Comprehensive Analysis of NBA Strategies and Performance

(National Basketball Association)

Supervisor Name: KINGSLEY IBOMO

Student Name: SHREYASH DHUMAL
Student ID: 10613796

Presentation includes following points:

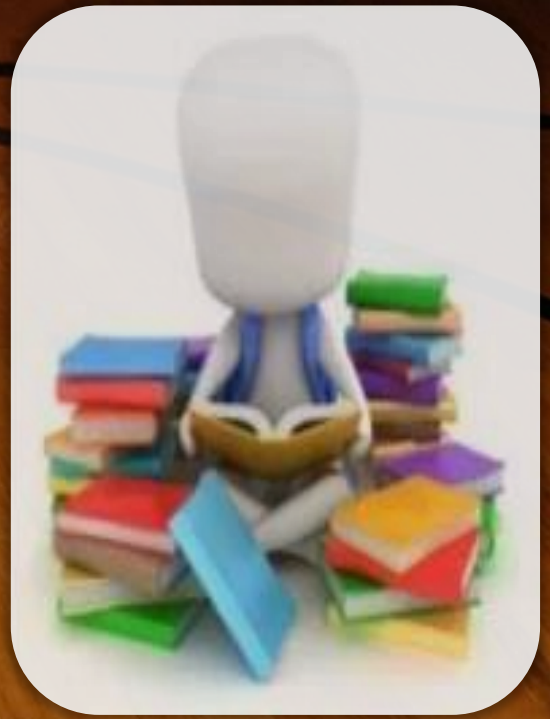
- Introduction to Problem Statement
- Literature Review
- Data Collection
- Methodology
- Data Analysis
- Feature Engineering
- Machine Learning
- Results
- Challenges
- Future Enhancement
- Conclusion

Introduction to Problem Statement

- NBA Overview:
  - Dynamic and Fiercely Competitive League
  - Continuous Excellence Requirement
- Evolving NBA Landscape:
  - Growing Demand for Detailed Examination
  - Exploration of Strategies and Performance Intricacies
- Analysis Objectives:
  - Providing Actionable Insights
  - Benefiting Teams and NBA Fans
- Team Strategies and Individual Player Performance:
  - Complexities Unraveled in Challenging Task

Literature Review

- Abundance of NBA Data
- Role of Machine Learning (Horvat et al., 2023)
- Shooting Trends Analysis (Zając et al., 2023)
- Effects of Dehydration on Performance(Louis et al., 2018)
- Contribution of Studies.



Data Collection

- Dataset Overview:
  - NBA Games Data" on Kaggle.com
  - Renowned Platform for Datasets and ML Competitions

- Value of the Dataset:
  - Comprehensive and Detailed Insight into NBA Games
  - Valuable Resource for Analytical Exploration

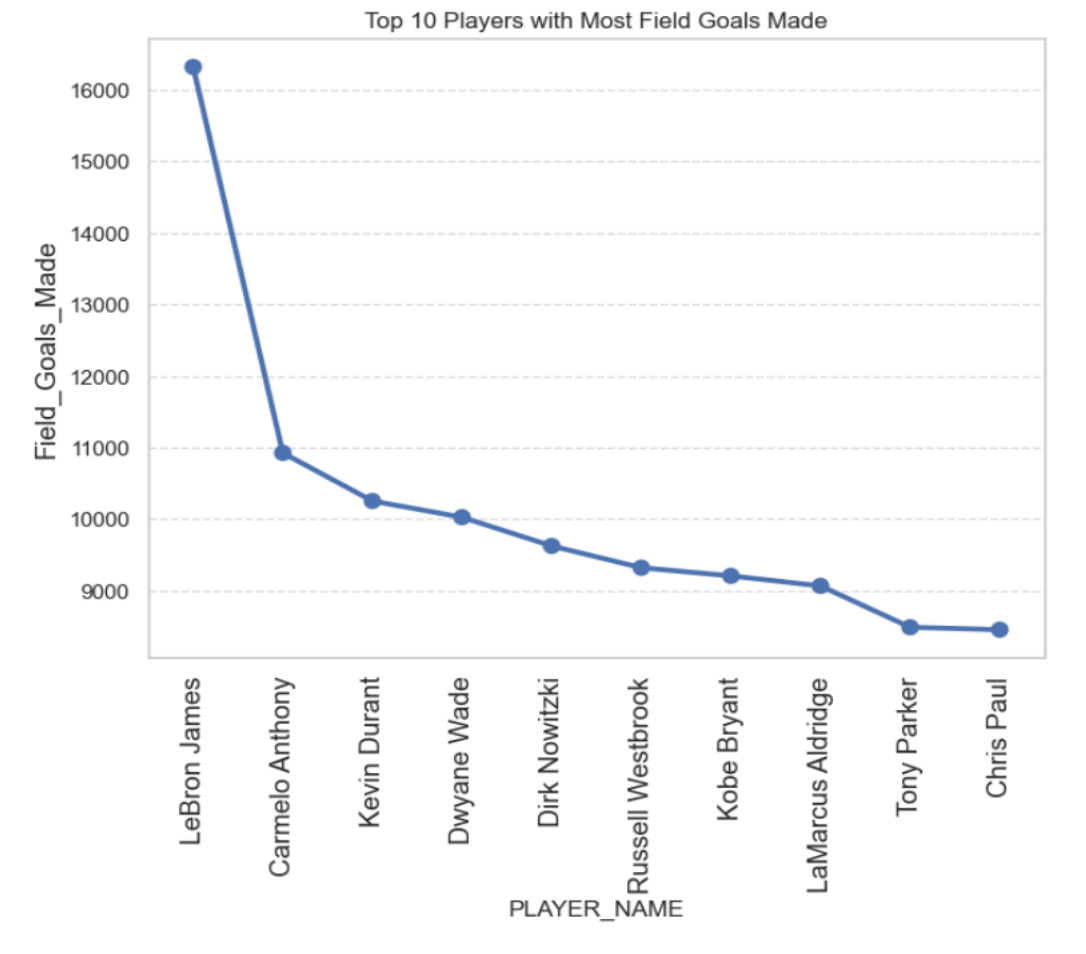
Methodology

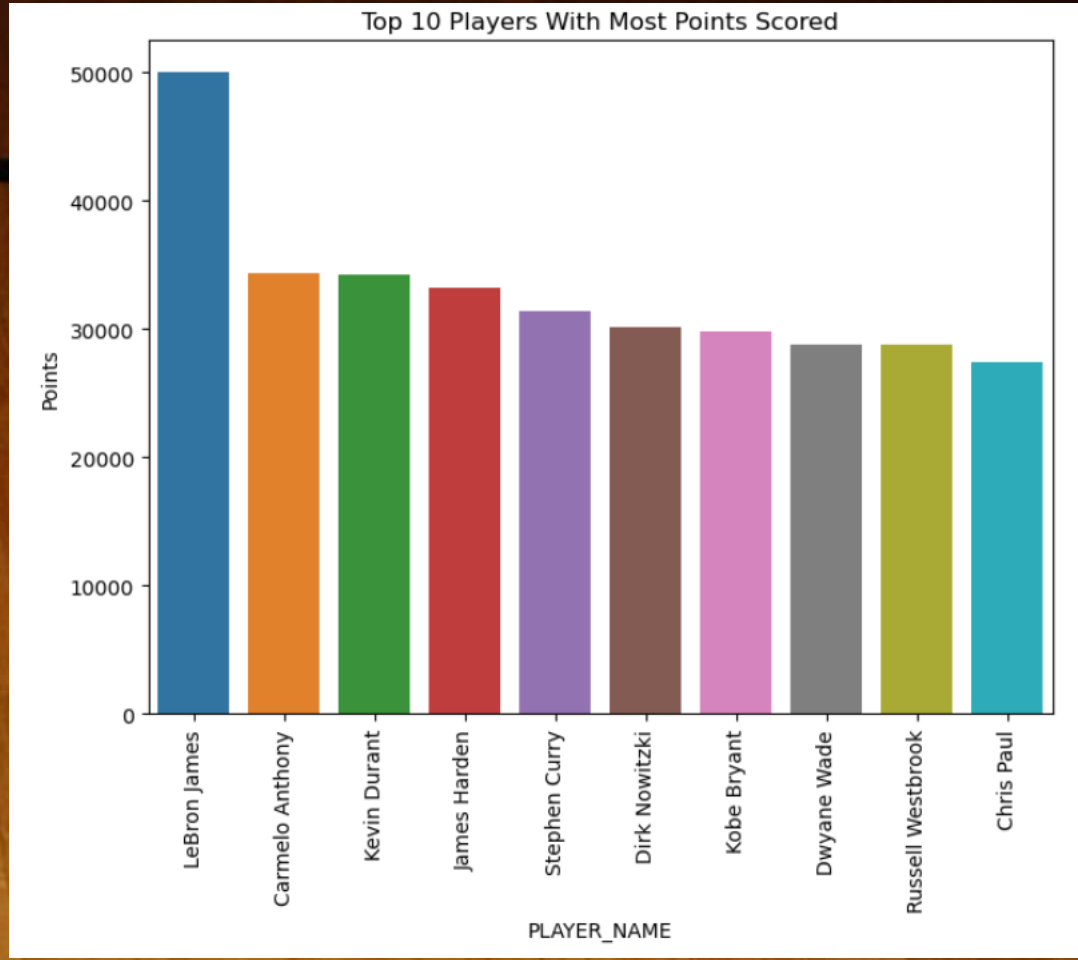
- Analysis of Crucial Features Impacting NBA Team and Player Performance.
- Identifying Factors Influencing Team Victory.
- Uncovering Intricate Link Between Player Performance (Plus/Minus Metric) and Team Success.
- Valuable Insights for Sports Enthusiasts, Analysts, and Teams.

Feature Engineering

- Removed unnecessary columns: 'NICKNAME,' 'START_POSITION,' 'COMMENT' for streamlined analysis.
- Created a new column for Total Points using a calculation formula based on field goals, three-pointers, and free throws.
- Converted the 'MIN' column from string to float for more effective numerical analysis.
- Removed the 'PLUS_MINUS' column and introduced a new column named 'MODIFIED_PLUS_MINUS' based on a binary classification.

Data Analysis





Machine Learning

- Logistic Regression in NBA Analysis:
  - Chosen as Primary Modeling Technique
  - Effective for Binary Classification Problems

- Prediction Process:
  - Utilized "predict" Function for Model Predictions
  - Feature Matrices: X_test for Test Dataset, X_train for Training Dataset

- Accuracy Assessment:
  - Leveraged "accuracy_score" Function
  - Comparison of Model Predictions to Actual Labels (y_test and y_train)

- Accuracy Rates:
  - Achieved 62% Accuracy for Both Training and Testing Datasets

```
accuracy_score(y_test,y_pred)
```

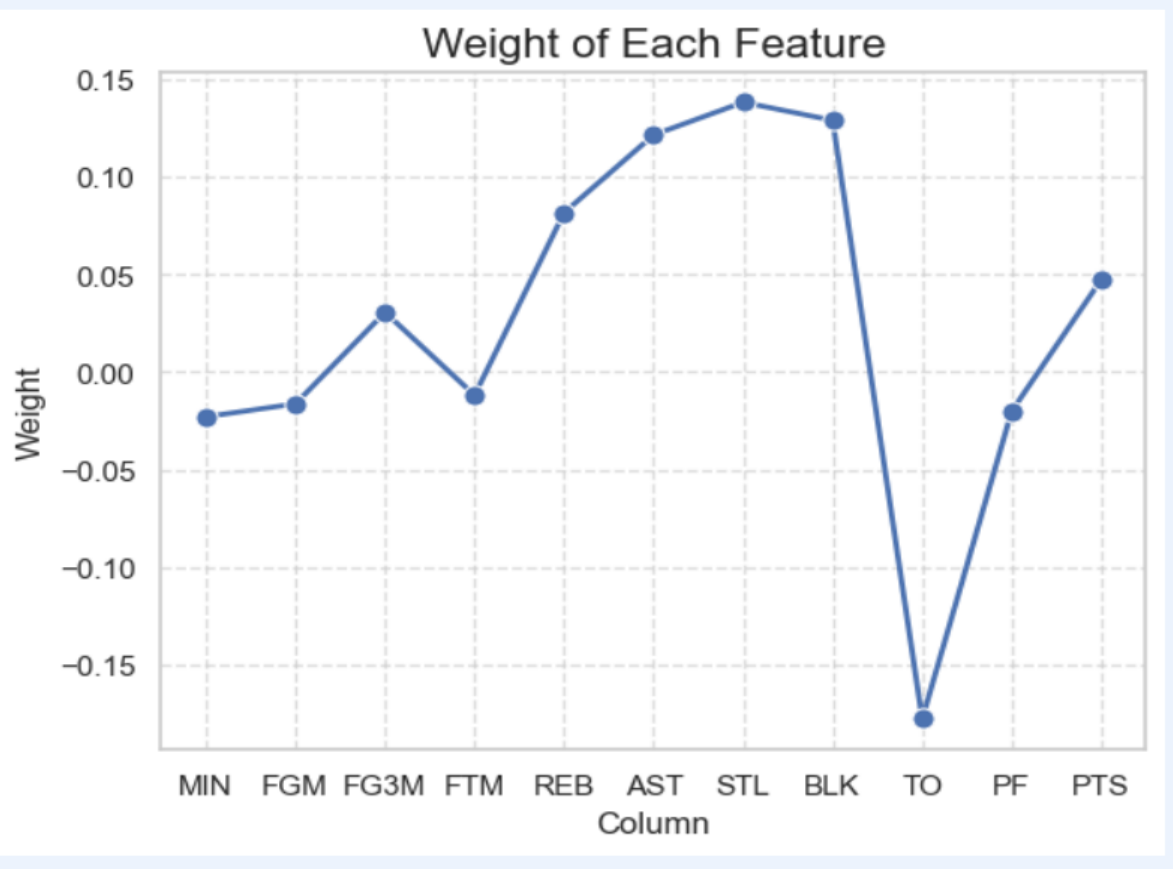
```
0.6184051336153079
```

```
y_pred_train = log_model.predict(X_train)
```

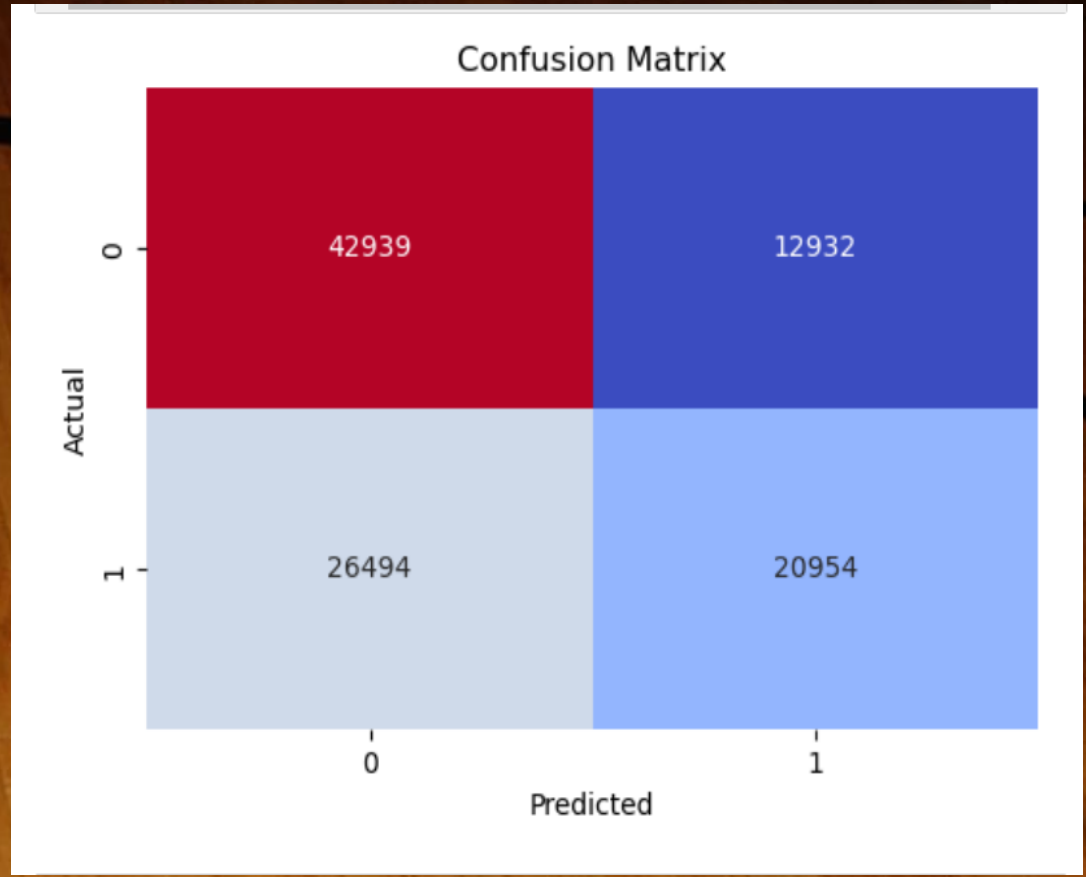
```
accuracy_score(y_train,y_pred_train)
```

```
0.6171202779740416
```

Weight Of Each Feature



Confusion Matrix



Results

- Model Predictions:

  - Utilizes loaded machine learning model on sample DataFrame
  - If Predicted Result is 1: Positive impact on position based on player's performance
  - If Predicted Result is not 1: Suggests negative impact based on player's performance

- Model Persistence with Pickle:

  - Leverages pickle library for saving and loading machine learning model
  - Common practice in data science for persisting trained models

- Serialization and Deserialization:

  - Model saved to file using pickle.dump
  - Loaded back into memory using pickle.load

```
import pickle
```

```
pickle.dump(log_model, open('Player_Ability.pkl','wb'))
```

```
predict_output = pickle.load(open('Player_Ability.pkl','rb'))
```

```
dataframe = pd.DataFrame([{
    'MIN':24.00,'FGM':3,'FG3M':2,'FTM':2,'REB':2,'AST':3,'STL':2,'BLK':1,'TO':2,'PF':1,'PTS':14 }])
```

```
dataframe
```

	MIN	FGM	FG3M	FTM	REB	AST	STL	BLK	TO	PF	PTS
0	24.0	3	2	2	2	3	2	1	2	1	14

```
predicted_ans = predict_output.predict(dataframe)[0]
```

```
if predicted_ans==1:
    print('based on player performace, position impact')
else:
    print('based on player performance, negative impact')
```

```
based on player performace, position impact
```

Challenges

- Project Challenges Overview:
  - Limited Enthusiasm for NBA in Home Country as Initial Complexity

- Data Gathering Struggles:
  - Obtaining Comprehensive and High-Quality NBA Data
  - Searched for Specific Dataset Criteria

- Feature Selection and Consideration:
  - Thorough Evaluation of Valid Features
  - Selection Based on Relevance and Significance

- Machine Learning Enhancement Efforts:
  - Focus on Improving Accuracy
  - Challenging Phase During the Analysis

Future Enhancement

- Real-time Predictions and In-Game Analysis:
  - Adoption of Real-time Analytics
  - Continuous Monitoring of Player Statistics and Game Dynamics
  - Timely Interventions and Strategic Adjustments
  - Live Tracking of Player Movements, Shot Accuracy, and Defensive Plays

- Fan Engagement Platforms:
  - Development of Interactive Platforms
  - Leveraging Advanced Analytics for Enhanced Fan Experience
  - Integration of Real-time Analytics Dashboards
  - Instantaneous Updates on Player Performance, Team Statistics, and Key Match Dynamics
  - Visually Appealing and User-friendly Interfaces for Deeper Fan Understanding

- Incorporating Advanced Metrics:
  - Exploration of Metrics Beyond FGA and FGM

Conclusion

In summary, these studies have successfully explored the intricate relationship between individual player performance and overall NBA team success through the application of machine-learning algorithms. Investigating the significance of the plus/minus metric, the research contributes to a deeper understanding of player contributions and the determinants of success in the NBA. The use of a logistic regression model achieved a commendable 62% accuracy rate, highlighting the effectiveness of data-driven approaches. Additionally, the comprehensive knowledge gained about NBA sports and rules overcame challenges in data gathering, providing valuable insights for future advancements in basketball analytics.

THANK YOU!