

Introduction:

The following is a general data science project for which I will be posting updates indefinitely. I am an avid basketball fan, and as a scientist and analyst by education I was intrigued by the increasing use of statistics and metrics **beyond counting totals and averages**. At a glance, some metrics are quite clear and straightforward, shooting percentages, points per game averages etc. while others are more obscure, subtle or cryptic even.

In this project the outline as follows:

1. Gather data using Python, VBA, NBA Stats APIs, and any other tools or sources deemed useful or necessary.
2. Store and organize the data in a SQL database (PostgreSQL and SQLite)
3. Design a series of mini projects centered around the assessment of various analytic tools and metrics used to interpret the sport as it is played and rate performance. It is also likely that I will be able create new metrics, or new versions of accepted metrics for either a fresh take or further investigative analysis.

Background and Interest Lead:

For most of the history of the NBA fans simply watched the games to support their favorite teams and players in their goal to **win games, awards or break standing records**.

In the last 15-20 years some advanced statistics and metrics in the background, and in a few cases previously nonexistent, have become fundamental measurement tools for teams, players and franchises.

A few of these advanced metrics, plus/minus (+/-), Player Efficiency Rating (PER) and True Shooting Percentage (TS%) among others eventually landed in the foreground as advanced statistics.

Despite this there has been mixed response from fans, notably several current and former players who criticized advanced metrics as a disruption to the natural evolution of the game as it is played.

By habit, I look over historical and current NBA statistics to gauge the evolution of the sport, comparing trends in stats such as 3pt shooting percentage, shooting volume etc.

One group of stats in particular, Win Shares, **intrigued me initially as one with the potential to wholly defining a player's value**. After looking over some win share data for current and former standout players I thought to the limitation of such a measure in a game so intrinsically dynamic and ever changing.

How close could one come to deriving a single metric that represents a player's contribution to their team? What is the shape of this idea? How reductive or imprecisely weighted might it be for the game as it is played in a particular era? What might it lack? What story does this statistic tell?

This last question is certainly substantial in the scope of the NBA and the association we have with each player. Where can metrics be derived that tell the story of a player's career? Where do trusted metrics fail to account for a player's abilities and contribution to their team and the game itself?

First Analysis Project: Win Shares

This document serves as an introduction to the full scope of the project. Documents specific to this phase of the project will be added

My first goal for analysis will be to formally investigate the statistics group "Win Shares." I will investigate at least 3(?) commonly used variations of each WS formula and assess them for bias, imprecisions in the weight of certain factors, and overall usefulness towards quantifying a player's contribution.

The three formulas for win shares are as follows:

- 1.

*** At a glance, or as a hypothesis there seems to be a bias in fav of players who play on winning teams. This shades the metric away from a complete measure of a players contribution, and towards a much more conditional or reduced measure. A great player on a team with fewer wins is penalized, disallowing for a more general measure for contribution. ***

Second Analysis Project: Win Shares Reformulation (contingent upon whether or not the assessment is made that there is another interesting way to derive a measure of this type)

Data Gathering/Wrangling

1. API calls _____ with Python 3

- gathering player names and ids

** searches and Web scraping for complete list of Active and Historical Players, Player IDs for the website: sportsreference.com_ (script 1a)

- Player Data with script 1b