Introduction

The following is a data science project for which I will be posting updates indefinitely. I, like many, have become intrigued by the increased use of statistics and metrics in the NBA over the past 15+ years. Casual fans and analysts have gone from tracking simple averages such as points, rebounds and assists per game to more complex, and sometimes cryptic formulas like Player Efficiency Rating (PER) and Win Shares (WS).

Naturally, a debate emerged which questioned the influence of advanced metrics and mathematical models on players, coaches, organizations and casual fans. There has been mixed response from fans and analysts. Some embrace the changes and influence while others criticize advanced metrics as a disruption to the natural evolution of the game.

Here, I will explore changes in the NBA using game data as well as some of the more obscure advanced stats and metrics to assess their usefulness as well as limitations. For this project the general outline is as follows:

---

1. **Gather data** via Python, Excel VBA, APIs, and any 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** to assess or create various analytic tools and metrics used to interpret the sport and rate performance with Python, R, SQL, Tableau and Excel VBA.

Background and Interest Lead

By habit I look over current and historical NBA statistics to check trends in the sport. A simple example would be comparing trends in '3-point shooting %' and '3-point shooting volume' across decades for teams and players.

In recent years, one group of stats in particular, 'Win Shares', intrigued me as having the potential to wholly defining a player's value. A summary, when searched for, implies that it represents the normalized amount of contribution to winning a player has logged.

However, in my first attempts to examine the formula, I found 'Win Shares' to be obscured by highly dependent percentage variables and subjective choices for several metric components. The formula seemed to favor players on winning teams. It also seemed to arbitrarily penalize players with less usage.

There are layers of percentage-based calculations which are both interdependent and highly conditional, rendering 'Win Shares' something I will refer to as a "meta statistic." In fact, there are several versions of 'Win Shares' formulas.

Background Conclusion

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? There seems to exist an almost proverbial give and take between precision and reduction. What is the shape of this idea? Can metrics and statistics tell a story? What story would this statistic tell?

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

In my experience as a thinker who once was and, still is in many ways a student mathematics and physics, the ideated ends of many interests often rest in the philosophy of life itself. It is the journey, the path taken, the story, the imprint across time that matters.

I look forward to the journey of this project; a lead from one of life's passions, sports and sports fandom, through to the essence of the passion and life itself - discovery.

The first analysis project will be a dive into 'Win Shares' as a family of statistics. Subsequent projects will be logged accordingly.

Please follow along using the *Table of Contents* document!