Raef Hany, 900213194

Mostafa El-Fagal, 900211128

Bemen Gergis, 900213066

Project 1

MACT 3223

Dr. Salah Eldeen, Rana

Outline

1. **Introduction:**
    a. Description of the NBA player dataset.
    b. Outline of the research objectives related to player performance metrics.
    c. Defining the methodology, including using Python with Pandas, NumPy, Matplotlib, Seaborn, and SciPy for loading, analysis, and visualization.

2. **The statistics produced by the script**

3. **Inferential Statistics**

4. **Conclusion**

Introduction

A)

The dataset pertains to the NBA players during the 2023-2024 regular season. It is derived from Basketball Reference, ensuring accuracy and reliability. The dataset covers several players from different teams, compiling their performance in various statistical categories. The statistics include minutes played, field goals, assists, steals, and other performance metrics used to evaluate basketball players.

Following is a brief description of each statistic in the dataset:

1. Rank (Rk): Player ranking.
2. Player: The name of the player.
3. Position (Pos): The player's designated role on the court.
4. Age: The age of the player.
5. Team (Tm): The team affiliation of the player.
6. Games Played (G): The number of games the player participated in.
7. Games Started (GS): The number of games where the player started.
8. Minutes Played per Game (MP): The average minutes played per game.
9. Field Goals per Game (FG): The average number of successful field goals per game.
10. Field Goal Attempts per Game (FGA): The average number of attempts at field goals per game.
11. Field Goal Percentage (FG%): The percentage of successful field goals.
12. 3-Point Field Goals per Game (3P): The average number of successful 3-point field goals per game.
13. 3-Point Field Goal Attempts per Game (3PA): The average number of attempts at 3-point field goals per game.
14. 3-Point Field Goal Percentage (3P%): The percentage of successful 3-point field goals.
15. 2-Point Field Goals per Game (2P): The average number of successful 2-point field goals per game.
16. 2-Point Field Goal Attempts per Game (2PA): The average number of attempts at 2-point field goals per game.
17. 2-Point Field Goal Percentage (2P%): The percentage of successful 2-point field goals.
18. Effective Field Goal Percentage (eFG%): A composite measure of field goal accuracy.
19. Free Throws per Game (FT): The average number of successful free throws per game.
20. Free Throw Attempts per Game (FTA): The average number of attempts at free throws per game.
21. Free Throw Percentage (FT%): The percentage of successful free throws.
22. Offensive Rebounds per Game (ORB): The average number of offensive rebounds per game.

23. Defensive Rebounds per Game (DRB): The average number of defensive rebounds per game.
24. Total Rebounds per Game (TRB): The average total number of rebounds per game.
25. Assists per Game (AST): The average number of assists per game.
26. Steals per Game (STL): The average number of steals per game.
27. Blocks per Game (BLK): The average number of blocks per game.
28. Turnovers per Game (TOV): The average number of turnovers per game.
29. Personal Fouls per Game (PF): The average number of personal fouls committed per game.
30. Points per Game (PTS): The average number of points scored per game.

B)

Our research objectives are:

1. Assess the potential impact of player aggressiveness on their overall performance in basketball games: We would focus on parameters that measure a player's aggressiveness, such as the number of fouls committed. We would relate these measures to performance measures and shed light on whether a more aggressive style negatively or positively affects their performance.
2. Investigate the relationship between age and statistics, such as games started or fouls committed. This could involve looking at how a player's age correlates with the number of games they start or the number of fouls they make. For example, do older players start fewer games or make more fouls?
3. Investigate the relationship between a player's position (specifically, being a center) and aggression: This could involve comparing the aggressiveness of players in the center position with the aggressiveness of players in other positions.

C) Methodology:

We wrote two scripts in Python, one for the data visualization and the other for performing statistical analysis on a dataset of NBA players, and here is how we applied our methodology using the code:

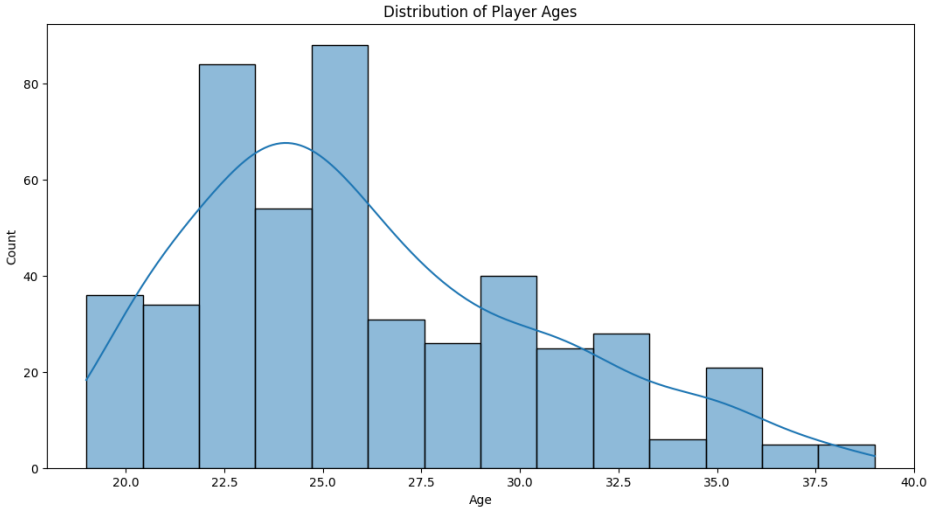
1. **Data Loading and Preprocessing:** The code reads a CSV file containing data about NBA players. It then splits the data into rows and columns, and converts appropriate columns to floating-point numbers.
2. **Confidence Intervals:** using the script, we calculated 95% confidence intervals for the mean points, variance of age, and proportions of aggressive players and Center players.
3. **Hypothesis Testing:** using the script, we tested several hypothesis:
    - Older players start more games and are more aggressive than younger players.
    - Aggressive players score more points than less aggressive players.

- Less than half of the players always start the games.
- Center players are more aggressive than non-Center players.

Descriptive statistics

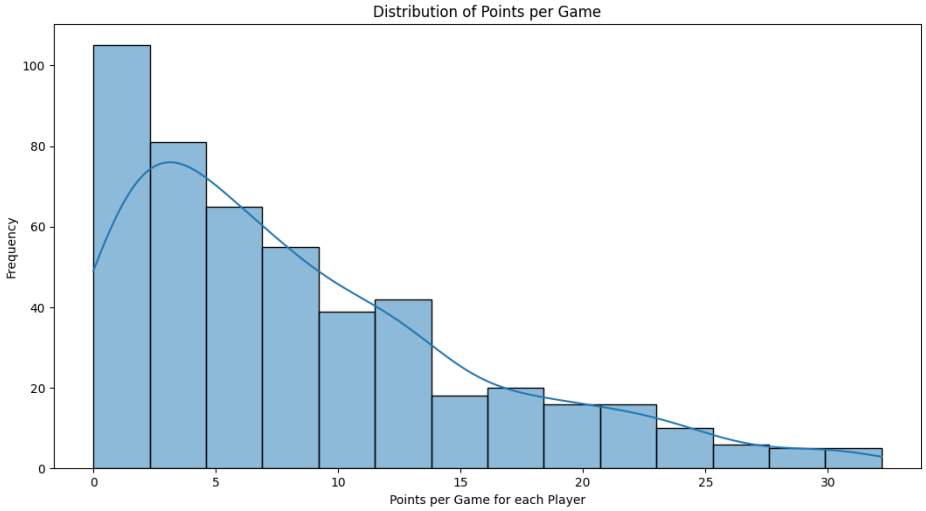
Using the script, we calculated the mean, standard deviation, and variance for the age, games started (GS), personal fouls (PF), and points (PTS) of the players. It also calculates the proportion of aggressive players (PF >= 2) and the proportion of players who are Centers (POS = "C").

- Ages (Age) distribution:
  The average age of the players is approximately 26.1 years, with a standard deviation of approximately 4.43 years. This means that most players' ages fall within a range of 26.1 ± 4.43 years.

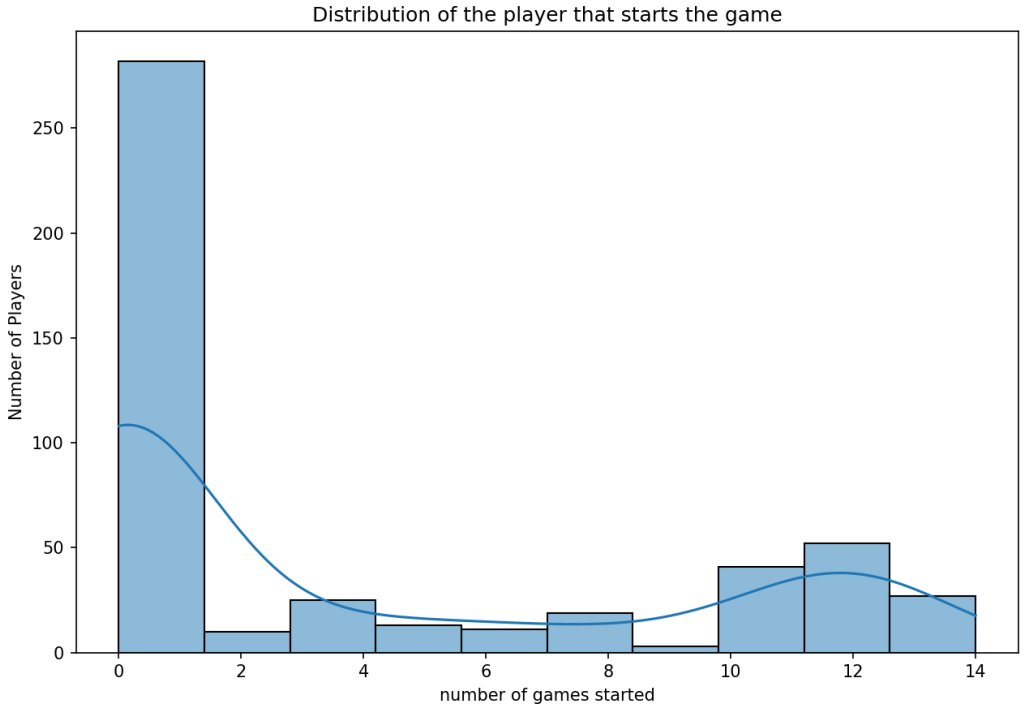


- Points per game (PTS) distribution:
  On average, a player scores approximately 8.53 points, with a standard deviation of roughly 7.29. This means the number of points scored by most players falls within a

range of 8.53 ± 7.29.

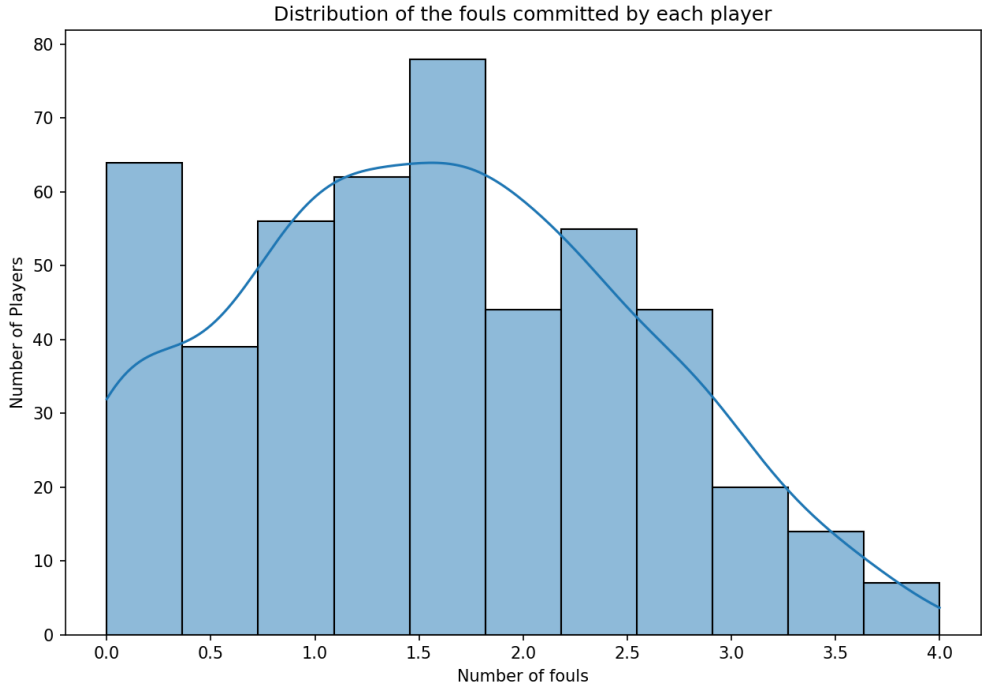


- Games started (GS) distribution:
  On average, a player starts approximately 3.83 games, with a standard deviation of approximately 5.01. This means the number of games started by most players falls within a range of 3.83 ± 5.01.

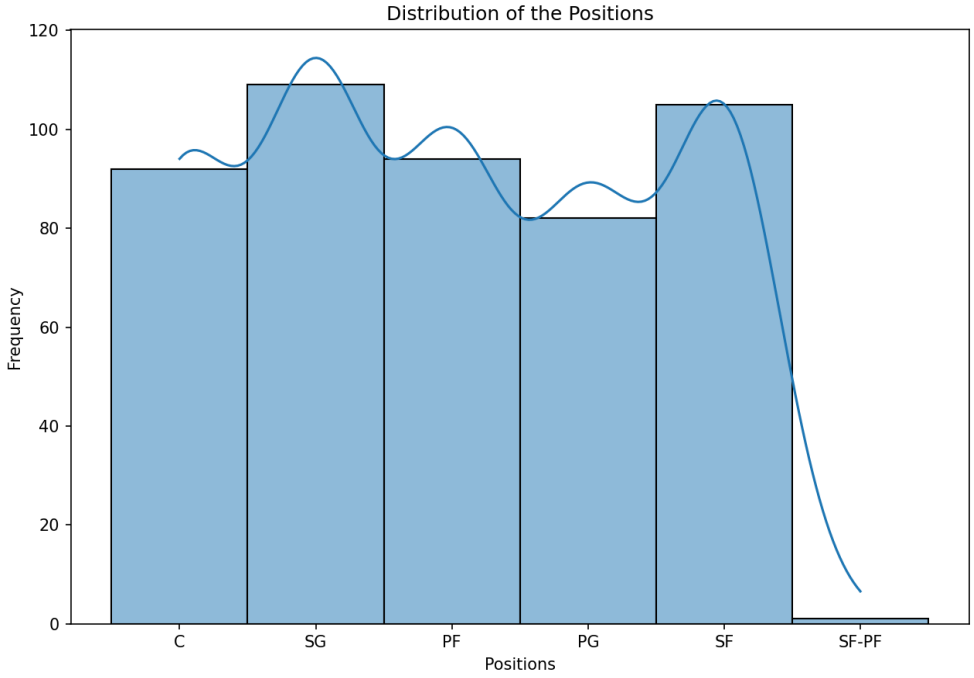


- Fouls committed (PF) distribution:

On average, a player commits approximately 1.57 fouls, with a standard deviation of approximately 0.96. This means the number of fouls committed by most players falls within a range of 1.57 ± 0.96.



- Positions (Pos) distribution :



Inferential Statistics

```
a)
Means: {'Age': 26.099378881987576, 'GS': 3.8302277432712217, 'PF': 1.5726708074534135, 'PTS':
8.530434782608703}
Stds: {'Age': 4.430943113471762, 'GS': 5.013690525612316, 'PF': 0.9566747981502055, 'PTS': 7.2
89047460478719}
Proportion of aggressive players(PF>=2): 0.35196687370600416
Proportion of Center players: 0.19047619047619047
b)
Intervals: {'mean PTS': [7.880375021374289, 9.180494543843116], 'variance Age': [17.3724823638
38614, 22.368410995684517], 'proportion PF': [0.30937448427158865, 0.39455926314041967], 'prop
ortion POS': [0.15545606743652418, 0.2254963135158567 5]}
c)
Point estimator for difference young and old, in terms of means games started: -1.533566989449
3429
Interval for difference of young and old, in terms of mean games started: [-2.463167471484589,
 -0.6039665074140966]
since the range is negative, then older players start more games
Point estimator for difference young and old, in terms of proportion of aggressive players: -0
.11502746061569591
Interval for difference of young and old, in terms of proportion of aggressive players: [-0.20
246792927225699, -0.027586991959134846]
since the range is negative, then older players tend to be more aggressive
d)
Relating aggression to points per game:  10.558829641498033 True
Since Z_cal is in the Rejection Region, then we reject and thus more aggressive players score
more points
Considering the proportion of players that start games:  -0.9555330859059086 True
Since Z_cal lies in RR, then we reject and so less than half the players always start the game
s
Relating aggression and position: 3.875132349870415 True
Since Z_cal is in the RR, then we reject and so Center players are more aggressive
```

a)

Age:

- Parameter of Interest: Mean Age
- Point Estimator: Sample Mean (given in the output)
- Value: The sample mean, 26.1
- Explanation: The sample mean is an unbiased and stable estimate of the average age of players in the dataset. It's reliable across different samples and becomes more precise with larger sample sizes (consistent).

Games Started (GS):

- Parameter of Interest: Mean Games Started
- Point Estimator: Sample Mean (given in the output)
- Value: The sample mean, 3.83
- Explanation: The sample mean is an unbiased estimator for the average number of games started by players. It minimizes errors and becomes more reliable as the dataset size increases (consistent).

Points (PTS):

- Parameter of Interest: Variance of Points Scored
- Point Estimator: Sample Variance (calculated in the script)
- Value: The sample variance (calculated in the script)

- Explanation: The sample variance serves as an unbiased and consistent estimate for the variability in points scored by players. It provides a reliable measure of how scores vary within the dataset, becoming more accurate with larger samples.

Personal Fouls (PF):

- Parameter of Interest: Proportion of Players with PF >= 2
- Point Estimator: Sample Proportion (given in the output)
- Value: The sample proportion, 0.351
- Explanation: The sample proportion provides an unbiased, efficient, and consistent estimate for the proportion of players with PF >= 2. It gives a reliable measure of aggression in the dataset, improving with larger samples.

b) We calculated 95% confidence intervals for the mean points,the proportion of aggressive players (PF>=2),the proportion of Center players, and the age variance. These intervals give an estimated range where the true population parameter lies with 95% confidence.

c) We compared two groups: players aged 26 or less and older than 26. We calculated point and interval estimates for the difference in mean games started and the proportion of aggressive players between these two groups. The results suggest that older players start more games and tend to be more aggressive.

d) We tested three hypotheses:

1. More aggressive players score more points. The null hypothesis is that aggressive players do not score more points; the alternative is that they do. The test rejects the null hypothesis, suggesting that more aggressive players do score more points.
2. Less than half the players always start the games. The null hypothesis is that at least half the players always start the games, and the alternative is that less than half do. The test rejects the null hypothesis, suggesting that less than half the players always start the games.
3. Center players are more aggressive. The null hypothesis is that Center players are not more aggressive, and the alternative is that they are. The test rejects the null hypothesis, suggesting that Center players are indeed more aggressive.

Conclusion with Main Findings and Decisions

After analyzing the basketball players' dataset we came up with the following:

1. Older players tend to start more games and are generally more aggressive than younger players. This implies that experience and maturity play a role in a player's aggression and likelihood of starting games.
2. More aggressive players tend to score more points. This suggests that aggression might be a beneficial trait in basketball, possibly leading to more scoring opportunities.
3. Less than half of the players always start the games. This could indicate a strategy of rotating players to keep them fresh or it might be a reflection of team dynamics and roles.
4. Center players are more aggressive. This could be due to the nature of the position which often involves more physical contact and competition for rebounds.

These findings could have several implications for team strategy and player development. Coaches might want to consider encouraging controlled aggression in their players to potentially increase scoring. Additionally, understanding the tendencies of older players to start more games and be more aggressive could inform decisions about player rotation and game strategy.

For player development, it might be beneficial to provide younger players with opportunities to gain experience and develop their aggression in a controlled manner. Furthermore, the finding about Center players being more aggressive could inform training programs for these players to help them harness their aggression effectively.

In terms of areas for improvement, further research could investigate the reasons behind these findings. For example, why are older players more likely to start games? Is it due to their experience, their skills, or other factors? Understanding these underlying reasons could provide more nuanced insights for decision-making.