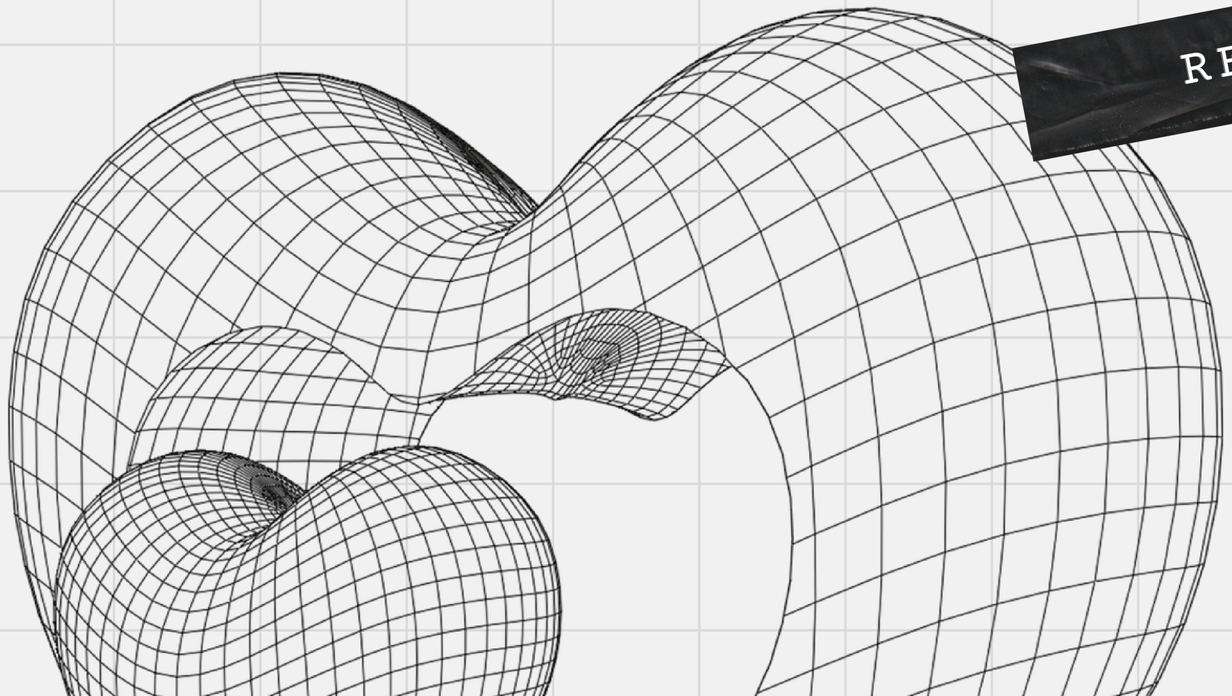
@LOU MEZIERE
@ANH VO

SUMMER 2024

Codédex

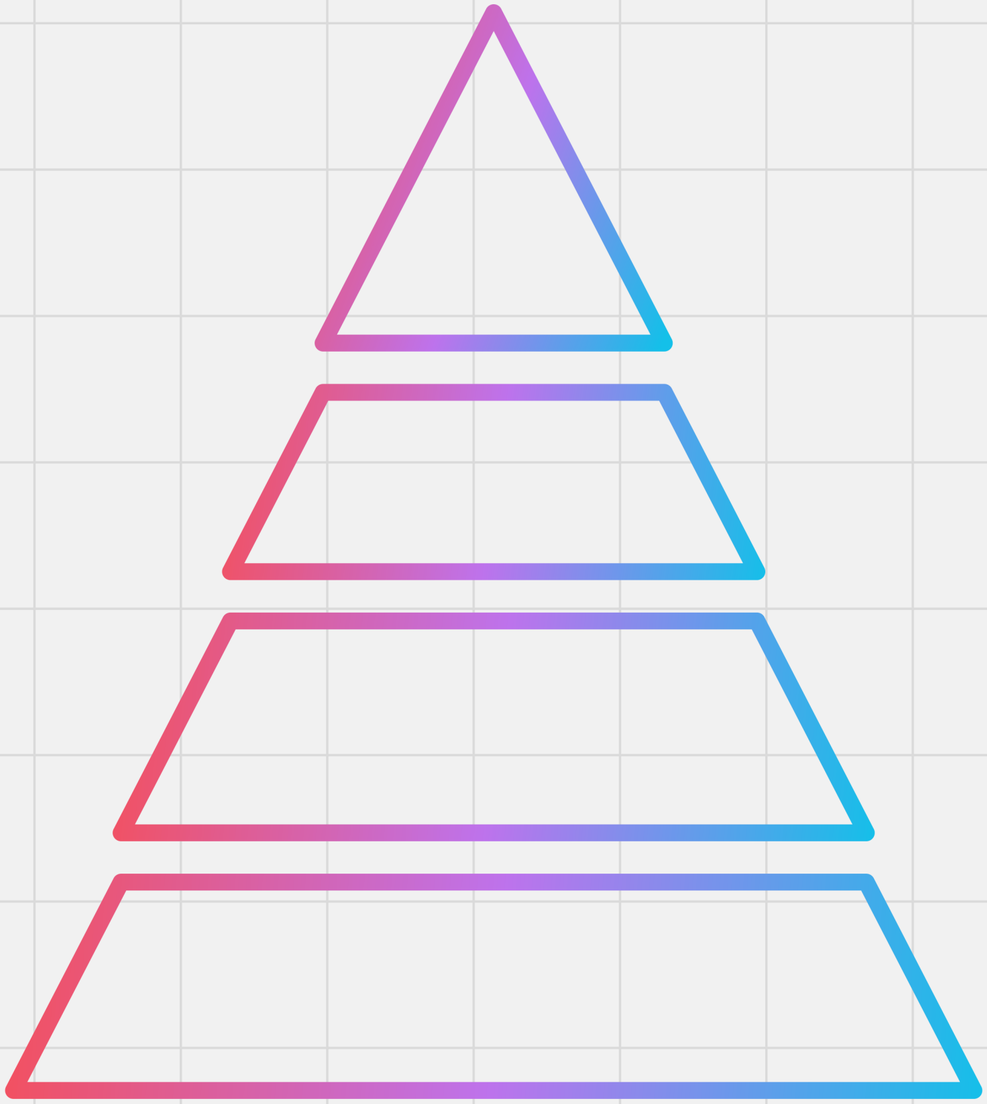
OLYMPIC 2024?

REPORT



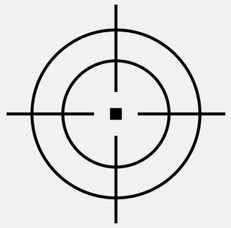
DS - TRACK 3

TABLE OF CONTENT



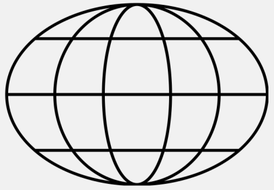
INTRODUCTION

This is not guessing, we’re trying to push the boundary of sport analysis



OBJECTIVES

Dig into the historical data and uncover clues to predict which country will come out on top!

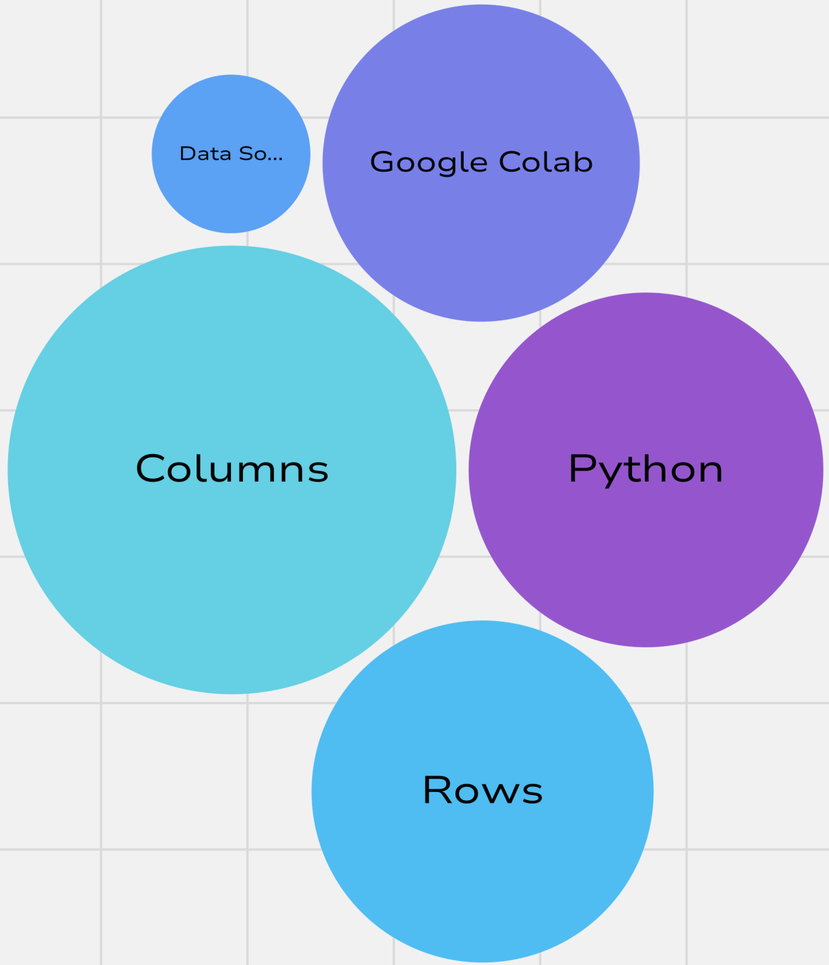


IMPORTANCE

Strive to deepen our understanding of what drives athletic success + highlights the potential of data-driven insights

OUR GOALS

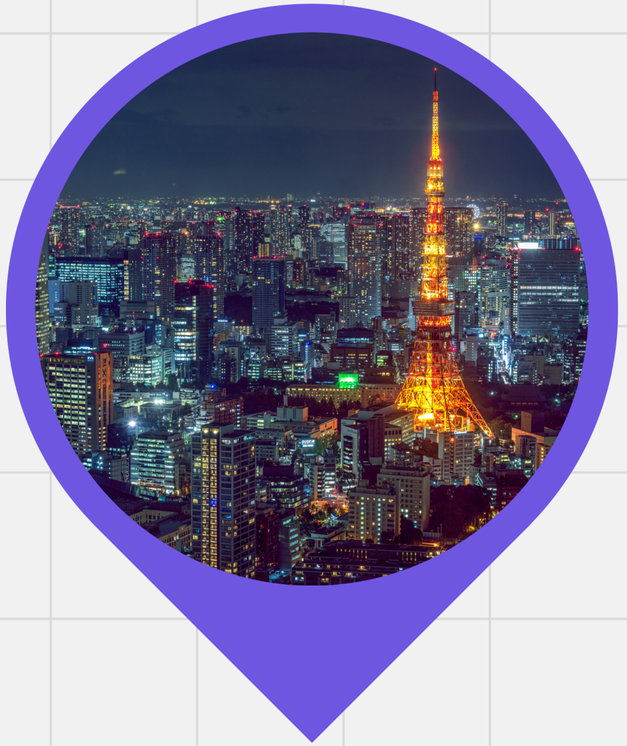
Challenge ourselves and let the wheel spin again!



DATA OVERVIEW

Tokyo 2021 Olympics dataset and 2020 GDP dataset

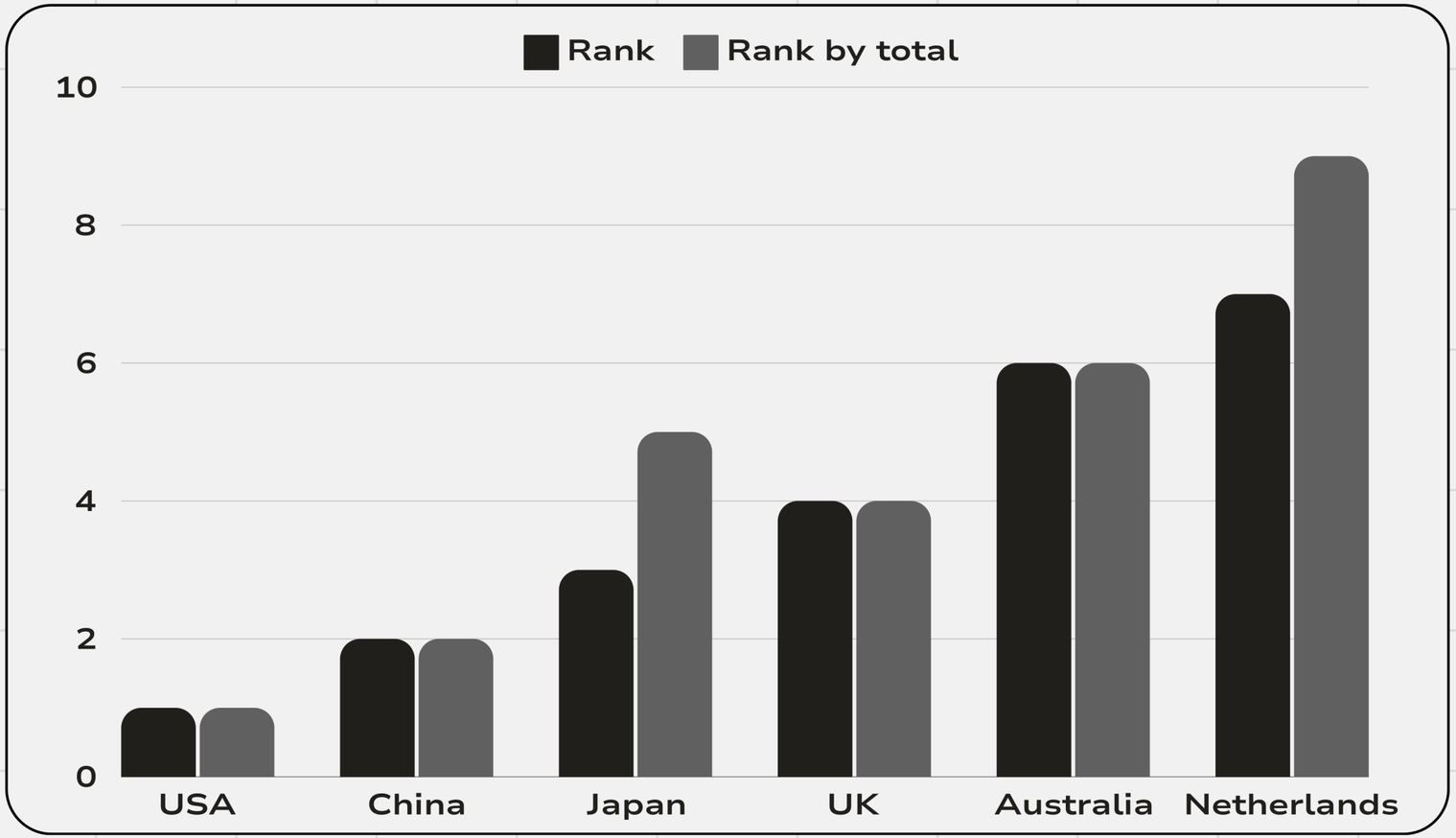
- Crucial data: country, rank, medals, team, gdp
- Top 5 countries



@REALLYGREATSITE

DATA OVERVIEW

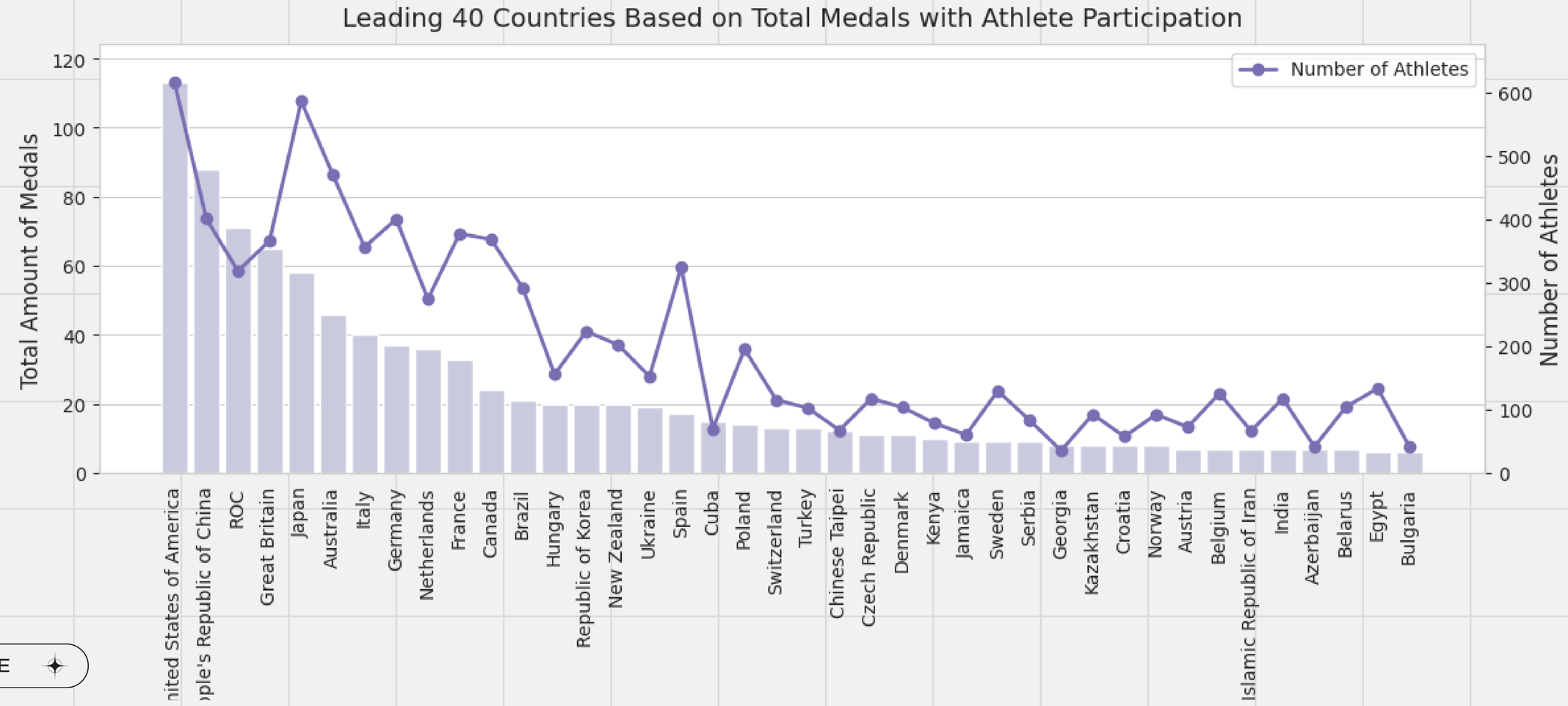
Top 5 countries



BORCELLE

DATA OVERVIEW

Countries having higher amounts of athletes tend to win more medals



DATA ENGINEERING

FEATURES

1. TEAMS COUNT
2. ATHLETES COUNT
3. COACHES COUNT
4. GDP

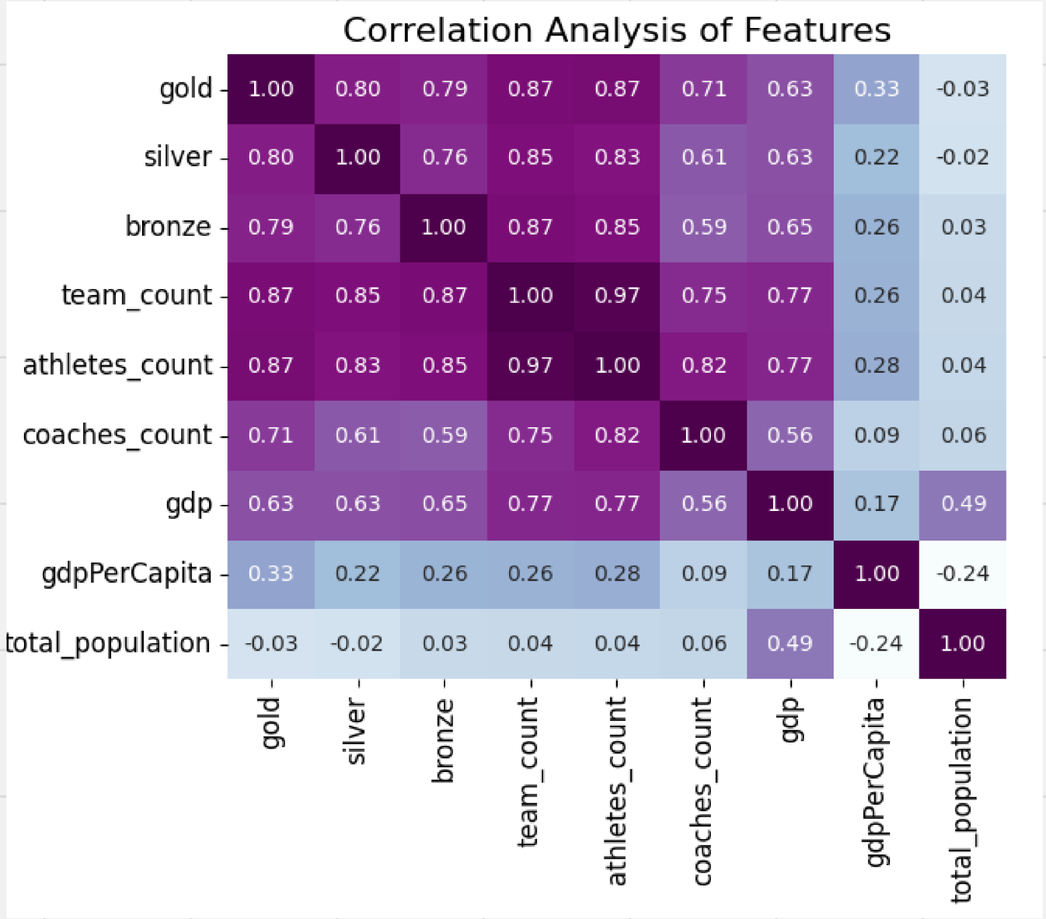
athlete	athlete_id	pos	medal	isTeamSport	year
Ernest Hutcheon	64710	DNS	NaN	False	1908
Henry Murray	64756	DNS	NaN	False	1908
Harvey Sutton	64808	3 h8 r1/2	NaN	False	1908
Guy Haskins	922519	DNS	NaN	False	1908
Joseph Lynch	64735	DNS	NaN	False	1908

	year	country	noc	gold	silver	bronze	total
0	1896	United States	USA	11	7	2	20
1	1896	Greece	GRE	10	18	19	47
2	1896	Germany	GER	6	5	2	13
3	1896	France	FRA	5	4	2	11
4	1896	Great Britain	GBR	2	3	2	7

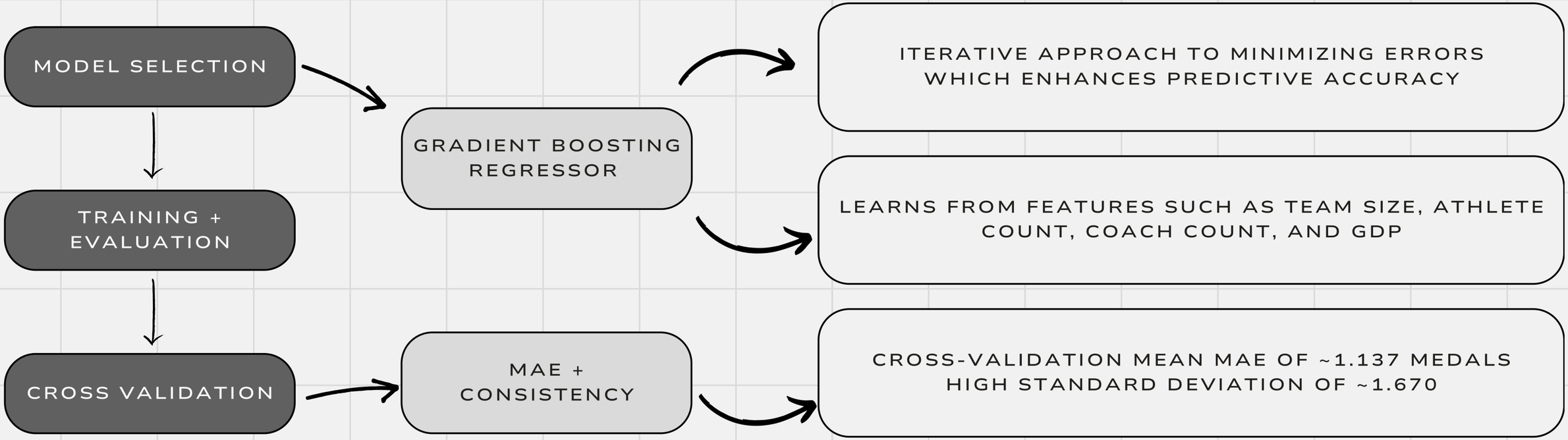
PREDICTIVE MODEL

FEATURES

1. TEAMS COUNT
2. ATHLETES COUNT
3. COACHES COUNT
4. GDP



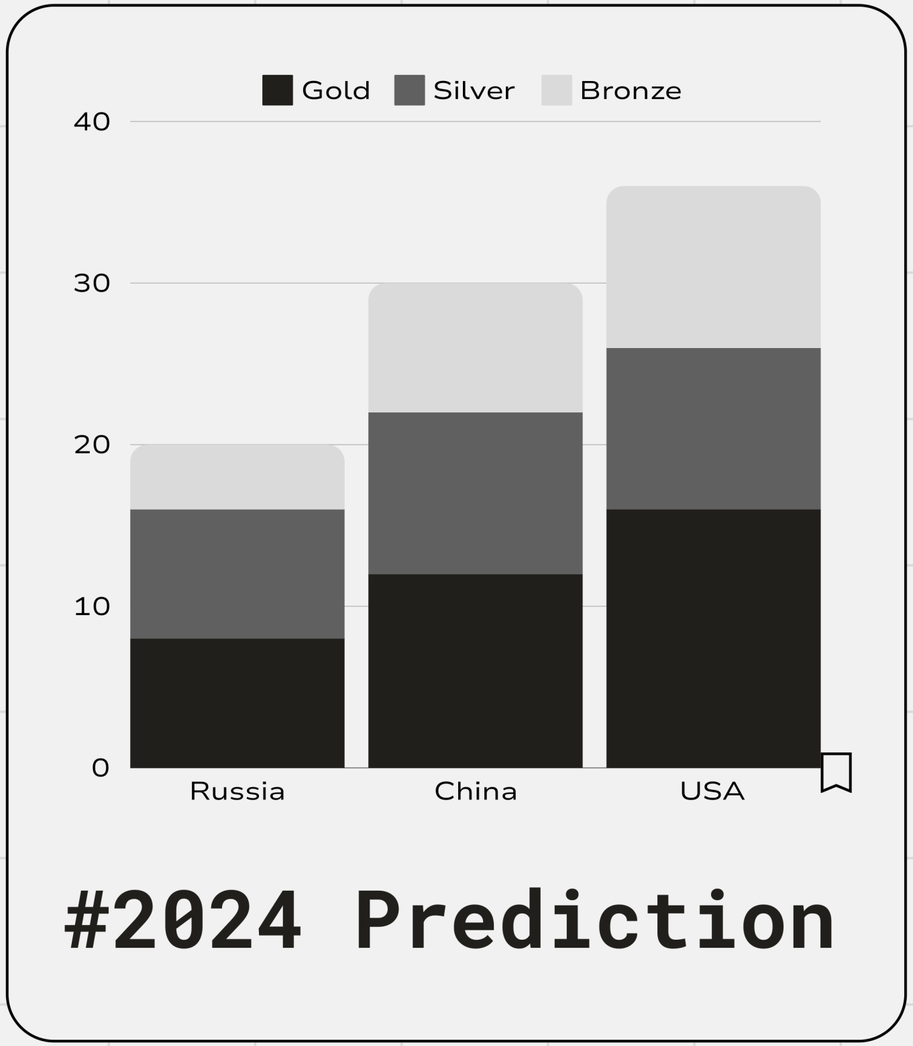
MODELLING APPROACH



RESULT

BASED ON THE 2021 OLYMPIC DATASET, OUR PREDICTIVE MODEL FORECASTS THAT THE USA WILL EMERGE AS THE TOP PERFORMER WITH THE HIGHEST TOTAL MEDAL COUNT. FOLLOWING THE USA, CHINA AND RUSSIA ARE PREDICTED TO SECURE THE NEXT POSITIONS BASED ON THEIR PERFORMANCE.

OUR MODEL WAS TRAINED ON THE 2021 DATA AND IS NOW PREPARED TO MAKE PREDICTIONS FOR THE UPCOMING EVENTS. IT ACHIEVED MEAN ABSOLUTE ERRORS OF 4.0, 4.7, AND 3.1 FOR GOLD, SILVER, AND BRONZE MEDALS RESPECTIVELY. THESE METRICS REFLECT THE AVERAGE MAGNITUDE OF ERRORS IN OUR PREDICTIONS, INDICATING OUR MODEL'S CAPABILITY IN FORECASTING MEDAL OUTCOMES WITH REASONABLE ACCURACY BASED ON HISTORICAL DATA.



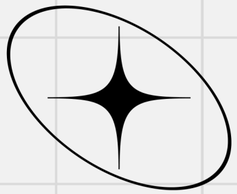
#2024 Prediction

CHALLENGES & LIMITATIONS

Acquiring legitimate data. Finding the source of where the data comes from and determining whether it is trustable.

Data inaccessibility arose when attempting to construct a 2024 Olympic file mirroring the 2021 dataset, where we faced a shortfall of 2 columns.

Merging the different files into one coherent dataframe was a challenge as we constantly had to make sure important data was not disregarded.



THANK YOU

- https://worldpopulationreview.com/countries/by-gdp
- https://www.kaggle.com/datasets?search=2021+tokyo+olympics