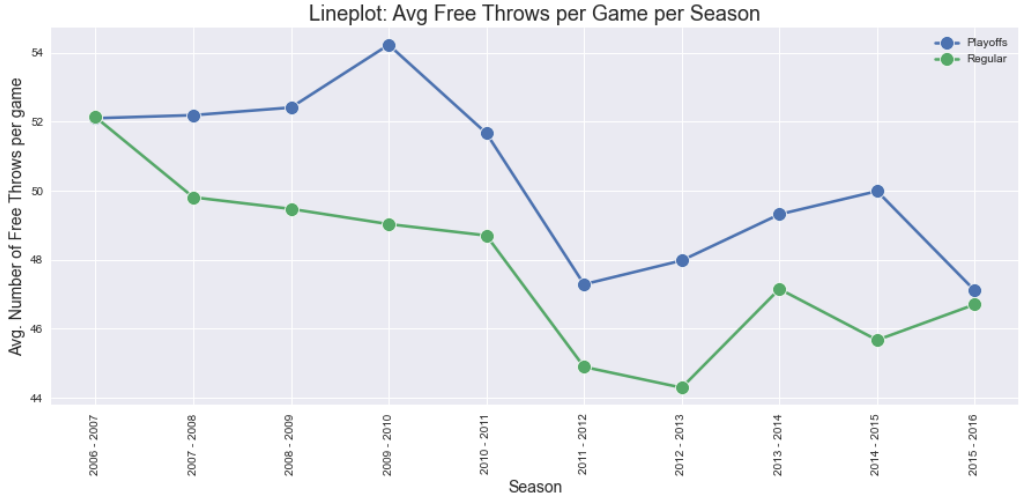


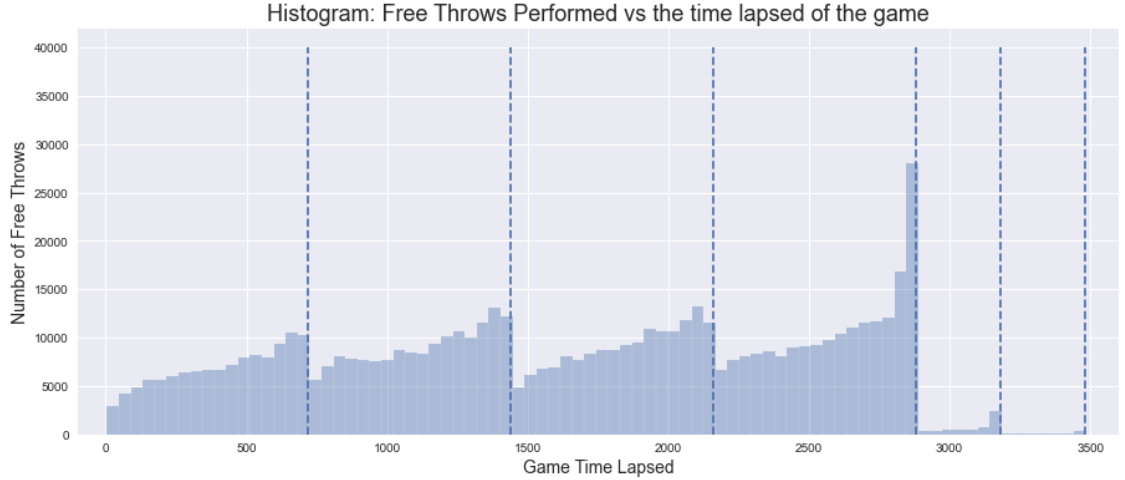


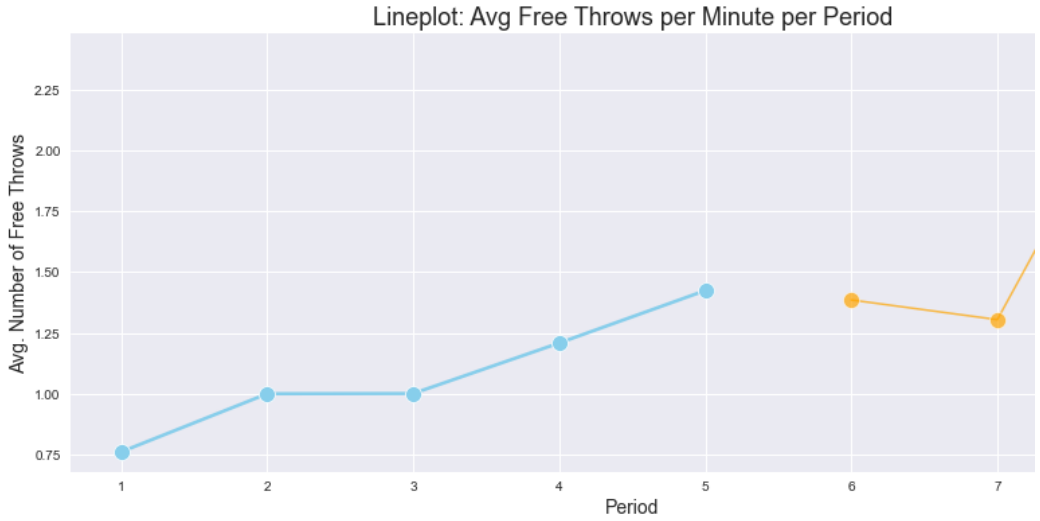


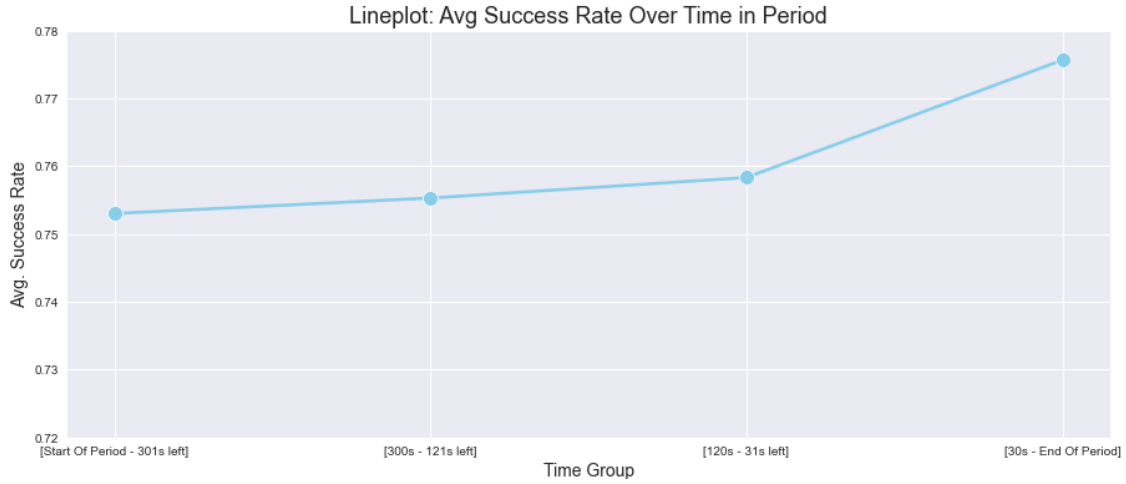


Exploratory Data Analysis: Free Throw Volume









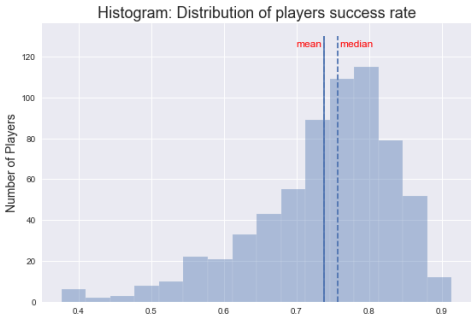
Exploratory Data Analysis: Success Rate

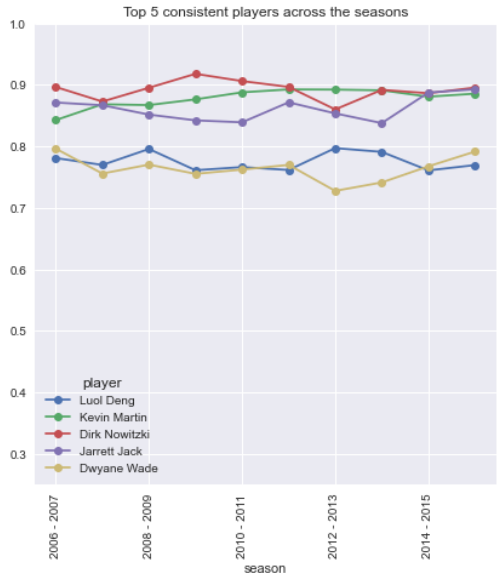
TOP5 Avg. Performers

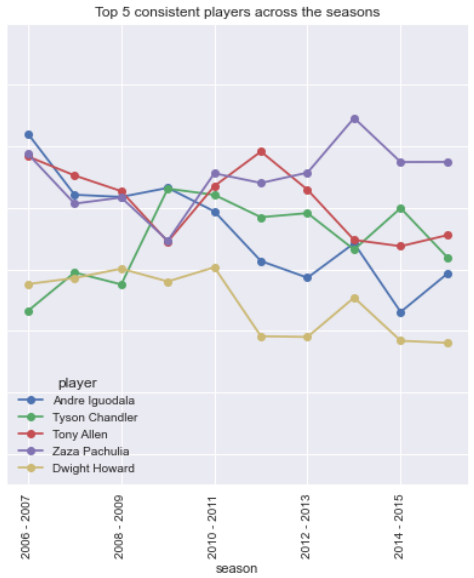
player	ft_count	success_rate
Steve Nash	1591	0.913891
Brian Roberts	337	0.910979
Ray Allen	2045	0.903178
Chauncey Billups	2793	0.901540
Peja Stojakovic	455	0.901099

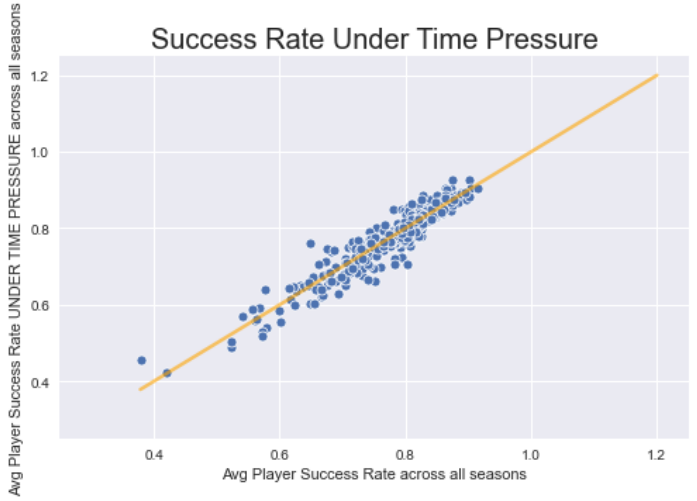
Bottom5 Avg. Performers

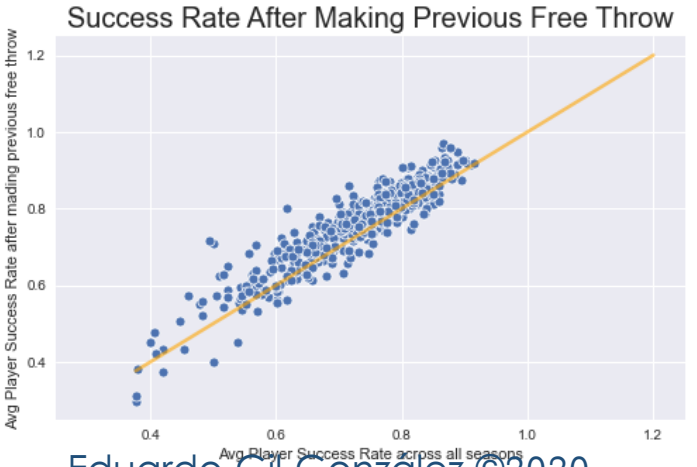
player	ft_count	success_rate
Joey Dorsey	162	0.376543
Clint Capela	268	0.376866
Andre Drummond	1459	0.378341
Kyrylo Fesenko	155	0.400000
Ben Wallace	837	0.406213











Further Analysis

- This dataset could be eventually a great tool to do predictions over which free throw will go in or not regarding several factors. Given that, further analysis would include the collection of further features of the shot such as:
  - Enviromental conditions (temperature oven the court, light intensity, noise level…)
  - Player Posture Conditions
  - Player Body Features (height, weight…)
  - Time where game was played
  - Timezone travelling
  - And many more

- Given this dataset is also a great tool to obtain insights of the success rates of players under certain conditions, I would have invest time discovering trends on:
  - Success rates consistency among players in the same team
  - Success rates consistency when player plays home or away
  - Success rates trends when pressure environment includes not only time but also home/away conditions and current game scoring (and its difference).

Q&A

Thank You

Eduardo Gil González-Madroño