

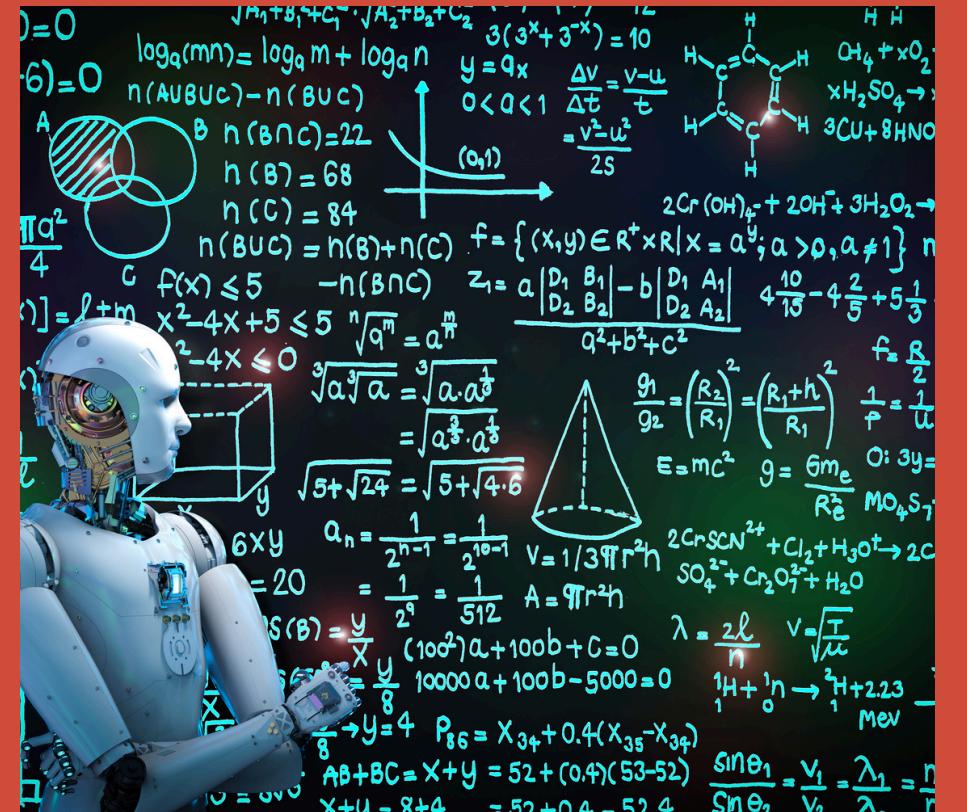


# Olympic Medal Predictor

A Machine Learning Project to predict  
number of medals a country is likely to win

# WHAT'S USED?

Machine learning, Python, Jupyter Notebooks were used for this project.



# 3 step process

01

## ETL

Extract, Transform, Load data.  
Once we've naturalised data we need to  
start building the model.

02

## Build ML model

Select the correct model, train it and  
extract insights.

03

## Improve model accuracy

Improve accuracy and lower the error rate  
of our model

[The files are available on my Github:  
Linked on last slide]

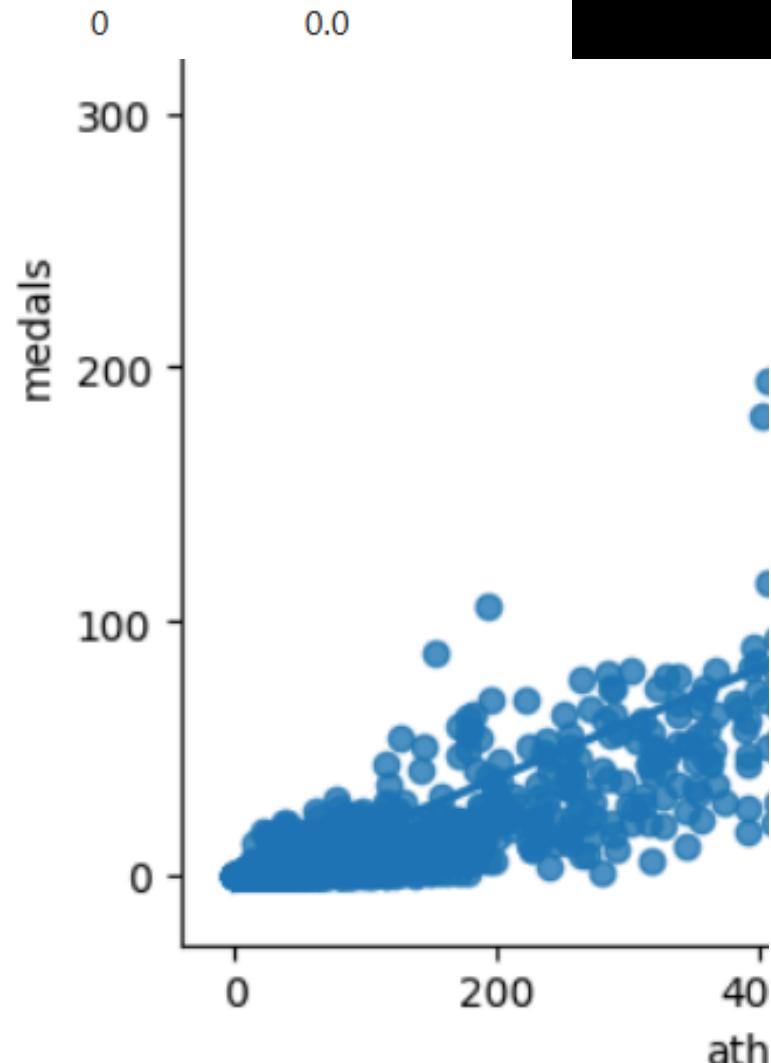
# How it all looks

team	country	year	events	athletes	age	height	weight	medals	prev_medals	prev_3_medals
0	AFG	Afghanistan	1964	8	8	22.0	161.0	64.2	0	0.0
1	AFG	Afghanistan	1968	5	5	23.2	170.2	10.0	0.0	0.0
2	AFG	Afghanistan	1972	8	8	29.0	168.3	10.0	0.0	0.0
3	AFG	Afghanistan	1980	11	11	23.6	168.4	10.0	0.0	0.0
4	AFG	Afghanistan	2004	5	5	18.6	170.8	10.0	0.0	0.0
...	...	...	...	...	...	...	...	...	...	...
9	ZIM	Zimbabwe	2000	19	26	25.0	179.0	10.0	0.0	0.0
10	ZIM	Zimbabwe	2004	11	14	25.1	177.8	10.0	0.0	0.0
11	ZIM	Zimbabwe	2008	15	16	26.1	171.9	10.0	0.0	0.0
12	ZIM	Zimbabwe	2012	8	9	27.3	174.4	10.0	0.0	0.0
13	ZIM	Zimbabwe	2016	13	31	27.5	167.8	10.0	0.0	0.0

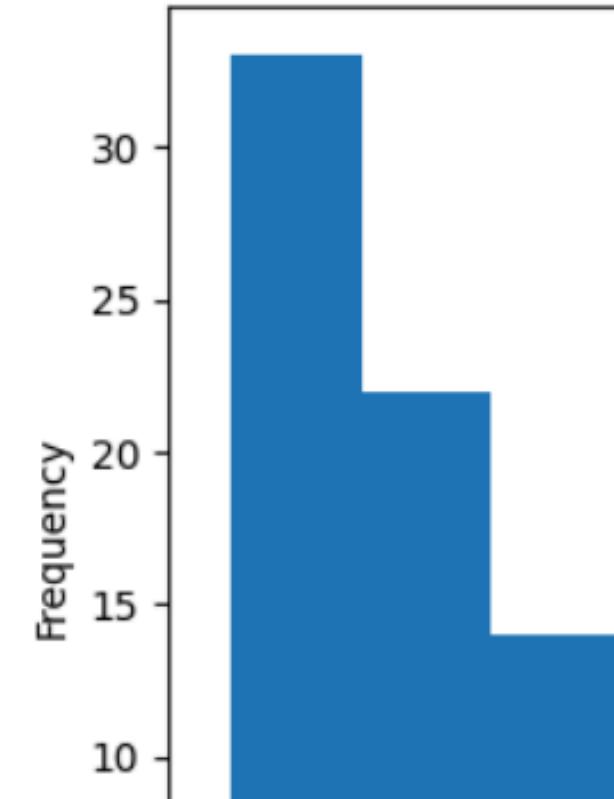
rows × 11 columns

ETL

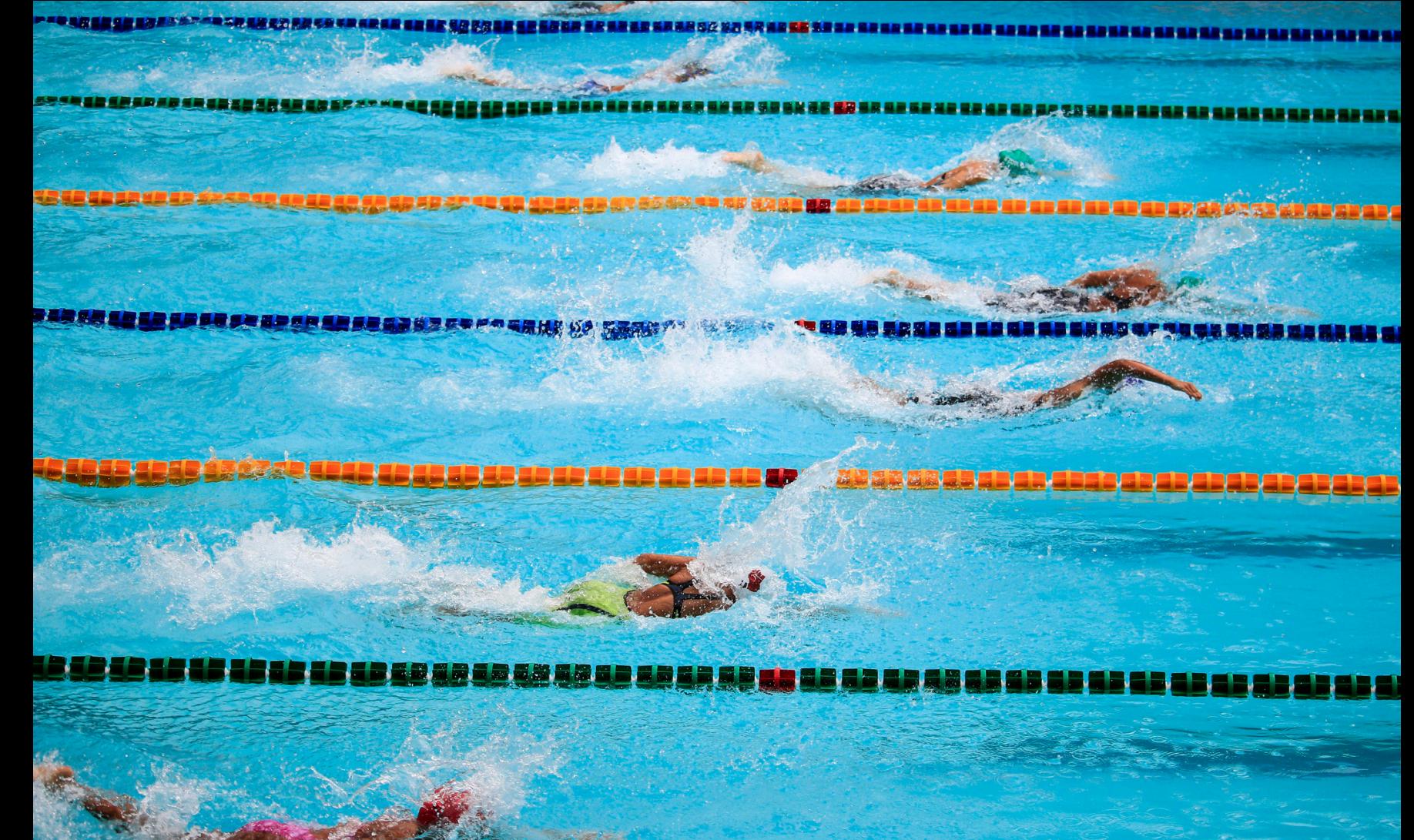
Build ML model



Manage model accuracy



```
[192]: error_ratio.plot.hist()  
[192]: <Axes: ylabel='Frequency'>
```



# In conclusion

More data revealed more accurate predictions.  
More athletes entered into Olympics increase the chances of winning medals



# Where to find my work?

I created a github repo for you to look at. This is found under the Data Science repo

<https://github.com/ThatoSeluku/DataScience>



# Like my work?

Feel free to follow me on Linkedin or Github  
I make Data Analytics, Data Science, Software dev and career-oriented  
content.



Thato Seluku



Github.com/ThatoSeluku