

FİNAL RAPORU:

YAREN SU SAATÇI

DataSet Seçimi:

Spora ilgim olduğu için olimpiyatları çok severek takip ederim. Ülkemizin de bu sene aldığı başarılarla gündemde Paris olimpiyatları konuşuluyordu. Ben de ülkelere göre olimpiyat başarılarını kıyaslamak istedim. Yaz ve kış olimpiyat verilerini içeren csv kullandım. Ayrıca ülkeleri ve kodlarını içeren csv ile araştırma alanımı genişlettim.

Csv Birleştirme:

Yaz ve kış olimpiyatları için csv leri summer ve winter olarak adlandırdım. Bunları merge ederek olympics adı altında yeni bir dataset oluşturdum. İki ayrı dataset yerine daha işlevsel hale geldi.

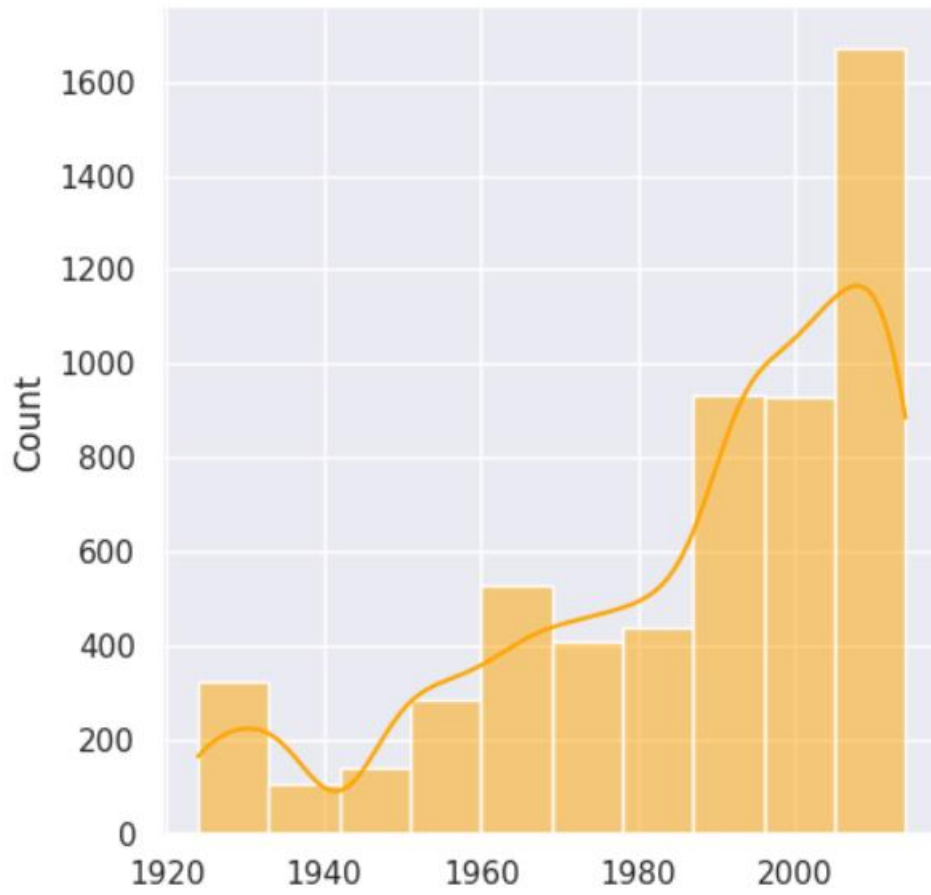
Data Temizliği:

Dataları daha iyi inceleyebilmek için null verileri ayıkladım. Ülkeleri kodları ile map ederek diğer adımlar için kolaylık oluşturdum. Örneğin, kullandığım yaz ve kış datasetlerinde ülke kodları bulunuyordu ama çok açıklayıcı değildi. Ülke adlarını ekleyerek daha anlaşılır hale getirdim. Sonrasında kod kolonunu dropladım.

Data içeriği inceleme:

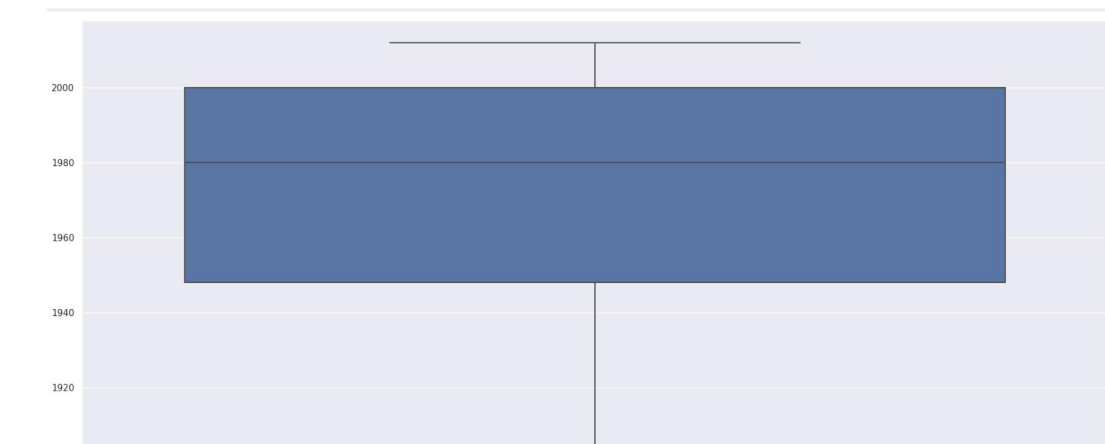
Birçok kod kullanarak datasetin içeriğini inceledim. Kolonlar hakkında daha çok fikrim oldu. Bu bilgileri nasıl kullanabileceğimi düşündüm.

Görselleştirme:



Temizlediğim verileri görselleştirmeye başladım. Bu örnekte kış olimpiyatlarında yıl bazında inceleme yaptım. Yıllara göre genelde bir artış olduğunu gözledim. Küçük dalgalanmaların sebebi o yıllardaki olaylar (savaş vb) incelenerek bulunabilir.

Yıllara göre boxplot:



Türkiye'yi inceledim:

Kadın sporcu sayısı: 9

Erkek sporcu sayısı: 77

En çok madalya kazanan sporcumuz: SULEYMANOGLU, Naim

Madalya sayısı: 3

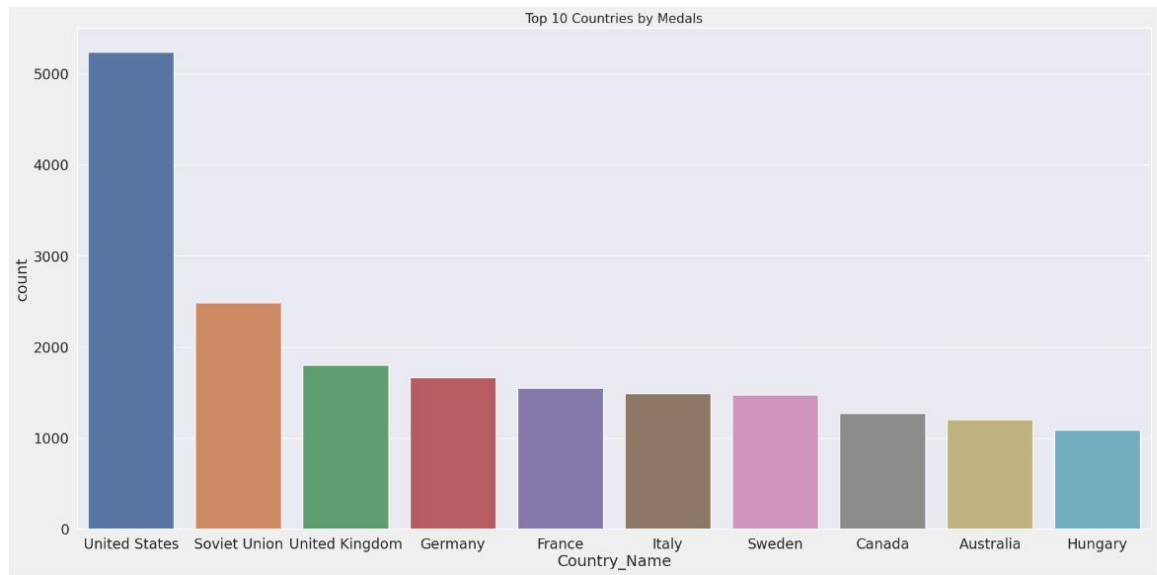
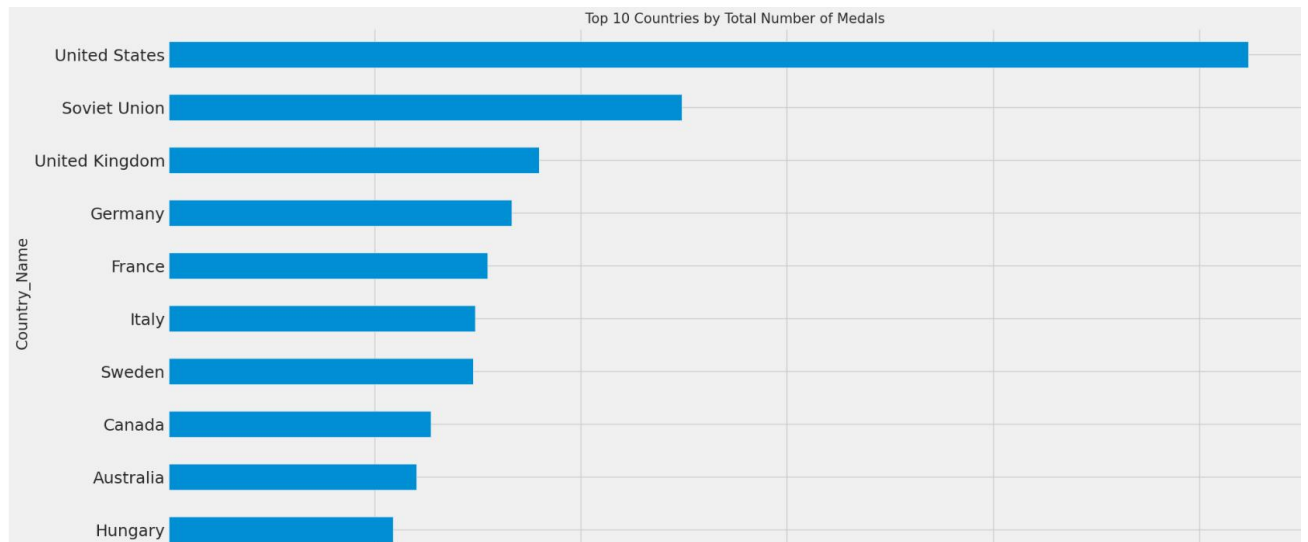
En çok sporcu bulunan spor dalı: Wrestling

Sporcu sayısı: 58

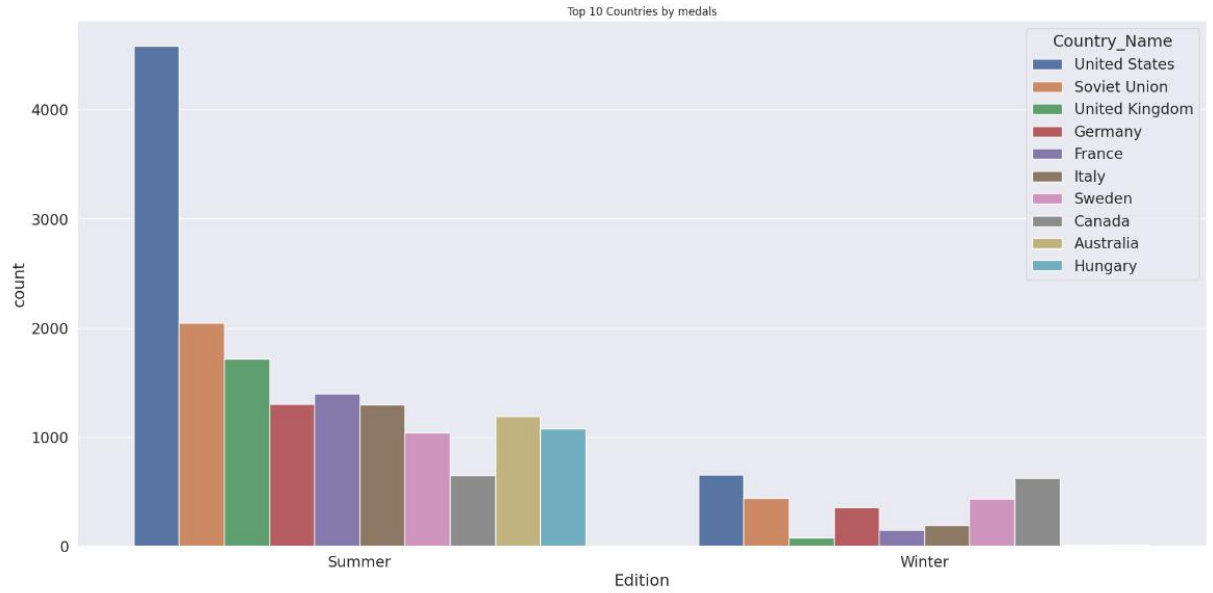
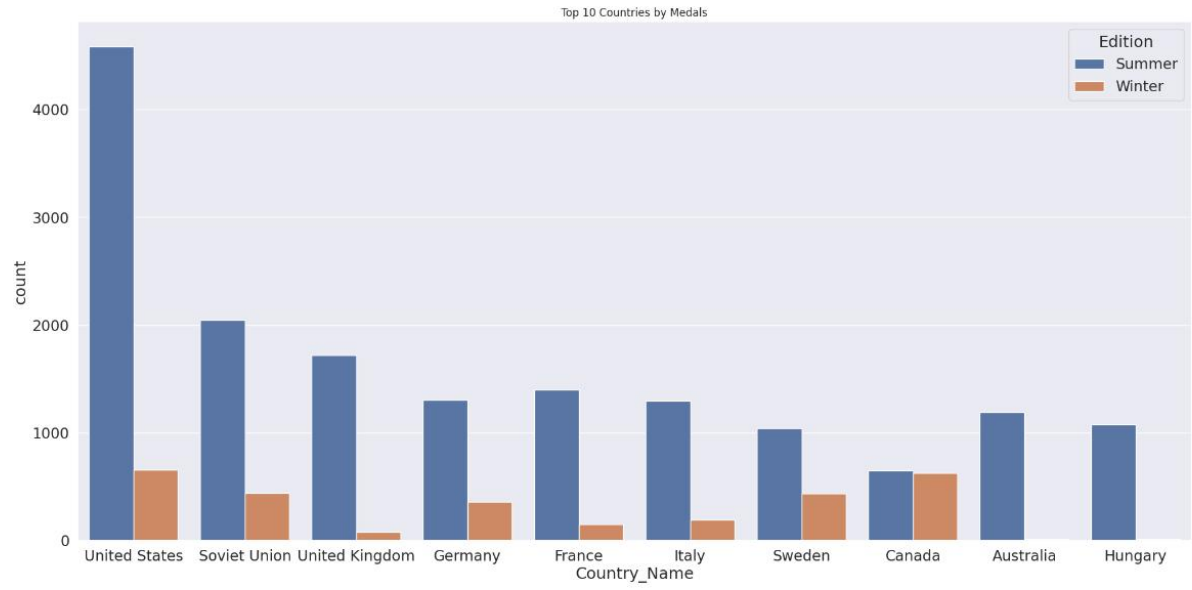
Şehir ve Yıllara göre regplot:



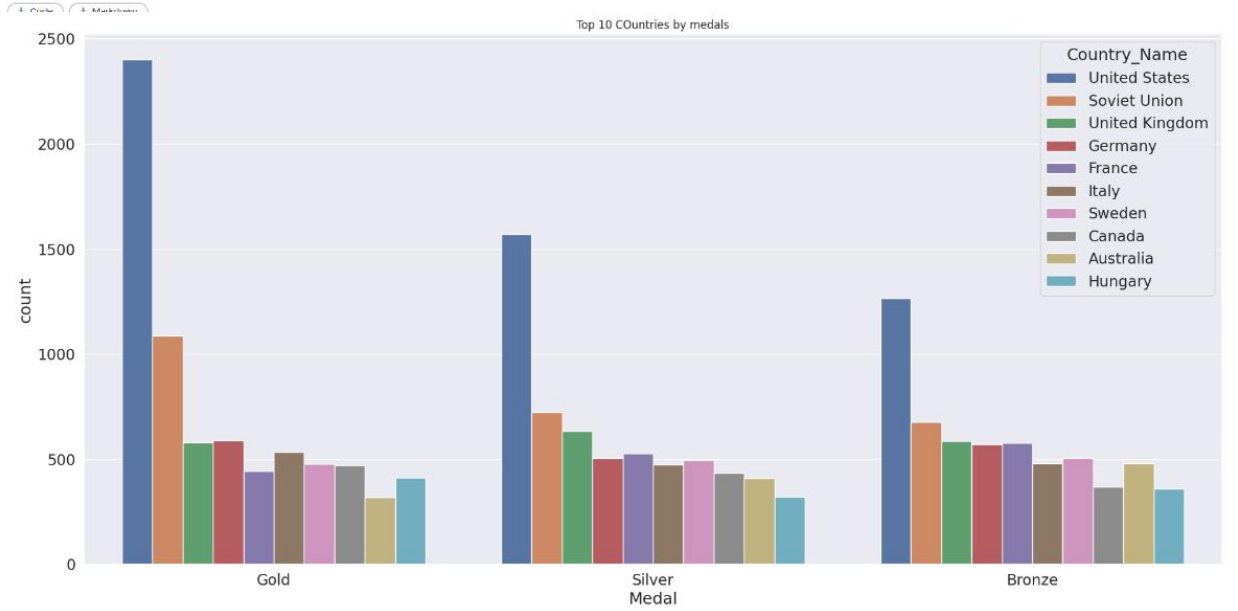
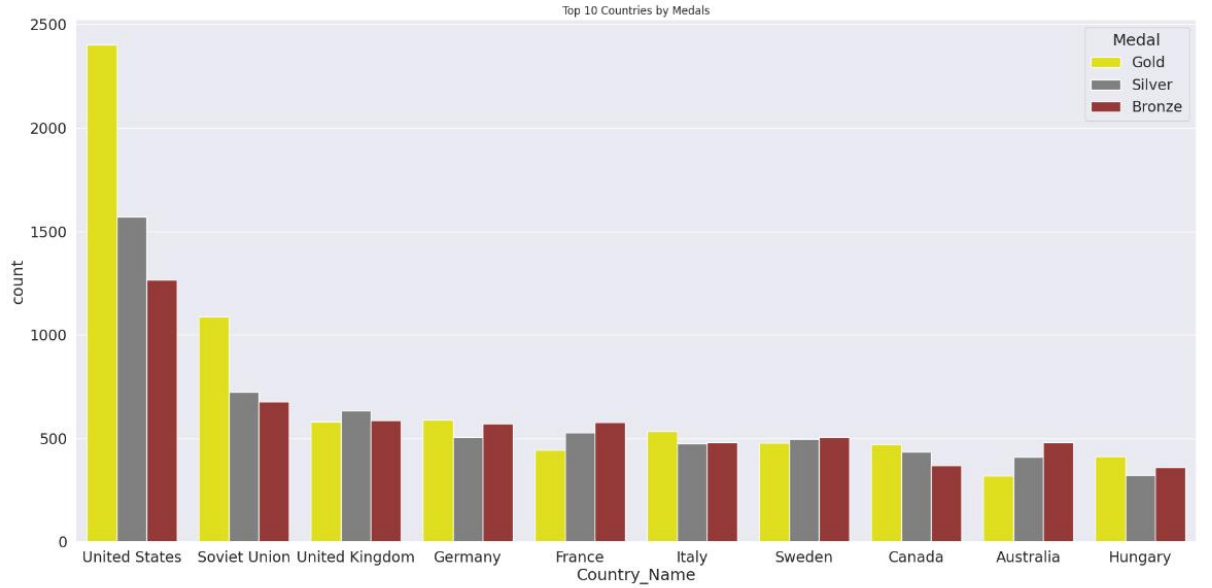
En başarılı ülkeleri inceledim:



Mevsimlere ayırarak:



Madalya çeşitlerine ayırarak:



Grafiklerde en başarılı ülke olarak Amerika'yı görebiliriz. Sonrasında Rusya, İngiltere ve Almanya da başarılı ülkeler olarak sıralanmaktadır.

[37_

	Total	Gold	Silver	Bronze	Code	Population	GDP per Capita
Country							
United States	5238	2402	1571	1265	USA	321418820.0	56115.718426
Soviet Union	2489	1088	724	677	NaN	NaN	NaN
United Kingdom	1799	580	632	587	GBR	65138232.0	43875.969614
Germany	1665	589	504	572	GER	81413145.0	41313.313995
France	1548	444	526	578	FRA	66808385.0	36205.568102
--	--	--	--	--	--	--	--
Iraq	1	0	0	1	IRQ	36423395.0	4943.760388
Macedonia	1	0	0	1	MKD	2078453.0	4852.657848
Mauritius	1	0	0	1	MRI	1262605.0	9252.110724
Niger	1	0	0	1	NIG	19899120.0	358.958152
Togo	1	0	0	1	TOG	7304578.0	559.635877

GDP Per Capita ‘ya ve nüfusa göre inceleme yaptım.

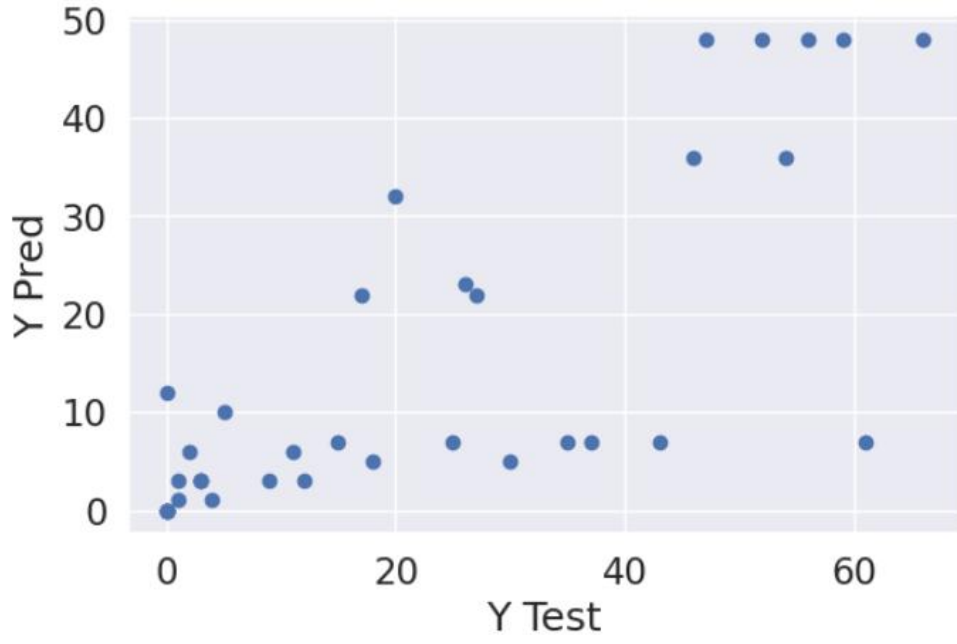
Train ve Test split yaptım.

Düşündüğüm gibi popülasyonun etkisi vardı. Ama tahmin ettiğim kadar doğrusal bir ilişki yokmuş. Türkiye’yi örnek verebilirim.

acc

0.8888888888888888

Aslında GDP Per Capita'yı ekleyerek **train** test yaptım. Acc: 0.77 çıktı. Negatif etkisi olduğunu gördüğüm için GDP yi teste dahil etmedim.



	Coefficient
Population	1.910835e-08
Total_Games	1.477425e+00

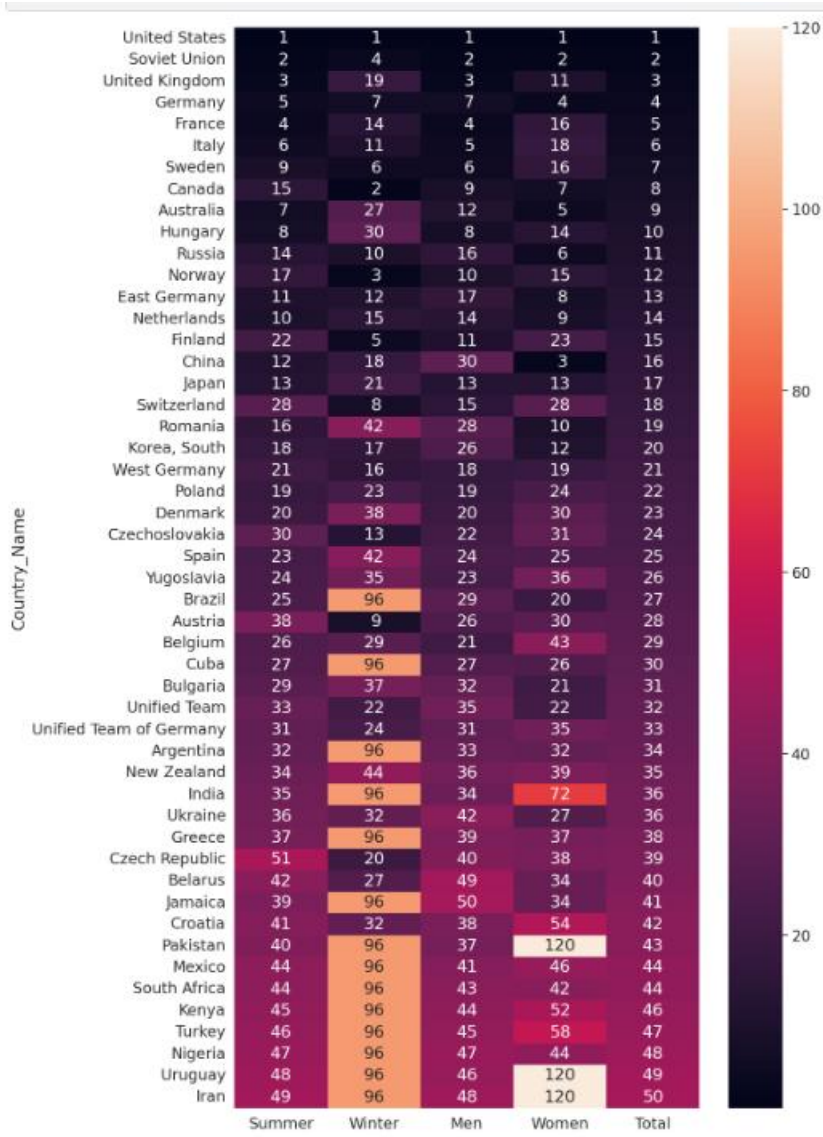
Hipotezler sunarak bunları inceledim.

Bulduğum sonuçlardan bazıları ise;

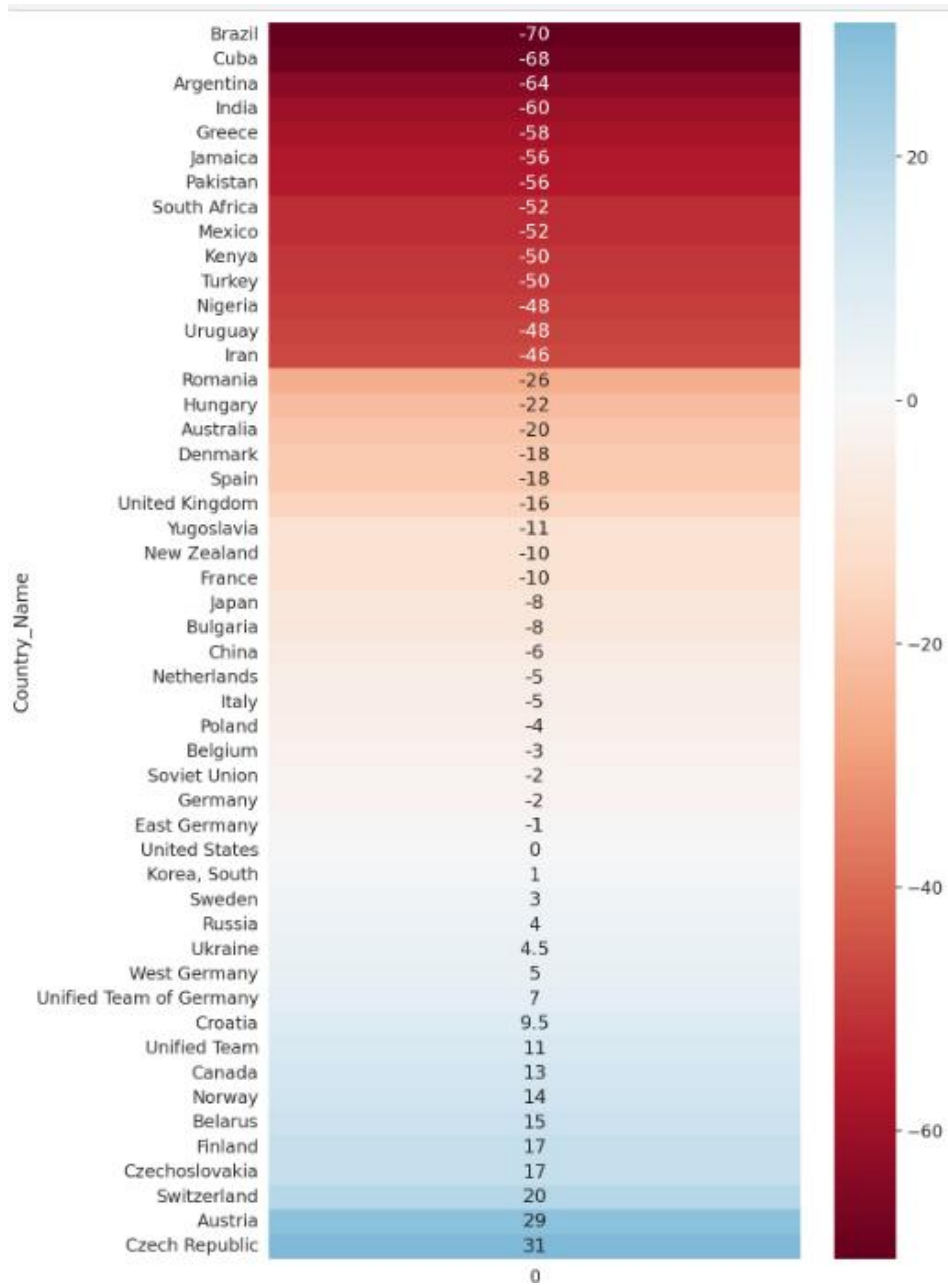
Popülasyon ve toplam madalya arasında önemli bir pozitif relation olduğunu gördüm: USA en iyi örneği diyebilirim.

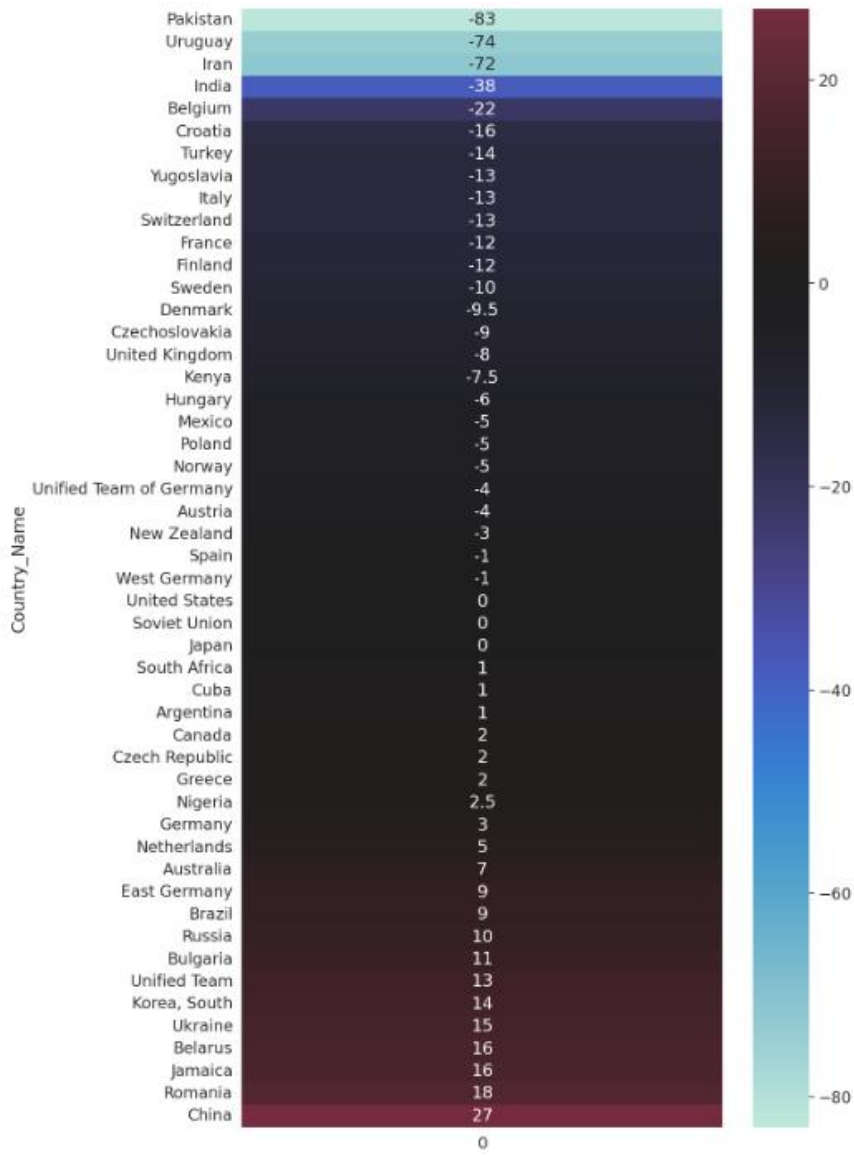
Aynı şekilde GDP Per Capita ve toplam madalya sayısı arasında da pozitif relation olduğunu görmekteyiz.

Cinsiyete göre de ayırdım:



Mevsimlere göre ayırdım. En sıcak ve soğuk ülkelere göre grafik çıkardım:





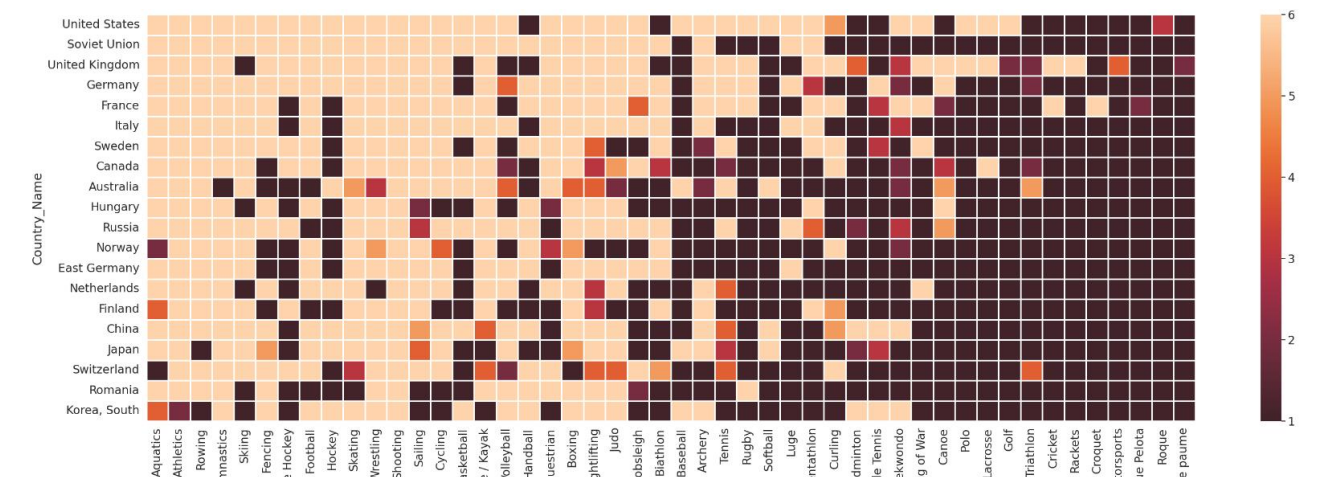
Ülkelerin gelişmişliği ve gelenekselliğini cinsiyetlere yansıdığını düşünüyorum. Yukarıdaki örnekte tepeden başlayan kısım daha çok erken sporcu bulunduran ülkelerdir. Grafiğin en altı daha çok kadın sporcu bulundurur. Ortalar ise birbirine yakın sayıyı sporcu bulundurur.

Türkiye’de ise erkek sporcuların daha çok olduğunu görüyoruz. Ama kadın sporcu sayımız gittikçe artmaktadır :)

Crosstab kullandım.

[illegible]

Heatmap ile geleneksel sporlarında ilişkili olduğunu inceledim.



Sonuç olarak, birçok etkenin olimpiyat sporlarını etkilediğini görmüş oldum. Bunların hipotezini kendimiz de yapabiliriz ama önemli olan datayı inceleyip analiz ederek gerçek sonuçlara ulaşmaktır. Böylece tahminlerimizin ne kadar doğru ne kadar yanlış olduğunu incelemiş olduk.