**NBA Dataset Assignment 1**

Kael Villa

Sheridan College

PROG25211 AI and ML - Python

Soleimani, Ahmad

June 9, 2025

# **Will Chris Paul make the Hall of Fame (HOF)?**

1. CHOOSE A DATASET. You can either go to Kaggle or to another site that contains a set of data. You may want to choose a dataset from a topic that you like, as it will make the assignment easier and more enjoyable. Provide a link in your report to your dataset.

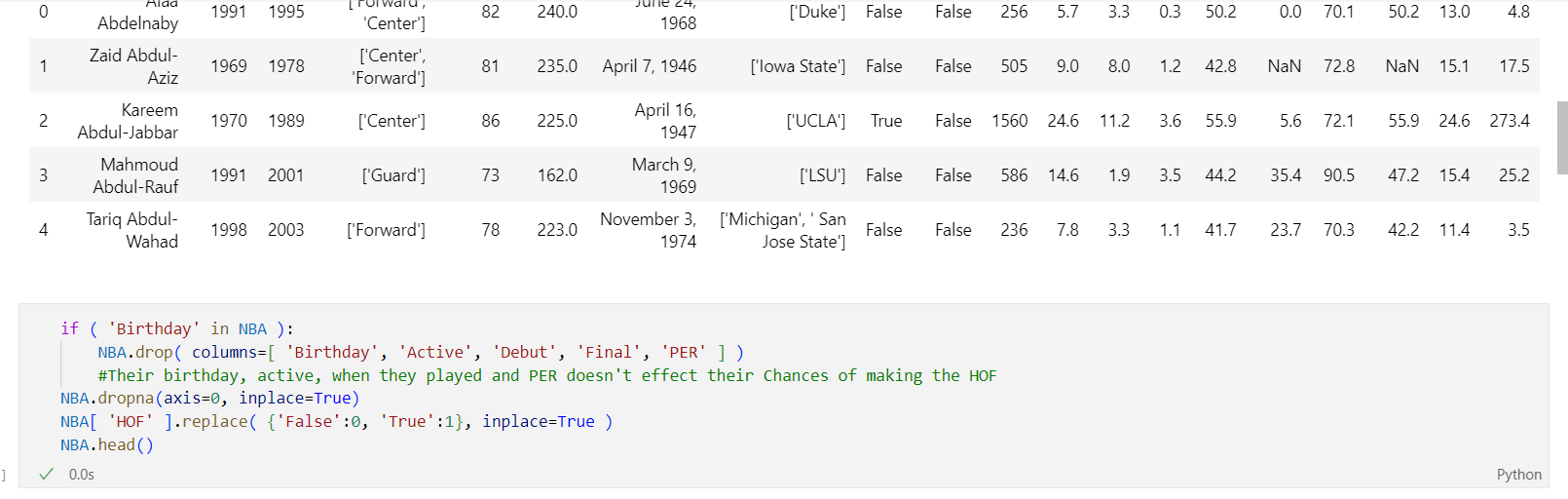
[*https://www.kaggle.com/code/devraai/nba-player-analysis-hof-prediction*](https://www.kaggle.com/code/devraai/nba-player-analysis-hof-prediction)

# ***V2: NBA Player Database***

1. ASK A QUESTION. In your report, ask a logistic question about your data. If you think about what we did in class, the question might be “Would I survive the Titanic?”. The question you ask will be the focus of your assignment. It should be part of the title of your assignment, and what your end goal is.

**Will Chris Paul make the Hall of Fame (HOF)?**

1. CLEAN YOUR DATA. You will need to go through your data removing null or empty values, removing columns that are not relevant to your question, and changing data so that the ML algorithm can process it. You will need to provide an explanation for each row you are deleting or altering.



if ( 'Birthday' in NBA ):

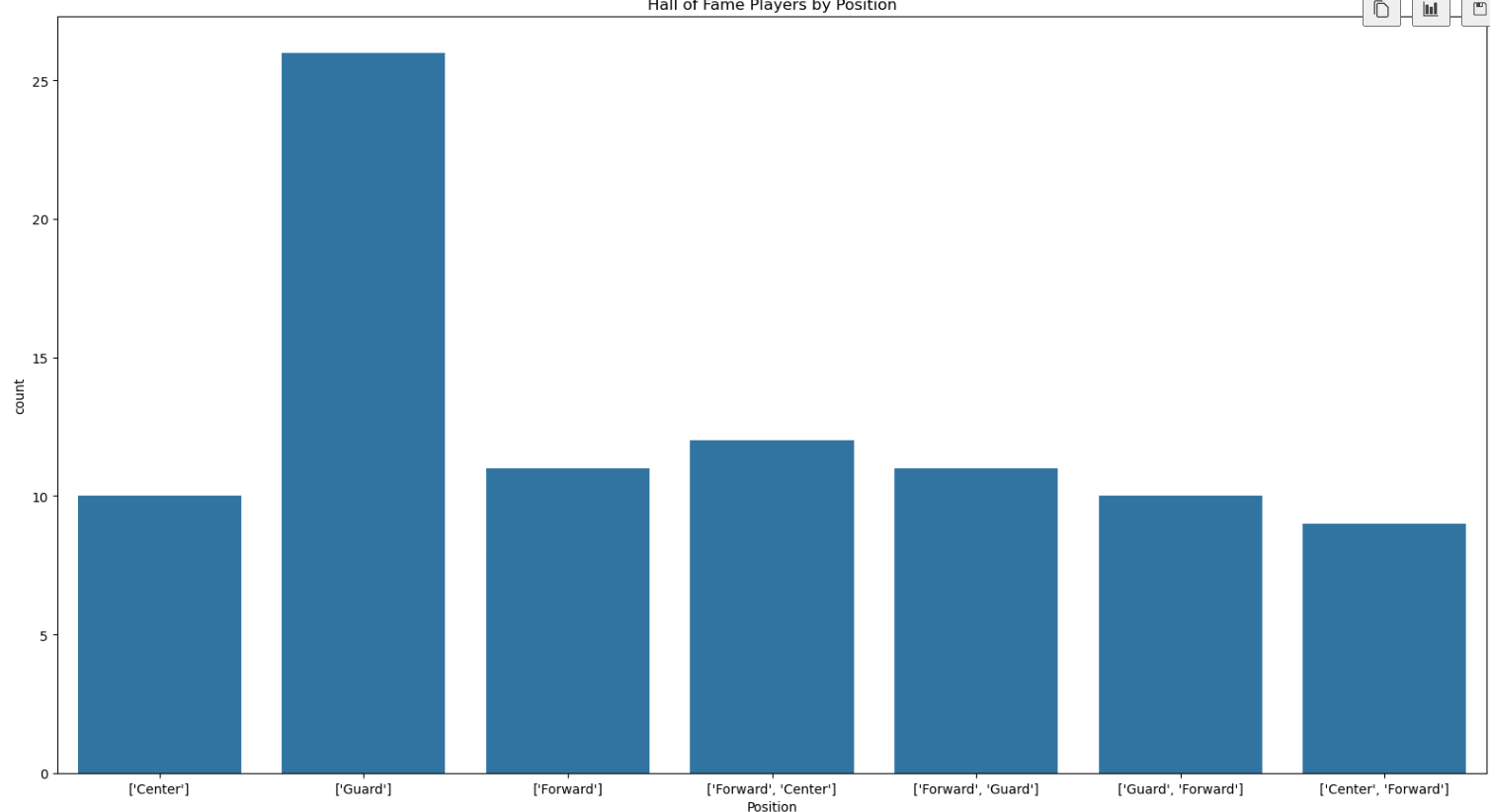
NBA.drop( columns=[ 'Birthday', 'Active', 'Debut', 'Final', 'PER' ] )HOF

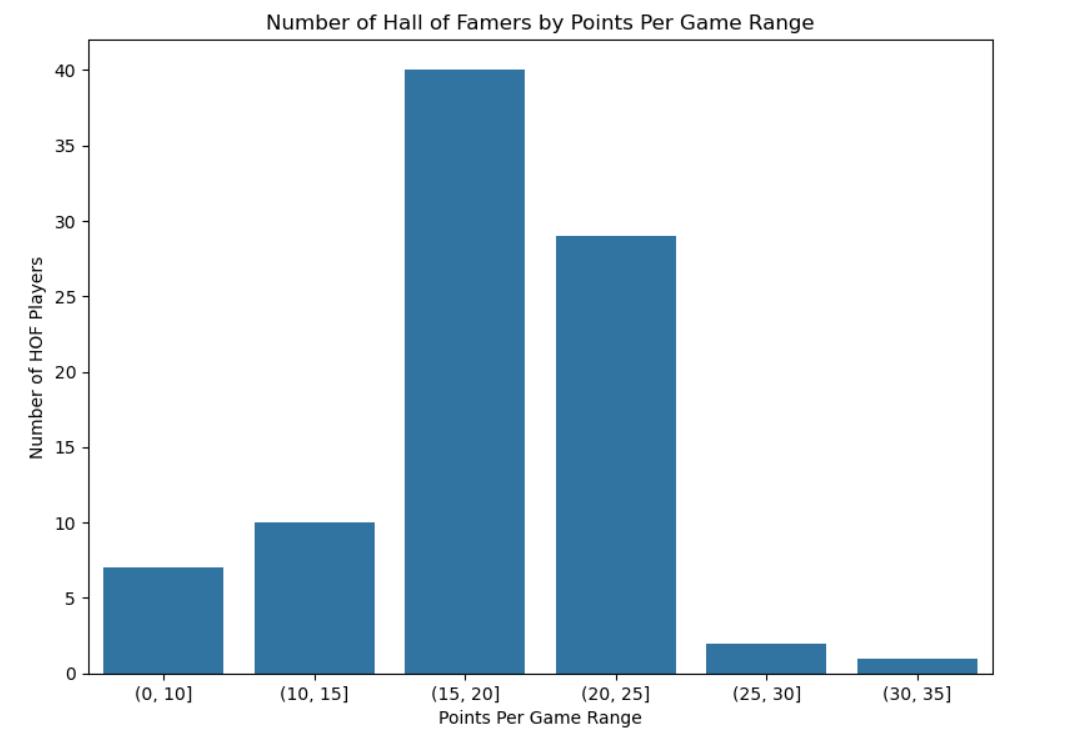
NBA.dropna(axis=0, inplace=True)

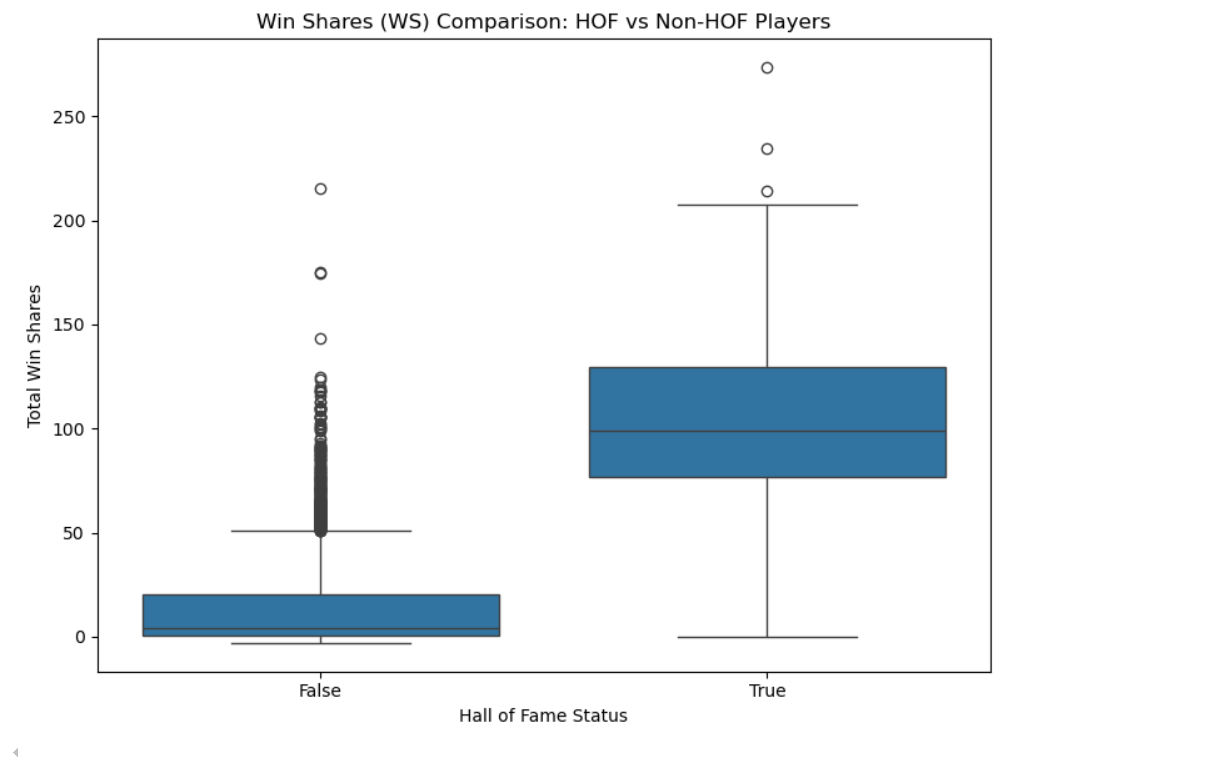
NBA[ 'HOF' ].replace( {'False':0, 'True':1}, inplace=True )

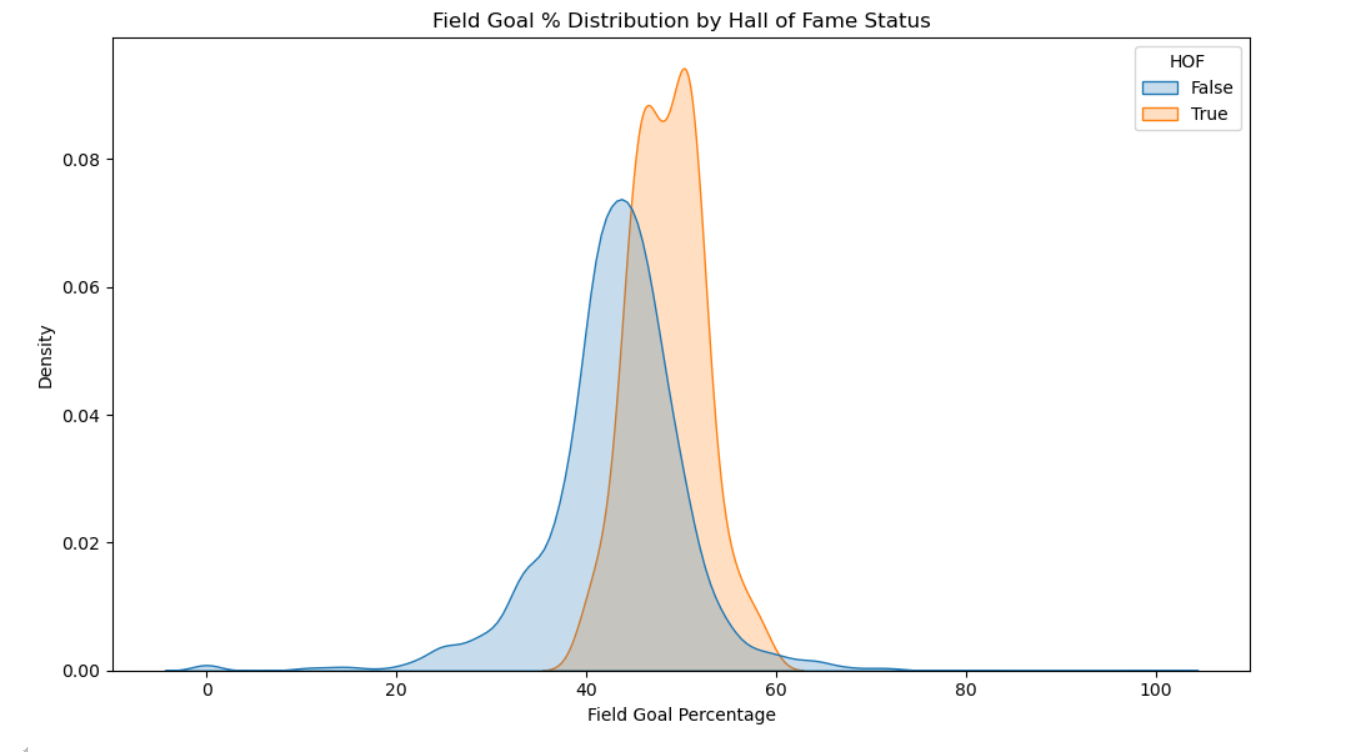
NBA.head()

1. GRAPH YOUR DATA. You need to provide at least 5 different graphs of your data in your report. Include the graphs in both your code and your report. Make sure all graphs are properly labelled with an x and y axis as well as a title. Also add a brief description about what each graph is representing in the report.

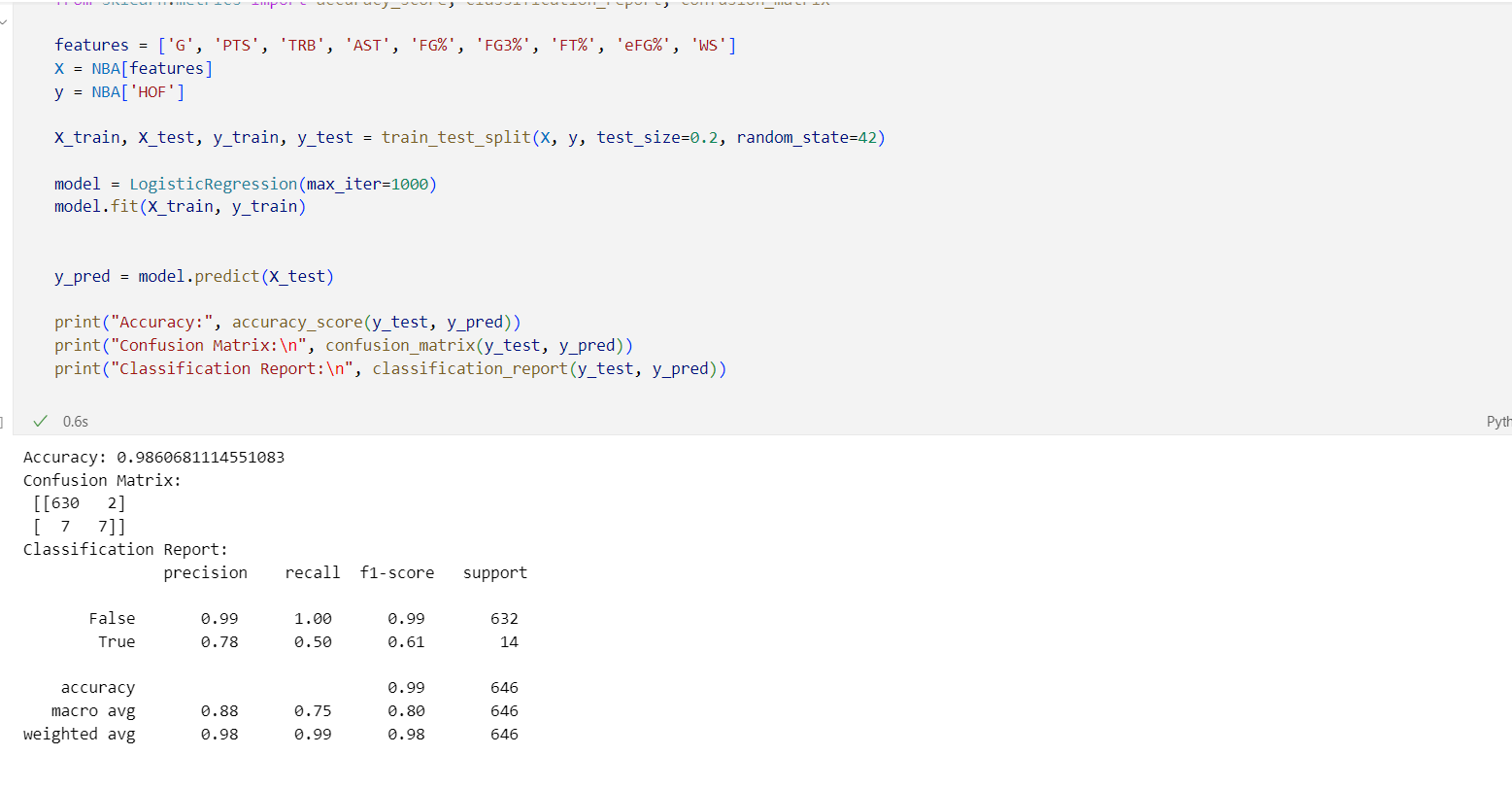




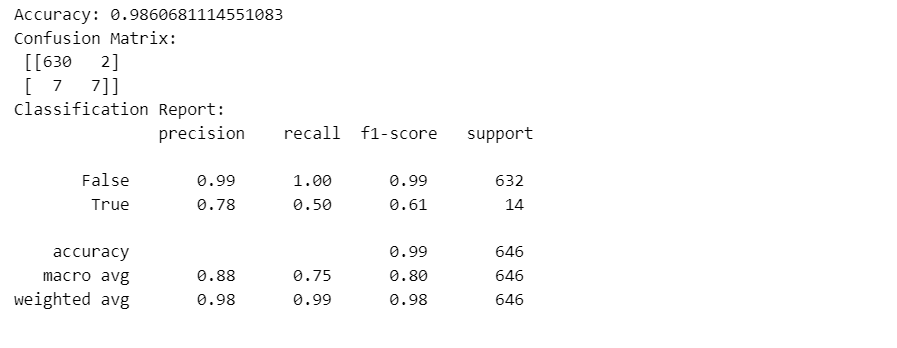




1. TRAIN AND TEST YOUR ALGORITHM. Use the data you collected to train your algorithm with a logistic regression. Make sure to split your data into a training and testing set of appropriate sizes. Include the code (documented) you used to train your data in your report.



1. EVALUATE YOUR MODEL. Provide an evaluation of your model. Use what you have learned in class to decide if the model is well trained based on your test data. If your model is not within the 70%-90% accuracy range explain why you think it is not accurate. Discuss ideas of how you can improve the accuracy of your model.



1. ANSWER YOUR QUESTION AND CONCLUSION. Now that you have trained and tested your algorithm, try to answer your question from part 3. Enter data relevant to the question you asked. Use this part to also provide a conclusion to your report.

****

**THEREFORE Chris Paul will make the HOF**