Pace and Efficiency: The Modern NBA Revolution

Karanbir Singh Grewal
Ameya Lambat
Aidan Dubel

Project Pitch

Problem: How has NBA playing style fundamentally changed in terms of pace and efficiency over 75+ years?

Approach: Mine historical NBA game data (64,000+ games, 1946-present) to quantify and predict the evolution of basketball strategy

Goal: Build a data mining system that:

- Identifies distinct playing style eras
- Predicts team success based on pace/efficiency metrics
- Quantifies the "three-point revolution"

Impact: Provides coaches, analysts, and teams with data-driven insights into optimal playing strategies across different eras

Data Exploration & Key Patterns

Dataset:

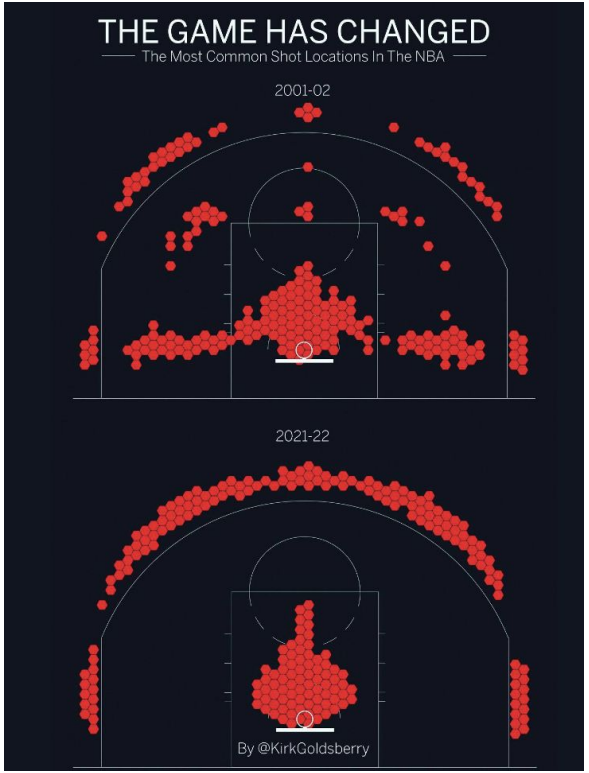
- **Source:** Kaggle NBA historical dataset (1946–present)
- **Size:** ~64,000 games, 5,000+ players
- **Key Features:** FG%, 3PA, REB, AST, TOV, PTS, Game Date

Visualizations Created:

- **Pace over time:** U-shaped trend (fast 1960s to slow 1990s to fast 2020s)
- **3-Point revolution:** Exponential growth post-2014
- **Efficiency vs Pace:** Modern teams cluster in high-pace, high-efficiency quadrant

Discovered Patterns:

- **Clear inflection point:** 2014-15 season
- Correlation between pace and efficiency reversed (negative to positive)
- Four distinct playing style clusters emerge
- Steady increase in 3-point attempts since 1980
- Assist-to-turnover ratio improving over time to smarter offense



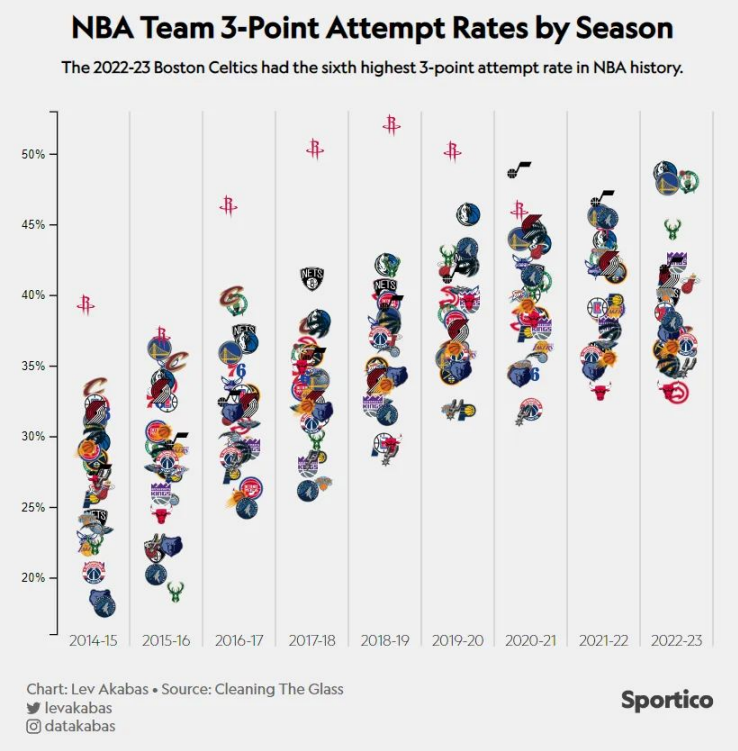
Preliminary Pipeline & Results

Pipeline:

1. Data preprocessing: cleaned 64K games, calculated pace/efficiency metrics
2. Feature engineering: possessions/48min, points/100 possessions, shot distribution
3. Baseline models: Logistic Regression, Linear Regression, K-Means

Experimental Results:

- **Pace Classification:** 78% accuracy (fast vs slow teams)
- **Win% Prediction:** $R^2 = 0.62$ (offensive efficiency most predictive)
- **Style Clustering:** 4 clusters identified (fast-efficient, slow-efficient, etc.)



Issues & Future Plans

Identified Issues:

- Missing data pre-1980 (~25% of advanced stats)
- Rule changes create non-stationary data (3-point line added 1979)
- Pace calculations require possession estimates (not always available)
- Player injuries

Improvements for Checkpoint 2:

- Implement ensemble models (Random Forest, XGBoost)
- Era-specific feature engineering and normalization
- Change point detection algorithms
- Time series forecasting for trend prediction
- Interactive visualization dashboard

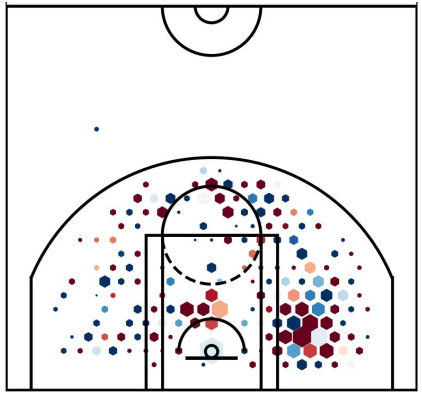
Conclusion

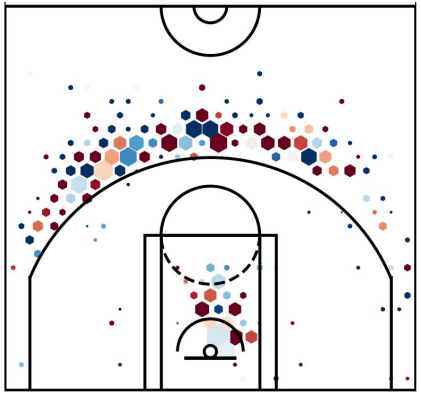
Current Progress:

1. Topic selected: Pace & Efficiency evolution in NBA
2. Dataset identified: Kaggle NBA Database (64K+ games, 1946-2025)
3. Initial data exploration completed
4. Pipeline framework designed

Next Steps for Checkpoint 2:

1. Implement complete preprocessing pipeline
2. Build and train baseline models (Logistic Regression, K-Means)
3. Generate comprehensive visualizations
4. Evaluate results and iterate on approach





References

- Datasets
  - https://www.kaggle.com/datasets/wyattowalsh/basketball
  - https://www.kaggle.com/datasets/sumitrodatta/nba-aba-baa-stats
- NBA Official Stats
  - https://www.nba.com/
- Additional Stats
  - https://www.statmuse.com/