# **5 Year Survival of NBA Rookies**

## **Muthukumar Kadhirvel**

* What percentage of players lasted 5 years in the league?

Chart, pie chart

Description automatically generated

We can see that we have around 60% of the surveyed rookies lasted 5 years in the league.

* Which year had the most rookies that lasted 5 years in the league?

Chart, bar chart

Description automatically generatedAs we can see that 2009 had most number of rookies who lasted 5 years in the league with 33 with 1988 and 1989 being next with 32 rookies.

* What is the relationship between Points per game and if the player lasts 5 or more years?

Chart, bar chart

Description automatically generated

We can see that for rookies that score less than 5 points per game they do not last 5 years in the league.

* What is the relationship between Games Played and if the player lasts 5 or more years?

Chart, bar chart

Description automatically generated

We can see that for rookies that play less than 60 games they do not last 5 years in the league.

* What is the relationship between Minutes Played per game and if the player lasts 5 or more years?

Chart, bar chart

Description automatically generated

We can see that for rookies that play less than 15 minutes per game they do not last 5 years in the league.

* What is the relationship between Field Goals Percentage and if the player lasts 5 or more years?

Chart, bar chart

Description automatically generated

We can see that Field goals percentage does not impact target variable.

* What is the relationship between 3 Pointers Percentage and if the player lasts 5 or more years?

Chart, bar chart

Description automatically generated

We can see that 3 Pointers percentage does not impact target variable.

* Important Feature that determines if the player lasts 5 or more years?

Chart, bar chart

Description automatically generated

We can see that Games Played has the most feature importance and then comes Minutes Played.

* Accuracy of the various models?
  + Logistic Regression 72.63
  + KNN 67.72
  + Support Vector Machine 67.72
  + Random Forest Classifier 72.98
  + Random Forest Classifier 69.47 (Scaled Data)

Random Forest Classifier model has the best accuracy and the best confusion matrix as shown below and so that can be used for further usage over the other models.

* What is the best predictive model that can be used for prediction of the target variable?

Since Random Forest Classifier has the best accuracy and the confusion matrix, we choose that one and go ahead with implementation.