Charity Smith

Email: smit689@regis.edu

Phone: 303.517.6085

MSDS69 Data Science Practicum II Project Proposal

**Title:**

Too Many Threes? A Data-Driven NBA Debate

**High Level Description and Key Questions:**

The NBA has changed dramatically over the past decade, with teams relying on three-point shooting more than ever, but has this trend made the game better, or is it hurting competitiveness and entertainment value? As Heck points out in USA Today (2024), "On average, NBA teams are taking 37 3-pointers per game this season... an increase of 10 from just eight years ago and 15 more threes than teams were taking 10 years ago."

In this project, I want to consider the following key questions:

* Are there more blowouts and fewer close games because of the three-point explosion?
* Has fan interest dropped as teams shoot more threes?
* Do teams that rely on three-pointers actually win more?

In this project, I will analyze historical NBA data, team performance, game outcomes, and fan sentiment to see if the three-point revolution has improved or hurt the league. Using machine learning, time series forecasting, and sentiment analysis, I’ll take a data-driven approach to one of the NBA’s current debates.

**Data Science Tasks:**

This project will use data visualization, statistical modeling, and machine learning to analyze the NBA’s shift toward three-point shooting. I will examine historical trends, predict future patterns, classify teams based on playstyle, and explore relationships between three-point attempts, competitiveness, and fan engagement. Additionally, I will analyze social media sentiment to understand how public perception of the three-point era has evolved.

**Data:**

This project will require a mix of historical NBA statistics, TV ratings data, and fan sentiment analysis to evaluate whether three-point shooting has had a positive or negative impact on the game.

* NBA Statistics: Three-point attempts per game, mid-range shot frequency, shooting percentages, game outcomes (score differentials, blowouts, overtime games, competitiveness), and player shot selection trends; sources: Basketball Reference, NBA statistics
* TV Ratings & Fan Engagement: Yearly NBA Finals, regular-season, and playoff ratings, attendance data and ticket sales over time; sources: Sports Business Journal, ESPN, NBA revenue reports
* Social Media Sentiment Data: Social media and Reddit discussions on three-point shooting, game excitement, and competitiveness, text analysis of fan opinions on whether the game has become less entertaining; sources: Reddit, YouTube comments on NBA debates, other social media platforms.

I expect the structured data (NBA stats, TV ratings) to be a smaller dataset, while the social media data could be 500MB+ depending on the volume of scraped text. Data collection and preprocessing should take 1-2 weeks.

**Data Analysis:**

I will analyze the data using a combination of statistical modeling, machine learning techniques, and data visualization to uncover trends and relationships.

* Data Visualization – I will create visual representations of key trends, such as changes in three-point shooting over time, team strategies, and game outcomes.
* Time Series Analysis – I will examine historical data to identify patterns and predict future trends in shooting tendencies, game competitiveness, and fan engagement.
* Regression Analysis – I will explore whether certain variables, such as three-point attempts, have a measurable impact on other factors like game margins or TV ratings.
* Clustering and Classification – I will group teams based on different playing styles to determine if reliance on three-pointers affects overall success.
* Sentiment Analysis – I will analyze public discussions on social media to gauge how fan opinions on three-point shooting have evolved over time.

**Anticipated Challenges & Solutions:**

One challenge will be finding reliable TV ratings and fan engagement data. To address this, I’ll use sources like Sports Business Journal and Google Trends for audience interest. Another challenge is defining "competitiveness.” I’ll quantify this by analyzing point differentials, close-game frequency, and playoff upsets. Additionally, social media data can be noisy, full of sarcasm and irrelevant discussions, so I’ll clean it using NLP preprocessing techniques to extract meaningful insights.

**Timeline:**

Below is a brief timeline with week-by-week deliverables:

|  |  |
| --- | --- |
| **Week** | **Deliverables** |
| Week 1 | Project proposal and research plan. |
| Week 2 | Data collection: begin web scaping and find relevant data sources. Clean and preprocess data (handle missing values, structure datasets). Start exploratory data analysis (EDA). |
| Week 3 | Continue EDA, create initial visualizations of three-point trends. |
| Week 4 | Build first machine learning models (regression for TV ratings, clustering for team styles). |
| Week 5 | Expand machine learning analysis: Time series forecasting for future three-point trends, sentiment analysis of fan opinions. |
| Week 6 | Complete data analysis and create visualizations. |
| Week 7 | Finalize analysis and visualizations; start project summary and presentation. |
| Week 8 | Final review, polish visualizations, record video presentation, and submit project. |

**GitHub Repository:**

<https://github.com/charityasmith/MSDS696_Practicum2_Project>

**References**

Heck, J. (2024, October 30). NBA teams are on pace to set another three-point record, but is that a problem? USA Today. <https://ftw.usatoday.com/story/sports/nba/2024/10/30/nba-three-pointers-record-attempts-problem/75948024007/>