# Statistical Exploration of the NBA

An Analytical Approach to Understanding Basketball Dynamics and the player revenue model


TEAM 6

Divyanshu Ahuja Huy Dinh Tejas Athreya Vaishnavi Kadale Akshita Gajengi


# Understanding Game Dynamics

1

## Number of Players

5 players on the court at a time during gameplay. NBA teams can have up to 12 players on their roster.

2

## Positions

Positions include

- Point Guard (PG)
- Shooting Guard (SG)
- Small Forward (SF)
- Power Forward (PF)
- and Center (C)

each with specific roles.

3

## Scoring

Scoring methods in basketball include Two-Point and Three-Point Field Goals, Free Throws, each contributing to a team's score.


# *Business Overview*

## Optimizing Player Investment for Enhanced Team Performance

To enhance team performance within financial limits, basketball clubs need a data-driven strategy for accurately optimizing investment in talent

**Objectives**

Develop a data-driven strategy to:

1. Align player salaries with performance.
2. Identify undervalued talent for cost-effective team enhancements.
3. Analyzing player’s salaries across timelines.

# *Data Sources*

| | |
|---|---|
| NBA Players Stats | This dataset comprises per-game statistics for NBA players spanning from the year 2000 to 2024. |
| | |
| NBA Salary Stats | This dataset contains salary information for NBA players for the same period as the player stats dataset |

# *Data Quality and Preparation*

| | |
|---|---|
| **Initial Dataset Size:** | Both datasets combined initially contained approximately 5,000 rows, covering multiple seasons and a wide range of players. |
| **Data Merge:** | To integrate performance with financial data, we merged the two datasets using 'Player', 'Year' and ‘Salary’ as key indices. |
| **Column Pruning:** | Retained only performance points and salary data. All other columns were dropped to focus our analysis on the core aspects of player performance and financial investment. |
| **Data Cleaning:** | Ensured accuracy in the merged dataset by correcting mismatches and filling in missing values where necessary, enhancing the reliability of our subsequent analyses. |

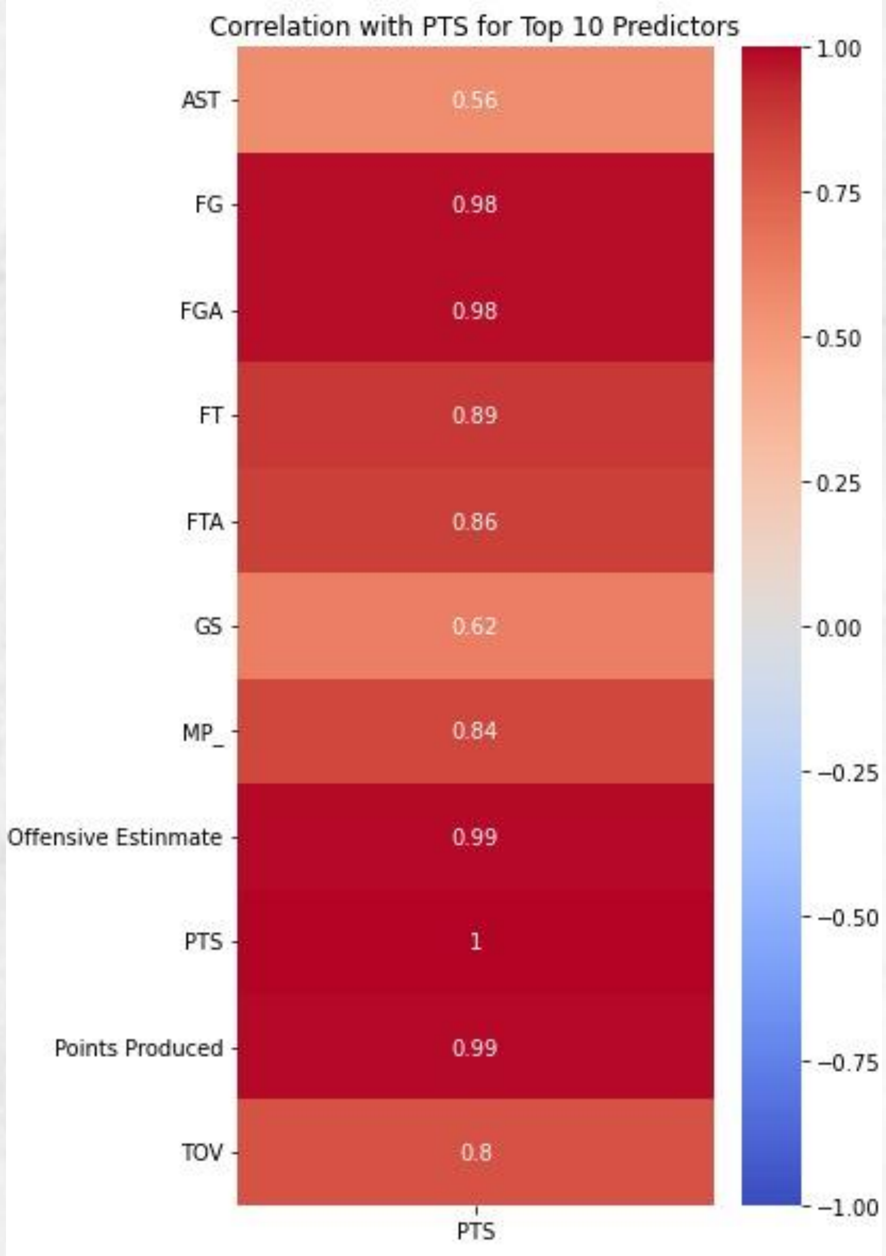
*Enhanced Metrics for Analysis: Introduced 'Calculation' columns for nuanced insights.*

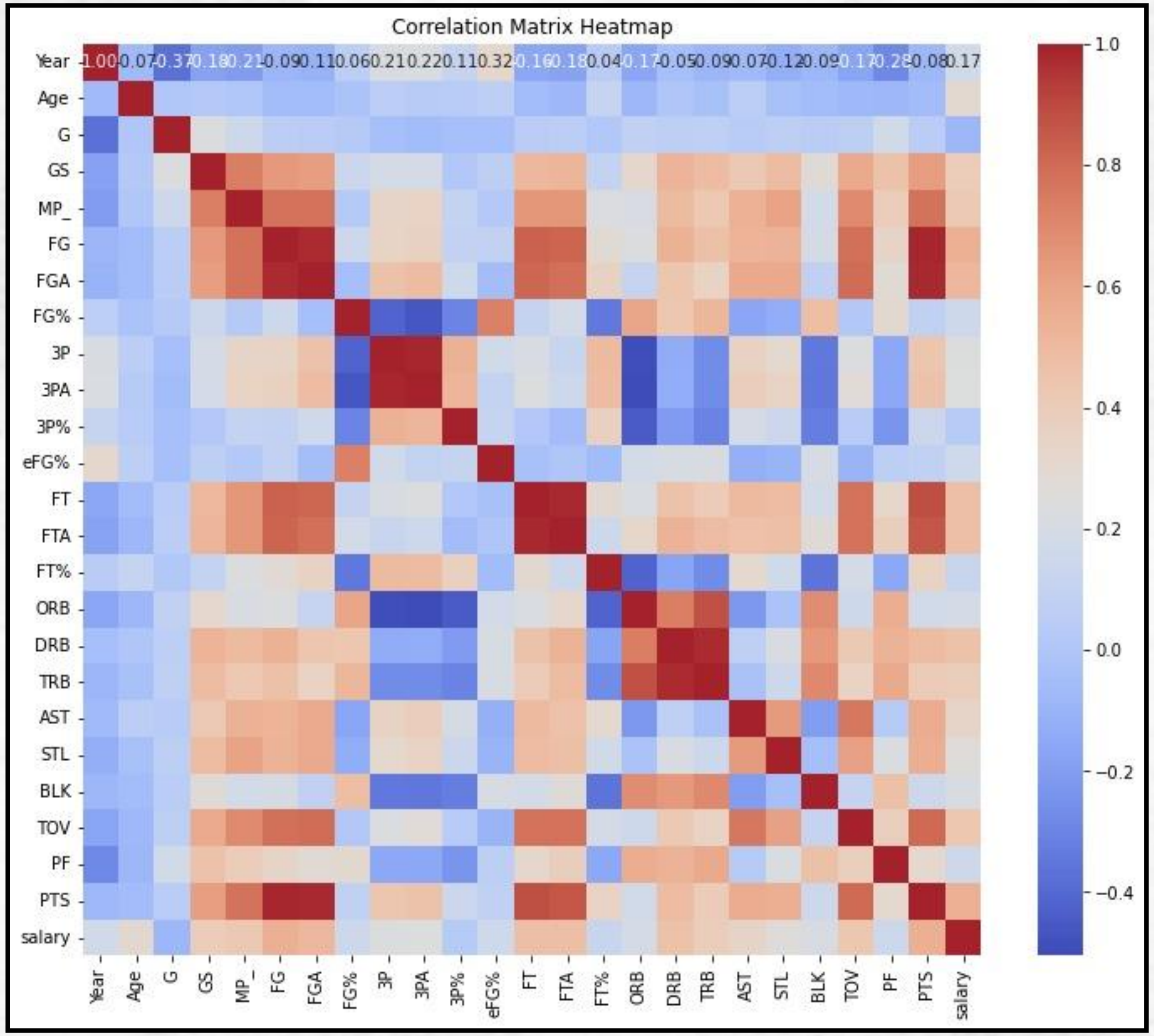
| | |
|---|---|
| **Awards & Achievements:** | Tracked accolades like MVP, DPOY, and Champion status. |
| **Performance Indicators:** | Calculated Points Produced, AST/TOV ratio, Offensive, and Defensive Estimates... |
| **Formulae Application:** | Applied precise formulae to estimate each player's contribution to the game. |

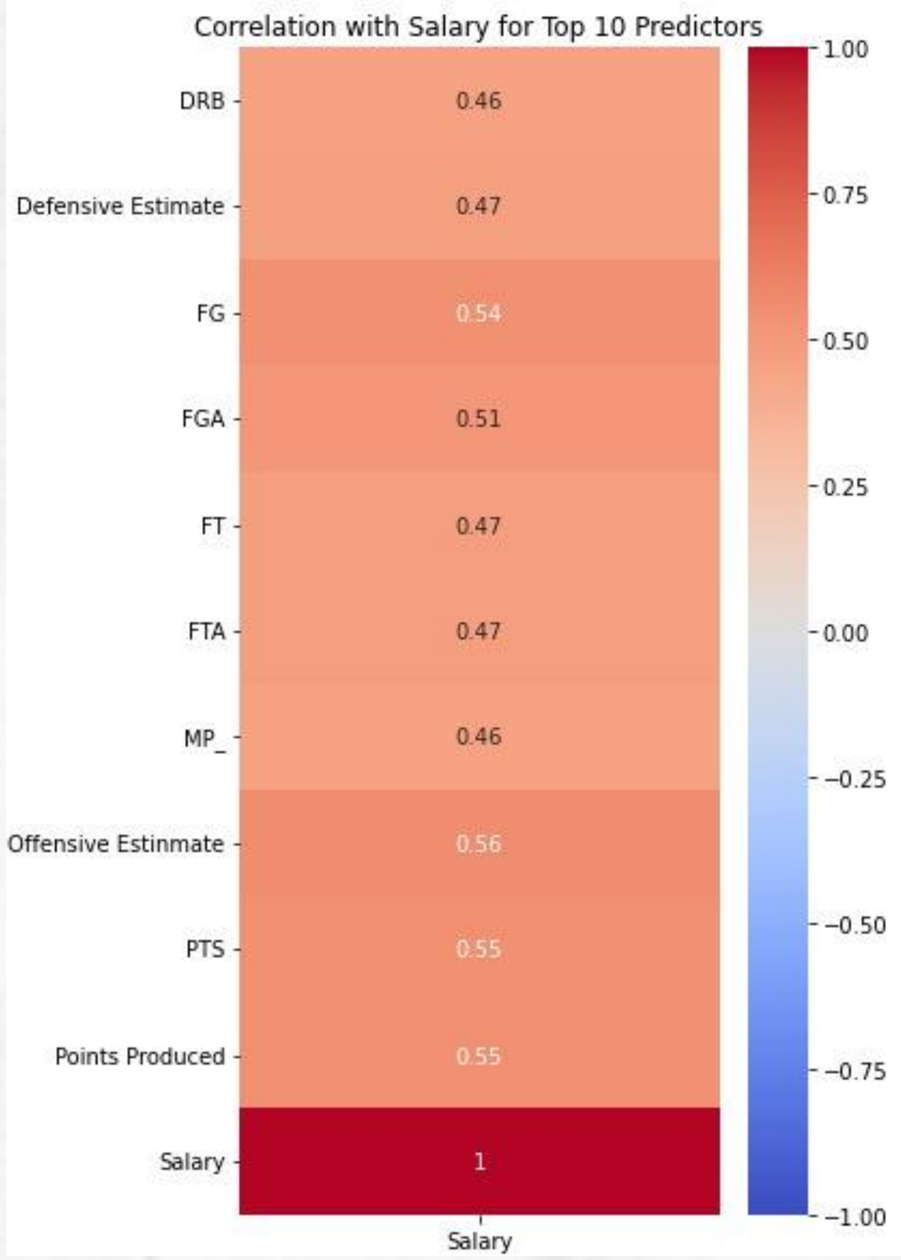
```
# ==========Our Defensive Players Each Year==========

       Year                Player   Defensive Estimate
29     2000      Dikembe Mutombo                 8.090
258    2001           Ben Wallace                 8.104
431    2002           Ben Wallace                 9.692
588    2003           Ben Wallace                10.190
793    2004           Ben Wallace                 9.080
986    2005        Kevin Garnett                 8.025
1168   2006         Shawn Marion                 7.908
1410   2007         Marcus Camby                 8.831
1598   2008         Marcus Camby                 9.452
1797   2009        Dwight Howard                 7.984
2003   2010        Dwight Howard                 7.765
2172   2011        Dwight Howard                 8.323
2366   2012        Dwight Howard                 8.714
2632   2013           Tim Duncan                 7.317
2787   2014       DeAndre Jordan                 7.617
3019   2015        Anthony Davis                 8.161
3366   2016     Hassan Whiteside                 7.978
3560   2017        Anthony Davis                 8.147
3831   2018        Anthony Davis                 8.356
4112   2019       Andre Drummond                 8.024
4378   2020       Andre Drummond                 8.470
4663   2021          Rudy Gobert                 8.023
4886   2022          Rudy Gobert                 7.872
5102   2023         Nikola Jokic                 6.460
5395   2024    Victor Wembanyama                 8.732
```

```
# ========ACTUAL Defensive Players Each Year========

       Year                   Player   Defensive Estimate
48     2000           Alonzo Mourning                6.374
257    2001          Dikembe Mutombo                7.202
431    2002              Ben Wallace                9.692
588    2003              Ben Wallace               10.190
1005   2005              Ben Wallace                7.806
1214   2006              Ben Wallace                7.840
1410   2007             Marcus Camby                8.831
1613   2008            Kevin Garnett                6.123
1797   2009            Dwight Howard                7.984
2003   2010            Dwight Howard                7.765
2172   2011            Dwight Howard                8.323
2403   2012           Tyson Chandler                4.825
2586   2013               Marc Gasol                4.617
2779   2014              Joakim Noah                5.931
3065   2015            Kawhi Leonard                6.155
3319   2016            Kawhi Leonard                5.573
3599   2017           Draymond Green                6.354
4128   2019              Rudy Gobert                7.014
4411   2020    Giannis Antetokounmpo                7.206
4663   2021              Rudy Gobert                8.023
4884   2022             Marcus Smart                3.258
5173   2023        Jaren Jackson Jr.                5.466
```

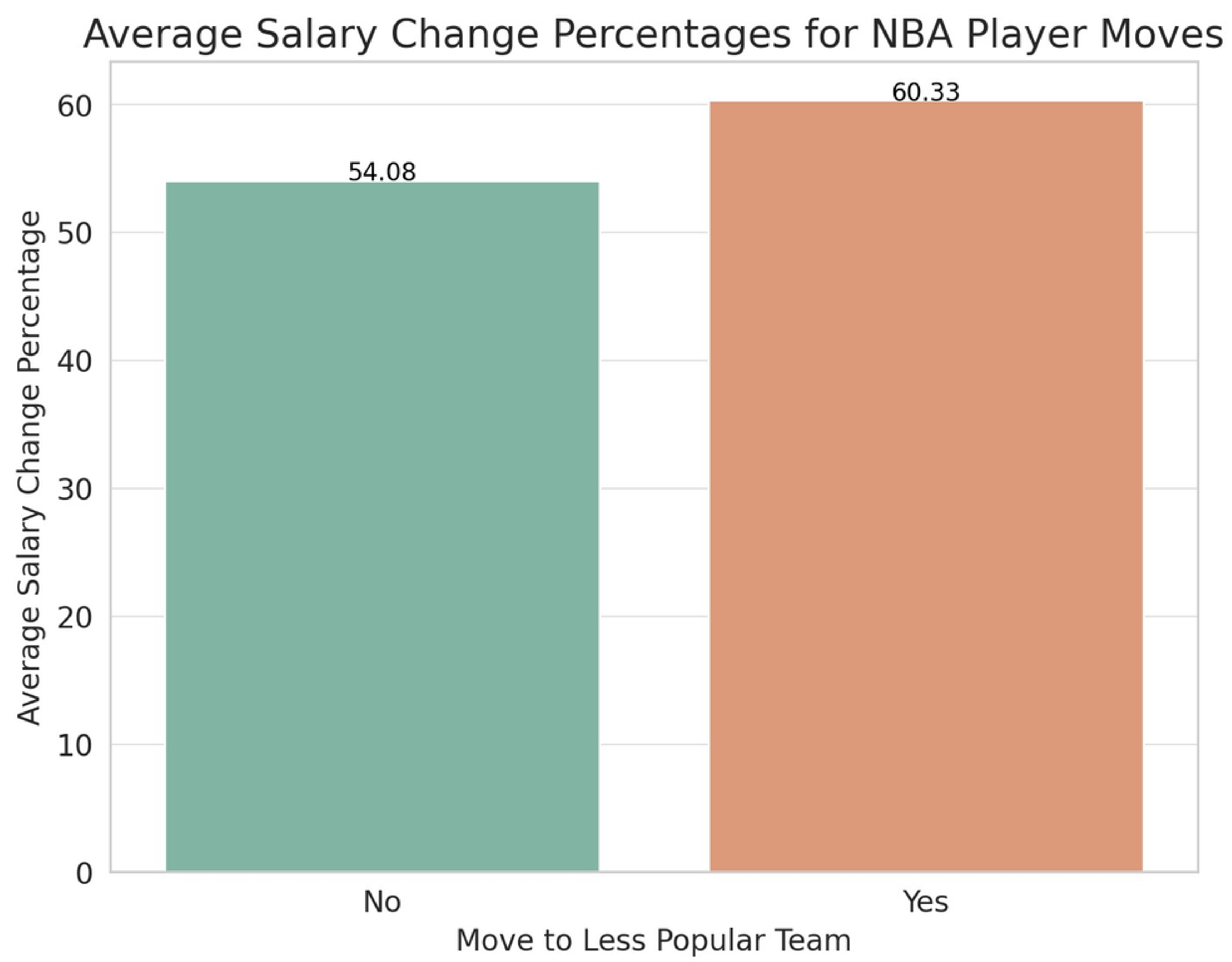
# Exploratory Data Analysis

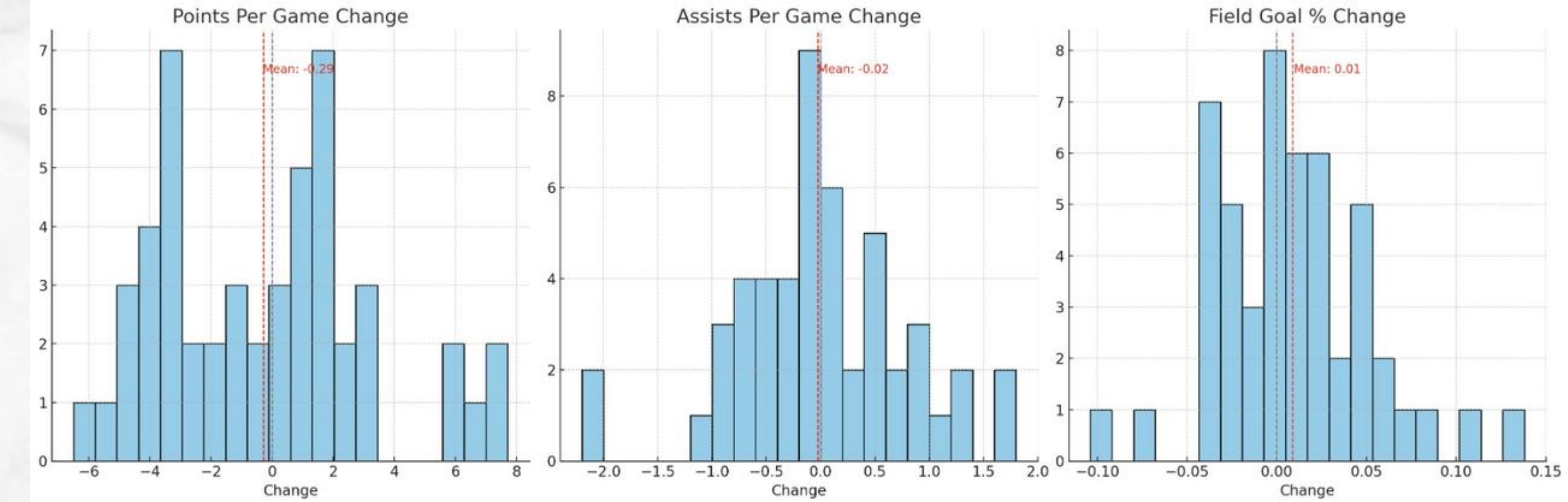


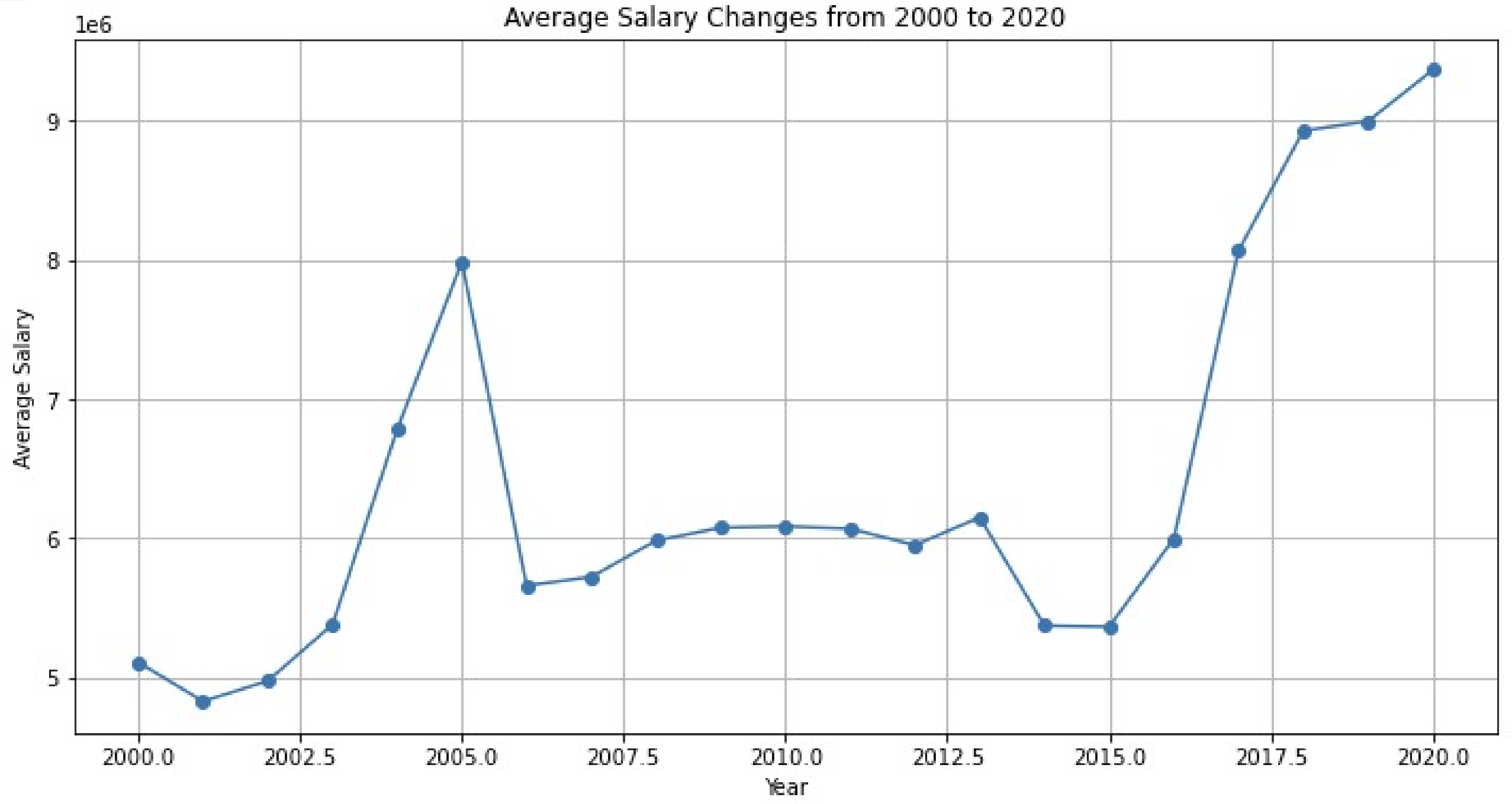



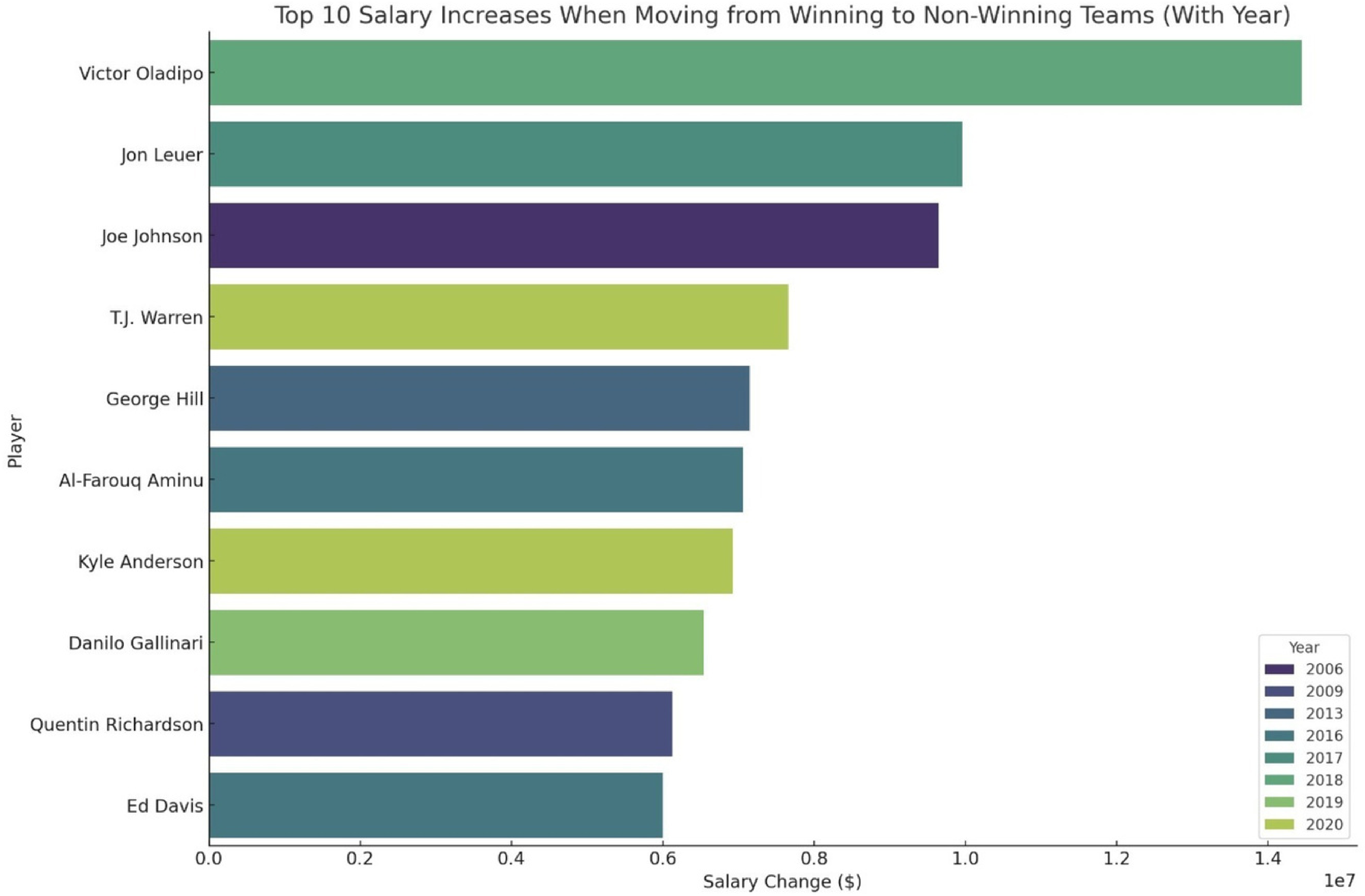




*Insight: The visual suggests that points scored (PTS) are a key predictor of a basketball player's salary and are strongly associated with offensive contributions, as evidenced by the correlation heatmaps.*


Average Salary Change Percentages for NBA Player Moves
54.08
60.33
Average Salary Change Percentage
No
Yes
Move to Less Popular Team


Points Per Game Change
Mean: -0.29
Change
Assists Per Game Change
Mean: -0.02
Change
Field Goal % Change
Mean: 0.01
Change


Average Salary Changes from 2000 to 2020
1e6
Average Salary
Year


Top 10 Salary Increases When Moving from Winning to Non-Winning Teams (With Year)
Victor Oladipo
Jon Leuer
Joe Johnson
T.J. Warren
George Hill
Al-Farouq Aminu
Kyle Anderson
Danilo Gallinari
Quentin Richardson
Ed Davis
Player
Year
2006
2009
2013
2016
2017
2018
2019
2020
Salary Change ($)
1e7

# Data Wrangling Steps

01 *Data Discovery*

02 *Data Formatting*

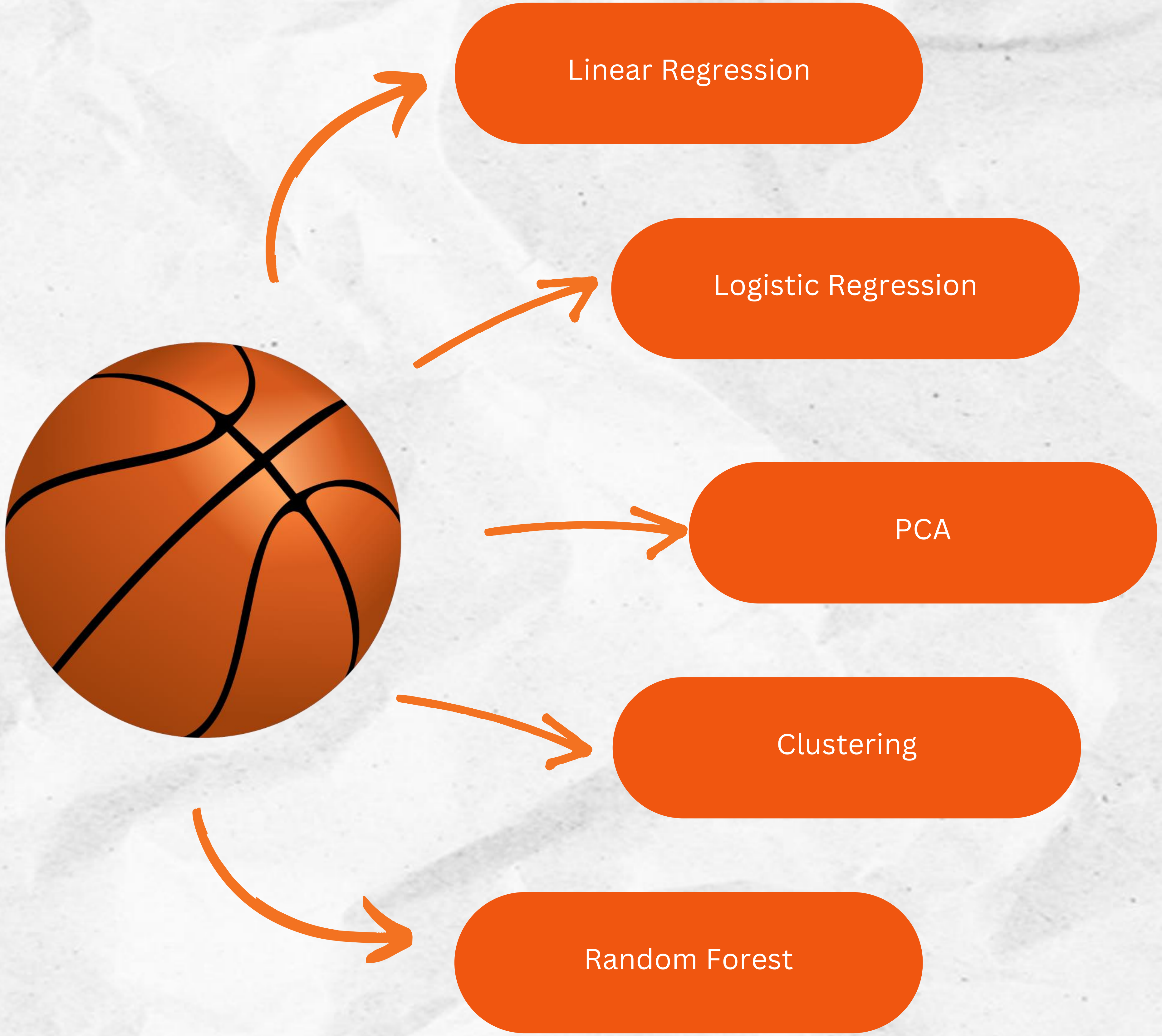
03 *Data Profiling*

04 *Data Preprocessing*

05 *Enrichment and Validation*

# *Methodology: Analytical Framework*

Employed a suite of statistical techniques to analyze and predict NBA player performance and salary structures, ensuring a multifaceted understanding of data patterns and trends.

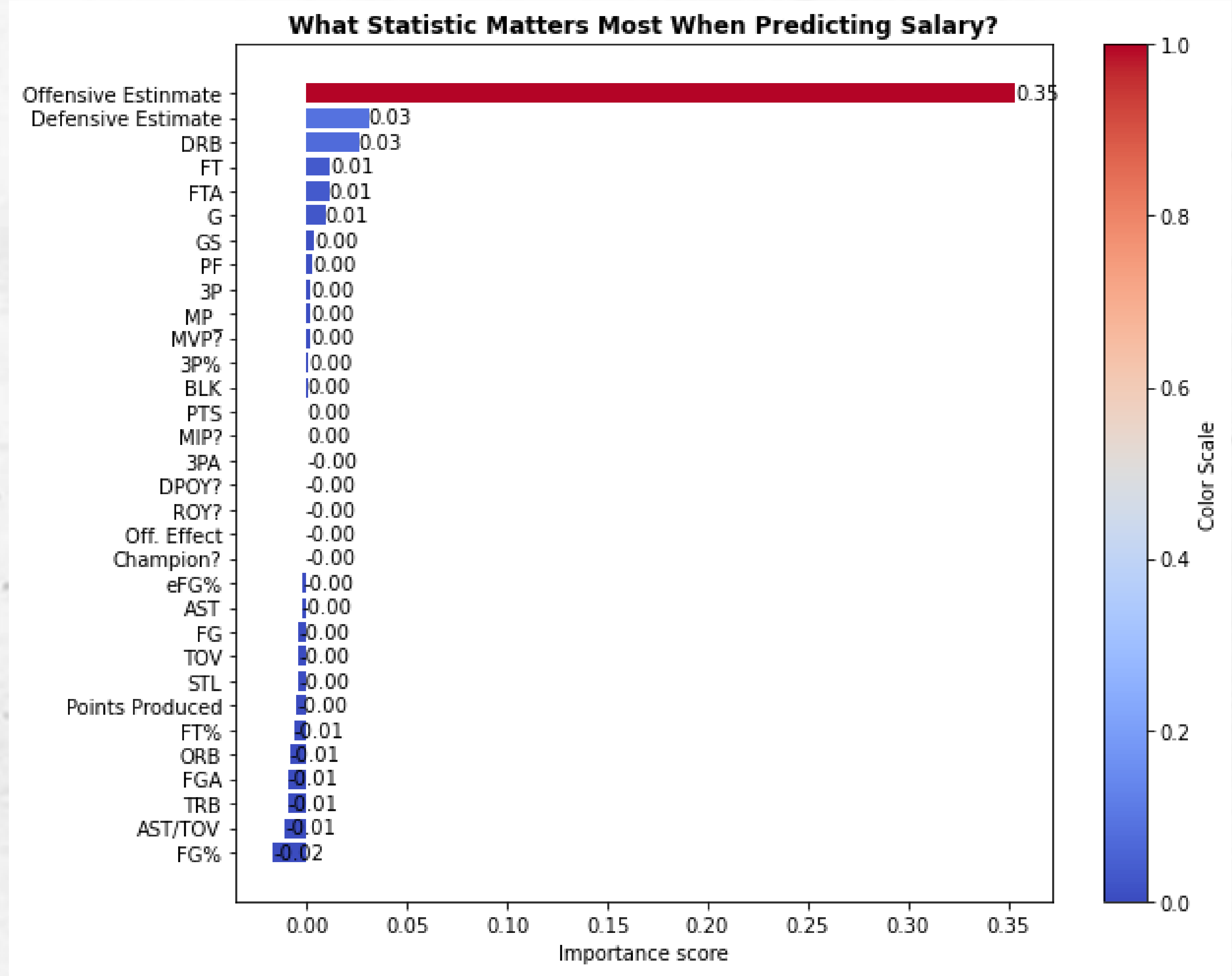


# Model Application

## *Random Forest:*

- **Offensive Estimate**: Most predictive of salary, highlighting the value of offensive skills.
- **Defensive Estimate**: Second most important, emphasizing defense as a key salary determinant.

```
In [51]: print("R2 Score:", r2)
R2 Score: 0.58689574992027

In [52]: print("Mean Absolute Error:", mae)
Mean Absolute Error: 1921791.5065119213
```

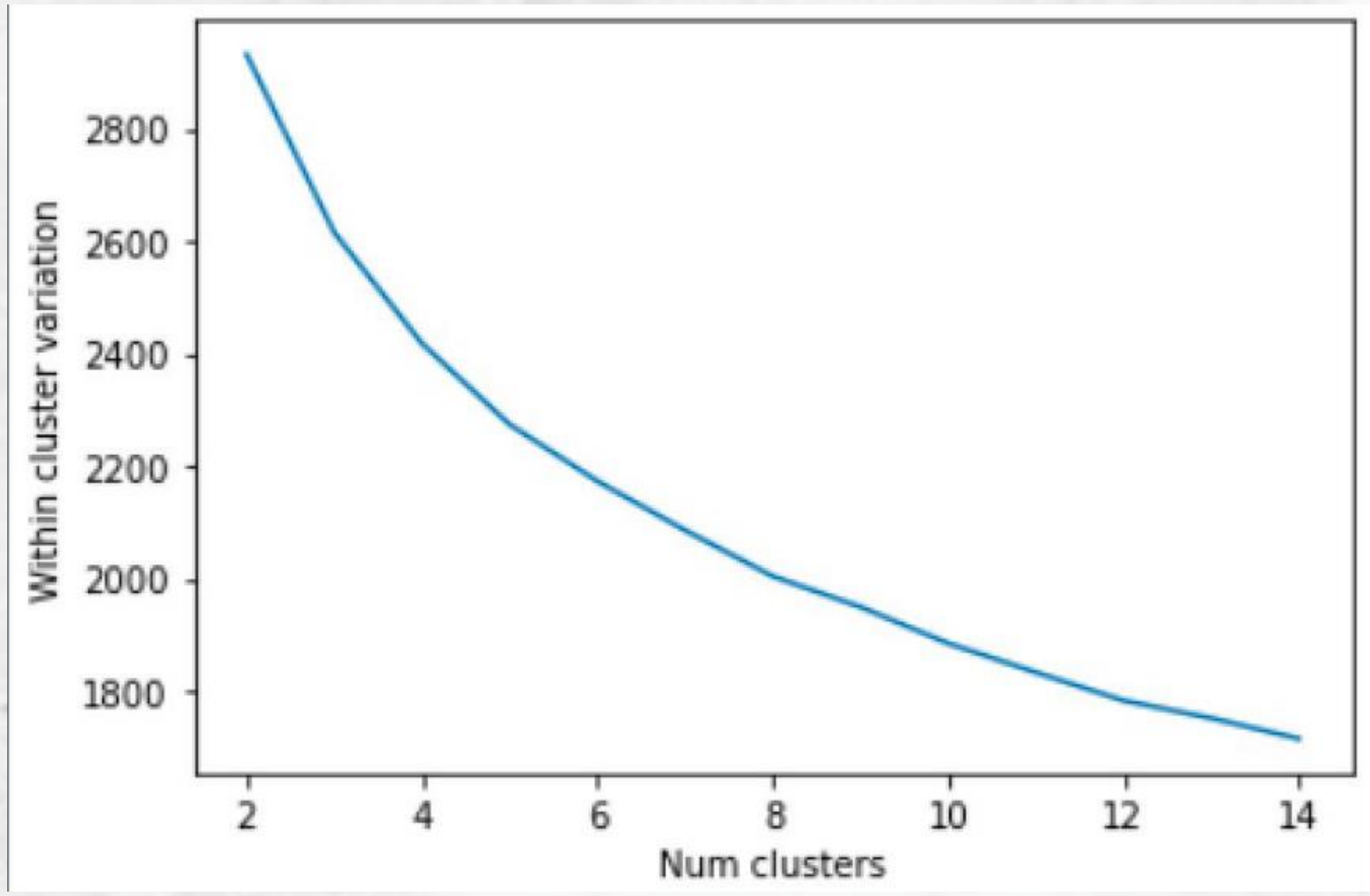



Insight: Player's offensive and defensive contributions, along with their playing time, are the most significant factors influencing their salary in professional basketball.

# Model Application

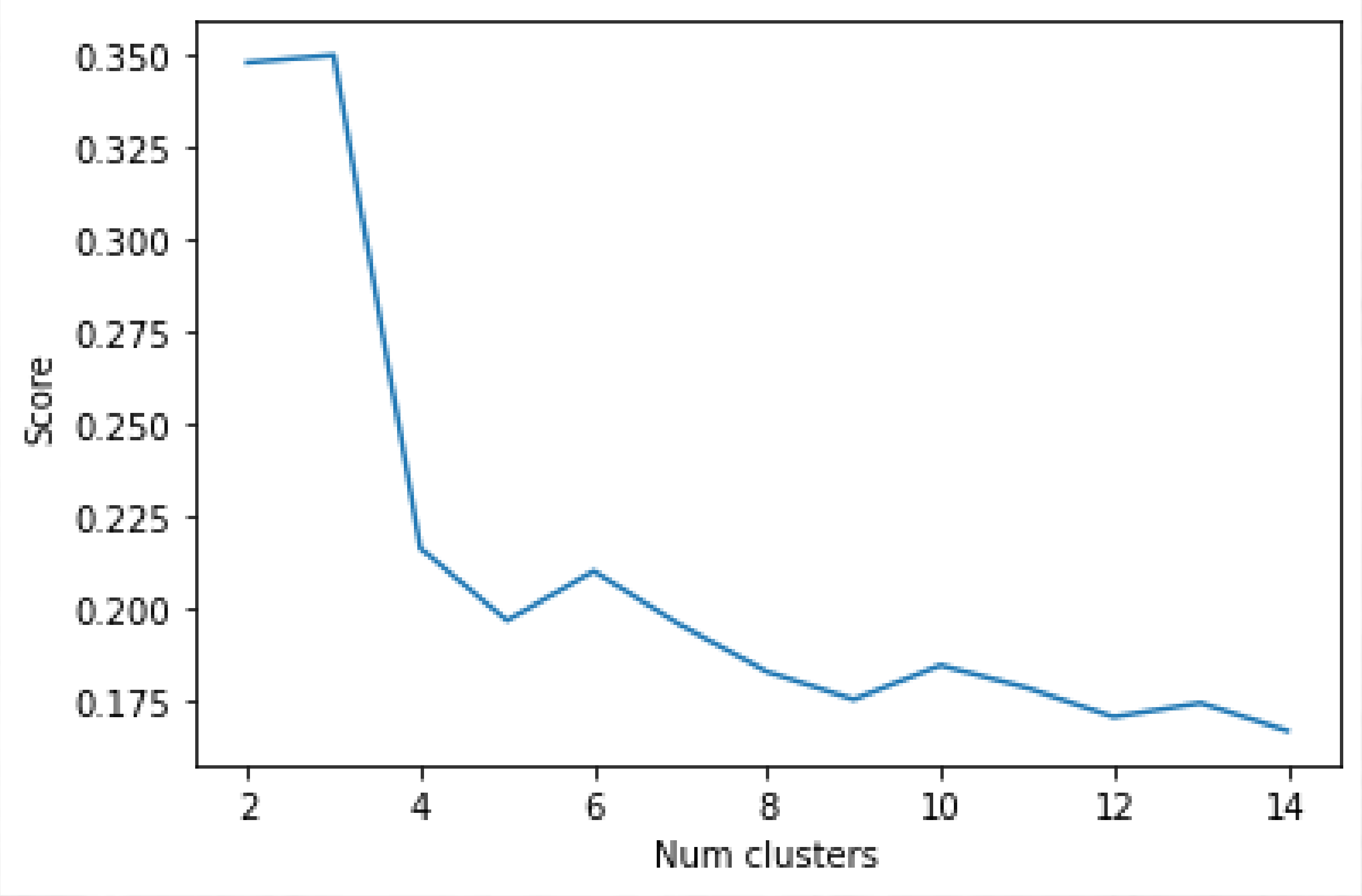
## Clustering Analysis: Player Segmentation

### Elbow Method




*The curve shows a smooth decline in within-cluster variation, which suggests a slight curve at the 3rd cluster marker suggesting optimal number of clusters within the range examined.*

### Silhouette Score




*The silhouette score maximizes in between 3-4 clusters, indicating this as the optimal number for distinct and well-separated data grouping.*

Insight: Salary trends correlate with the defined roles of players, underlining the value placed on defense and all-round capabilities in salary determinations.

# Model Application

## K-mean Clustering

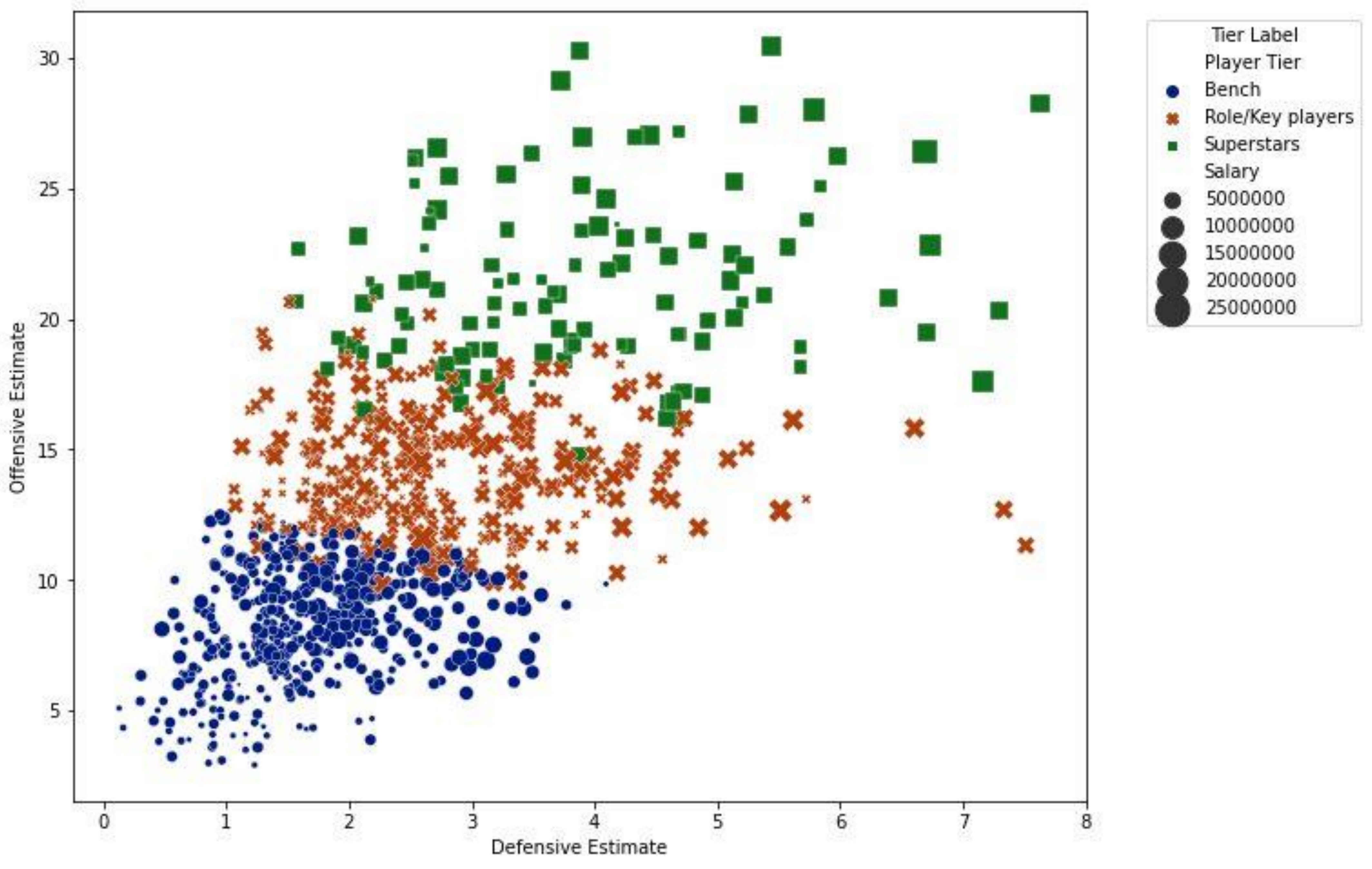
**Identified Clusters:**

Cluster 0 - The Bench Players: Characterized by low offensive and defensive contributions; typically, this group commands the lowest salaries.

Cluster 1 - Key Role Players:
Key Offensive Players: Take on a significant offensive role within their team.

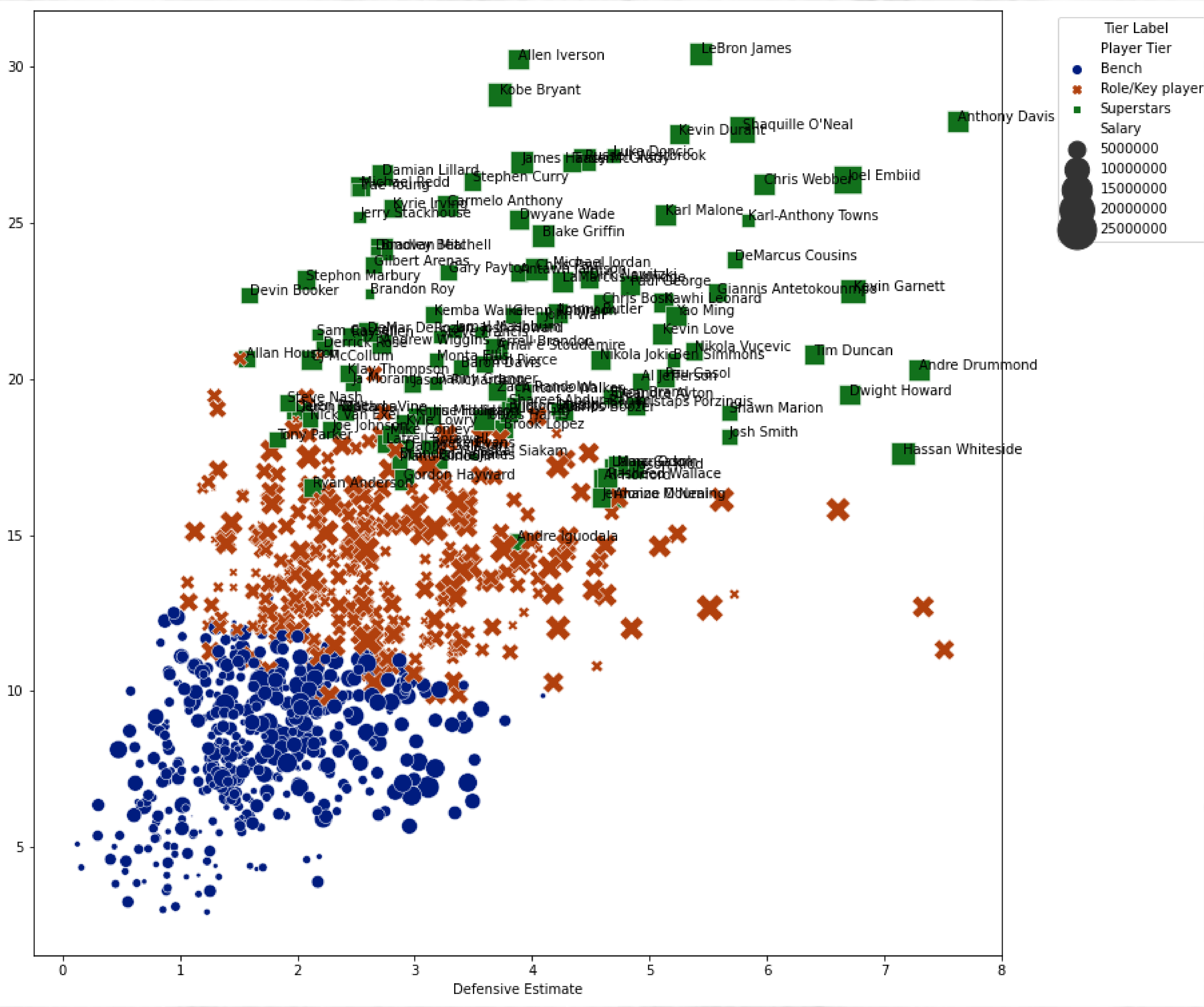
Cluster 3 - The Superstars: Responsible for the majority of the team's offense.
High volume offensive players who are also good at defense command the most money.
High variance in roles played

Superstars: Impact both ends of the court, with increased salaries reflecting their comprehensive contributions.



# *Model Application*

## *K-mean Clustering*



# Model Application

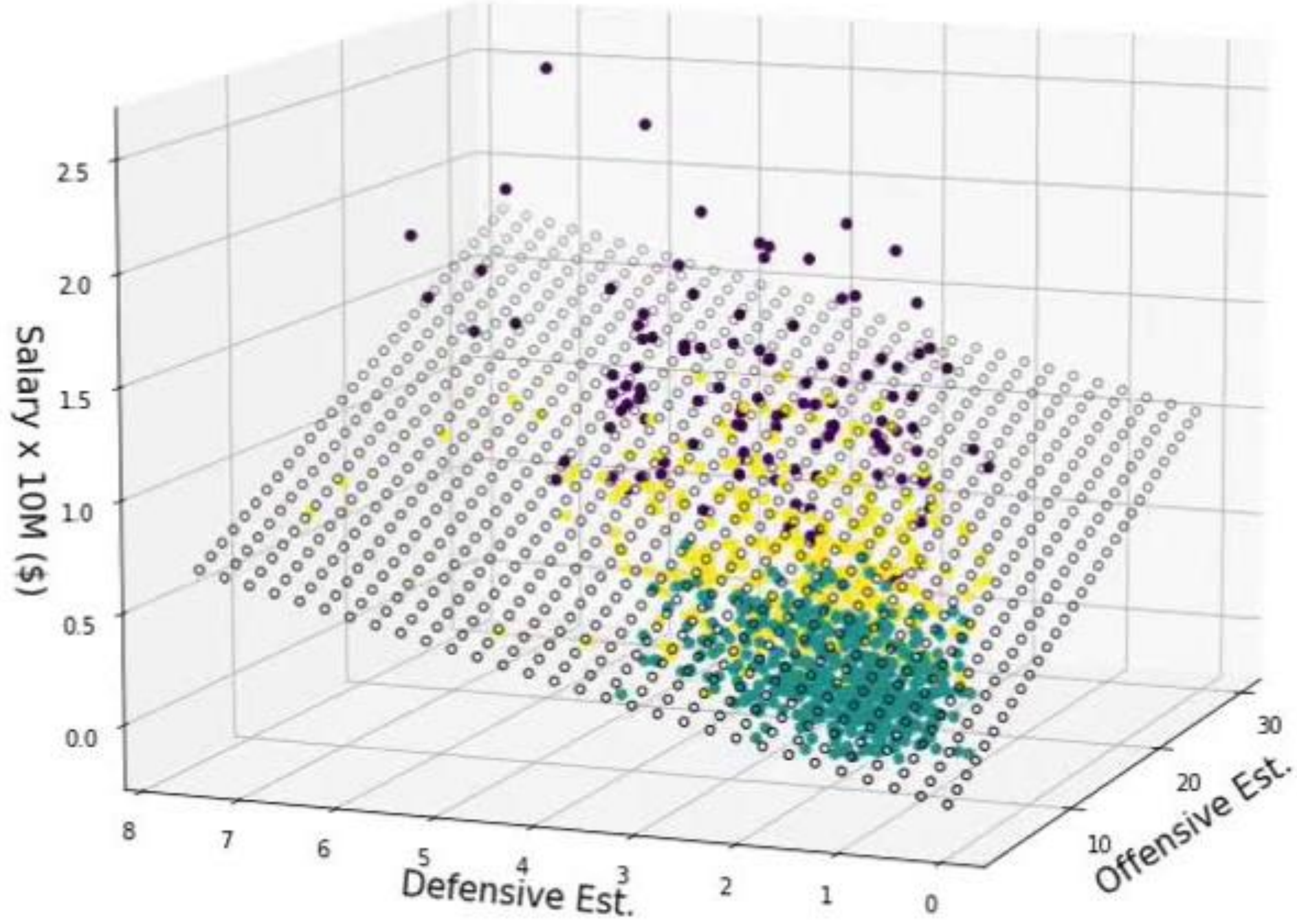
## *Linear Regression*

Model Metrics (as shown in code output):

- *RMSE (Root Mean Square Error): 2637059.161985856 This metric indicates the average error in the predicted salary, with a lower value signifying better model performance.*
- *R^2 (R-squared): 0.62992638848565 This value explains how much of the variability in player salaries our model can predict, with 1 being perfect prediction.*

```
In [319]: print("RMSE:", rmse)
RMSE: 2673905.1619588356

In [320]: print("R²:", r2)
R²: 0.6299263804856505
```

$$R^2 = 0.63$$




Insight: Insight: This model helps clubs pinpoint undervalued players to target and overvalued players to trade, while also categorizing them as Superstars, Key Players, or Benched Players for optimal team composition and salary cap strategy.

# Conclusion

KEY LEARNING

- **Offensive Dominance:** Offensive skill is the top salary driver, even more than accolades or defense.
    - Entertainment Value: Players who excel offensively boost viewership particularly for smaller market teams.
    - **i.e.**: John Wall, De'Aaron Fox, Gilbert Arenas...
    - Have to pay for high volume, low efficiency players to maintain fanbase
- **League Trend:** Scoring ability isn't just key to winning games—it's a strategic business asset for teams.
    - Avg Salary Inflation coincide with Scoring Inflation

The analysis underscores a league-wide trend: offensive talent isn't just a game winner—it's a business decision. Teams are willing to pay a premium for players who can increase offensive output, enhance the fan experience, and potentially drive revenue, even if those players haven't amassed a collection of personal awards.

# THANK YOU!

Any Questions?