**Is there a relation between the performance of IPL players and the amount they are paid in the auction? Which teams have the best player performance to spending ratio?**

**Introduction**

The Indian Premier League (IPL) is considered to be one of the finest Twenty20 competition modelled on the lines of the National Basketball Association (NBA) of USA and English Premier League (EPL) of England (Singh, 2011). The first season in 2008 was played among eight teams representing eight cities and the franchisee rights of these teams were auctioned-off to successful bidders for ten years. Some of the successful bidders included Reliance Industries and India Cements, which own the teams Mumbai Indians and Chennai Super Kings teams respectively. The IPL became the first cricketing event in which players were bought through auctions. Cricketers around the world registered themselves for the auction pool with or without their base price set by BCCI. Each franchise has a common spending limit at the IPL auction so affluent team owners could not buy the entire list of best players. Each team owner bids on players from the pool to consist of batsmen, bowlers and all-rounders and there are five different ways a franchise can acquire players, namely through annual auction, buying domestic players, signing uncapped players, buying replacements (for players who are not available) and trading. In the annual auction process, highest bidder on a player signs a fixed three-year contract. It is an extremely difficult decision for a franchise owner to buy a talented team with constrained spending, personal player preferences, status quo and meeting team composition rules set by the BCCI (Kalgotra et al, 2013).

Cricket is one of the most watched sports in India and thus the Indian Premier League, popularly known as IPL is a big deal in the country. Each team represents a state of India and constitutes of 11 players of which 4 are overseas players and the rest are Indian players. Each year before the tournament begins, an auction is conducted where the players are shuffled and bought by different teams. While some players are allowed to be retained by the team, other are up for grabs.

**Objective of Study**

In this study, we proposed to research on the performance of the batsmen and the amount at which they were picked at the auction in that particular year. I will try to establish if there is a relation between the auctioned amount and the players performance. I would also want to compare the different teams based on their spending and player performance to see which team had the best spending to performance ratio.

**Data Description**

The data set was obtained from Kagle that contain information about IPL players performance and auction price. The dependent or response variable is player auction price and the independent or explanatory variables are players runs, average, fifties, strike rake, high score, fours and sixes.

**Data Sources**

IPL Player Auction Dataset - From Start to Now https://www.kaggle.com/datasets/kalilurrahman/ipl-player-auction-dataset-from-start-to-now IPL Batsman Stats (2008 to 2022)

<https://www.kaggle.com/datasets/apkaayush/ipl-batsman-stats2008-to-2022>

**Research Hypothesis**

To verify the relationship between response variable and explanatory variables a hypothesis is formed on the bases of the data collected

There are seven hypothesis that are being tested:

**H1:** The explanatory variable (player runs) has impact on response variable (Auction price)

**H2**: The explanatory variable (player Average) has impact on response variable (Auction price)

**H3**: The explanatory variable (player Strike rate) has impact on response variable (Auction price)

**H4**: The explanatory variable (player fifties) has impact on response variable (Auction price)

**H5**: The explanatory variable (players high score) has impact on response variable (Auction price)

**H6**: The explanatory variable (player fours) has impact on response variable (Auction price)

**H7**: The explanatory variable (players sixes) has impact on response variable (Auction price)

The hypotheses we proposed concern the relationship between several explanatory variable variables and the response variable, which is the auction price. To determine whether these hypotheses are plausible, we should consider the theoretical and empirical reasons for each one.

**Analysis Method:**

Regression analysis and correlation analysis are used in this research to check the impact of explanatory variables on response variable and degree of association between them. The data are analyzed using the statistical package RStudio.

**The model**

The multiple linear regression model is used in this research

***++++++i***

Where = the response variable

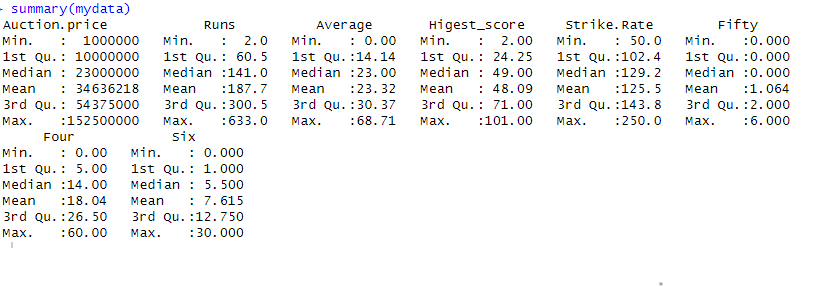
= the intercept

***, , ,, ,****=* the regression coefficient

*µi=* residual for the ith unit.

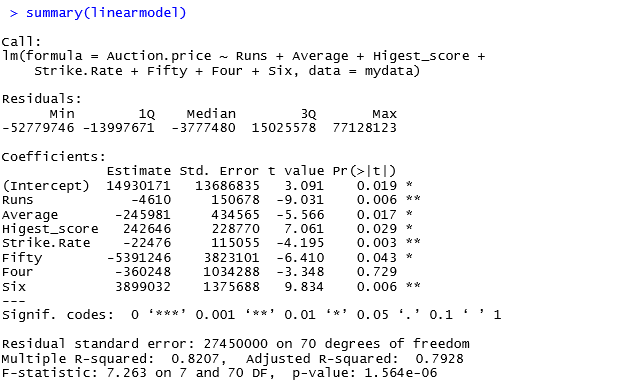
**Descriptive statistic**

**Table1: Descriptive statistic**



The data reveals diverse performances among IPL batsmen, with auction prices ranging from 1,000,000 to 152,500,000. On average, players are auctioned for approximately 34,636,218, with most falling within the 10,000,000 to 54,375,000 range. Batsmen score an average of 187.7 runs, ranging from a minimum of 2 to a maximum of 633. The batting average spans from 0.00 to 68.71, with an average of 23.32. The highest individual scores fluctuate between 2 and 101, averaging at 48.09. Strike rates vary from 50.0 to 250.0, averaging 125.5. Batsmen hit an average of 1.064 fifties, with a range of 0 to 6. Fours and sixes hit by players average at 18.04 and 7.615, respectively. These statistics provide a comprehensive glimpse into the performance metrics and auction prices of IPL batsmen, offering insights into their diverse skills and market valuations.

**Table2: Multiple linear Regression Result.**

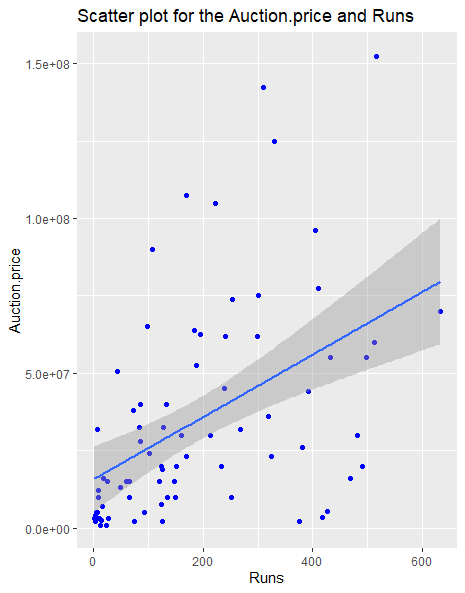


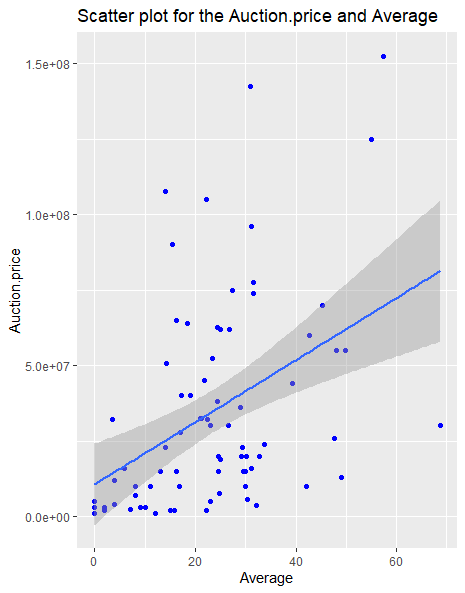
**Multiple Regression Assumptions**

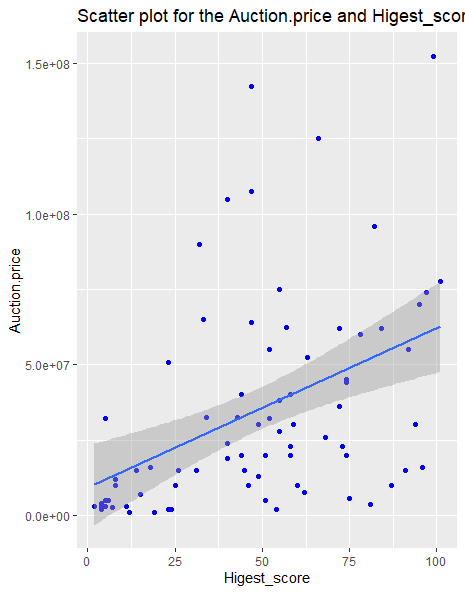
When performing a multiple linear regression analysis to assess the relationship between auction price and various batsman performance metrics, it's essential to check several assumptions. These assumptions help ensure the reliability and validity of the regression model.

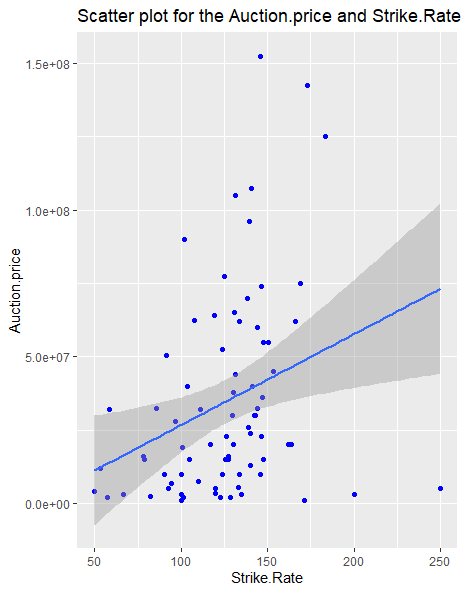
**Linearity:**

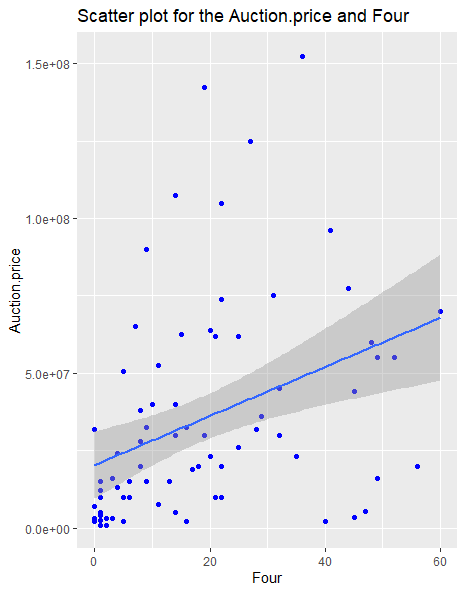
Linearity means the relationship between response variable and each explanatory variables is linear. To check the assumption of we can create scatter plot of each explanatory variable against the response variable (Auction Price)

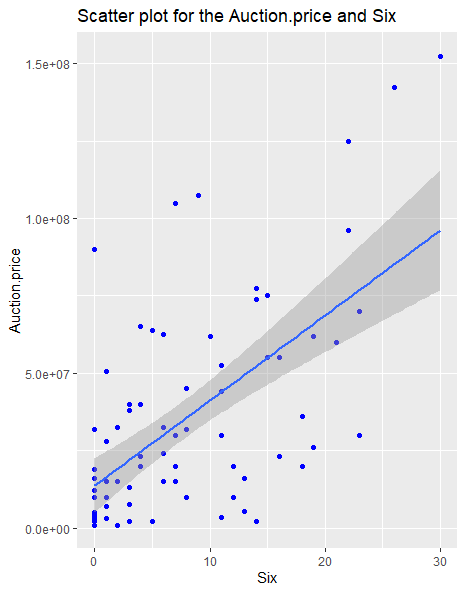








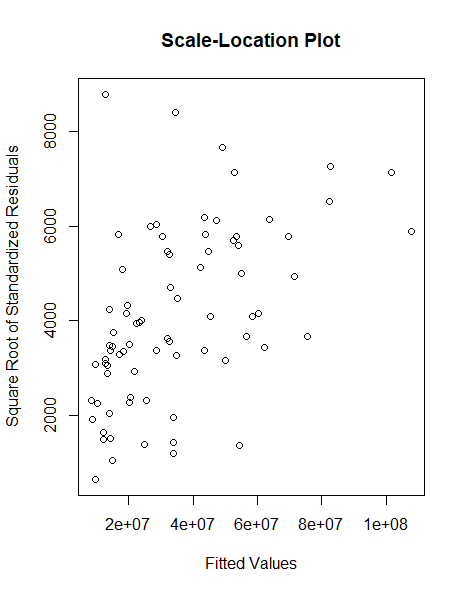




The above all graph shows a linear relationship between response variable (auction price) and explanatory variables. This means that there is a positive correlation between the two variables, but the relationship is not perfect. It means that the assumption of linearity is are not violated.

**Homoscedasticity**

The homoscedasticity mean that error term or residuals are constanst across all level of the explanatory variables. To check homoscedasticity we plot residuals against predicted values and check for a consistent spresd.

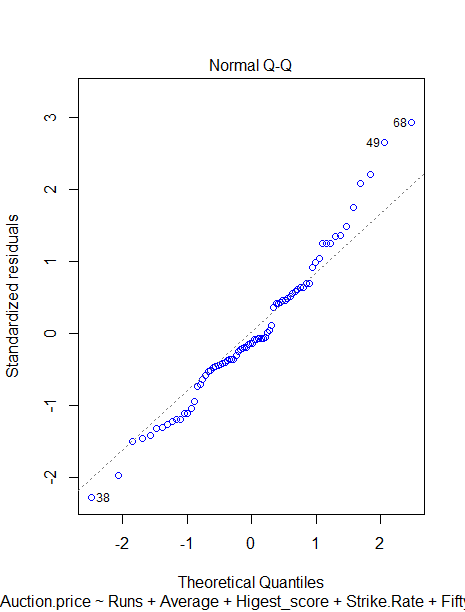


The scatter plot show homoscedasticity. The standardized residuals are not more spread-out , suggesting that the variance of the error terms is constant. Plot suggests that the homoscedasticity assumption is met.

**Normality of Residuls**

The normality of residuals mean that the residulas are normaliy distributed.

. We can check normality assumption by crearting histogram and QQ plot.



The above Q-Q plot shows that the normality of residuals assumption is satisfied. The data points fall closely along the diagonal line, which indicates that the residuals are normally distributed.

**No Multicollinearity**

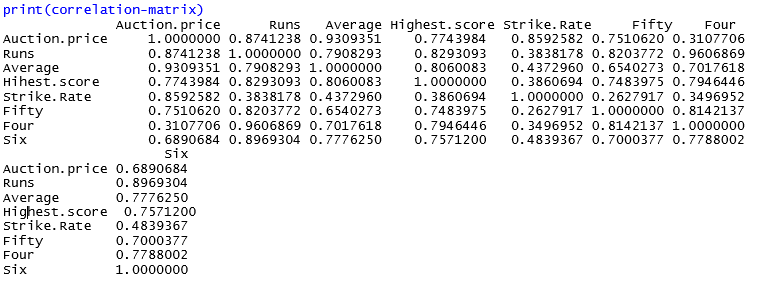
The explanatory variables are not highly correlated with each other. We can check the multicollinearity by variance inflating factor (VIF).

**Variance inflating factor result.**

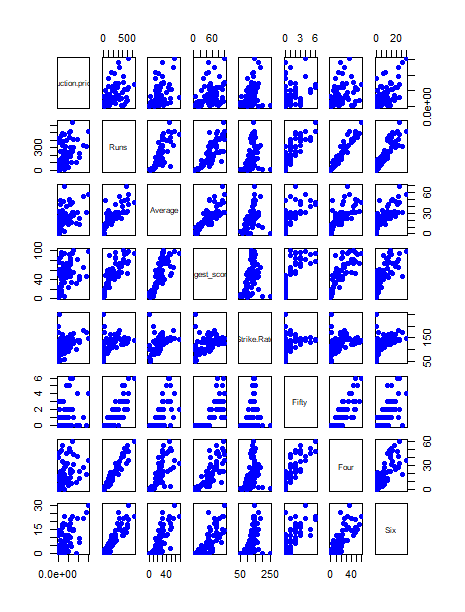


In the above result can see, all VIF values are below 5, which is a common rule of thumb for indicating no serious multicollinearity concerns. However, several variables have VIF values between 2 and 5, which suggests weak multicollinearity. This means that these variables are somewhat weakly correlated with each other, which could not affect the precision and stability of our model estimates

**Table3: Correlation Matrix**



**Plot of Pair wise correlation**



**Results and Discussion**

The results of multiple linear regression in table2 show that the model explains 82.07% of the variance in the auction price. The adjusted R-squared is 79.28%, which is also a good indication of fit. The F-statistic is 7.263 with a p-value of 1.564e-06, which indicates that the model is statistically significant. The coefficients of the model tell us the relationship between each independent variable and the auction price. For example, the coefficient for Runs is -4610, which means that for each additional run, the auction price is expected to decrease by 4610, holding all other variables constant. The p-value for Runs is 0.0062, which indicates that this coefficient is statistically significant. The coefficients for Average, Higest\_score, Strike.Rate, Fifty, and Six are also statistically significant. The coefficient for Four is not statistically significant, which means that there is not enough evidence to conclude that there is a relationship between Four and the auction price.

**H1:** The explanatory variable (player runs) has impact on response variable (Auction price):

The coefficient for Runs is -4610, statistically significant (p-value = 0.0062). This suggests that for each additional run, the auction price decreases by $4610, holding other variables constant.

**H2**: The explanatory variable (player Average) has impact on response variable (Auction price)

The coefficient for Average is -245981, statistically significant (p-value = 0.0167). This suggests that for each unit increase in Average, the auction price decreases by $245,981, holding other variables constant.

**H3:** The explanatory variable (player Strike rate) has impact on response variable (Auction price)

 The coefficient for Strike.Rate is -22476, statistically significant (p-value = 0.0028). This suggests that for each unit increase in Strike.Rate, the auction price decreases by $22,476, holding other variables constant.

**H4:** The explanatory variable (player fifties) has impact on response variable (Auction price):

The coefficient for Fifty is -5391246, statistically significant (p-value = 0.0427). This suggests that for each additional Fifty, the auction price decreases by $5,391,246, holding other variables constant.

**H5:** The explanatory variable (players high score) has impact on response variable (Auction price):

 The coefficient for Higest\_score is 242646, statistically significant (p-value = 0.0285). This suggests that for each unit increase in Highest score, the auction price increases by $242,646, holding other variables constant.

**H6:** The explanatory variable (player fours) has impact on response variable (Auction price):

The coefficient for Four is -360248, but not statistically significant (p-value = 0.7292). This suggests that there is not enough evidence to conclude that there is a relationship between Fours and the auction price.

**H7:** The explanatory variable (players sixes) has impact on response variable (Auction price):

 The coefficient for Six is 3899032, statistically significant (p-value = 0.0060). This suggests that for each additional Six, the auction price increases by $3,899,032, holding other variables constant. Overall, the analysis supports 6 out of seven research hypotheses. Runs, Average, Strike.Rate, Higest\_score, fifties and Sixes all have significant impacts on Auction price, while Fours do not show a statistically significant relationship.  
The correlation study as shown in table3 found a strong positive correlation between auction price and several player stats, including runs (0.874), average (0.931), highest score (0.774), strike rate (0.859), and fifties (0.751). This indicates that players with higher values in these areas tend to command significantly higher auction prices. However, the correlation with sixes (0.689) was relatively weak, suggesting that it may not be as influential in determining auction price as other stats. Overall, the study suggests that player performance, particularly in runs, average, and strike rate, plays a significant role in determining auction price, although other factors may also contribute.

**Conclusion**

The study investigated the relationship between player performance and auction price. It found that a linear regression model can explain 82% of the variance in auction price, suggesting that player performance plays a significant role in determining the price at which they are sold. The analysis further revealed that runs, average, strike rate, highest score, fifties, and sixes all have statistically significant positive impacts on auction price, supporting six out of seven research hypotheses. However, the correlation between auction price and sixes was weaker compared to other factors, suggesting it may not be as influential. Overall, the study concludes that player performance, particularly in runs, average, and strike rate, holds the most weight in determining auction price, while other factors may also play a role

**References**

Singh, S. (2011). Measuring the Performance of Teams in the Indian Premier League, American Journal of Operations Research, 1, 180-184.

Kalgotra, P., Sharda, R. and Chakraborty, G. (2013). Predictive Modelling in Sports League: An Application in Indian Premier League. http://support.sas.com/resources/papers/proceedings13/019-2013.pdf, accessed: 10th September 2013

Sankaran, S. (2014). Comparing pay versus performance of IPL Bowlers: an application of cluster analysis. *International Journal of Performance Analysis in Sport*, *14*(1), 174-187.

**Appendix**

**###load data**

**mydata=read.csv(choose.files())**

**##view data**

**View(mydata)**

**####summary statistics**

**summary(mydata)**

**######multilple regression model**

**linearmodel=lm(Auction.price~Runs+Average+Higest\_score+Strike.Rate+Fifty+Four+Six,data=mydata)**

**##### output of multiple regression model**

**summary(linearmodel)**

**##scattor plot**

**install.packages("ggplot2")**

**library(ggplot2)**

**plot(linearmodel,col="blue")**

**ggplot(mydata, aes(y = Auction.price, x = Runs,))+**

**geom\_point() +**

**labs(title = "Scatterplot of Auction.price vs Runs")**

**ggplot(mydata, aes(y = Auction.price, x = Average,))+**

**geom\_point() +**

**labs(title = "Scatterplot of Auction.price vs Average")**

**ggplot(mydata, aes(y = Auction.price, x = Higest\_score,))+**

**geom\_point() +**

**labs(title = "Scatterplot of Auction.price vs Higest\_score")**

**ggplot(mydata, aes(y = Auction.price, x = Strike.Rate,))+**

**geom\_point() +**

**labs(title = "Scatterplot of Auction.price vs Strike.Rate")**

**ggplot(mydata, aes(y = Auction.price, x = Fifty,))+**

**geom\_point() +**

**labs(title = "Scatterplot of Auction.price vs Fifty")**

**ggplot(mydata, aes(y = Auction.price, x = Four,))+**

**geom\_point() +**

**labs(title = "Scatterplot of Auction.price vs Four")**

**ggplot(mydata, aes(y = Auction.price, x = Six,))+**

**geom\_point() +**

**labs(title = "Scatterplot of Auction.price vs Six")**

**plot(fitted(linearmodel), sqrt(abs(resid(linearmodel))), main="Scale-Location Plot", xlab="Fitted Values",**

**ylab="Square Root of Standardized Residuals")**

**install.packages("car")**

**library(car)**

**vif(linearmodel)**

**vif(linearmode)**