NBA Player Analysis

A Data Science Capstone Project,
by Christos Magganas

Objective

The goal of this project is to create an accurate predictor for the metric RPM. Real-Plus-Minus is an unbiased comparative metric to evaluate player value.

Specifically, a player's estimated on-court impact on team performance, measured in net point differential per 100 offensive and defensive possessions.

In 2015, Stephen Curry had a real-plus-minus of 9.34

(on average his team would score 9.34 more points when he was playing compared to when he was out)

Data Collected to be analyzed

Data from ESPN and Basketball-Reference from the years 2014 - 2018 was collected in the manor as seen below

HTML Data
Website
(HTML)

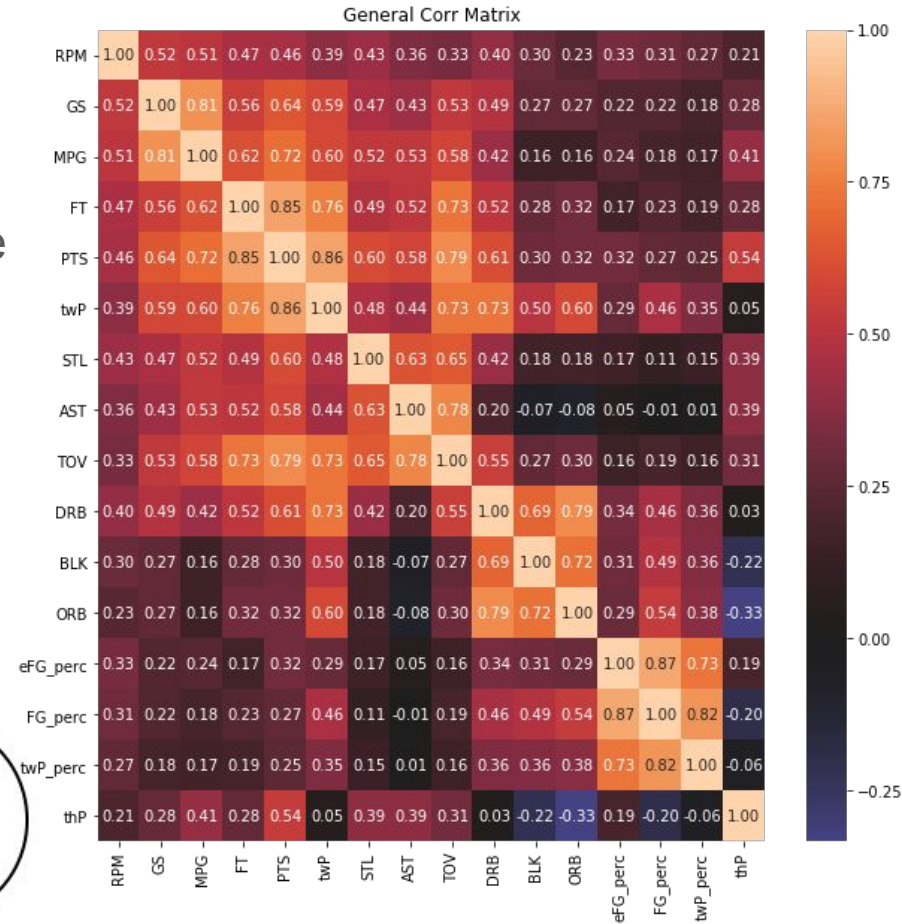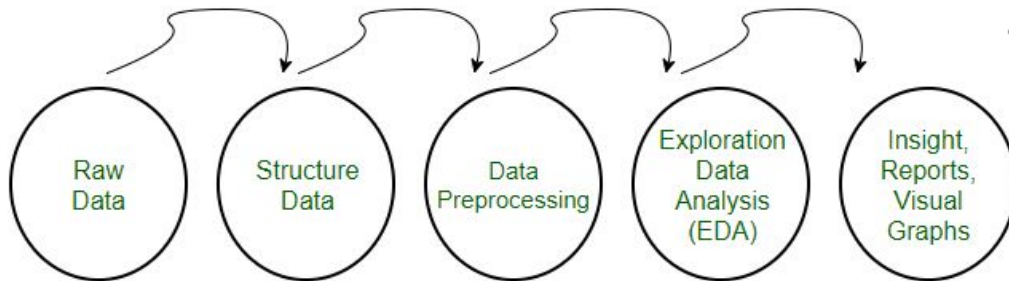Collect Data
Web Scraper
(Python, PHP, etc)

Saved Data
(JSON)

Parse Data
(Python)

Feature Selection

Determining which features have the strongest relationships to the target variable and each other is critical when building a model. Separating the correlated features from each other is important to understand how each component will determine prediction.

General Corr Matrix

Raw Data → Structure Data → Data Preprocessing → Exploration Data Analysis (EDA) → Insight, Reports, Visual Graphs

Applying Machine Learning Techniques
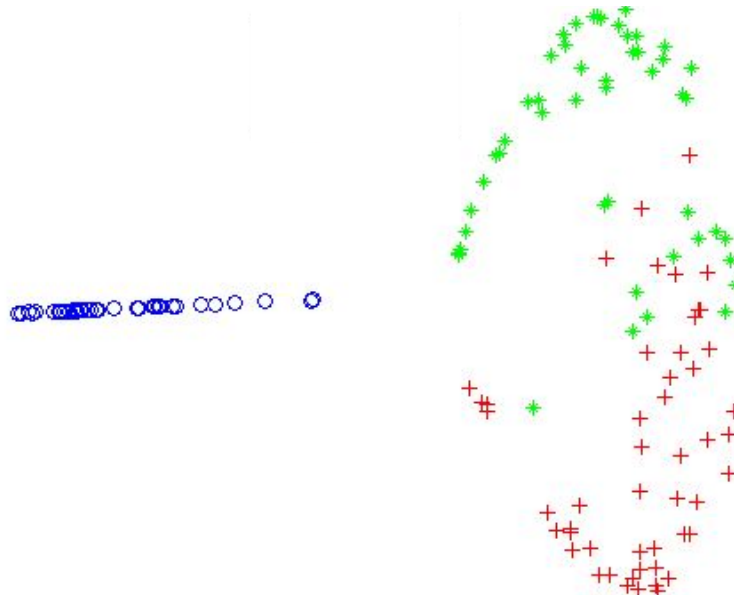
Building a model with regression:

- Linear, Ridge, Lasso, Decision Tree

Cross Validating and Testing:

- Ensuring model is best for unseen data

Partial Component Analysis & Reduction:

- Tuning parameters
- Reducing Model by using only essential components

Secret to Success

Although, to the best of our ability, we created a model that can predict the RPM of an individual player fairly accurately with only their individual stats (points, assists, etc.)

The best players have intangible qualities that are undervalued by the individual stats used to predict RPM.

We see the best players have been underrated because they contribute more to team than their individual stats would suggest.

Year	Player	Pos	TEAM	Age	RPM	predRPM	predRPM_resid
2016	LeBron James	SF	CLE	31	9.79	1.76	8.03
2015	LeBron James	SF	CLE	30	8.78	1.71	7.07
2014	LeBron James	PF	MIA	29	9.08	2.55	6.53
2017	LeBron James	SF	CLE	32	8.42	2.03	6.39
2015	Stephen Curry	PG	GS	26	9.34	3.36	5.98
2016	Stephen Curry	PG	GS	27	8.51	2.97	5.54