Evidence vs Eminence: A Machine Learning Based Approach to IPL Auction Price Prediction

Sarang Deb Saha
Dept. of Computer Science and Engineering
The LNM Institute of Information Technology
Jaipur, India
20ucs173@lnmiit.ac.in

Abstract:
The Indian Premier League (IPL) has become a global phenomenon, with astronomical sums involved in player auctions, understanding the factors that influence player prices is of paramount importance. In this study, we explore the IPL auction market, employing statistical analysis and machine learning models to uncover insights into player valuation and predict auction prices. Our analysis reveals intriguing patterns in player prices, including the influence of player performance metrics, age, captaincy experience and more. We observe correlations between specific player attributes and their auction prices, shedding light on the IPL's pricing dynamics. Moreover, we utilize regression models to predict auction prices, allowing stakeholders to make informed decisions during player auctions. By bridging the gap between tradition and data-driven decision-making, this study offers a new perspective on the IPL auction ecosystem.

Introduction:
The Indian Premier League (IPL) has emerged as one of the most popular and lucrative cricket leagues globally since its inception in 2008. With a fusion of cricketing talent, electrifying matches, and fanatical team loyalties, the IPL has transformed the cricketing landscape. A significant aspect of this league is the player auction, where teams bid for players with diverse cricketing backgrounds, skills, and experiences. In this comprehensive analysis, we delve into the intriguing world of IPL auctions between 2008 and 2013. Our primary focus is to understand the factors influencing the sold prices of players and to develop predictive models that can estimate a player's auction price based on their attributes and performance metrics. Our journey begins with data collection, where we compile detailed statistics about each player, including their age, playing role, international experience, past IPL performance, and more. Armed with this dataset, we embark on an exploratory data analysis (EDA) to uncover trends, correlations, and valuable insights to train our own model for a fairer price tag for these athletes.

Throughout this analysis, we investigate several key questions:

- Player Nationality: Does a player's nationality impact their auction price? Are certain nationalities more sought after in IPL auctions?

- Playing Role: How do different playing roles (batsmen, bowlers, all-rounders, and wicket-keepers) influence player prices? Are all-rounders more highly valued?

- Performance Metrics: Which performance metrics, such as batting averages, strike rates, or bowling economy, correlate most strongly with sold prices?

- Captaincy Experience: Does captaincy experience in domestic or international cricket play a role in a player's auction price?

- Age and Experience: How do age and experience levels affect the prices players command in the auction?

Using statistical models, including Linear Regression, Ridge Regression, and Random Forest Regression, we aim to predict player auction prices. By employing these techniques, we strive to enhance our understanding of the IPL auction dynamics and provide valuable insights to team management, cricket analysts, and enthusiasts. Our analysis is not only an exploration of data but also a journey into the fascinating world of cricket economics, where data-driven decisions are transforming the way teams assemble their squads in one of the world's most exciting sporting events.

Literature Survey:

The modeling of various dimensions of sports and auctions in sports, especially within the context of cricket and the Indian Premier League (IPL), has garnered considerable attention in recent years. Some of the work offer valuable perspectives and methodologies for understanding the dynamics of IPL cricketers' auctions and the factors influencing auction prices. Rastogi and Deodhar (2009) delves into IPL auctions, presenting a comprehensive analysis of the bidding behavior among different franchisees. The authors employ a Hedonic Price Modeling approach, considering both cricketing and non-cricketing attributes as determinants of final bid prices. Notably, they highlight several non-cricketing attributes, such as age, nationality, and franchisee behavior, which significantly impact players' final bid prices. For instance, Indian players command a premium over their Pakistani counterparts, and specific franchisees consistently place higher bids than others. Cricketing attributes like batting strike rate in One-Day Internationals (ODIs), half-centuries, and wickets taken also contribute to players' auction values. While their pioneering work provides insights into IPL auction dynamics, the model includes a multitude of variables, which may affect its predictive power.

Another relevant study focused on player compensation and quality attributes reveals the significance of product attributes and their influence on consumer preferences and market prices. This concept, rooted in economic theory, treats products or services as collections of attributes that differentiate them in the marketplace. It has gained prominence through various researchers, including Waugh (1928) and Rosen (1974). These scholars emphasized that consumers evaluate product quality attributes and are willing to pay prices that reflect the implicit values of these attributes. Furthermore, the research underscores the influence of reputation on prices, highlighting that reputation-building can be considered an investment in achieving a price premium.

The insights from these sources serve as a foundation for the present research. By leveraging a unique dataset and methodology, we aim to recalibrate existing models and improve predictive power in understanding IPL cricketers' auction prices. We consider attributes that significantly affect players' final bid prices, thereby contributing to a more comprehensive understanding of this intriguing phenomenon.

About the dataset:

The dataset used in this analysis is a comprehensive collection of player statistics and auction-related information for the Indian Premier League (IPL) seasons from 2008 to 2013. This dataset provides a detailed and granular view of the players who participated in the IPL during this period and includes a wide range of attributes and metrics that are crucial in understanding player performance and market valuation.

The dataset contains 25 attributes like name, age, nationality, role, ODI runs, strike rate, economy, wickets etc. for 130 players.

Note: Age has been represented as a group, namely 1,2 & 3 and not in years and Sold Price is in USD ($) and not INR (₹).

Exploratory Analysis:

1) Focusing on Player Trends

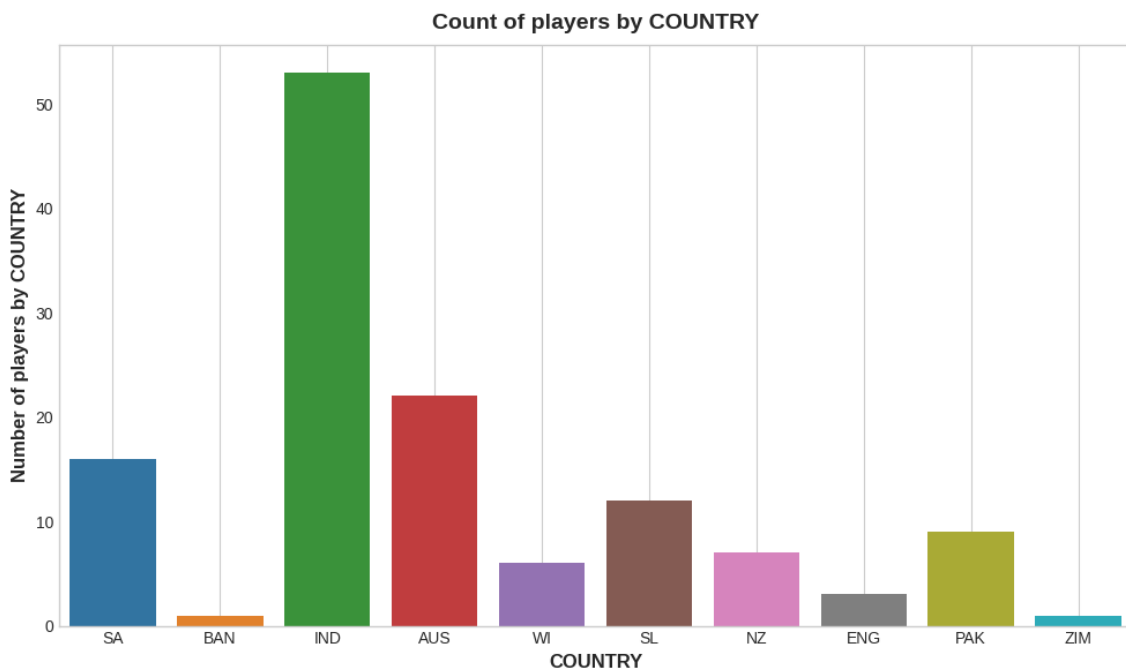


Fig. 1. Number of Players by Country in the IPL

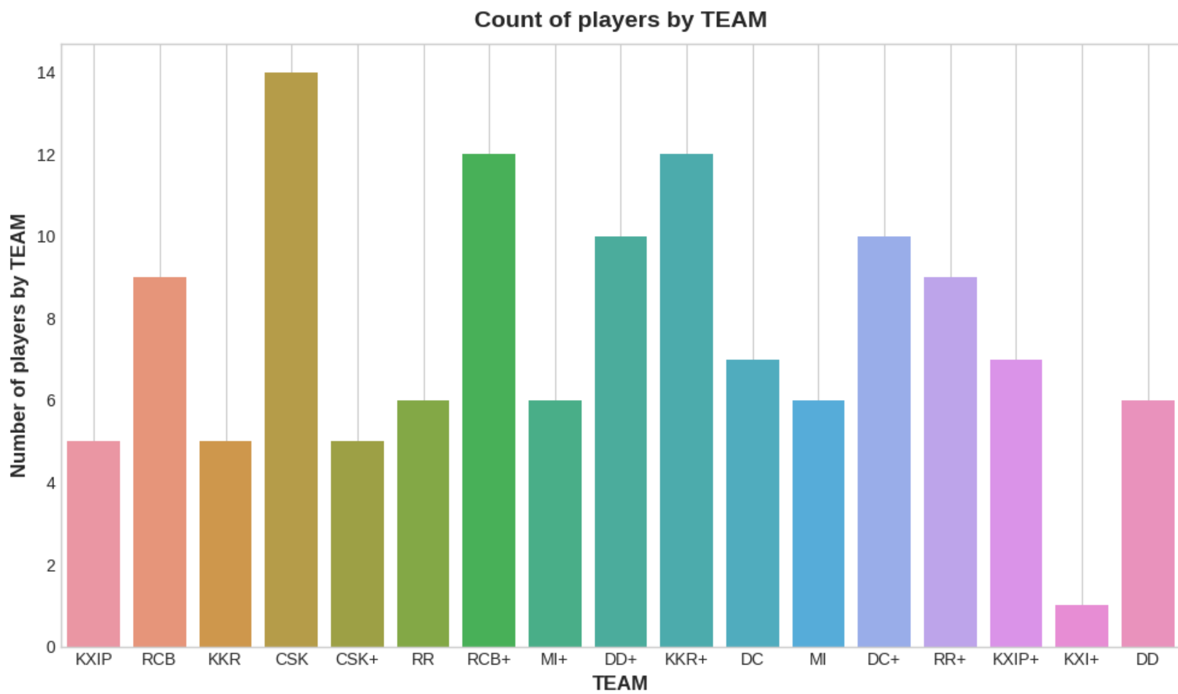


Fig. 2. Number of Players by Team in the IPL

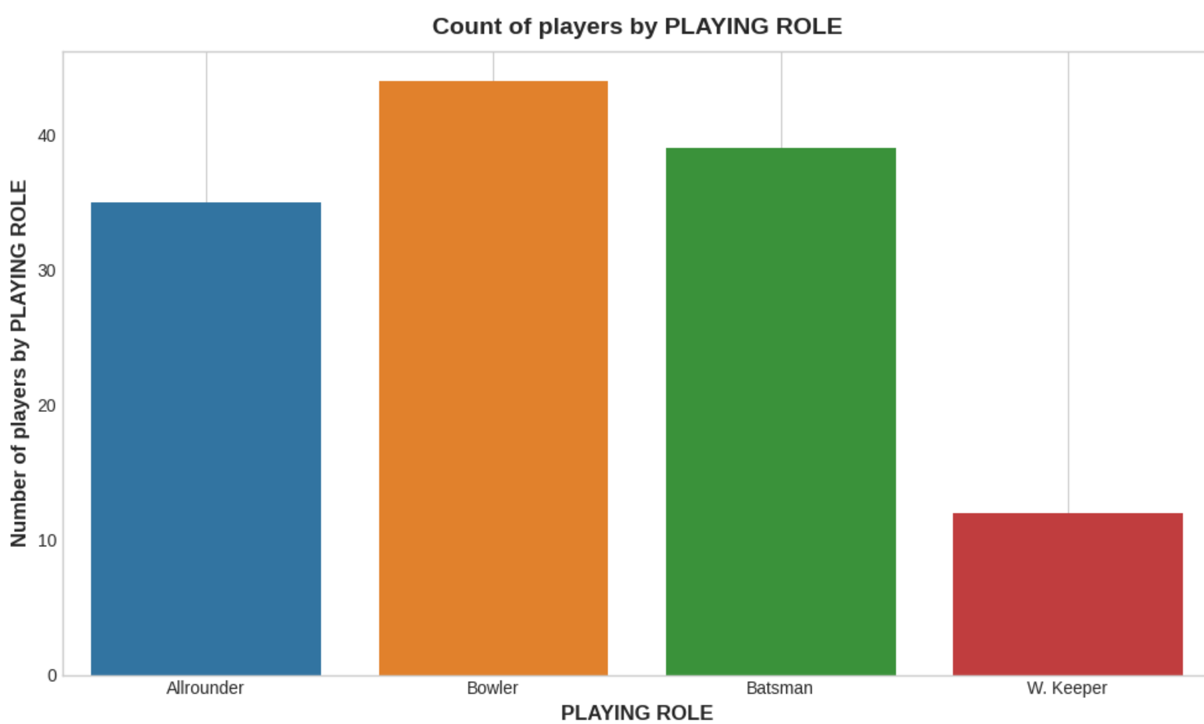


Fig. 3. Number of Players by Role in the IPL

Key Insights from the EDA:

- ***Nationality Influence on Sold Prices:***

1) Indian players were the most sought-after in IPL auctions, with the highest number of players in the league.
2) Australian and South African players also commanded significant attention from IPL teams, indicating their popularity and performance in the tournament.

- ***Preferred Playing Role:***

1) Among playing roles, Bowlers emerged as the favorites among IPL teams followed by Batsmen and Allrounders. This preference highlights the importance of strong bowling line-ups in T20 cricket.

- **Team Composition:**

1) Chennai Super Kings (CSK) had the largest representation of players during the analyzed period. Their consistent presence showcased the team's stability and success in the IPL.
2) Royal Challengers Bangalore (RCB) and Kolkata Knight Riders (KKR) also had substantial player representation, reflecting their competitiveness and popularity among fans.

2) Focusing on *Sold Prices*

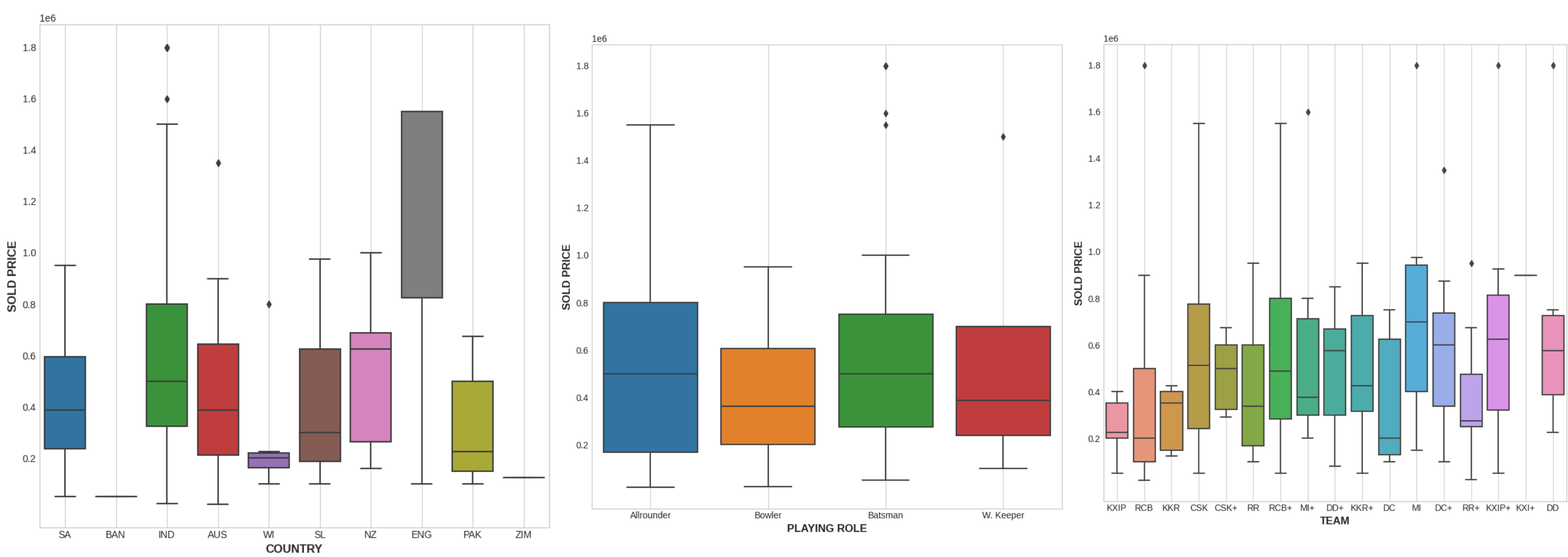


Fig. 4 Range of *Sold Price* by Country, Role and Team

Insights from Boxplots (Price Ranges)

- ***Mumbai Indians Premium:***
  The Mumbai Indians (MI) team appears to have paid significantly higher prices for players compared to their competitors in the IPL.

- ***Premium for English Players:***
  English players have fetched premium prices, with their price range starting at a significantly higher point compared to players from other countries.

- ***Economical West Indian Players:***
  In contrast, West Indian players have been relatively more affordable in the IPL auctions, with their price range starting at a lower point.

- ***All-rounders Command High Prices:***
  Players identified as all-rounders tend to have the widest price range, indicating that teams are willing to invest heavily in versatile players who can contribute both with bat and ball.

- ***Batsmen's Price Premium:***
  Batsmen have been sold at higher price ranges compared to wicket-keepers (Wk) or bowlers. This suggests that teams highly value strong batting performance and are willing to pay a premium for skilled batsmen.

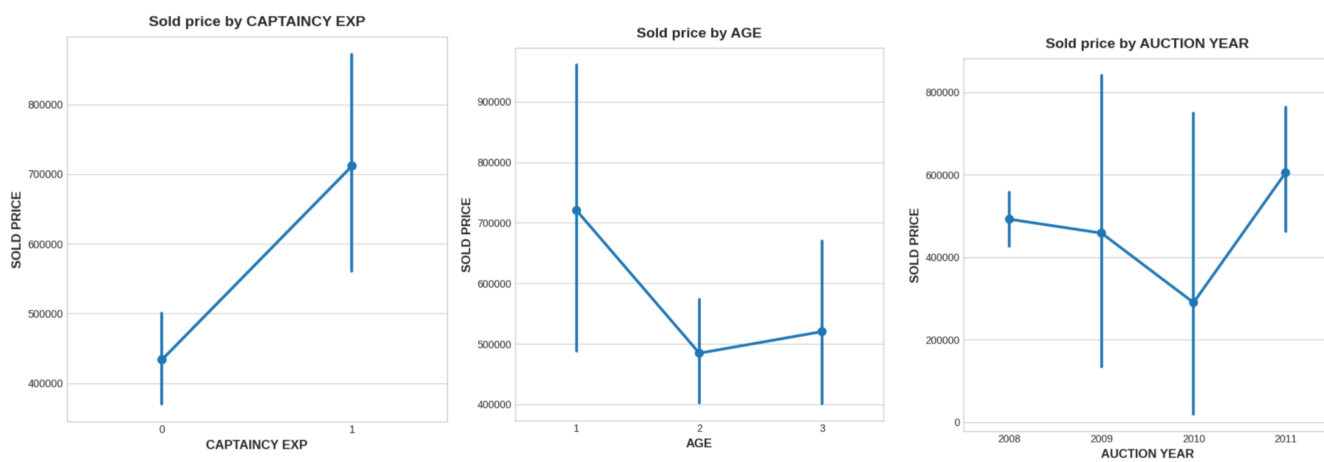


Fig. 5 Central tendencies of *Sold price* by CAPTAINCY EXP, AGE, AUCTION YEAR

Captaincy Experience Impact on Sold Prices:
Players with prior captaincy experience tend to be sold at higher prices in the IPL auctions. This suggests that leadership skills and experience in leading a team may contribute to a player's perceived value.

Age Group Influence on Sold Prices:
The age group of a player appears to have an impact on the sold price. Players in different age groups are likely to be sold at varying price ranges. This indicates that teams consider the age of players as a factor when determining their value in the auctions.

Auction Year's Limited Impact:
In contrast to age and captaincy experience, the auction year does not seem to have a significant impact on the sold prices of players. This implies that the year of the auction is not a crucial determinant of a player's value, and teams primarily focus on factors like performance and experience when making their bids.

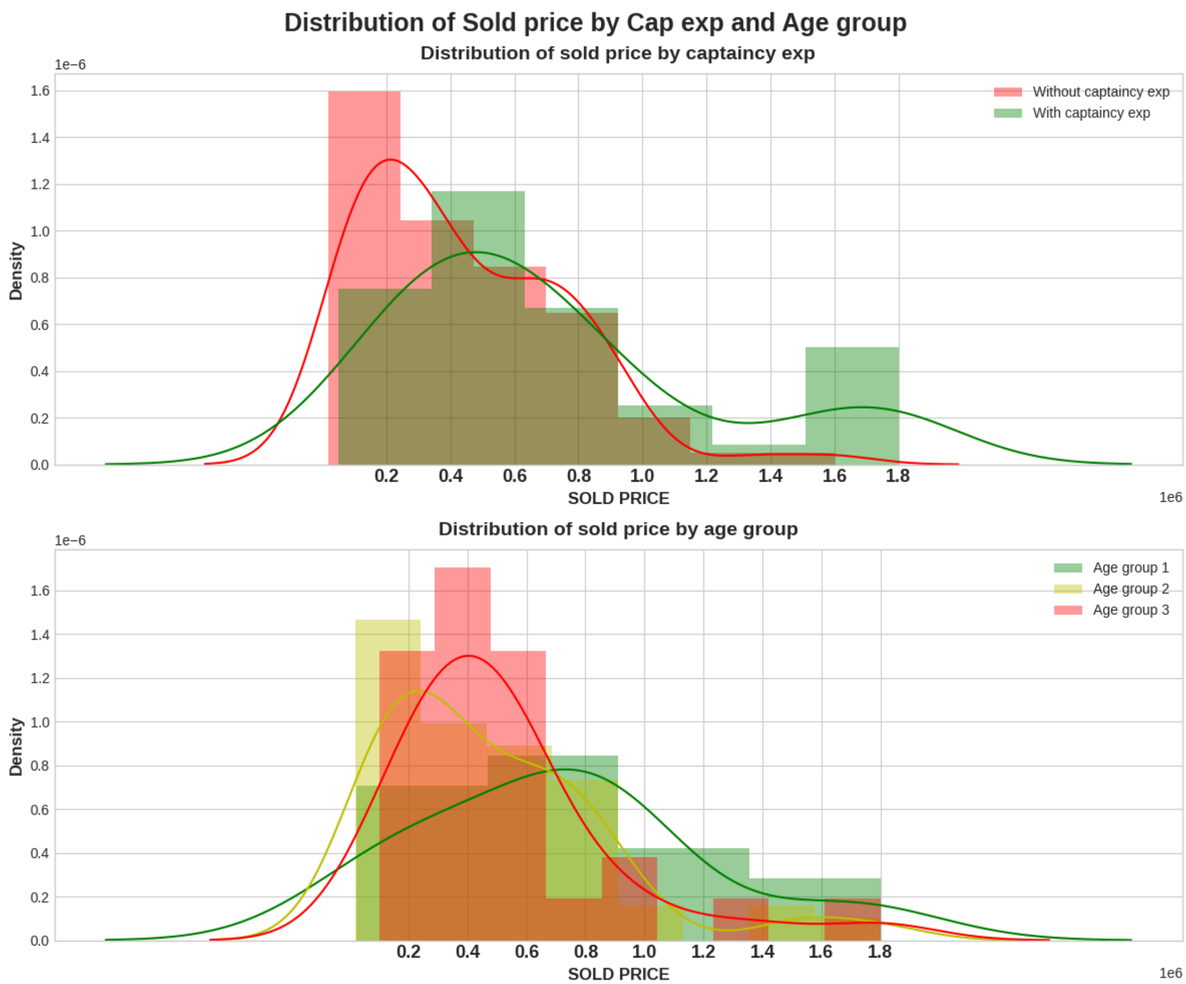
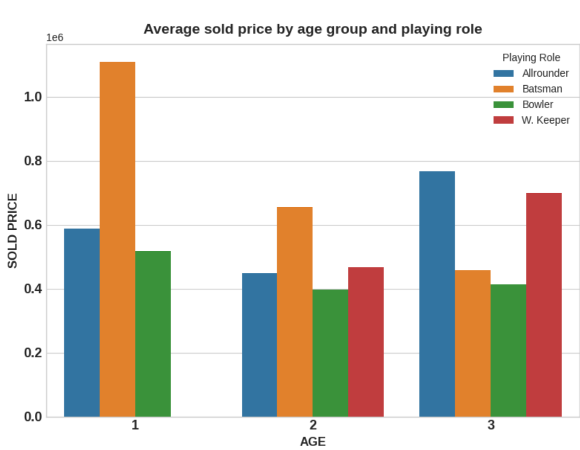


Fig. 6 Distribution of *Sold Price* by Captaincy exp and Age group



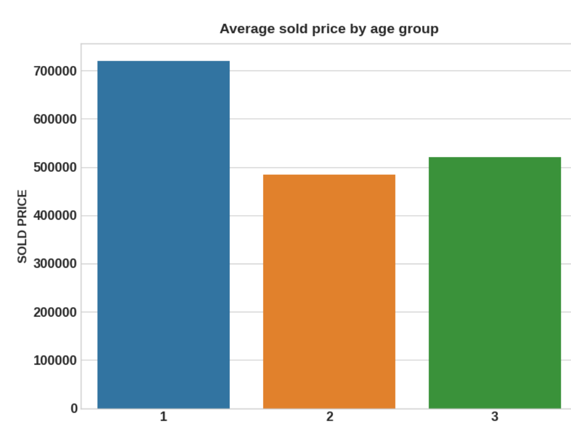


Fig. 7 Average *Sold Price* by Age group and Role **Fig. 8 Average *Sold Price* by Age group**

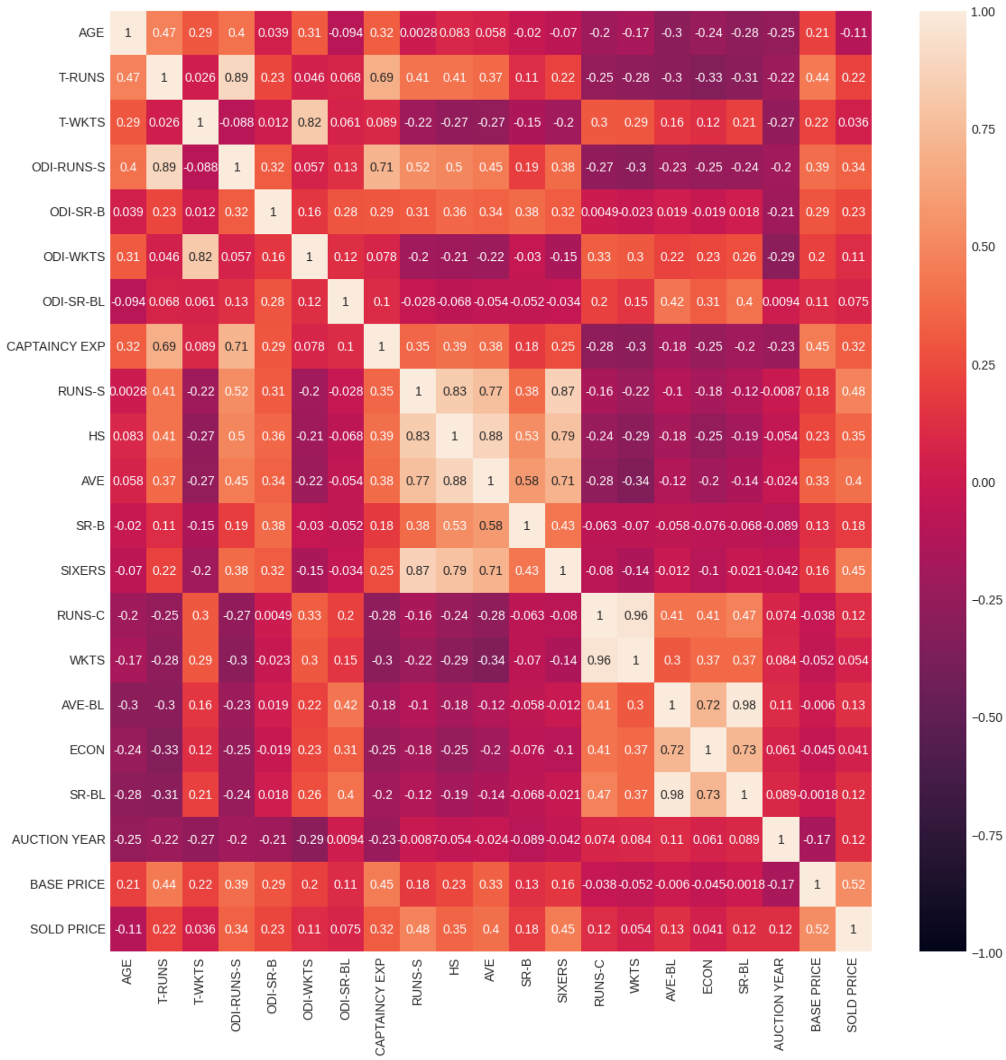


Fig. 9 Correlation plot

Key Correlation Inferences:

- ***Strong Positive Correlations:***
    1) There exists a strong positive correlation between several batting performance metrics, including 'Run-s' (Runs Scored), 'HS' (Highest Score), 'AVE' (Average), 'SR-B' (Batting Strike Rate), and 'SIXERS' (Number of Sixes).
    2) 'SIXERS' and 'Run-s' exhibit strong positive correlations with the 'Sold Price' of players. This suggests that players who score more runs and hit more sixes tend to command higher auction prices.

- ***Base Price Impact:***
    1) The 'Base Price' of players shows a positive correlation with their 'Sold Price.' This implies that a higher base price generally results in a higher final selling price during the auction.
- ***Moderate Correlations:***
    1) 'ODI-SR-BL' (Bowling Strike Rate in ODIs), 'Captaincy Exp' (Captaincy Experience), and 'ODI-runs-S' (Runs scored in ODIs) exhibit moderate correlations with the 'Sold Price' of players.

Identifying the Outliers:

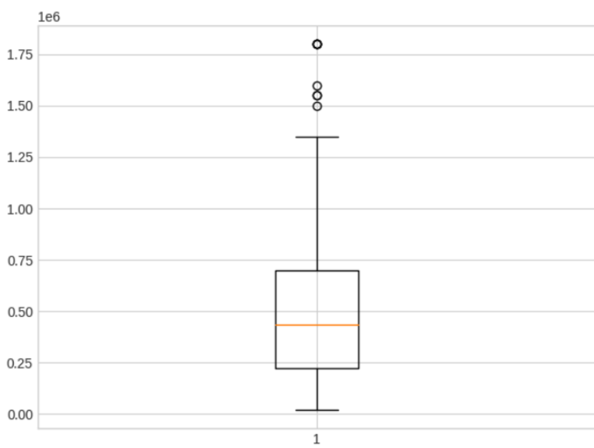


Fig. 10 Boxplot of *Sold Price*

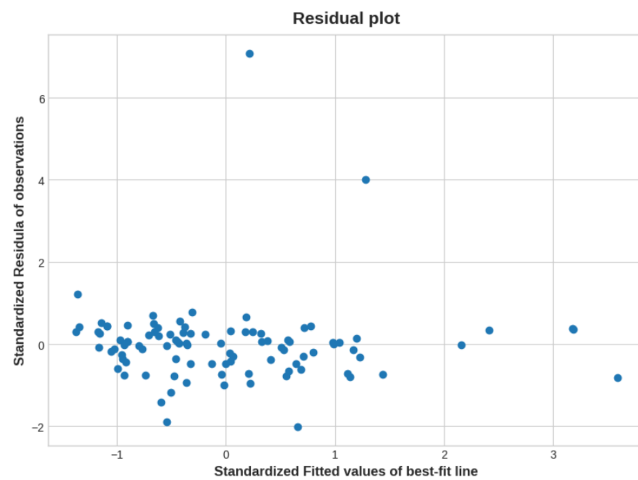


Fig. 11 Residual Plot

The 25th quartile is 225000.0 and 75th quartile is 700000.0 with the median of the sold price at 437500.0, minimum sold price is at 20000 and maximum sold price is at 1350000. Thus we set the players upwards of this value as outliers. As per the residual-plot there are a few outliers that might affect the training of our model.

Tabulating these outliers:

Sl. No.	Player Name	Age	Playing Role	Sold Price
51	Kohli, V	1	Batsman	1800000 USD
94	Sehwag, V	2	Batsman	1800000 USD
112	Tendulkar, SR	3	Batsman	1800000 USD
128	Yuvraj Singh	2	Batsman	1800000 USD
114	Tiwary, SS	1	Batsman	1600000 USD

24	Flintoff, A	2	Allrounder	1550000 USD
84	Pietersen, KP	2	Batsman	1550000 USD
16	Dhoni, MS	2	W. Keeper	1500000 USD

Table 1 Outlier data

Kohli was sold at the highest price of his age group back in 2013.
Note- These outliers were removed from our training corpus.

Model Inferences and Results:

1) Linear Regression: This model assumes a linear relationship between the independent variables and the dependent variable (sold prices, in this case). It aims to find the best-fitting linear equation to make predictions.

   $Y=\beta_0+\beta_1 X_1+\varepsilon$

   Where:
   ***Y** is the predicted sold price*
   ***β_0** is the intercept*
   ***β1** is the coefficient for the independent variable X_1*
   ***X_1** is the value of the independent variable (feature)*
   ***ε** is the error term*

2) Ridge Regression: Ridge regression is a variant of linear regression that introduces regularization to prevent overfitting. It adds a penalty term (L2 regularization) to the linear regression equation.
3) Random Forest: Random forest regression is an ensemble learning method that combines multiple decision tree regressors to make predictions. It is robust and capable of capturing complex relationships between features and the target variable.

Evaluation metrics:

1) RMSE: It measures the average deviation between the predicted auction prices and the actual auction prices. Lower RMSE values indicate better model performance.
2) R2 Score: R2 measures the proportion of the variance in the dependent variable (auction prices) that is predictable from the independent variables (player attributes). It ranges from 0 to 1, where higher values indicate a better fit of the model to the data.

The selected features and their values for our sample players P1 and P2 are:
P1->
AGE (Age Group): 1
HS (Highest Score): 73
AVE (Average Runs Scored by Batsman in IPL): 28.86

SIXERS (Number of Six Runs Scored by Player in IPL): 49
RUNS-S (Number of Runs Scored by Player): 1639
SR-B (Batting Strike Rate in IPL): 119.29
CAPTAINCY EXP (Captaincy Experience): 1 (Indicating the player has captaincy experience)
ODI-RUNS-S (Number of Runs Scored in ODIs): 3590
BASE PRICE: 150,000

P2->
AGE (Age Group): 3
HS (Highest Score): 116
AVE (Average Runs Scored by Batsman in IPL): 39.92
SIXERS (Number of Six Runs Scored by Player in IPL): 25
RUNS-S (Number of Runs Scored by Player): 958
SR-B (Batting Strike Rate in IPL): 120.65
CAPTAINCY EXP (Captaincy Experience): 1 (Indicating the player has captaincy experience)
ODI-RUNS-S (Number of Runs Scored in ODIs): 5262
BASE PRICE: 250,000

The feature values of these players might be recognizable to some!

Model Name	**RMSE**	**R2 Score**	**Predicted *Sold Prices* (P1)**	**Predicted *Sold Prices* (P2)**
Linear Regression	87,902.76	96.371	1,008,542.52 USD	417590.92 USD
Ridge Regression	84,222.51	96.655	1,007,628.89 USD	417807.77 USD
Random Forest Regression	127,589.38	91.522	1,354,261.90 USD	453190.48 USD

Table 2 Comparison of our models

Player ID	**Player Name**	**Actual *Sold Price***
P1	Virat Kohli	1,800,000 USD
P2	Michael Hussey	250,000 USD

Table 3 Player (P1 & P2) Information

Inferences:

- The Linear Regression and Ridge Regression models have similar RMSE values, indicating that they provide relatively accurate predictions of sold prices. The R2 scores for both models are also quite high, indicating a strong correlation between the predicted and actual prices.

- The Random Forest Regression model, while having a lower RMSE than the other models, has a slightly lower R2 score, suggesting that it may not fit the data as closely as the linear models. However, it still provides reasonably accurate predictions.

- For the specific player, Virat Kohli, the models' predictions fall short of the actual auction price. The models predicted a sold price in the range of approximately 1,008,000 to 1,008,500, while Virat Kohli's actual auction price was 1,800,000.
  This suggests that the models may not fully capture the premium price often due to sentiment, ego-clashes and/or bidding wars associated with high-profile players like Virat Kohli. Also a problem that the present/traditional method suffers from is that opponents are often aware of the team's favorite player due to regional popularity. This knowledge has often been abused by the opponents to inflate the price of the desired players. Also since we excluded player sales of such stars from our training corpus, we have now observed the performance of our model on outlier data as well.

- For the specific player, Michael Hussey, the models' predictions fall short of the actual auction price. The models predicted a sold price in the range of approximately 417,000 to 417,500, while Hussey's actual auction price was 250,000. Mike Hussey's story in the IPL serves as a compelling example of an underdog purchase. Despite being acquired at a relatively lower price compared to some star players, he went on to become one of the most successful and consistent performers in the league's history. This underlines the fact that in cricket and the IPL, it's not always the highest-priced players who make the most significant impact.

  Probable reasons for Hussey's *Actual Sold Price* Being lower than predicted:

  - ❖ Baby steps for IPL: In the initial seasons of the IPL, the league was still evolving, and players' valuations were not as precise as they are today. Mike Hussey's lower sold price in the 2008 auction can be attributed to the relative uncertainty and less mature player valuation processes during that time.

  - ❖ Lack of International Stardom and Marketability: While Hussey was a respected international cricketer for Australia, he may not have had the same level of stardom or recognition as some other players in the IPL and T-20 cricket. This could have influenced his auction price downward.

  - ❖ Team Strategy: The Chennai Super Kings (CSK), who eventually purchased Mike Hussey, might have had a strategic approach that prioritized value-for-money acquisitions. Their decision to sign Hussey at a lower price could have been a tactical move to optimize their squad composition.

  - ❖ Relative Unknown: Sometimes, players who are not widely recognized but possess exceptional skills become underrated gems. Hussey's initial lower valuation may have been a result of teams not fully recognizing his potential and the impact he could have in the IPL.

  - ❖ It's noteworthy that Hussey was originally bought for this price in the 2008 IPL auction. Subsequently, he was retained by the Chennai Super Kings and achieved significant success in the league, notably winning the prestigious Orange Cap, awarded to the highest run-scorer of the season, during the 2013 edition of the IPL. Hussey's journey with CSK showcased the value of astute team management, scouting, and recognizing talent that might be undervalued by the auction process. It reinforces the unpredictability and excitement that the IPL brings, where players like Hussey can emerge as legends from under the radar.

These results demonstrate the predictive capabilities and need for different regression models in estimating player prices in IPL auctions. While they provide valuable insights, factors like player reputation and bidding dynamics can significantly impact the final auction price, as seen in the case of Virat Kohli.

Conclusion:

Our exploration of the dataset revealed intriguing insights into player distribution, team preferences, and trends in player auctions. Notably, we observed that Indian players are highly sought after, followed closely by Australians and South Africans, underlining the diversity of talent in the IPL. Moreover, the prominence of bowlers as a favorite choice for teams echoes the game's ever-changing dynamics. Captaincy experience emerged as a valuable attribute affecting player valuation, while age and auction year exhibited varying degrees of influence. These findings underscored the importance of a player's cricketing journey in determining their worth in the IPL. Our regression models, including Linear Regression, Ridge Regression, and Random Forest Regression, provided predictions that often aligned closely with actual auction prices. The Ridge Regression model performed the best according to our evaluation metrics. However, our study also revealed the limitations of these models, particularly when valuing star players like Virat Kohli or Mike Hussey. Such discrepancies emphasized the role of market sentiment, emotional factors, and team strategies that go beyond pure statistical modeling. The Virat Kohli and Mike Hussey examples illustrated the underdog nature of certain purchases, highlighting the influence of external factors in determining auction outcomes. This demonstrated that IPL auctions are not merely driven by data but are a product of the unique blend of cricketing knowledge, market dynamics, and team strategies. Future work would build upon this work and use data from the more recent seasons for more context and relatable economics. Also a recommendation engine for building a scouting network is currently in the pipeline.

In conclusion, the need for predictive models in IPL player auctions is undeniable, as they provide valuable insights and serve as decision support tools. However, they must be viewed as complementary to the art of player valuation, rather than replacements. The evolving nature of the IPL auction market necessitates adaptable models that can incorporate both data-driven insights and market expertise. Our study emphasizes the importance of data quality, feature engineering, and continuous model refinement to enhance predictive accuracy. Ultimately, the IPL auctions remain a fascinating arena where the worlds of cricket and business converge, creating a captivating spectacle that continues to evolve with each passing season. As we look forward to the next IPL auction, we acknowledge that while data and algorithms play a significant role, the heart of the game still lies in the prowess of the players on the field and the enthusiasm of fans in the stands.

References:

1. A. Gupta, "India and the IPL: Cricket's Globalized Empire. The Round Table," 2009.
2. C. Barrett, "Big Bash League jumps into top 10 of most attended sports leagues in the world," The Sydney Morning Herald, 10 January 2016. [Online]. Available: https://www.smh.com.au/sport/cricket/big-bash-leaguejumps-into-top-10-of-most-attended-sports-leagues-in-theworld-20160110-gm2w8z.html. [Accessed 7 January 2018].
3. Rastogi, S.K. and Deodhar, S.Y. (2009) "Player Pricing and Valuation of Cricketing Attributes: Exploring the IPL Twenty-Twenty Vision", Vikalpa, 34(2), pp. 15-23.
4. Saikia, H. and Bhattacharjee, D. (2011) "On Classification of All-rounders of the Indian Premier League (IPL): A Bayesian Approach", Vikalpa: The Journal for Decision Makers, 36 (4), pp. 51-66.
5. V. Staden, "Comparison of cricketers' bowling and batting performances using graphical displays," 2009.
6. Swartz, T. B. (2011) "Drafts versus Auctions in the Indian Premier League: Theory and
7. Methods", South African Statistical Journal, 45 (2), pp. 249-272.
8. Montgomery, Douglas C., Elizabeth A. Peck, and G. Geoffrey Vining. Introduction to linear regression analysis. John Wiley & Sons, 2021
9. McDonald, Gary C. "Ridge regression." Wiley Interdisciplinary Reviews: Computational Statistics 1.1 (2009): 93-100.
10. Segal, Mark R. "Machine learning benchmarks and random forest regression." (2004).
11. A. A. A. Rupai "Predicting Bowling Performance in Cricket from Publicly Available Data," 2020.
12. A. Adhikari "An innovative super-efficiency data envelopment analysis, semi-variance, and Shannon-entropybased methodology for player selection: evidence from cricket," 2020.