# STATISTICAL ANALYSIS OF THE PERFORMANCE OF ROYAL CHALLENGERS BANGALORE IN THE INDIAN PREMIER LEAGUE

**Ryanston Rodrigues**

Department Of Statistics, V.P.M.'s B.N. Bandodkar College of Science, Thane – 400601

ryanstonrodrigues2000@gmail.com

## ABSTRACT

The Indian Premier League (IPL) is a professional Twenty20 cricket league in India usually contested between March and May of every year by eight teams representing eight different cities or states of India. The league was founded by the Board of Control for Cricket in India (BCCI) in 2007. The IPL is the most attended cricket league in the world. 13 seasons of the IPL have been contested so far. The Royal Challengers Bangalore (RCB) is a franchise cricket team based in Bangalore, Karnataka, that plays in the IPL. Founded in 2008, the team has played its home matches at the M. Chinnaswamy Stadium. Virat Kohli has been the team's captain since 2014. RCB remains as one of the most popular IPL teams despite their lack of success. RCB has never won a single edition of the IPL but have finished as runners-up on three occasions (2009, 2011 and 2016). Even with the presence of various notable players, the team constantly underperforms and hence, they have been labeled as 'underachievers'. The team also holds the record for both the highest (263) and lowest (49) scores in the IPL. The aim of this research is to understand the poor performance of RCB in the IPL using various statistical and mathematical techniques and approaches, namely matrices, graphs, and also build a model to predict the outcome of a given match for the team using the technique of logistic regression, in R Software.

## INTRODUCTION

The journey of RCB in the IPL is a mixed bag. While they have had a fair share of wins, the team has performed very poorly over the years. With major losses against all the other teams and failing to win even a single IPL over the course of 13 years, one cannot help but look down upon RCB as a disappointment in the IPL. This is where logistic regression comes into play.

Logistic Regression is a statistical model where the response variable has only two outcomes, success and failure, usually denoted by 0 and 1. In this case, 0 denotes failure and 1 denotes success. The response is usually a count such as number of wins and losses (in this scenario). Since we have two possible outcomes, this model is termed as Binary Logistic Regression Model.

In this research, we are going to look at the performance of RCB in the IPL and try to find out the reasons for the poorresults in the tournament in recent years.

## DATASET

The dataset and the R code used for the research is provided at the end of the paper. It contains 191 observations in 5 variables. Each observation corresponds to a match played by RCB. The variables are Runs, Innings, Wickets. Taken,Wickets. Lost and Results. Note that this dataset does not include data for those matches which not have a conclusion (matches abandoned due to rain,etc.).

## PART 1: ANALYSIS USING GRAPHS

## LEGEND

For Batsmen                    For Bowlers

Kohli                          Ali
De_Villiers                    Siraj
Padikkal                       Saini
Gurkeerat                      Negi
Finch                          Dube
Phillippe                      Yadav
Parthiv                        Sundar
Ali                            Chahal
Sundar                         Ahmed
Morris                         Udana
Negi                           Zampa
Udana                          Steyn
Dube                           Morris

## ANALYSIS

From our given dataset, we consider the Results column, which shows the matches in which RCB won or lost (0 for loss and 1 for win). For our convenience, we have omitted the results of those matches which ended in NR (no result). The total wins and losses are as follows:

0   1

100  91

And their win loss percentage is given as follows:

0       1

0.5235602 0.4764398

Clearly, it is evident that RCB has a greater loss% then win% which shows that why RCB has severely underperformed in the IPL. The following graphs show various cricket statistics for RCB, with the help of which we shall analyze their performance over the years. For convenience, we pay close attention to the previous two seasons since the core team was more or less the same for these two seasons.

**Runs scored in 1st Innings**
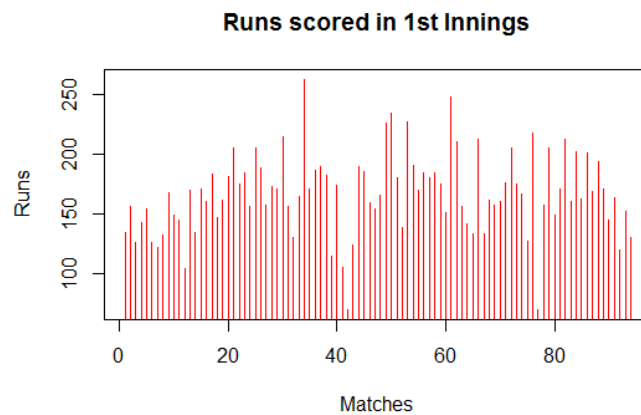
Graph of Runs scored by RCB in 1st innings in IPL matches

**Runs scored in 2nd Innings**
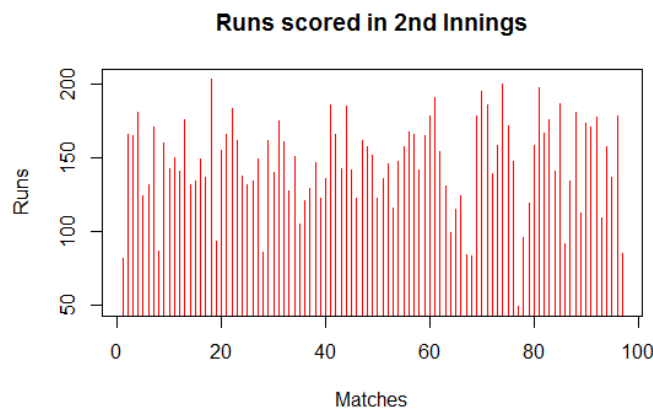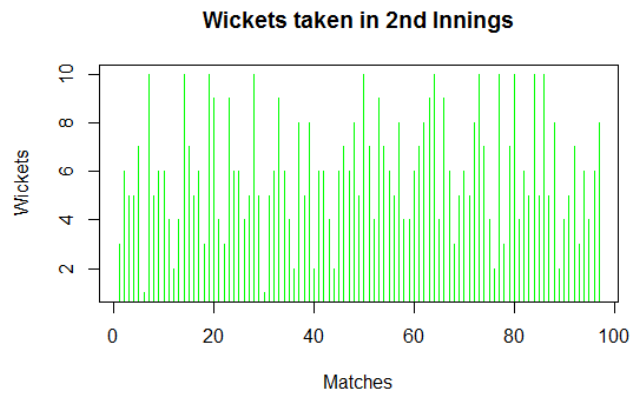
Graph of Runs scored by RCB in 2nd innings in IPL matches

The above graphs show the runs scored by RCB in both innings. From the given data, the average first and second innings total by RCB are 166.67 and145.74 respectively, which shows that RCB perform well with the bat and score more runs in the 1st innings as compared to the 2nd innings over the years. Hence, we can conclude that RCB are better in setting the target rather than chasing it.
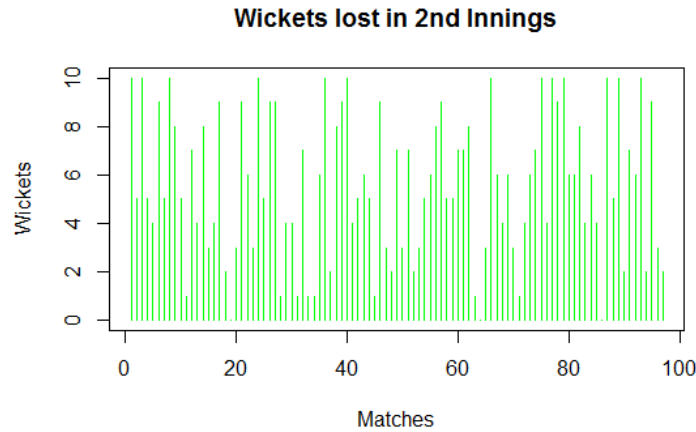
**Wickets taken in 1st Innings**
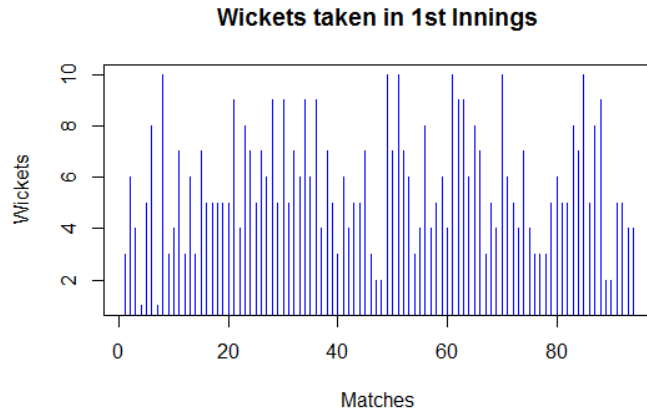


Graph of Wickets taken by RCB in 1$^{st}$ innings in IPL matches

**Wickets lost in 1st Innings**
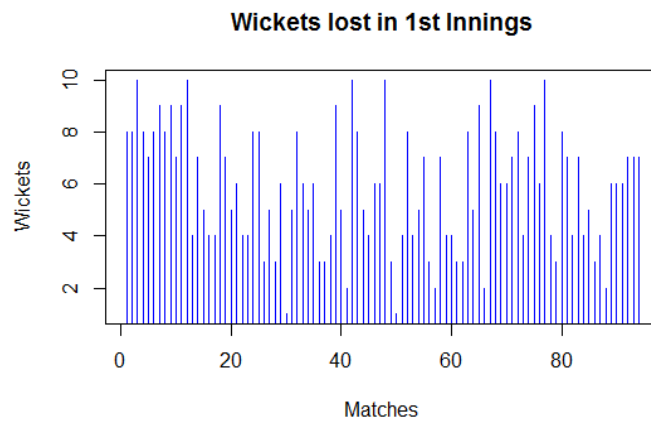


Graph of Wickets lost by RCB in 1$^{st}$ innings in IPL matches

**Wickets taken in 2nd Innings**



Graph of Wickets taken by RCB in 2$^{nd}$ innings in IPL matches

**Wickets lost in 2nd Innings**



Graph of Wickets lost by RCB in 2$^{nd}$ innings in IPL matches

The average number of wickets taken by RCB is 5.774869 and that of wickets lost is 5.706806, both of which are approximately equal to 6 wickets per match. This means that RCB takes as well as loses the same number of wickets per match. From a bowling perspective, picking up 6 wickets per match is actually quite good. However, losing 6 wickets per match is a sign of batsmen struggling on the pitch. This scenario is supported by the fact that RCB has a lower average total in the 2$^{nd}$ innings.

The purple line in each of the graphs below indicates the 12$^{th}$ edition of the IPL (2019)
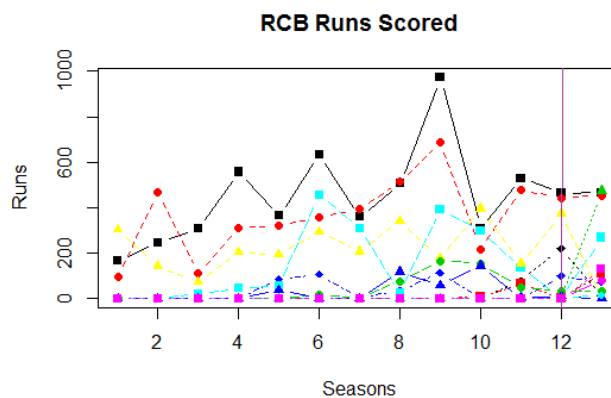
**RCB Runs Scored**



Graph of Runs scored by RCB batsmen per season in IPL
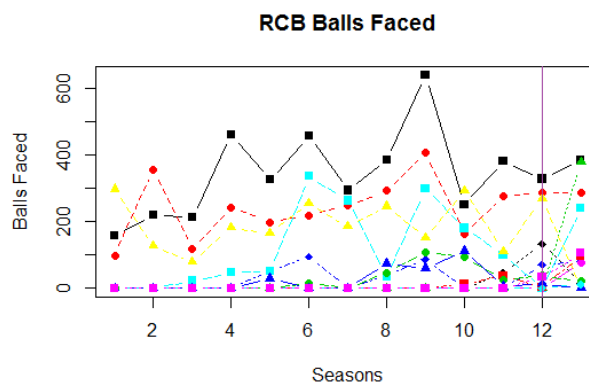
From the above graph, we can say that both Kohli and De Villiers, who have played for RCB for a long time (Kohli since 2008,De Villiers since 2011) and are upper order batsmen, have done

the majority of run scoring for RCB. Aaron Finch, who had decent outings in previous years for other teams, failed to make an impact this year. Devdutt Padikkal (indicated by the green triangles) who is an opening batsman for the team, had a very good outing in his debut year, scoring 473 runs this year. The rest of the batsmen did not score any considerable amount of runs. This shows that the team is totally dependent on their upper order for scoring runs and if the upper order fails, the rest of the team has a hard time to score runs.
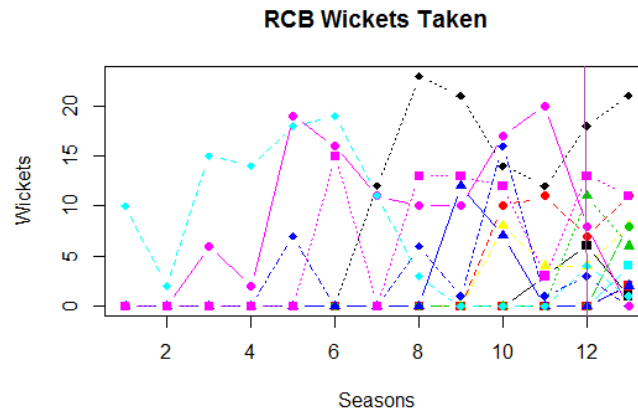
**RCB Wickets Taken**



Graph of Wickets taken by RCB bowlers per season in IPL

The above graph indicates that YuzvendraChahal has been the most consistent bowler for RCB in recent years, with an average of 17 wickets in 7 years of playing for the team. Umesh Yadav, who also used to pick wickets regularly failed to pick up a single wicket this year. The rest of the bowlers did not have significant contributions this year. We can conclude that one of the reasons that RCB struggle in their bowling department is that they are heavily dependent on a single bowler (Chahal) for picking up the bulk of wickets.
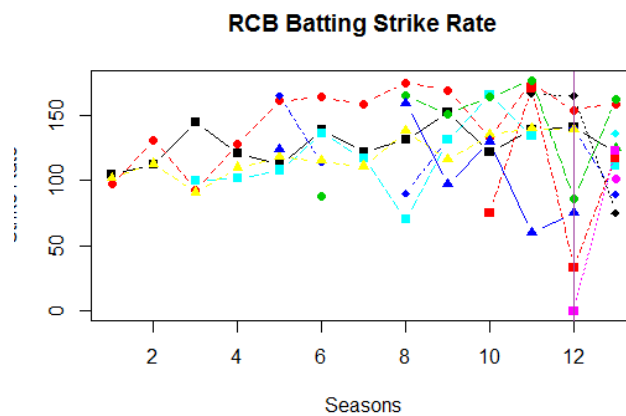
**RCB Balls Faced**



Graph of Balls faced by RCB batsmen per season in IPL

From the above graph, we can see that only 4 batsmen (Kohli,Padikkal,deVilliers,Finch) have played more than 200 bowls this year. Since, all of them are upper order batsmen,the majority of run scoring was on their shoulders. The rest of the batsmen, who usually bat at lower order, faced few balls this year.
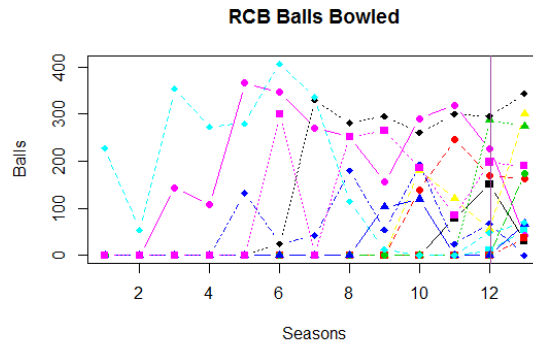


Graph of Balls bowled by RCB batsmen per season in IPL

From this graph, it is evident that Chahal bowled the most bowls this year, followed by Washington Sundar and Navdeep Saini. One key thing to observe is that even though Sundar and Saini have bowled many deliveries, they haven't picked up many wickets this year (from Graph of Wickets Taken). This again cements the fact that the team was heavily dependent of Chahal as their strike bowler.



Graph of Batting Strike Rate of RCB batsmen per season in IPL

The above graph indicates that Chris Morris had the highest strike rate among the batsmen this year, followed by de Villiers. However, Morris faced only 21 balls this year and scored 34 runs since he is a lower order batsman. Hence, a higher strike rate in this case does not imply that the

player has contributed significantly with the bat. In case of Kohli and de Villiers, they scored more runs and faced more balls, but still had lower strike rates. This is why the batting strike rate can be misleading sometimes.
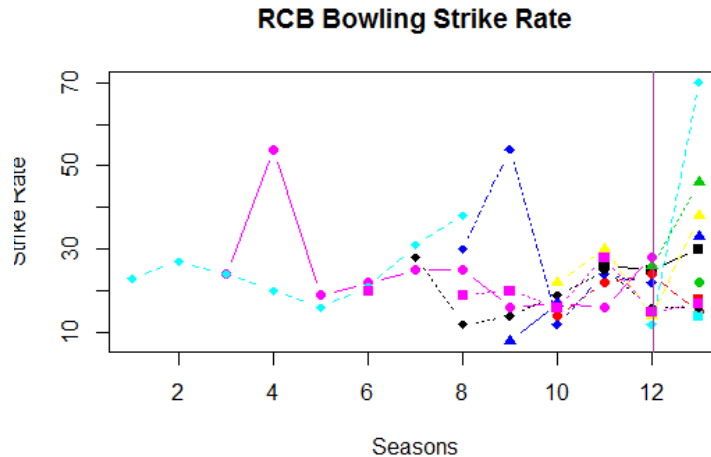


Graph of Bowling Strike Rate of RCB bowlers per season in IPL

Generally, in the T20 format, lower the bowling strike rate, the consistent the bowler is said to be. Chahal, Morris, Siraj and Dube have had lower bowling strike rates this year. However, one should also take their Wickets Taken and Bowls Bowled in this case. Chahal, who picked up the most wickets and also bowled the most balls this year for the team, has a low bowling strike rate. Even though ShivamDube has a lower bowling strike rate than Chahal, he neither bowled as much as Chahal nor picked up more wickets than him. So here too, the bowling strike rate can be misleading. Dale Steyn had the highest strike rate among bowlers, which means that he conceded more runs without much success.

In this way, the above graphs help us to analyze the main reasons for the team's lack of success. The over dependency of the team on only a couple of batsmen and bowlers had led to teams downfall in the tournament.

**PART 2: BUILDING LOGISTIC REGRESSION MODEL**

In the second part of the analysis, we try to fit a Logistic Regression Model to the given data. Our aim is to build a model that predicts the outcome of a given match based on variables such as Innings, Runs Scored, Wickets Taken and Lost, etc.

By using backward elimination technique, we conclude that the two most statistically significant variables are Wickets.Taken and Wickets.Lost. We divide the dataset into two parts, a training set and a test set. We build the model using the training set and use the test set for predicting new values. By using the summary() function, we get the information about the fitted model as

```
summary(classifier)

Call:
glm(formula = Results ~ Wickets.Taken + Wickets.Lost, family = binomial(link
= "logit"),
data = training_set)

Deviance Residuals:
     Min         1Q     Median         3Q        Max
-2.14891   -0.51678   -0.07435    0.46461    2.05324

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)     1.4041     0.9715   1.445    0.148
Wickets.Taken   0.6067     0.1482   4.094 4.23e-05 ***
Wickets.Lost   -0.8779     0.1522  -5.766 8.10e-09 ***
---
Signif.codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 197.897  on 142  degrees of freedom
Residual deviance:  97.608  on 140  degrees of freedom
AIC: 103.61

Number of Fisher Scoring iterations: 6
```

We use the predict() function to predict the results and create a confusion matrix for checking our predictions. The confusion matrix is given as follows,

```
>cm1
Confusion Matrix and Statistics

          Reference
Prediction  0  1
        0 66 11
1  9 57

Accuracy : 0.8601
               95% CI : (0.7923, 0.9124)
    No Information Rate : 0.5245
    P-Value [Acc > NIR] : <2e-16

Kappa : 0.7192

Mcnemar's Test P-Value : 0.8231

Sensitivity : 0.8800
Specificity : 0.8382
PosPredValue : 0.8571
NegPredValue : 0.8636
```

```
Prevalence : 0.5245
        Detection Rate : 0.4615
  Detection Prevalence : 0.5385
     Balanced Accuracy : 0.8591

       'Positive' Class : 0


>cm2
Confusion Matrix and Statistics

          Reference
Prediction  0  1
        0 19  4
1  6 19

Accuracy : 0.7917
                95% CI : (0.6501, 0.8953)
    No Information Rate : 0.5208
    P-Value [Acc > NIR] : 9.772e-05

Kappa : 0.5841

Mcnemar's Test P-Value : 0.7518

Sensitivity : 0.7600
Specificity : 0.8261
PosPredValue : 0.8261
NegPredValue : 0.7600
Prevalence : 0.5208
        Detection Rate : 0.3958
  Detection Prevalence : 0.4792
     Balanced Accuracy : 0.7930

       'Positive' Class : 0
```

Where cm1 and cm2 are the confusion matrices for training and test set respectively. The model has an accuracy of 86% and 79% for training and test set respectively, which is a pretty good result considering our dataset has only 191 observations. With RCB playing more matches in the future, the dataset can be increased and accordingly the accuracy can also be increased.

**CONCLUSION**

Even though RCB remains a popular team among IPL fans, it can be seen that why the team has consistently been underperforming. With only a few key players performing for the team, the success rate is low for the teams. This assumption of their poor success rate is verified by the analysis using graphs on various statistics. Logistic Regression helps us to build a model with a decent accuracy for predicting outcome for a given match played by the team.

**REFERENCES**

1. **Douglas C. Montgomery, Elizabeth A. Beck, G. Geoffrey Vining**(2012), Introduction to Linear Regression Analysis, Wiley, Fifth Edition,

2. **G.Sudhamathy and G.RajaMeenakshi**(2020), Prediction on IPL Data using Machine Learning Techniques in R Package, ISSN: 2229-6956 (Online)

3. **Indian Premier League** – Wikipedia

4. **Indian Premier League Official Website** www.iplt20.com

5. **Royal Challengers Bangalore** – Wikipedia

6. **LINK FOR DATASET AND CODE:** https://github.com/Ryanston/RCB-Research