Coláiste na Tríonóide, Baile Átha Cliath
Trinity College Dublin
Ollscoil Átha Cliath | The University of Dublin

How Do NBA Teams Plan All-Star Players' Salaries?

Assignment Submission Cover Sheet

Programme Title:	**MSc Business Analytics 2024/25**
Module Code and Title:	**BU7143 Business Data Mining**
Assessment Title:	**Group Assignment – Project Report**
Group Number:	**13**

Student Name and Contribution	**%**		**%**
1.Chenyu Wang		**4.Shanshan Tan**	
2.Geunju Park		**5.Xiaoxue Ji**	
3.Panagiotis Georgiadis		**6.**	

For group work – individual % contributions need to be stated **only** where they **are not equal.**

Please read the plagiarism provisions in the General Regulations of the University Calendar for the current year, found at: http://www.tcd.ie/calendar

Online Tutorial on avoiding plagiarism 'Ready, Steady, Write', is located at http://tcd-ie.libguides.com/plagiarism/ready-steady-write

1. Executive Summary

In modern sports management, achieving a balance between on-field performance and financial stability is crucial for long-term success. Professional basketball clubs must not only excel in competition but also manage their financial resources effectively, particularly in salary allocation. This project employs data mining techniques to help clubs make informed salary decisions based on players' historical performance data, ultimately supporting better financial outcomes and team success.

The data for this project was sourced from Kaggle, a well-known platform for data-driven projects. It includes detailed player performance metrics, salary records, and other relevant historical data from professional basketball leagues. This comprehensive dataset provides the foundation for building predictive models that address the club's primary business goal: optimizing salary allocations. The club's primary business goal is to optimize salary allocations.

Two key data mining models have been developed: an All-Star player prediction model and a salary prediction model. The All-Star model helps the club identify players with high potential based on their performance data, allowing for early salary adjustments that secure these players at reasonable rates. The salary prediction model builds on this by forecasting future salary needs for players based on historical performance and salary trends. Together, these models provide the club with the insights needed to align player compensation with their contribution to the team.

This data-driven approach offers several significant business advantages. First, it provides a predictive framework that allows for accurate planning of future salary needs, reducing the risk of financial missteps that could arise from overcompensating underperforming players. By proactively managing salaries, the club can maintain budget flexibility to pursue emerging talent and retain key players. Additionally, optimizing the salary structure helps the club remain competitive by ensuring resources are allocated efficiently, thereby preventing the financial strain that often accompanies poorly structured player contracts.

While the models provide strong predictive capabilities, they are not without limitations. External factors, such as injuries or changing team dynamics, may affect player performance in ways that the models cannot fully predict. Furthermore, player market value is influenced by factors beyond performance, such as commercial appeal, which may not be fully captured in the models. Despite these limitations, the overall benefits of incorporating data mining into salary planning are clear. By using predictive analytics, the club can mitigate financial risks, ensure competitive salaries are paid only to deserving players, and maintain the long-term operational sustainability needed for success.

In conclusion, data-driven salary management is a valuable business proposition for any basketball club looking to maintain both financial health and on-field performance. By leveraging data mining models, clubs can better align player compensation with actual performance, reduce overpayment risks, and retain key talent—all while ensuring that the team remains flexible and competitive in a rapidly evolving market. This approach provides a clear path toward smarter financial planning and sustained success.

2. Project Overview

1) Introduction

In modern sports management, teams must not only achieve excellent performance on the field but also ensure sound financial operations. The success of professional teams largely depends on effective salary management. All-Star players play a crucial role in both performance and generating significant commercial returns for the team. However, improper salary allocation can impose financial burdens and even weaken the team's long-term competitiveness.

Case Study: Miami Heat and Duncan Robinson's High-Pay Contract

In 2021, the Miami Heat signed Duncan Robinson to a five-year, $90 million contract. However, in the 2022-23 season, Robinson's performance did not justify his salary, with reduced playing time and limited impact, especially in the playoffs. This costly contract strained the Heat's salary cap, restricting their ability to recruit other key players and weakening overall team competitiveness. This case illustrates the drawbacks of improper salary allocation on team success.

2) Problem Statement

a. Business Goal

The team aims to make reasonable salary adjustments, avoiding overpayment to players with insufficient performance or potential, while ensuring that current star players are retained.

b. Data Mining Goals

- **All-Star Player Prediction Model:** Based on the performance data of NBA players in 2023, we built a prediction model to identify which players are likely to become All-Stars. With this model, the team can generate a list of potential star players.
- **Salary Prediction Model:** Building on the All-Star prediction model, we constructed a salary prediction model that incorporates players' historical performance and salary trends. This model enables the team to predict future player salaries and compare them with current salaries, facilitating informed salary adjustment decisions.

3. Data and Methodology Overview

1) Data Description

2021-2022 NBA Player Stats - Regular.csv, 2022-2023 NBA Player Stats - Regular.csv, 2023-2024 NBA Player Stats - Regular.csv: Regular-season stats for NBA players across 2021-2024 seasons. Each row represents a player's season performance, including Player, Team, Position, and performance metrics (Points, Assists, etc.). Dimensions: 811, 677, and 733 rows with 30 columns.

	Player	Pos	Salary.2021.2022.	Salary.2022.2023.	Salary.2023.2024.	G2021	GS2021	MP2021	FG2021	FGA2021
1	Joel Embiid	C	31579390	33616770	47607350	68	68	33.8	9.8	19.6
2	Giannis Antetokounmpo	PF	39344900	42492492	45640084	67	67	32.9	10.3	18.6
3	Luka Doncic	PG	10174391	37096500	40064220	65	65	35.4	9.9	21.6
4	DeMar DeRozan	SF	26000000	27300000	28600000	76	76	36.1	10.2	20.2
5	Donovan Mitchell	SG	28103500	30913750	33162030	67	67	33.8	9.2	20.5

Table 1. Sample of 5 Rows from Dataset

2021-2024PlayersSalaries.csv: Contains salary data for 200 players, with each row representing a player's full name and annual salary details for the past 3 seasons. (208 rows, 4 columns).

Code_2_15AllStarPlayersSalaries.csv: Salary details for top players we selected from Code1 over multiple seasons. Each row represents a player's full name, position, and salaries for the past 3 seasons. (15 rows, 5 columns).

2) Brief Data Preparation

Code1 focuses on cleaning data, analyzing player metrics, and identifying correlations (e.g., weak between age and points, strong between minutes and points). Using PCA, it highlights key performance features and develops a weighted composite score to classify All-Star players with Random Forest-based predictions.

Code2 prepares player and salary data (2021–2024) by cleaning, scaling, and aggregating key metrics. It computes averages, trends, and correlations, creating a refined dataset for analysis and modeling.

Code3 processes data for players under 25, merging stats and salaries (2021–2024). It employs PCA for potential assessment, Random Forest for salary prediction, and K-means for clustering. High-potential players with favorable salary ratios are identified.

3) Data Mining Solution

All detailed explanation of dataset and code have been updated on Github:
https://github.com/KristWangCY/BusinessDataMiningProject

Code1_All-Star Player Prediction Model.R: We built a Random Forest to predict if a player can become an All-Star using key performance metrics (PTS, AST, STL, BLK, TRB, G) from the latest season as predictors and a binary target variable (All_star). PCA was applied to weight variables into a composite score, and the model's performance was evaluated with ROC and Confusion Matrix against real-world All-Star lists.

Code2_Salary Prediction Model.R: We analyzed the relationship between performance and salaries using multi-season stats as predictors and Salary.2023.2024 as the target, employing a Random Forest model. Predicted salaries for All-Star players were compared to actual salaries, revealing insights such as Anthony Edwards' high salary ratio due to his potential. Specific predictions for players like James Harden and Ja Morant were made, with recommendations based on R^2 performance evaluation.

Code3_Potential Players Model.R: We focused on players under 25, using PCA to compute a Potential_Score and K-means clustering to group players into high, medium, and low potential categories. A Random Forest model predicted future salaries, with salary ratios identifying undervalued players (ratio > 1). These models provided actionable insights for contract decisions, highlighting undervalued talents and guiding team investments.

4. Detailed Explanation of Design Choices

1) All-Star Prediction Model: The analysis from the All-Star prediction model closely matches the actual All-Star selections for that year, demonstrating the model's ability to accurately reflect player potential. Therefore, the team can use these predictions to make early adjustments to contract terms and salary levels, ensuring future stars are secured at reasonable costs while avoiding unnecessary resource expenditure.

2) Performance-Salary Prediction Model: Case Studies of Anthony Edwards and James Harden Using the salary prediction model, we analyzed the salary situations of several All-Star players and found room for adjustments in some cases.

Anthony Edwards Case: The model predicted that Anthony's salary should be 1.6 times his current pay, indicating that his performance exceeds his current salary level. Further analysis revealed that although his "minutes played" has decreased, his scoring ability has steadily improved, showing significant potential. Therefore, the team should consider increasing his salary to ensure his continued contribution to the team and prevent other teams from signing him away.

James Harden Case: With a highly accurate model, we applied it to James Harden and found that his predicted salary is only 0.82 times his current actual salary. Analyzing his performance over the past three years, we observed that while his "minutes played" has increased, his "points" has steadily decreased. This may be due to injury, so that the team should consider lowering his salary to better align with his current contributions and performance.

3) Potential Players Model: Through the Performance-Salary Prediction Model, we suddenly identified young and highly promising players like Anthony Edwards. Based on this, the project is extended to discover other young players with significant potential and predict their salaries. This will help the team secure these young players early by providing them with fair compensation, thereby supporting the team's long-term competitiveness.

These case studies demonstrate that the model effectively helps the team identify which players' salaries need adjustments, providing a strong basis for optimizing the overall salary structure.

5. Conclusion

1) Advantages

Model Accuracy: The model can accurately predict future player performance and salary needs based on extensive historical data, helping the team develop sound salary strategies.

Mitigating Financial Risks: Through efficient salary planning, the team can avoid overpaying players, thus reducing financial pressure and ensuring the team's long-term operations.

Enhancing Competitiveness: Optimizing the salary structure allows the team to retain high-potential players while maintaining flexibility, ensuring competitiveness in future seasons.

2) Limitations

Data models have certain limitations in evaluating player value. In the Performance-Salary Prediction Model, the predicted salary is only based on the player's performance, while many other factors cannot be quantified, and ethical considerations are frequently overlooked. For example, in 2013, the Lakers signed Kobe Bryant to a two-year, $48.5 million contract, despite his declining performance. This decision was based on his historical contributions and emotional impact on the team and fans, which a data model can't capture.

Moreover, the All-Star Prediction Model cannot simulate or account for a player's long-term development and personal growth. In the All-Star Prediction Model, we only used the player's data from 2023, which represents their current state. Additionally, there are many unpredictable factors on the field, such as injuries, that cannot be accounted for.

3) Recommendations

Integrated Decision-Making: It is recommended to combine the model results with the observations of management and coaching staff to ensure more comprehensive salary planning decisions. For example, management's insights into a player's influence, reputation, and other intangible factors, as well as the coaching staff's assessment of possible short-term factors like injuries, should be integrated into the decision-making process. This approach would provide a more holistic view, as seen in the case of James Harden in the model.

Dynamic Adjustments and Updates: Regularly monitor and update the model to reflect the latest player performance and market trends, ensuring the team can quickly adjust salaries and optimize the structure to remain competitive in a constantly changing environment.