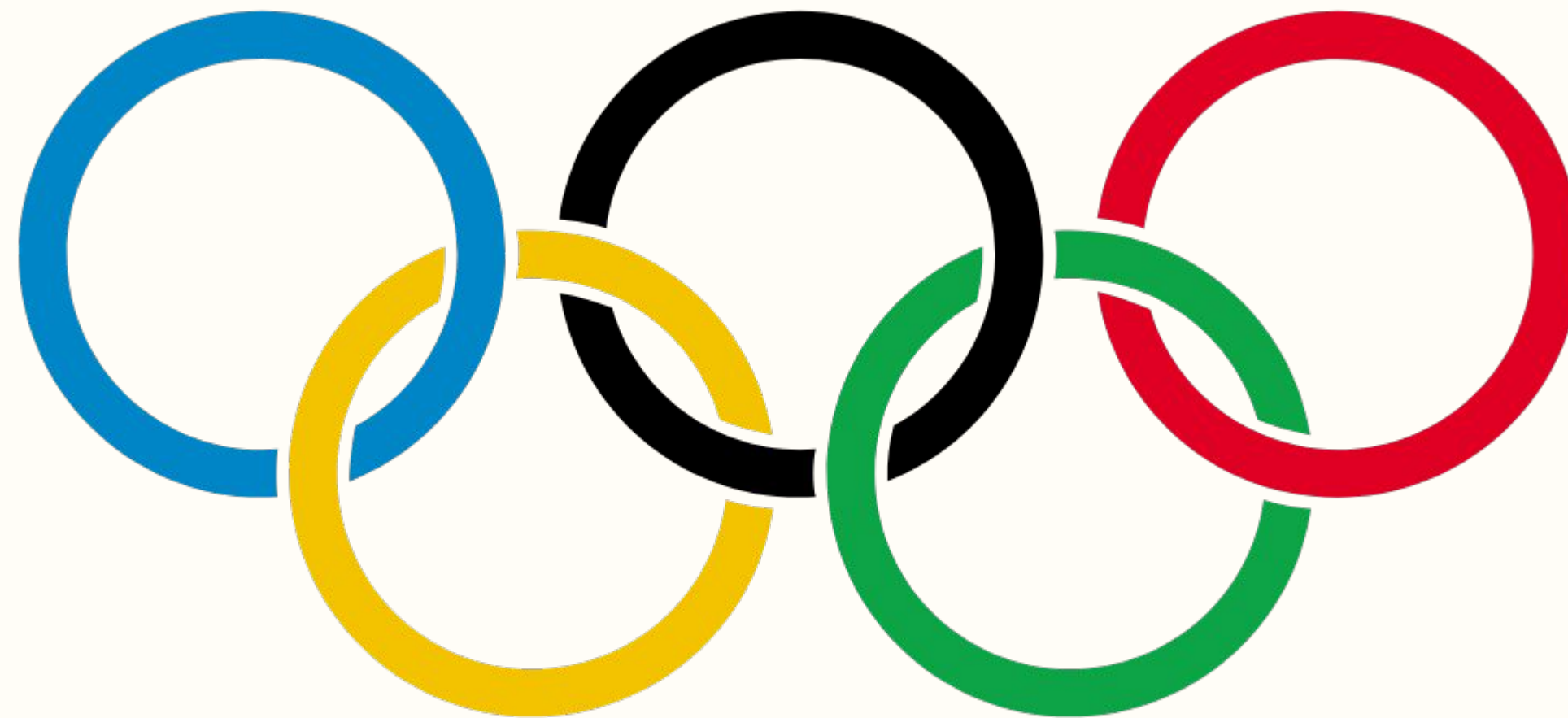


Predicting Olympic Results



By: Rohan Giri, Jennifer Gonzalez, Sameer Khan,
Kristen Lowe, María Laura Peña, Scott Stempak

Table Of Contents



1

Intro / Data
Cleaning and
Merging

2

Exploratory
Data Analysis

3

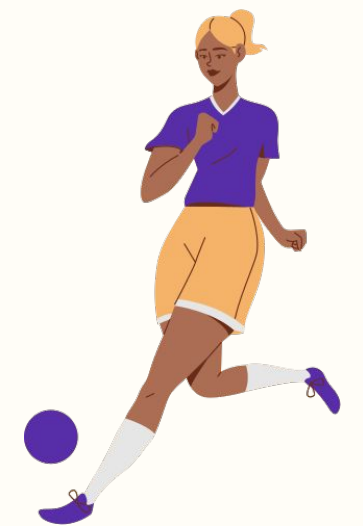
Machine Learning
Models



Intro / Data Cleaning and Merging



To what extent
does a country's
resources impact
its performance at
the Olympics?



Datasets from Kaggle



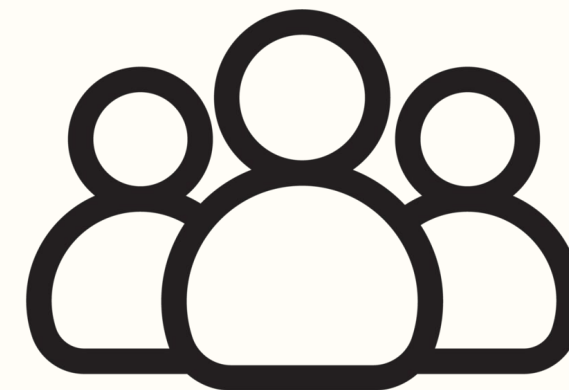
Olympic Summer & Winter
Games, 1896-2022

medal counts, host
country, etc.



World GDP by
Country: 1960-2022

GDP by year



Country Population
from 1960 to 2022

population by year



Countries of
the World

area in sq. km., region,
coastline ratio

Data Cleaning and Merging

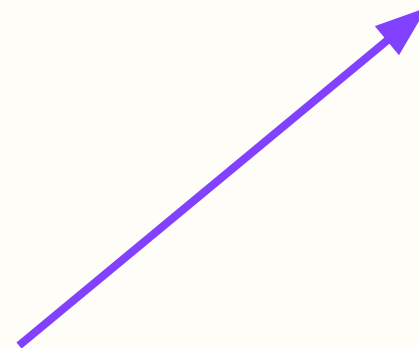
Clean

- mapped country names
- pivoted data frames
- removed missing values



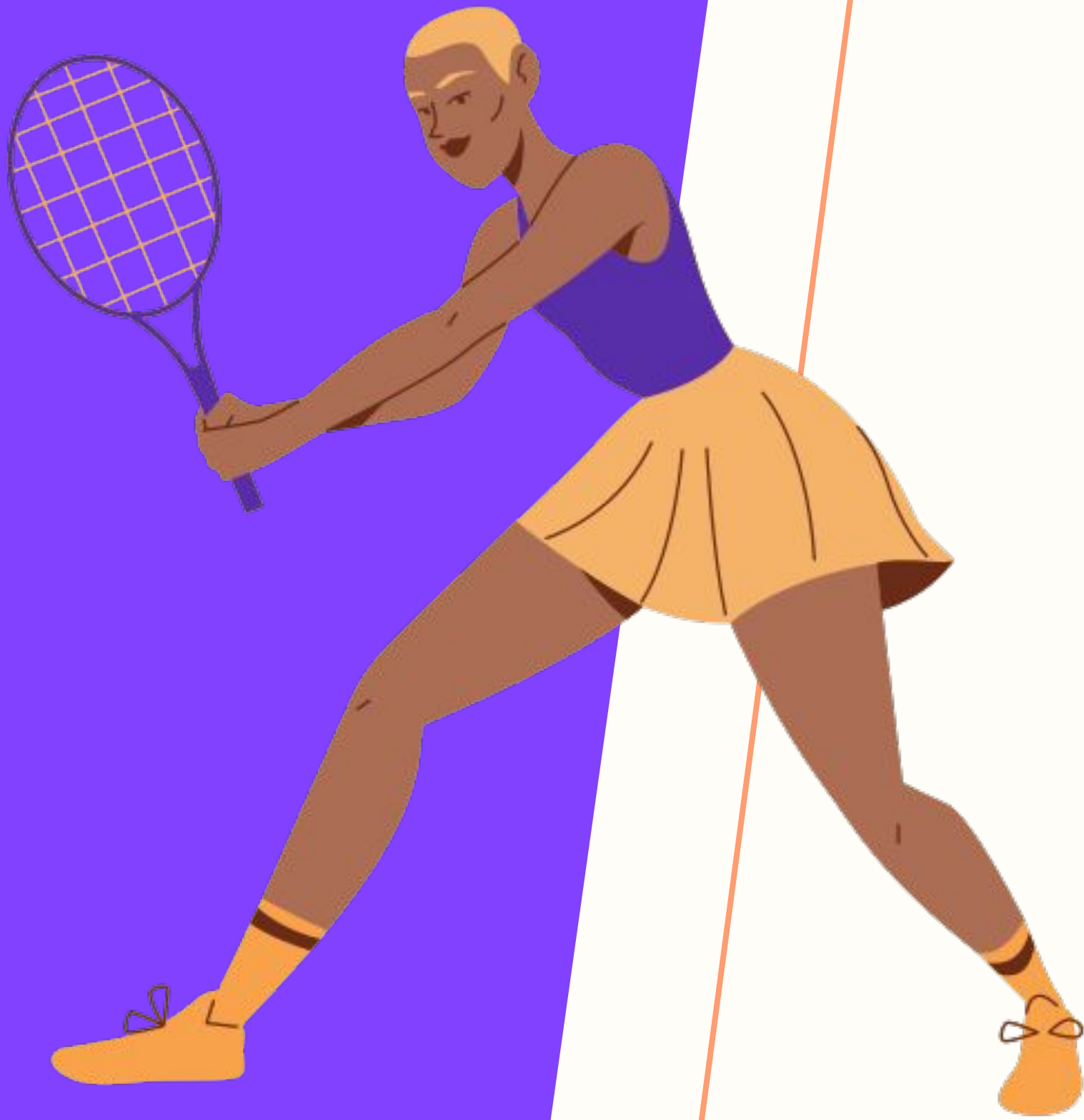
Merge

- olympics + GDP
- + population
- + other country resources
- feature engineering



Resulting Variables

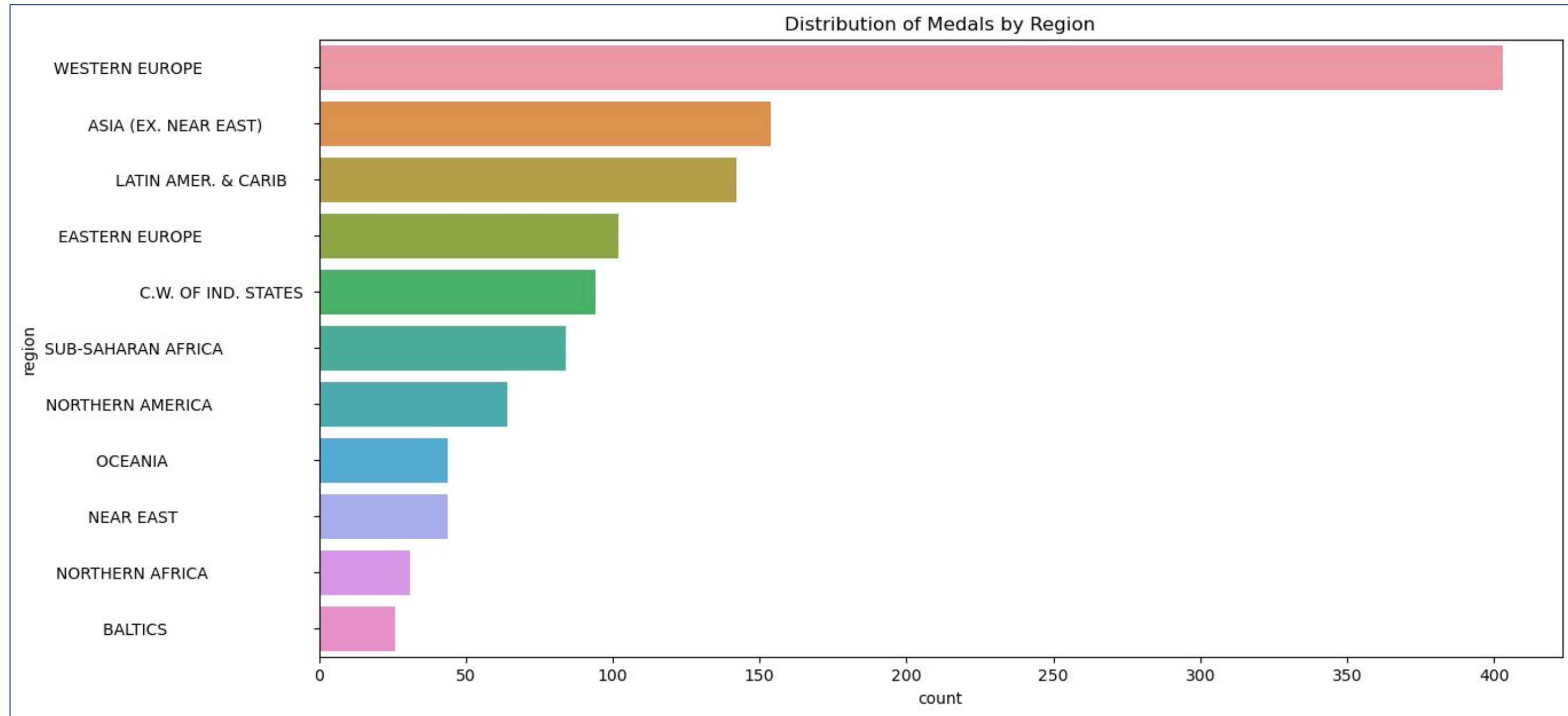
- country
- year
- medal count (gold, silver, bronze)
- host country
- game season (winter/summer)
- GDP
- population
- population density
- region
- area (sq. km.)
- coastline (coast-to-area ratio)
- GDP per capita
- host country status (binary)



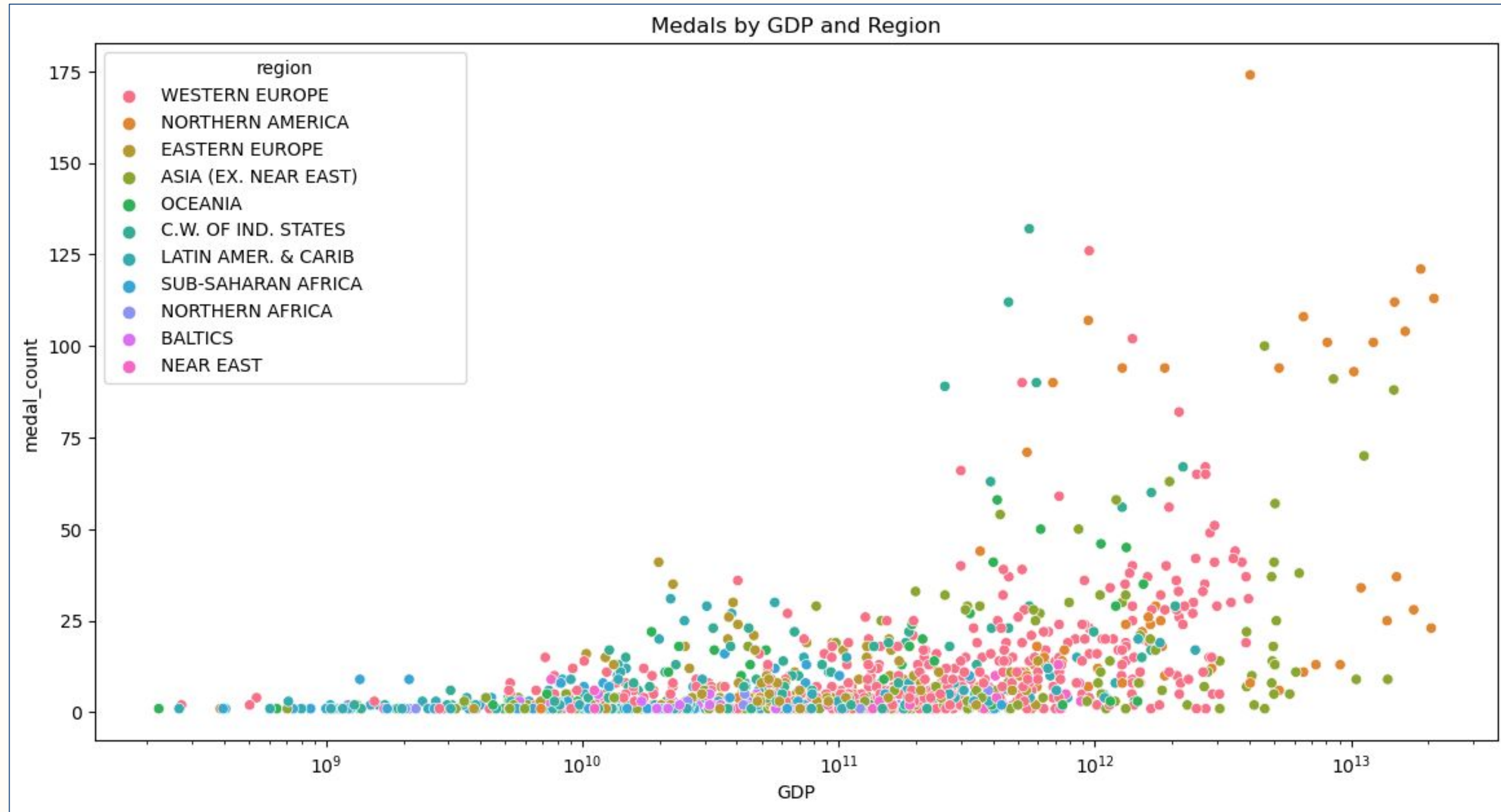
Exploratory Data Analysis



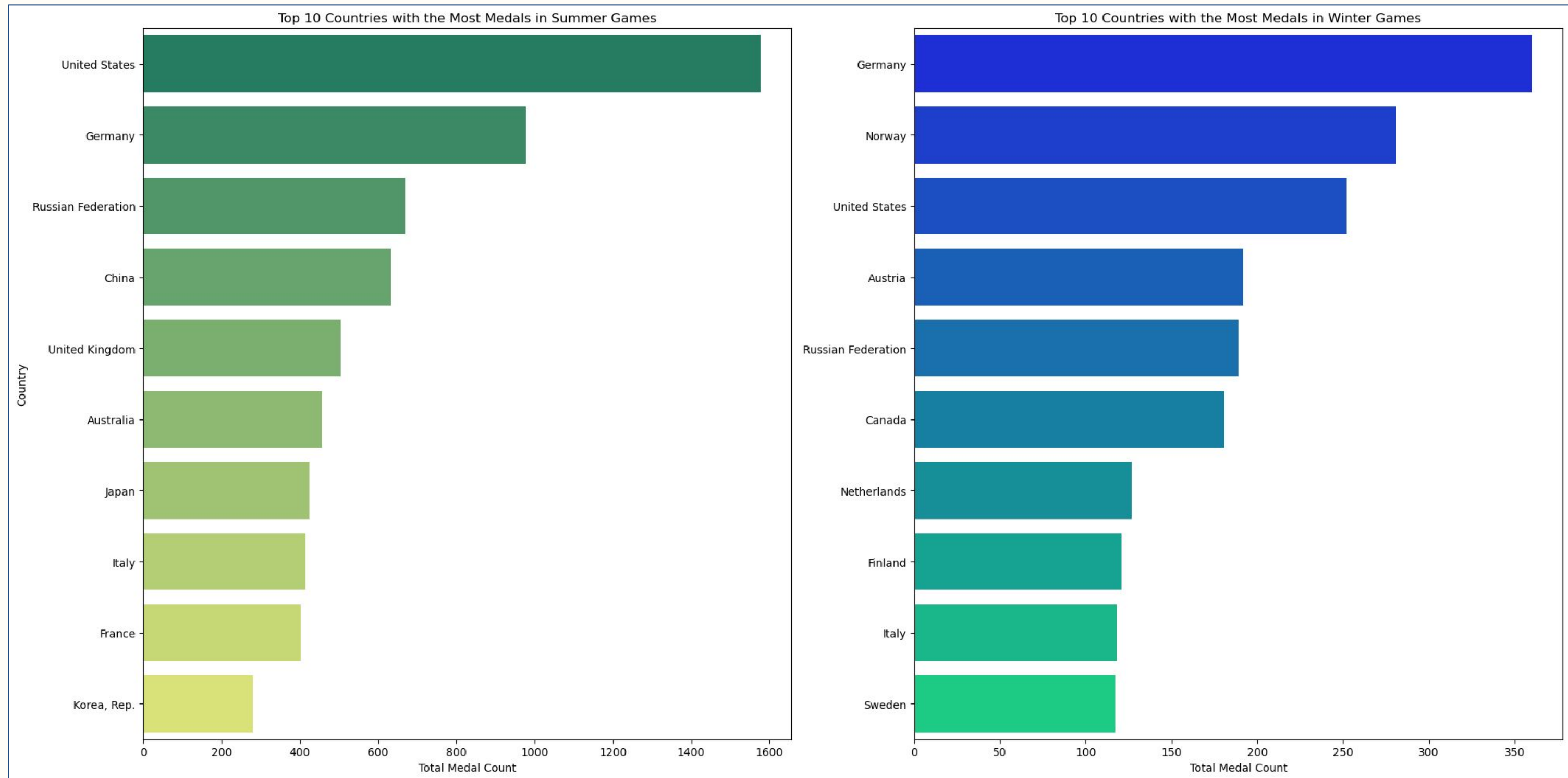
EDA



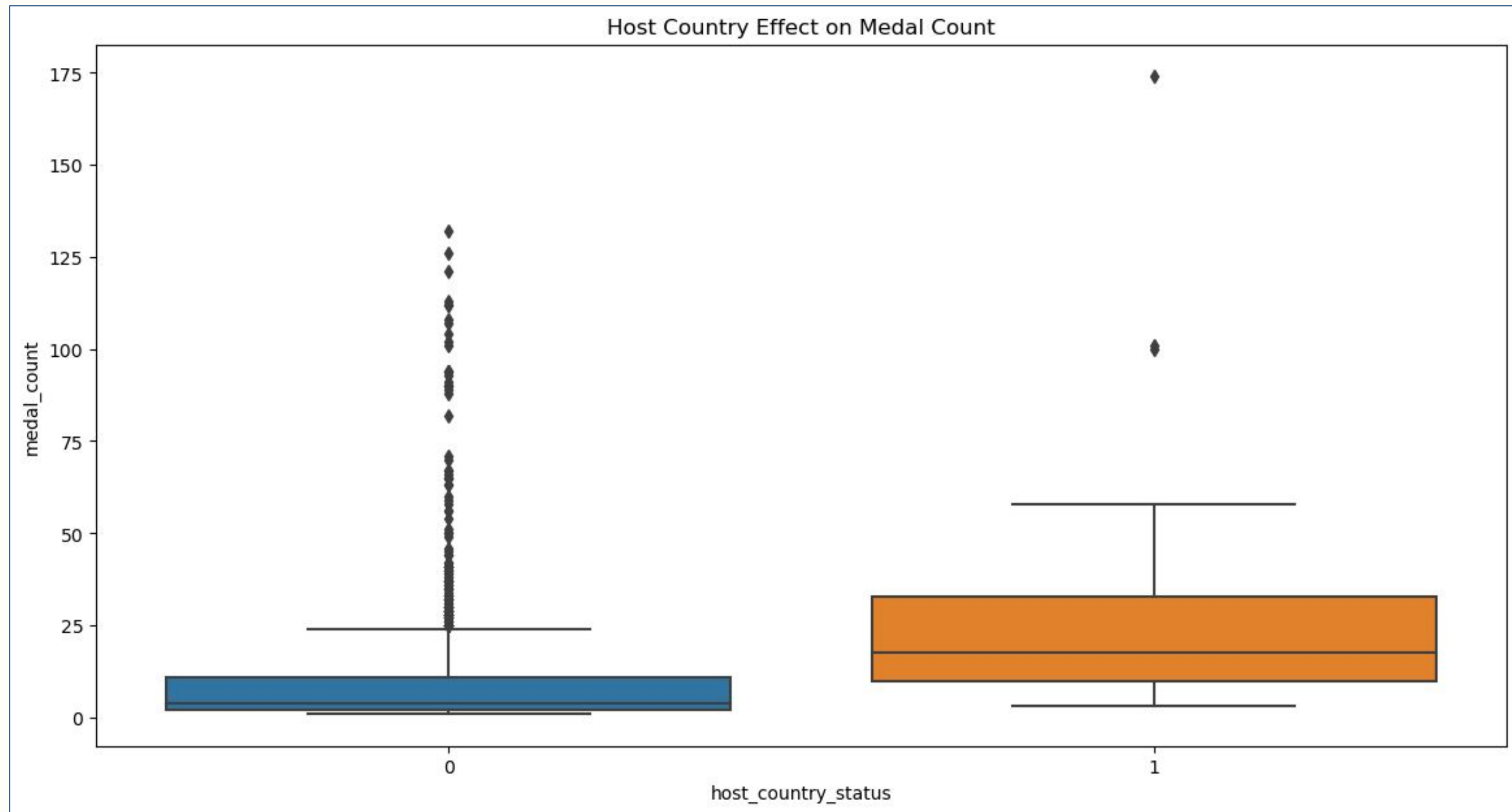
EDA



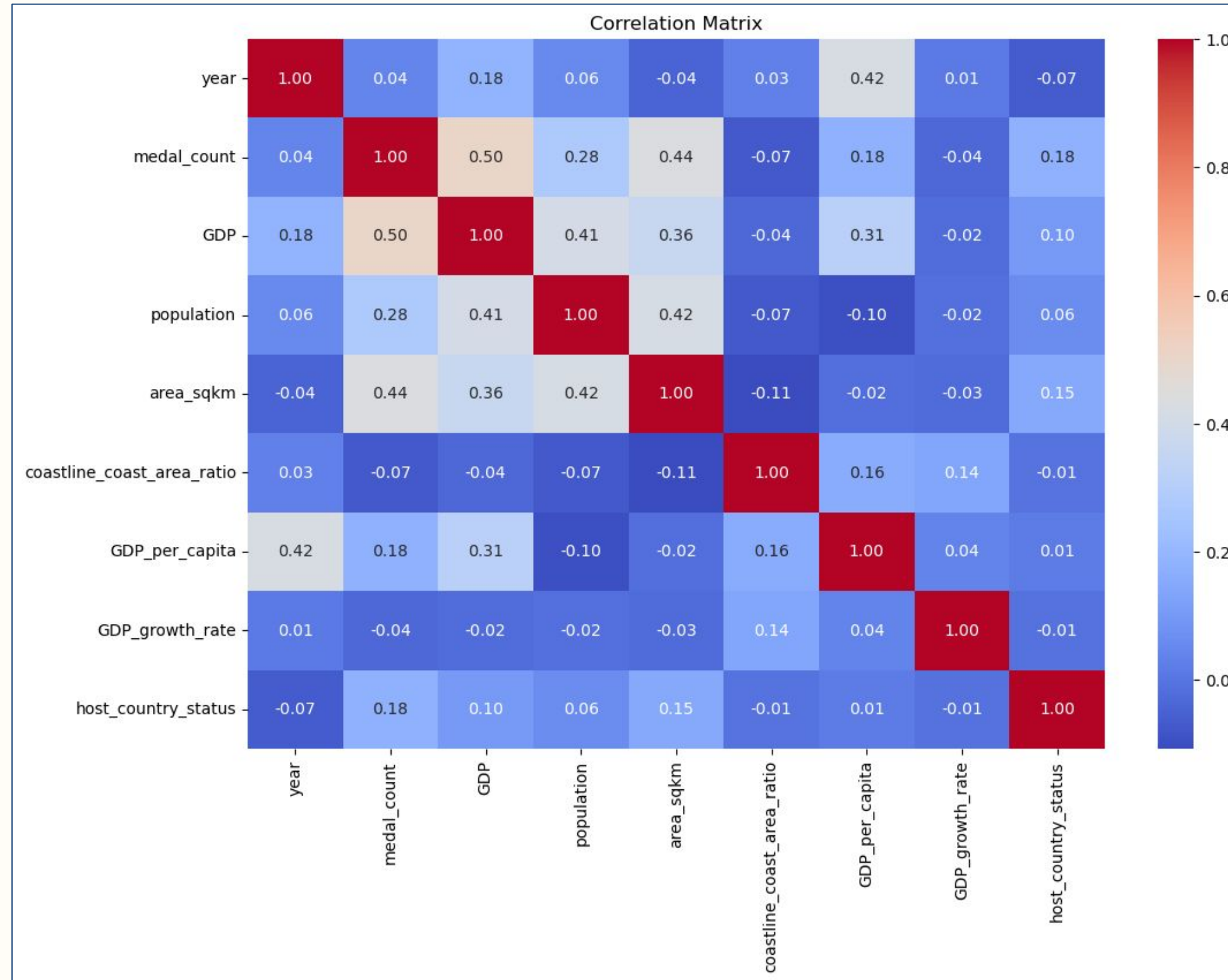
EDA



EDA



EDA





Machine Learning Models

Models Overview

Decision Tree

Bagging

Random Forest

Gradient Boosting

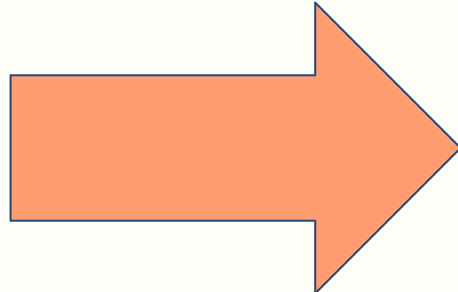


Summer Decision Tree Model

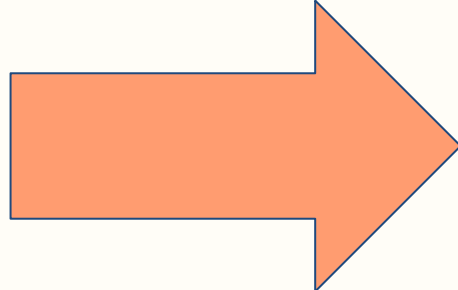
Evaluation Metrics
MAE: 7.785714
MSE: 224.551948
RMSE: 14.985057
R ² : 0.322729



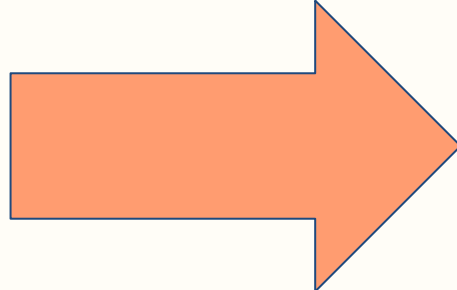
Summer Bagging Model

Evaluation Metrics	Optimize Hyperparameters using Grid Search	Optimized Evaluation Metrics
MAE: 5.201948	 <ul style="list-style-type: none">● bootstrap: False● bootstrap_features: False● max_features: 0.7● max_samples: 0.7● n_estimators: 200	MAE: 4.794478
MSE: 85.332792		MSE: 57.540310
RMSE: 9.237575		RMSE: 7.585533
R ² : 0.742628		R ² : 0.826453

Summer Random Forest Model

Evaluation Metrics	Optimize Hyperparameters using Grid Search	Optimized Evaluation Metrics
MAE: 4.674238	 <ul style="list-style-type: none">● bootstrap: True● max_depth: 20● max_features: sqrt● min_samples_leaf: 1● min_samples_split: 2● n_estimators: 100	MAE: 4.507012
MSE: 52.975255		MSE: 49.951581
RMSE: 7.278410		RMSE: 7.067643
R ² : 0.840221		R ² : 0.849341

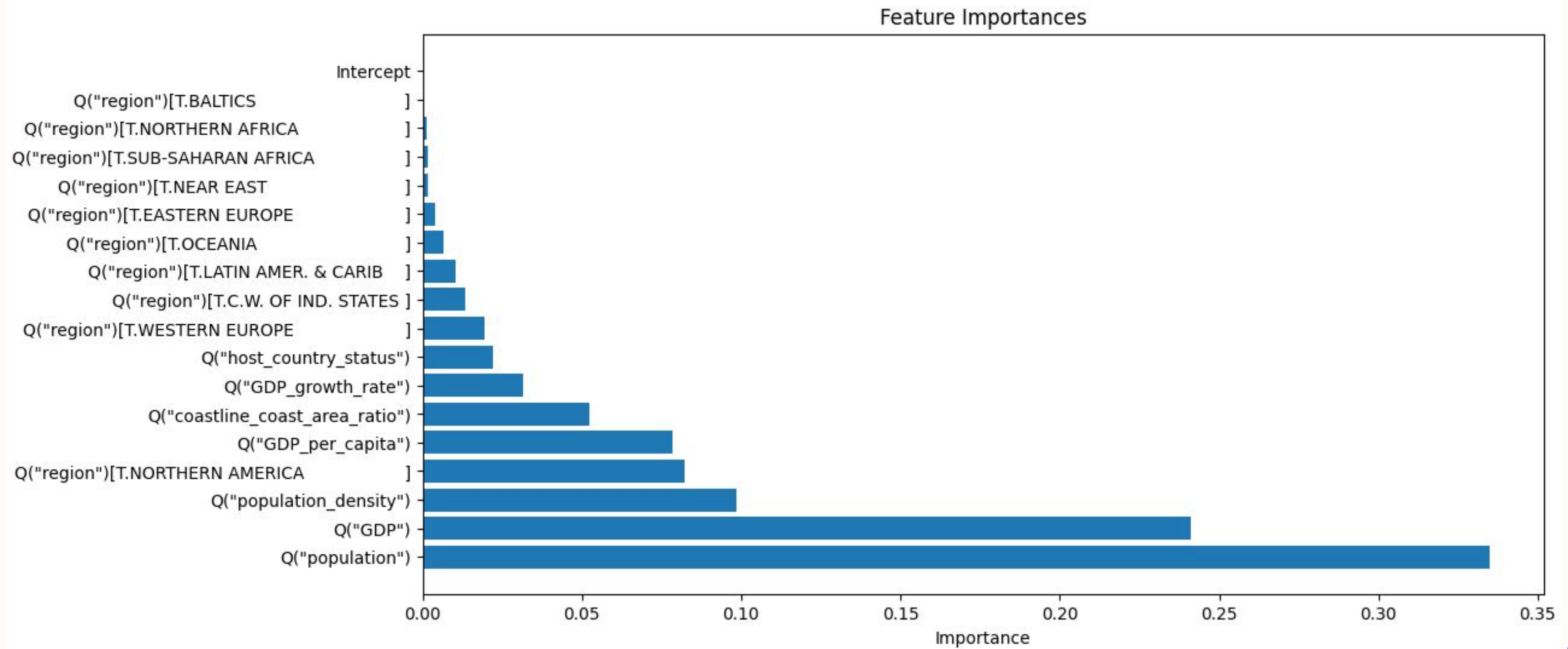
Summer Gradient Boosting Model

Evaluation Metrics	Optimize Hyperparameters using Grid Search	Optimized Evaluation Metrics
MAE: 5.013585	 <ul style="list-style-type: none">● learning_rate: 0.1● max_depth: 5● max_features: sqrt● min_samples_leaf: 2● min_samples_split: 2● n_estimators: 250● subsample: 0.8	MAE: 4.538984
MSE: 66.179258		MSE: 47.197012
RMSE: 8.135063		RMSE: 6.870008
R ² : 0.800397		R ² : 0.857649

Summer Olympics Models Summary

	MAE	MSE	RMSE	R ²
Decision Tree	7.786	224.552	14.985	0.323
Bagging	5.202	85.333	9.238	0.743
Random Forest	4.674	52.975	7.278	0.840
Gradient Boosting	5.014	66.179	8.135	0.800
Optimized Bagging	4.794	57.540	7.586	0.826
Optimized Random Forest	4.507	49.951	7.068	0.849
Optimized Gradient Boosting	4.539	47.197	6.870	0.858

Summer Olympics Models Feature Importance

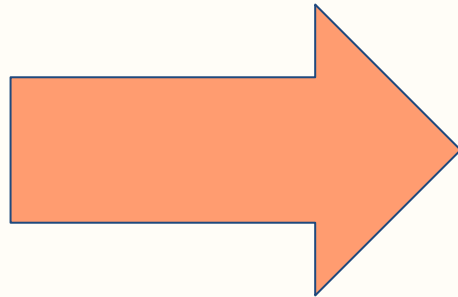


Winter Decision Tree Model

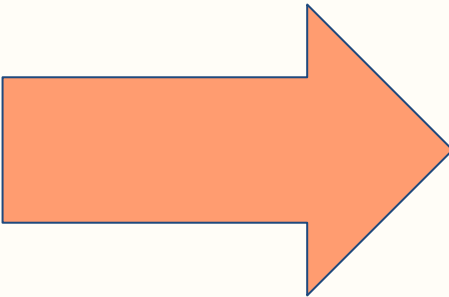
Evaluation Metrics
MAE: 5.224138
MSE: 49.396552
RMSE: 7.028268
R ² : 0.484955



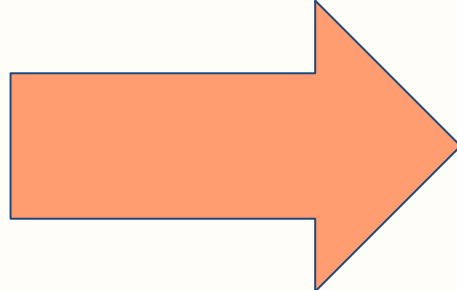
Winter Bagging Model

Evaluation Metrics	Optimize Hyperparameters using Grid Search	Optimized Evaluation Metrics
MAE: 3.950000	 <ul style="list-style-type: none">● bootstrap: False● bootstrap_features: False● max_features: 1.0● max_samples: 0.7● n_estimators: 100	MAE: 3.884483
MSE: 30.017759		MSE: 30.226231
RMSE: 5.478846		RMSE: 5.497839
R ² : 0.687013		R ² : 0.684839

Winter Random Forest Model

Evaluation Metrics	Optimize Hyperparameters using Grid Search	Optimized Evaluation Metrics
MAE: 4.059540	 <ul style="list-style-type: none">• bootstrap: False• max_depth: 20• max_features: sqrt• min_samples_leaf: 1• min_samples_split: 2• n_estimators: 200	MAE: 3.615690
MSE: 31.426216		MSE: 27.948496
RMSE: 5.605909		RMSE: 5.286634
R ² : 0.672327		R ² : 0.708588

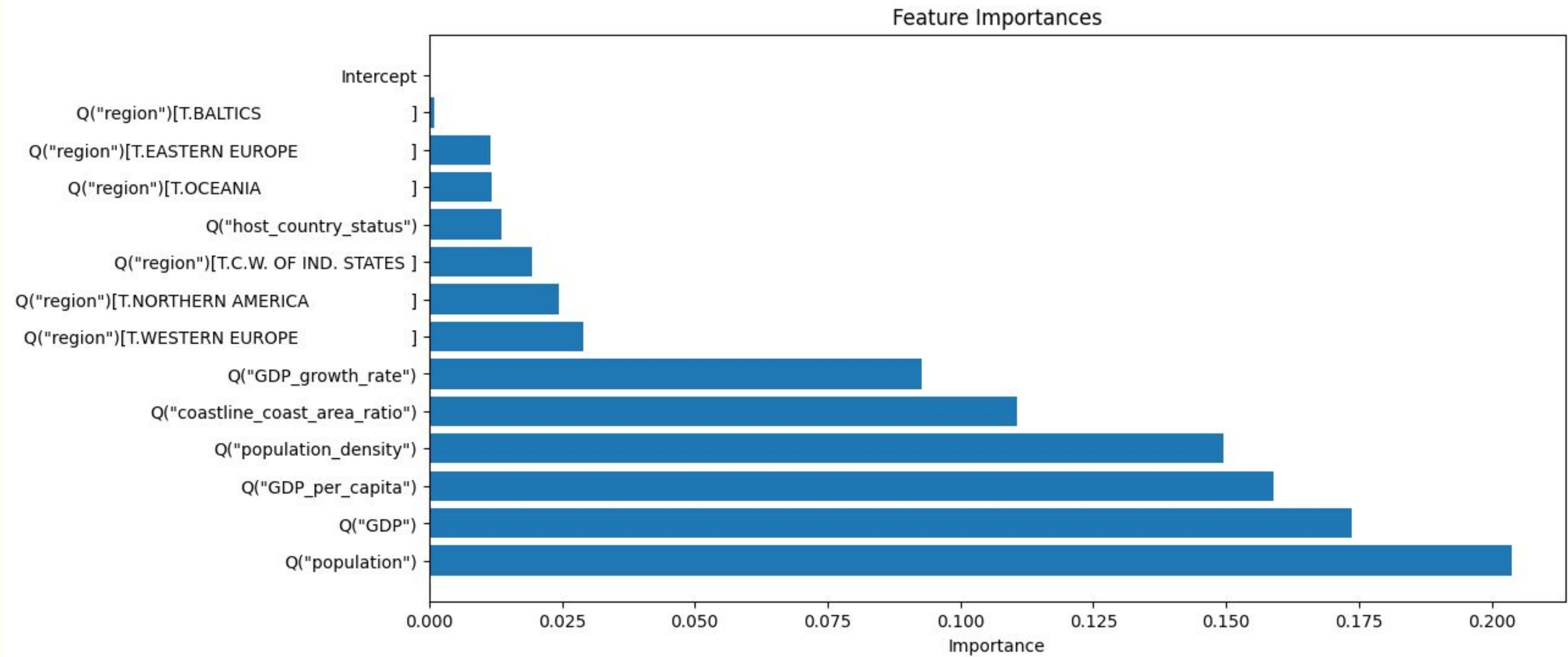
Winter Gradient Boosting Model


Evaluation Metrics	Optimize Hyperparameters using Grid Search	Optimized Evaluation Metrics
MAE: 4.152153	 <ul style="list-style-type: none">● learning_rate: 0.1● max_depth: 3● max_features: sqrt● min_samples_leaf: 4● min_samples_split: 2● n_estimators: 250● subsample: 0.8	MAE: 3.675121
MSE: 32.578276		MSE: 28.368045
RMSE: 5.707738		RMSE: 5.326166
R ² : 0.660315		R ² : 0.704214

Winter Olympics Models Summary

	MAE	MSE	RMSE	R ²
Decision Tree	5.224	49.397	7.028	0.485
Bagging	3.950	30.018	5.479	0.687
Random Forest	4.060	31.426	5.606	0.672
Gradient Boosting	4.152	32.578	5.708	0.660
Optimized Bagging	3.884	30.226	5.498	0.684
Optimized Random Forest	3.616	27.948	5.287	0.709
Optimized Gradient Boosting	3.675	28.368	5.326	0.704

Winter Olympics Models Feature Importance



The background features several abstract geometric shapes. On the left, a large purple parallelogram is partially visible, with an orange-outlined parallelogram overlapping it. On the right, there are two more purple parallelograms, one at the top and one at the bottom, both with orange outlines. The text is centered in a purple, sans-serif font.

Can our model
predict how many
medals the US will
bring home in
2024?

Our prediction:

(from the Summer optimized gradient boosting model)

```
[127]: usa2024 = pd.read_csv('USA2024.csv').head(1).drop(columns=['Unnamed: 0'])  
       grid_summer_gb.predict(usa2024)
```

```
[127]: array([113.81950708])
```

114 medals!



Conclusion

■ GDP and Population
are the most predictors

■ Limitations
we weren't able to take into
account a country's cultural
emphasis on sports or other
economic indicators besides
GDP



The background features several abstract geometric shapes. On the left, a large purple parallelogram is partially visible, with an orange parallelogram overlapping its right side. On the right side, there are three smaller shapes: a purple parallelogram at the top, an orange parallelogram below it, and another purple parallelogram at the bottom right, with an orange parallelogram overlapping its right side.

Thanks for listening!
Any questions?