

### Client 3: [SportsStats](#) (Olympics Dataset - 120 years of data)

SportsStats is a sports analysis firm partnering with local news and elite personal trainers to provide “interesting” insights to help their partners. Insights could be patterns/trends highlighting certain groups/events/countries, etc. for the purpose of developing a news story or discovering key health insights.

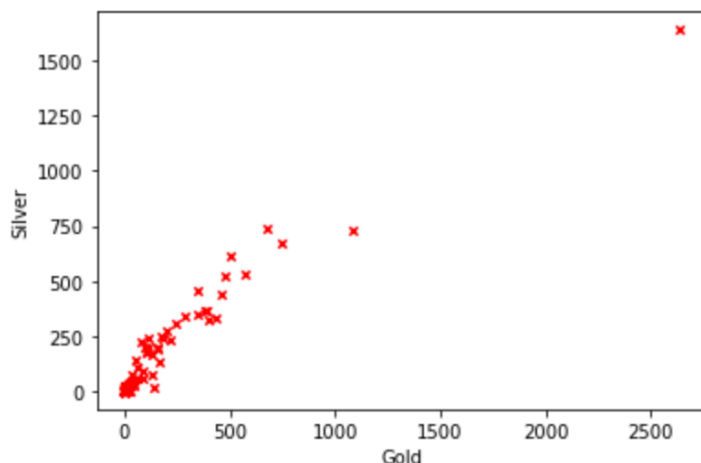
### Dive Deeper

Look deeper into the features you are investigating, consider:

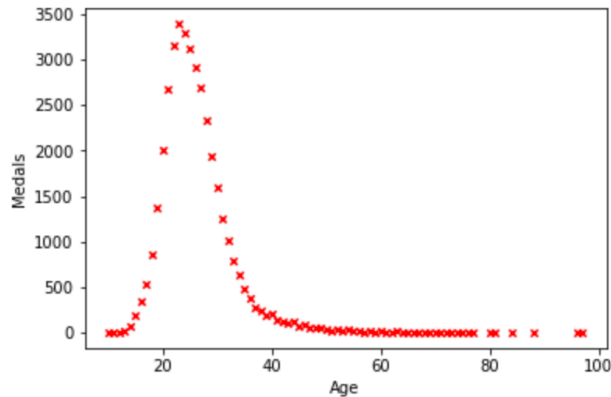
- Relationships / Correlation, Pearson Correlation
- Linear Regression for future prediction (if the relationship is linear)
- Textual Analysis for TF-IDF (Term Frequency-Inverse Document Frequency; Row-based and column-based, stop-word removal?)

Specify 1-2 correlations you discovered. List the fields that you found to be correlated and describe what you learned from these correlations.

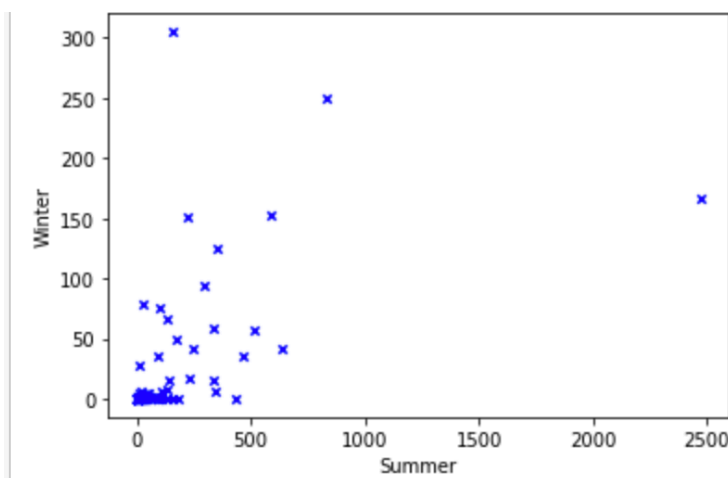
I first looked to analyze if there was a correlation between a country winning gold medals and its tally of silver medals and found a strong correlation (Pearson correlation of 0.95). This demonstrated that the more gold medals a country wins, the higher chances of it winning silver medals as well.



I also looked at a correlation between age and total number of medals and found an inverse correlation - Pearson correlation of -0.52. The number of medals goes up from around age 10 to mid-20's and then declines thereafter



Further analyzing if countries that participate in both summer and winter olympics whether there was a correlation between gold medals won in the summer olympics vs the winter olympics, we find there some correlation between the gold medals across season barring a few outliers (Pearson correlation of 0.57)



## Go Broader

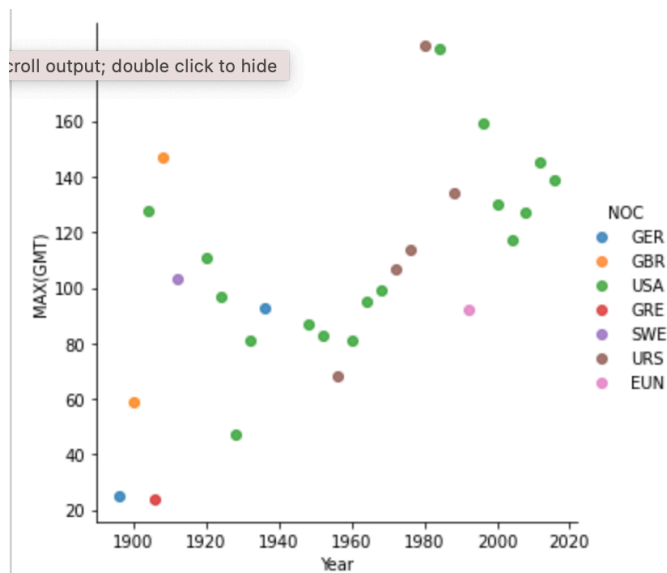
Expand the features you are investigating. Look for connections/relationships that you may have initially missed.

1. What jumps out at you now?
2. Use the descriptive stats to point you to features that you may now want to consider.

What key terms did you discover in any text analysis, for whom? Any themes? If you are not analyzing text, summarize what other things you are considering in your analysis?

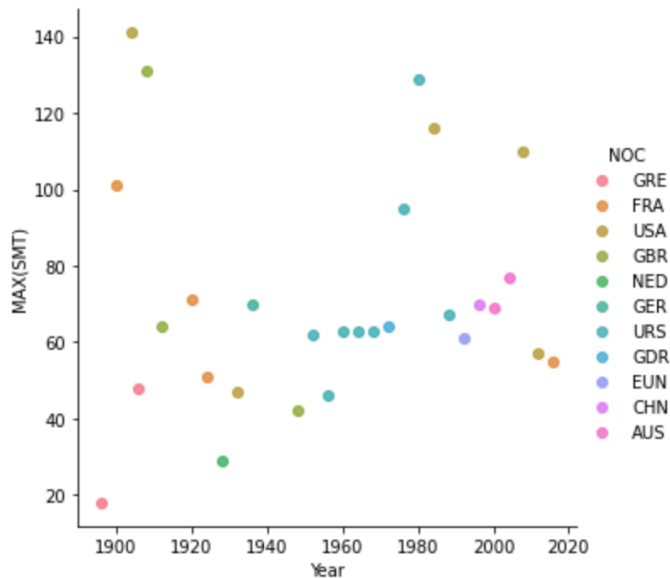
I want to understand how countries perform over time - so trying to understand which countries have had the most Gold, Silver medals over time (in the summer olympics).

Seven countries have been consistently winning the the maximum number of Golds in the summer olympics. It started with Germany and Great Britain performing the best in Gold medals and then USA has been consistently the performing nation in Gold medals (Chart 1).



*Chart 1 Countries with Maximum Gold Medals by Year*

For Silver Medals, only 11 countries consistently won most silver medals through the years in the summer olympics. It started with Greece performing the best in Silver medals and more recently the USA has been consistently performing well (Chart 2).



*Chart 2 Countries with Maximum Silver Medals by Year*

## New Metric

Create 1 or 2 new metrics to track relationships of data you discovered. Explain why you created them.

To understand the performance of country or importance of a sport, the 1st metric created was: number of medals by category (gold, silver, bronze)

To understand the effectiveness of a country's athletes in winning medals, I created the 2nd metric (normalized):

Medals (Gold, Silver, Bronze, Total)/Participant

In terms of overall medals by participant over the horizon, the top performing countries include the former Soviet Union, former East Germany, Australasia (although very few participants and outliers), an Unified Team and the United States

	NOC	Team	SUM(Year_Tot_Ath)	SUM(GMT)	SUM(SMT)	SUM(BMT)	SUM(Total_M)	Normalized_Tot_Medals
214	URS	Burevestnik	3937	1082	732	689	2503	0.635763
76	GDR	East Germany	1712	397	327	281	1005	0.587033
7	ANZ	Australasia	55	20	4	5	29	0.527273
66	EUN	Unified Team	603	127	71	81	279	0.462687
216	USA	United States	12850	2638	1641	1358	5637	0.438677