Analyzing 120 years (1896-2016) Olympic Dataset

# Team Members:

- 1. Rajath John Bosco
- 2. Manshi Shah
- 3. Chaitya Mehta



### **Data Cleaning**

Does team enjoy Home Advantage?

Analyzing number of athletes and events in Summer and Winter Olympics over the years

Does a country enjoy geographical advantage?

Does athlete enjoy a physiological advantage?

Analyzing the top performing country in Olympics

Does hosting an Olympic has any impact on GDP of hosting country?

Analyzing Female trend over the years in Olympics

Calculating TOP 3 performing events for different countries in summer vs winter

Analyzing trend of Medals won by athletes of USA in olympics

Can we predict the medal tally of a country?

Does Human Development Index play a role in deciding the medal tally of a country?



### **Olympics Dataset, NOC-**

https://www.kaggle.com/heesoo37/120-years-of-olympic-history-athletes-and-results¶

**GDP Dataset - https://datahub.io/core/gdp** 

Population Dataset - <a href="https://vizhub.com/celtic660/datasets/world\_pop">https://vizhub.com/celtic660/datasets/world\_pop</a>

**Human Development Index -**

https://data.world/brianray/gapminder-hdi-human-developmen

Olympics Host - <a href="https://www.kaggle.com/pavanraj159/olympic-host">https://www.kaggle.com/pavanraj159/olympic-host</a>

Cleaning data of the medal column by replacing NA with DNW(Did not Win)

Data Cleaning:

Categorizing events into Individual and Team events

We awarded 1 medal per team instead of awarding medal to each player in the team.

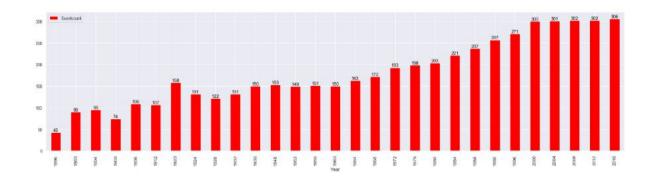
If a Country has won medal in soccer, it should count one single medal instead of 11 medals (since 11 players in a team)

# Does Team enjoy Home Advantage?

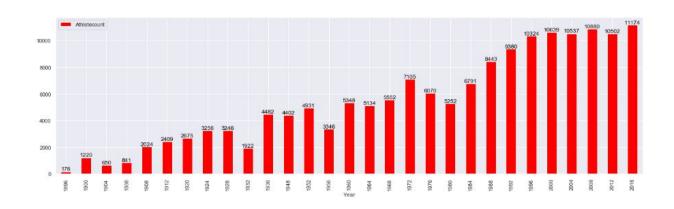
	Year_of_Hosting	Host_Country_code	Previous_Year_Tally	Tally_Overall	Next_Year_Tally
1	1900	FRA	11.0	95.0	2.0
2	1904	USA	48.0	228.0	21.0
3	1908	GBR	25.0	141.0	41.0
4	1912	SWE	24.0	63.0	63.0
5	1920	BEL	6.0	40.0	13.0
6	1924	FRA	42.0	38.0	25.0
7	1928	SUI	25.0	17.0	1.0
8	1928	NED	11.0	23.0	8.0
9	1932	USA	56.0	110.0	57.0
11	1948	SUI	17.0	22.0	13.0
12	1948	GBR	14.0	27.0	11.0
13	1952	NOR	7.0	5.0	3.0
14	1952	FIN	22.0	20.0	15.0
15	1956	ITA	20.0	24.0	35.0
17	1956	SWE	34.0	18.0	6.0
18	1960	USA	74.0	71.0	90.0
19	1960	ITA	24.0	35.0	27.0
21	1964	JPN	16.0	28.0	24.0
22	1968	FRA	15.0	15.0	12.0
23	1968	MEX	1.0	7.0	1.0
24	1972	JPN	24.0	29.0	24.0
25	1972	FRG	25.0	39.0	38.0
26	1976	AUT	3.0	1.0	4.0
31	1984	USA	93.0	169.0	91.0
32	1988	CAN	43.0	10.0	17.0
33	1988	KOR	18.0	32.0	26.0
34	1992	FRA	16.0	27.0	36.0
35	1992	ESP	4.0	21.0	15.0
37	1996	USA	106.0	97.0	90.0
43	2008	CHN	60.0	98.0	85.0
45	2012	GBR	46.0	63.0	65.0
47	2016	BRA	17.0	18.0	1.0

Analyzing number of athletes and events in Summer and Winter Olympics over the years

### No. of events over the years in Summer Olympics



### No. of athletes over the years in Summer Olympics



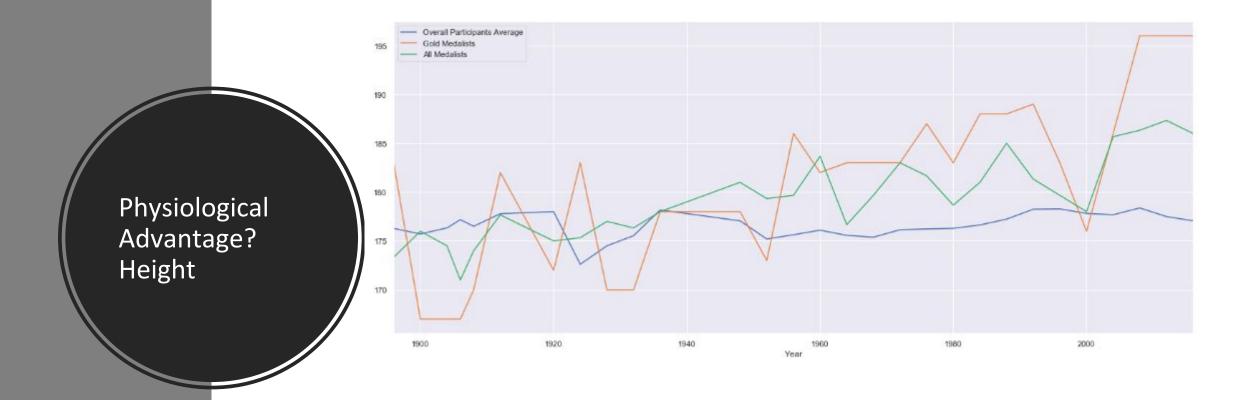
## Does country enjoy geographical Advantage?

### **Top Performing countries in Summer**

	region	Summer Gold Medal
17	USA	1846
11	Russia	617
5	Germany	493
8	Italy	221
1	China	136

### **Top Performing countries in Winter**

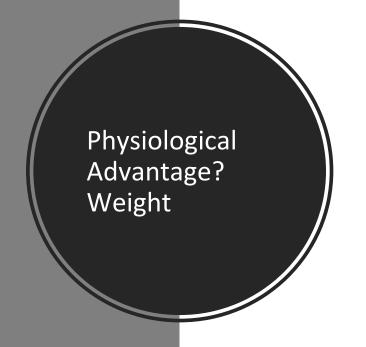
region	Winter Gold Medal
Canada	243
Russia	180
Germany	109
South Korea	45
Netherlands	40
	Canada Russia Germany South Korea



Height Comparison of Men's 100m Runners

Standards on height in Olympic running events for men.

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1859632/?page=2



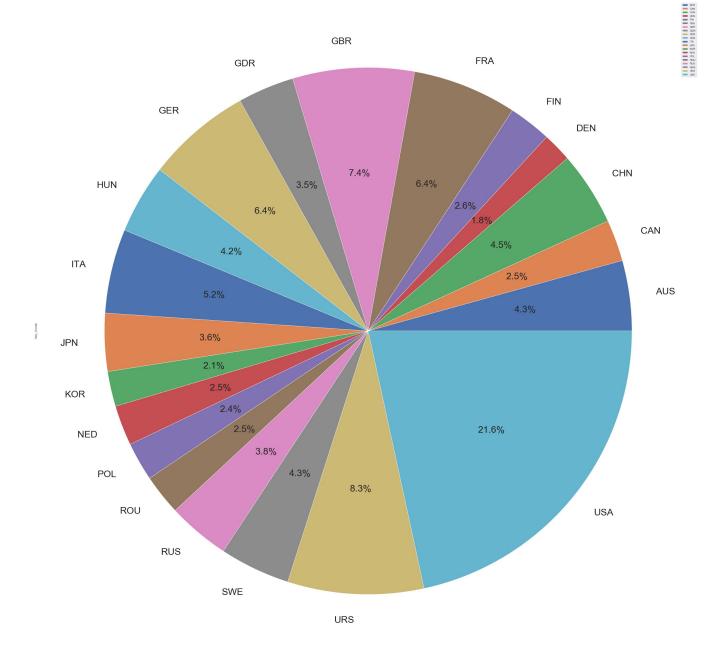


Weight Comparison of Men's 100m Runners

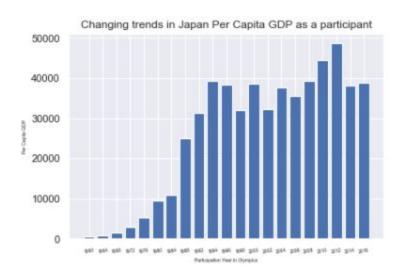
Standards weight in Olympic running events for men.

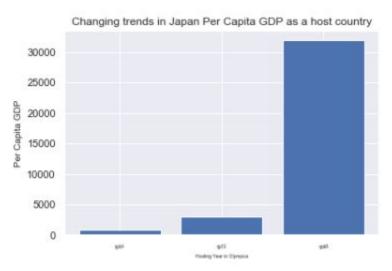
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1859632/?page=2

Top Performing country over the years in Olympics



Does hosting an Olympic has any impact on GDP of hosting country?

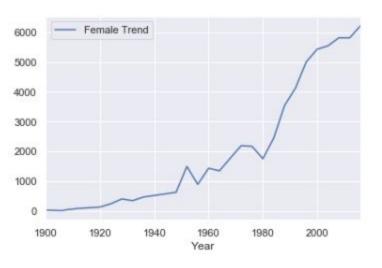




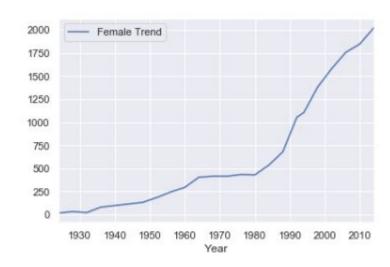
It can be observed that the 3 years where Japan hosted the Olympics, GDP per capita observed a higher increase in comparison to the GDP per capita in the participating years

# Changing Trends in women participation in Olympics

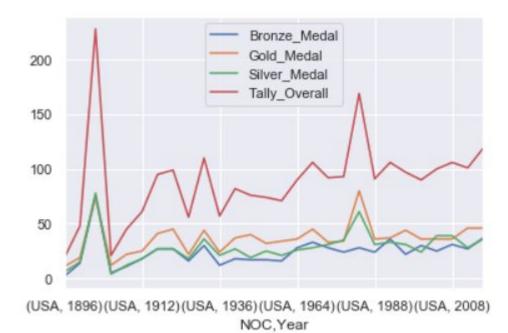
### **Female Trend - Summer Olympics**



### **Female Trend - Winter Olympics**



Analyzing trend of Medals won by athletes of USA in olympics



# Calculating TOP 3 performing events for different countries in summer vs winter

### Calculating top 3 PERFORMING events for USA,RUSSIA and UK in SUMMER OLYMPICS

	region	Sport Si	ummer Count	of Gol	d Meda	1		
53	USA	Swimming			64	19		
42	USA	Athletics			54	12		
43	USA	Basketball			28	31		
51	USA	Shooting			11	17		
46	USA	Football			6	56		
	region		Sport 5	Summer	Count	of	Gold	Medal
21	Russia	Gy	ymnastics					176
22	Russia	A	Handball					97
28	Russia	l.	Wrestling					97
26	Russia	Vo	olleyball					93
25	Russia	Synchronized	Swimming					54
	region	Sport	Summer Cour	nt of G	iold Me	da.	1	
41	UK	Tug-Of-War				10	6	
36	UK	Cricket				1.	2	
38	UK	Polo				1:	1	
37	UK	Motorboating				(	6	
39	UK	Racquets					3	

# Calculating top 3 PERFORMING events for USA, GERMANY and NORWAY in SUMMER OLYMPICS

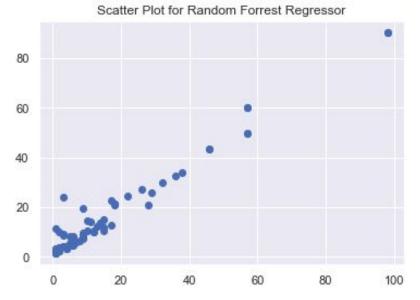
```
region Sport Winter Count of Gold Medal
17 USA Snowboarding 10
15 USA Freestyle Skiing 8
16 USA Skeleton 3
region Sport Winter Count of Gold Medal
4 Germany Bobsleigh 48
5 Germany Luge 44
6 Germany Ski Jumping 17
region Sport Winter Count of Gold Medal
8 Norway Nordic Combined 19
```



Dimension	Depende nt Variable:	Indicator	Hypot hesized Sign N/A	
Olympic Performance	Medal Count (M <sub>t</sub> )	The number of medal won by a country in a particular Olympics		
	Independ ent Variables			
Demographic Environment	Populatio n (N <sub>t</sub> )	The population size of a country at a particular Olympic year	+	
Economic Development	GDP per capita (Y <sub>t</sub> /N <sub>t</sub> )	The per capita GDP (measured in PPP current international dollars) of a country at a particular Olympic year	+	
Social Development	Human Developm ent Index HDI <sub>t</sub>	The Human Deployment Index of a country at a particular Olympic year	+	
Political Environment	Political System (Pol)	1 if the country is or used to be a socialist country or 0 otherwise	+	
Geographic Hosting country is the hosting country of the hosting co		+		

Factors that affect the medal tally of a country





R-Squared Value of Random Forest Regression is 0.96 (96% accuracy) and RMSE is 3.8 which means the predictor is predicting with an error of 4 medals.

Results from Random Forest Regression

Correlation of various factors that contribute towards the overall medal tally

	GDP	Population	Tally_Overall	Per_Capita_GDP	Human_Dev_Index
GDP	1	0.457163	0.642159	0.193817	0.169667
Population	0.457163	1	0.571119	-0.27975	-0.395806
Tally_Overall	0.642159	0.571119	1	0.0403584	0.132623
Per_Capita_GDP	0.193817	-0.27975	0.0403584	1	0.751846
Human_Dev_Index	0.169667	-0.395806	0.132623	0.751846	1



### **Books and Research Papers**

- 1. Jake VanderPlas. 2016. Python Data Science Handbook: Essential Tools for Working with Data (1st ed.). O'Reilly Media, Inc.
- 2. McKinney: "Python for Data Analysis", 2nd Edition, by Wes McKinney, ©2017.
- 3. "Going for Gold Medals: Factors affecting Olympic Performance"
- -"https://www.researchgate.net/publication/278585677 Going for Gold Medals Factors affecting Olympic Performance
- 4."Standards on age, height and weight in Olympic running events for men" <a href="https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1859632/?page=2">https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1859632/?page=2</a>

#### **Online References**

- 1. https://en.wikipedia.org/wiki/Olympic Games
- 2. https://stackoverflow.com/
- 3. <a href="https://www.olympic.org/">https://www.olympic.org/</a>
- 4. https://towardsdatascience.com/understanding-the-need-for-optimization-when-using-pandas-8ce23b83330c
- 5. https://www.kaggle.com/chadalee/olympics-data-cleaning-exploration-prediction
- 6. <a href="https://www.kaggle.com/pavanraj159/olympics-history-1896-2016">https://www.kaggle.com/pavanraj159/olympics-history-1896-2016</a>
- 7. <a href="https://www.kaggle.com/arunsankar/key-insights-from-olympic-history-data">https://www.kaggle.com/arunsankar/key-insights-from-olympic-history-data</a>
- 8. https://en.wikipedia.org/wiki/List of Olympic Games scandals and controversies

# THANK YOU!