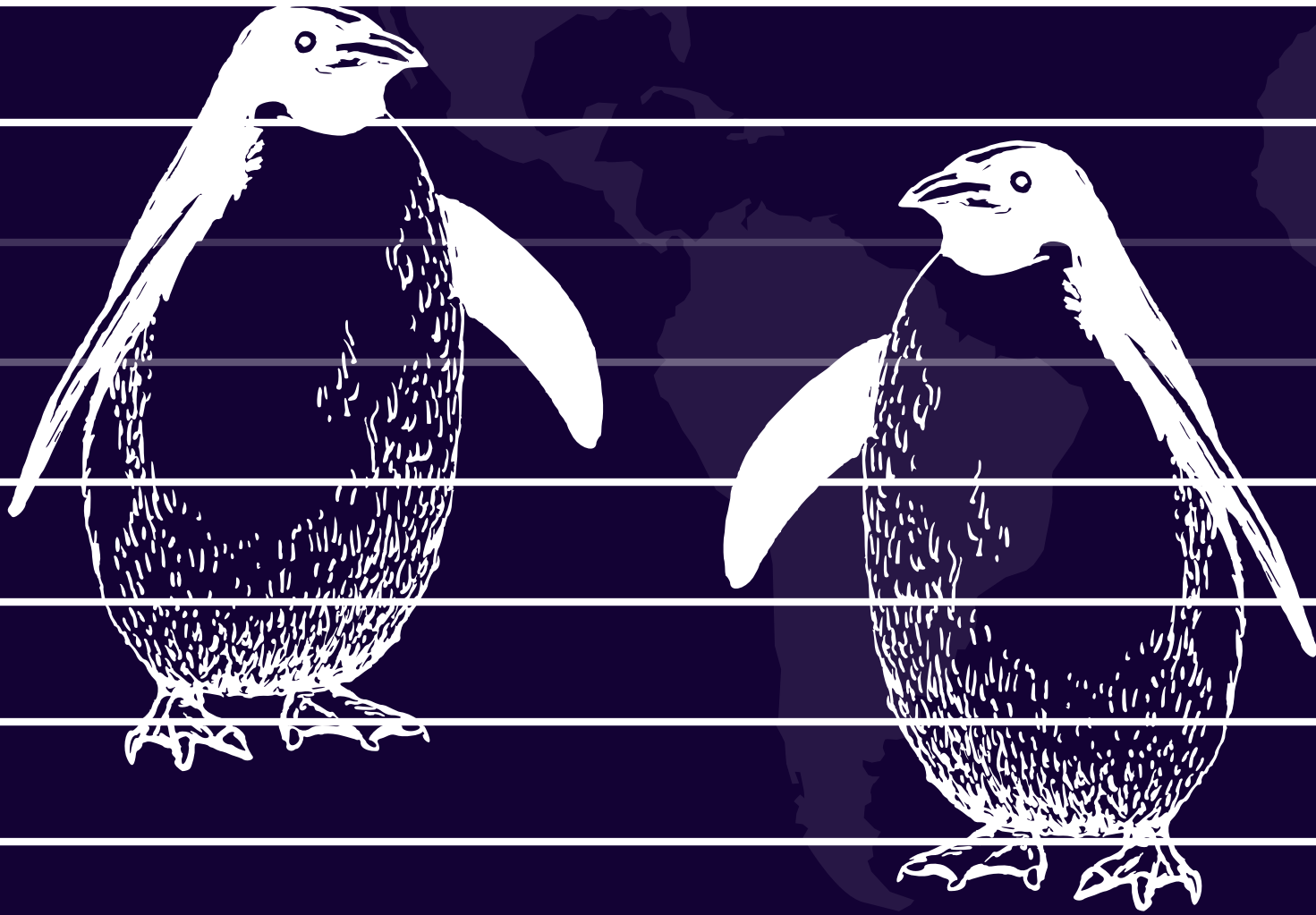


GR_DAILY_SEA_ICE3

HÜSNA NUR KAYA,19052080

YTÜ,VERİ MADENCİLİĞİNE GİRİŞ,MTM4641



Günlük Deniz Buzu Veri Seti Üzerinde Veri Madenciliği



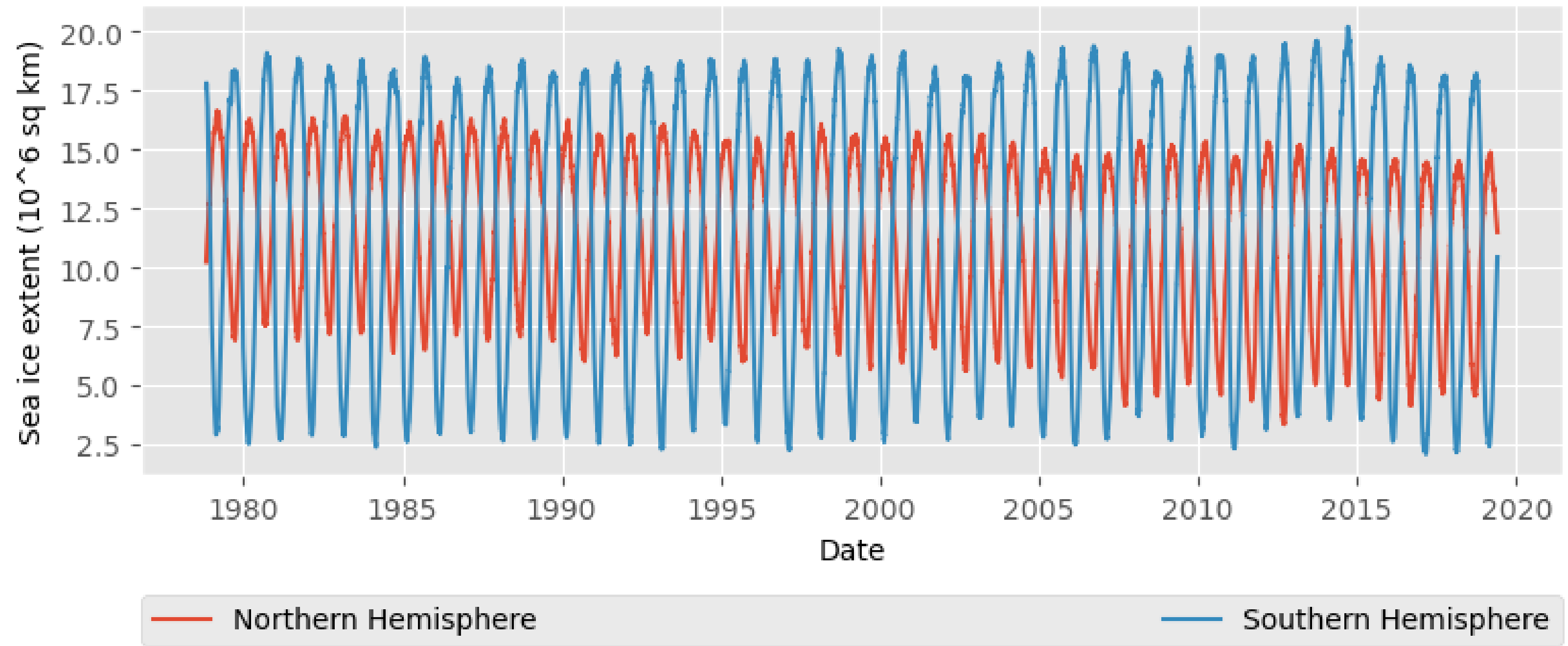
Veri Setini Tanima

	Year	Month	Day	Extent	Missing	Source Data	hemisphere
0	1978	10	28	10.231	0.0	[ftp://sidads.colorado.edu/pub/DATASETS/nsid...	north
1	1978	10	28	10.420	0.0	[ftp://sidads.colorado.edu/pub/DATASETS/nsid...	north
2	1978	10	30	10.557	0.0	[ftp://sidads.colorado.edu/pub/DATASETS/nsid...	north
3	1978	11	1	10.670	0.0	[ftp://sidads.colorado.edu/pub/DATASETS/nsid...	north
4	1978	11	3	10.777	0.0	[ftp://sidads.colorado.edu/pub/DATASETS/nsid...	north

	Year	Month	Day	Extent	Missing
count	28354.000000	28354.000000	28354.000000	28354.000000	28354.000000
mean	2000.591941	6.507399	15.740885	11.494986	0.000003
std	10.896821	3.451938	8.801607	4.611734	0.000227
min	1978.000000	1.000000	1.000000	2.080000	0.000000
25%	1992.000000	4.000000	8.000000	7.601000	0.000000
50%	2001.000000	7.000000	16.000000	12.217000	0.000000
75%	2010.000000	10.000000	23.000000	15.114000	0.000000
max	2019.000000	12.000000	31.000000	20.201000	0.024000

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 26354 entries, 0 to 26353
Data columns (total 7 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Year                  26354 non-null  int64
1   Month                 26354 non-null  int64
2   Day                   26354 non-null  int64
3   Extent                26354 non-null  float64
4   Missing               26354 non-null  float64
5   Source Data           26354 non-null  object
6   hemisphere            26354 non-null  object
dtypes: float64(2), int64(3), object(2)
memory usage: 1.4+ MB
```

Daily sea-ice extent



Kullanılan Veri Setleri

Climate Change Survey

- Total_Agriculture_forest
- Total_Fossil_Industry
- CO2_global_CEDS_emissions_by_fuel_2021_04_21

National Centers for Environmental Information, Global Time Series

- land temperatures -south
- land temperatures -north
- ocean temperatures -south
- ocean temperatures -north

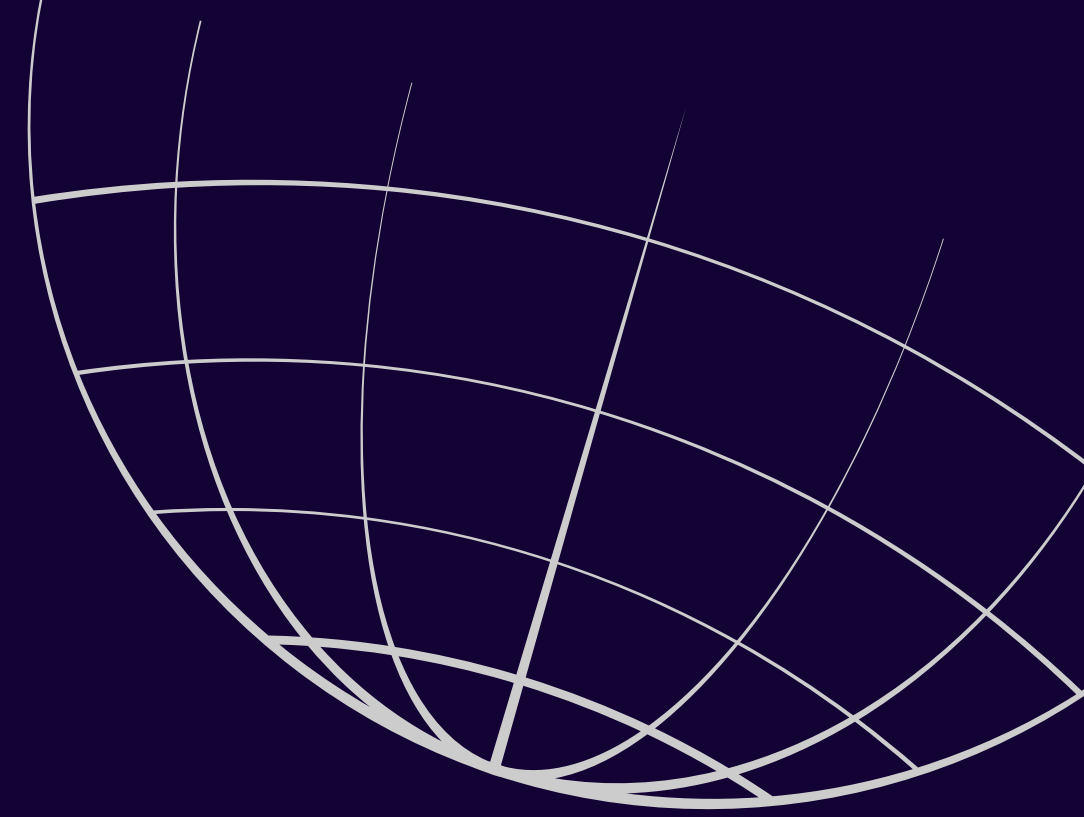
https://github.com/Omelettee/penguin_population_analysis/tree/master

- penguin population

VERİ ÖNİŞLEME

Neler Kullanıldı?

- Sütun isimlerini Değiştirilmesi
- Reset index
- Transpoz
- Biomass'ın silinmesi
- Concat ve Merge İşlemleri
- Yarım küre için değerlerin duplicate edilmesi vs.



Örnek ön işleme süreci-1

```
df_northern_land.head()
```

	Year	Anomaly
0	197801	0.10
1	197802	0.07
2	197803	0.37
3	197804	0.21
4	197805	-0.04

	Year	Anomaly
0	1978	0.010833
1	1979	0.042500
2	1980	0.259167
3	1981	0.714167
4	1982	-0.011667

	Year	AnomalyLand	hemisphere
0	1978	0.010833	north
1	1979	0.042500	north
2	1980	0.259167	north
3	1981	0.714167	north
4	1982	-0.011667	north

Öncelikle Year sütunu pandas Datetime modülü ile uygun hale getirildi. Daha sonra yıl ve küre bazlı veri elde etmek istediğimiz için Year sütununa göre(groupby) Anomaly değerlerinin ortalaması alındı. Son olarak hemisphere sütunu tüm değerleri north olacak şekilde eklendi. Aynı işlemler df_southern_land için de uygulanıp kara sıcaklarının olduğu bu dataFrameler axis=0 da concat edildi.

Örnek ön işleme süreci-2

```
df_co.head()
```

	fuel	em	units	X1750	X1751	X1752	X1753	X1754	X1755	X1756	...	X2010	X2011
0	biomass	CO2	ktCO2	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000e+00	0.000000e+00
1	brown_coal	CO2	ktCO2	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	7.473192e+05	7.867700e+05
2	coal_coke	CO2	ktCO2	366.177444	366.177444	366.320930	366.320930	366.464417	366.607903	391.861520	...	4.246643e+05	4.707760e+05
3	diesel_oil	CO2	ktCO2	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	3.876727e+06	3.971659e+06
4	hard_coal	CO2	ktCO2	6735.455519	6735.455519	6738.094804	6738.094804	6740.734089	6743.373374	7207.887548	...	1.057572e+07	1.105209e+07

5 rows × 273 columns

	Year	biomass	brown_coal	coal_coke	diesel_oil	hard_coal	\
0	1750	0.0	0.000000	366.177444	0.000000e+00	6.735456e+03	
1	1751	0.0	0.000000	366.177444	0.000000e+00	6.735456e+03	
2	1752	0.0	0.000000	366.320930	0.000000e+00	6.738095e+03	
3	1753	0.0	0.000000	366.320930	0.000000e+00	6.738095e+03	
4	1754	0.0	0.000000	366.464417	0.000000e+00	6.740734e+03	
...	
265	2015	0.0	720718.475341	531023.090905	4.231086e+06	1.123129e+07	
266	2016	0.0	715459.119445	546957.337511	4.197037e+06	1.092686e+07	
267	2017	0.0	718296.340601	543464.732025	4.279408e+06	1.103649e+07	
268	2018	0.0	741064.980372	548917.287608	4.339067e+06	1.122733e+07	
269	2019	0.0	698781.625261	553333.808495	4.372833e+06	1.118936e+07	

- melt() fonksiyonu
- transpoze
- hemisphere sütunu

Karar vermeliyim! Her küresel ısınma konulu veri setini dahil edebilir miyim?

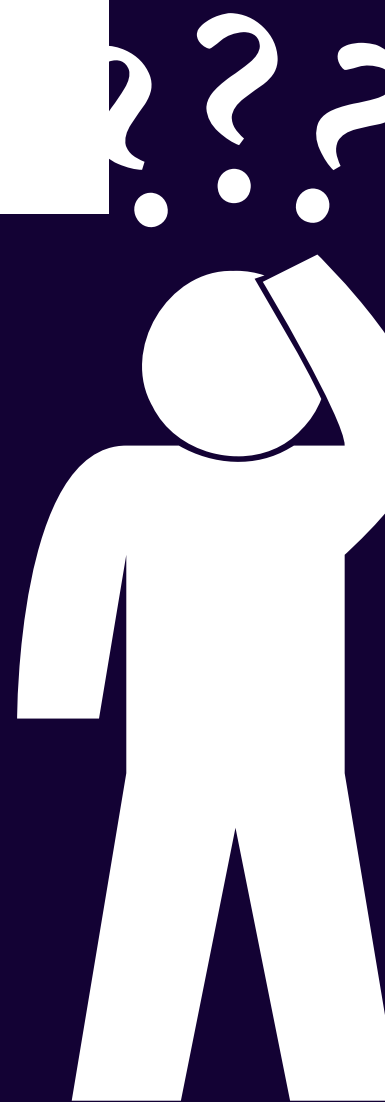
```
df_annual_afforest.head()
```

	Entity	Code	Year	Afforestation
0	Afghanistan	AFG	2000	70
1	Afghanistan	AFG	2010	70
2	Afghanistan	AFG	2015	70
3	Albania	ALB	1990	800
4	Albania	ALB	2000	800

```
df_annual_afforest["Year"].value_counts()
```

```
2010    98  
2000    97  
2015    98  
1990    81  
Name: Year, dtype: int64
```

Hayır. Mesela burada Entity sütunundaki ülke verisine göre hemisphere değerlerini eklemek için çabalıyordum. Sonra yıl bazlı veri sayısına baktım. Ve gördüm ki bizim ana veri setimiz olan sea_ice_extent in kapsam yılları ile çok fazla örtüşmüyor.



Veri setleri üzerindeki işlemlerimiz bitti, son hali ile merged_df e bakacak olursak;

```
merged_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
```

```
Int64Index: 84 entries, 0 to 83
```

```
Data columns (total 16 columns):
```

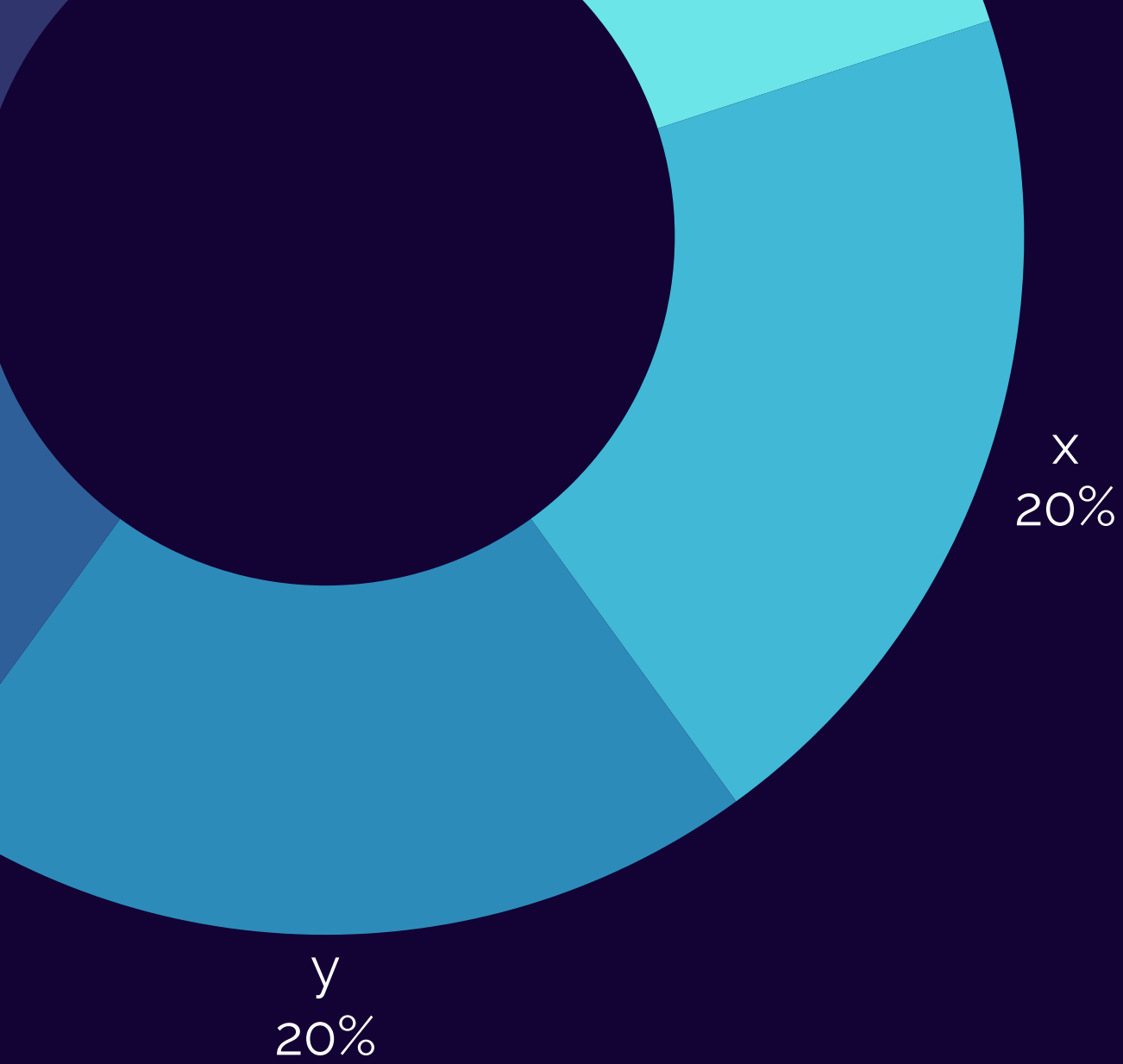
#	Column	Non-Null Count	Dtype
0	Year	84 non-null	int64
1	hemisphere	84 non-null	object
2	Extent	84 non-null	float64
3	AnomalyLand	84 non-null	float64
4	AnomalyOcean	84 non-null	float64
5	Fossil and Industrial	84 non-null	float64
6	Agriculture Waste and Forest Burining	84 non-null	float64
7	biomass	84 non-null	float64
8	brown_coal	84 non-null	float64
9	coal_coke	84 non-null	float64
10	diesel_oil	84 non-null	float64
11	hard_coal	84 non-null	float64
12	heavy_oil	84 non-null	float64
13	light_oil	84 non-null	float64
14	natural_gas	84 non-null	float64
15	process	84 non-null	float64

```
dtypes: float64(14), int64(1), object(1)
```

```
memory usage: 11.2+ KB
```

- 15 sütun
- 84 satıra indirgedik.
- hemisphere haricinde kalan değerlerimiz sürekli(continuous) değerlerdir. LabelEncoder kullanacağız.





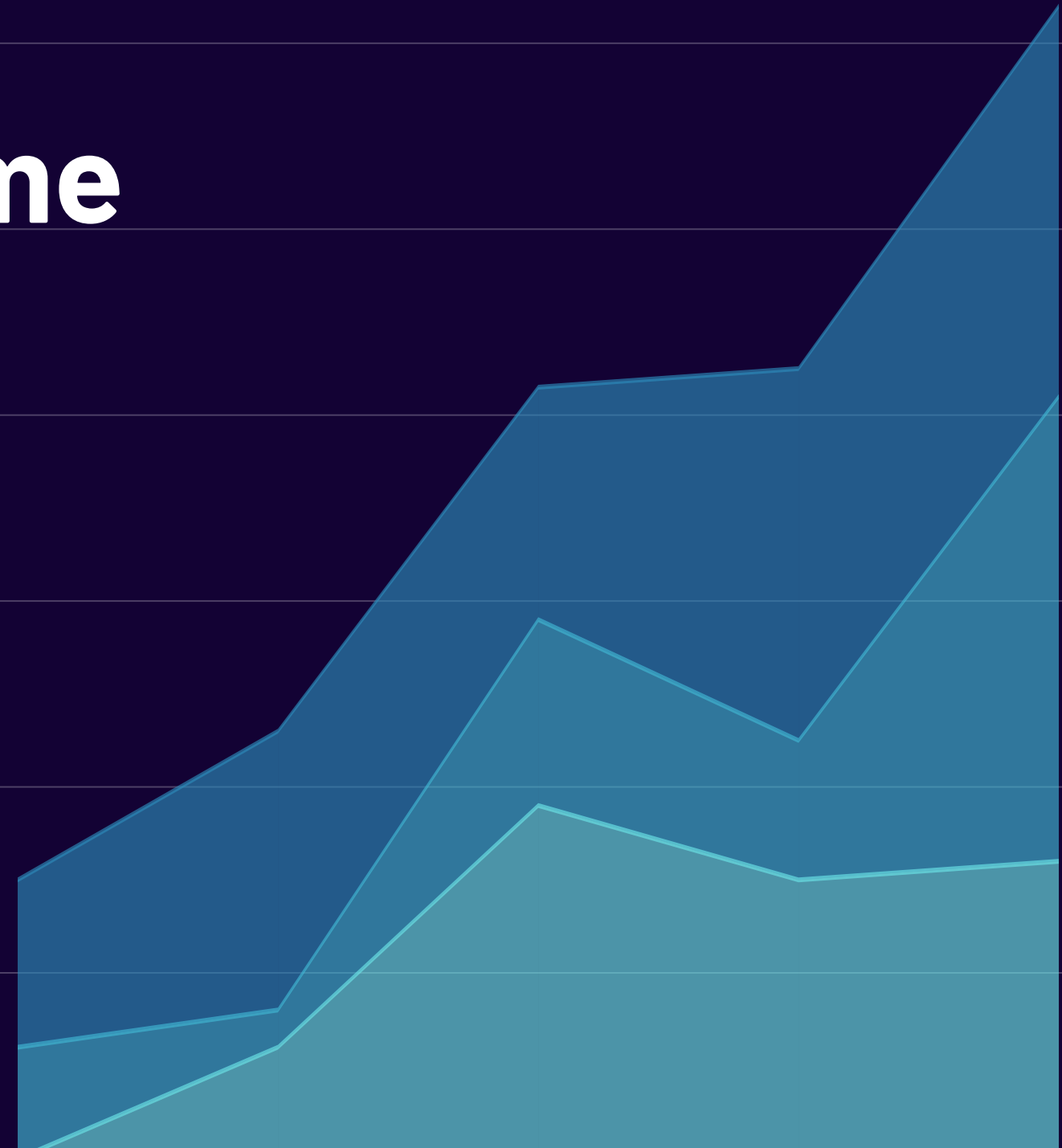
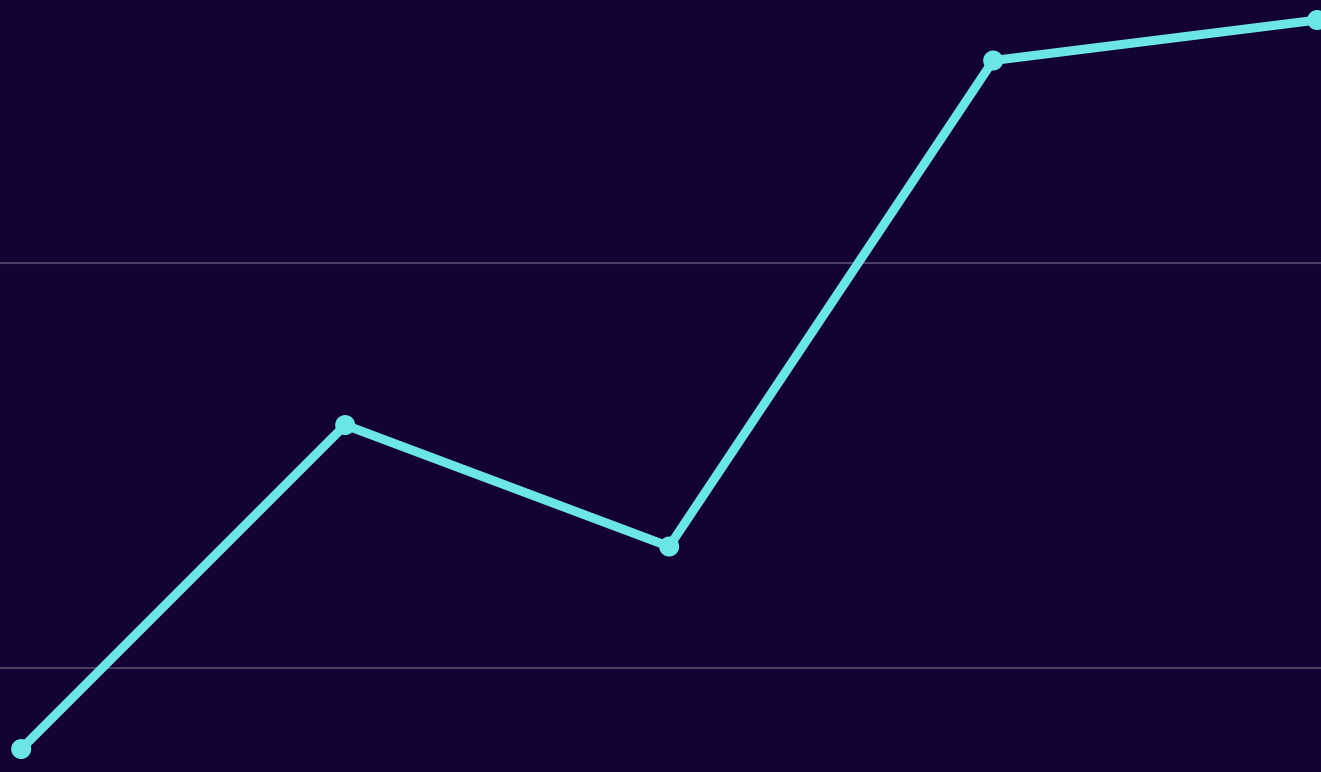
Veri Görselleştirme

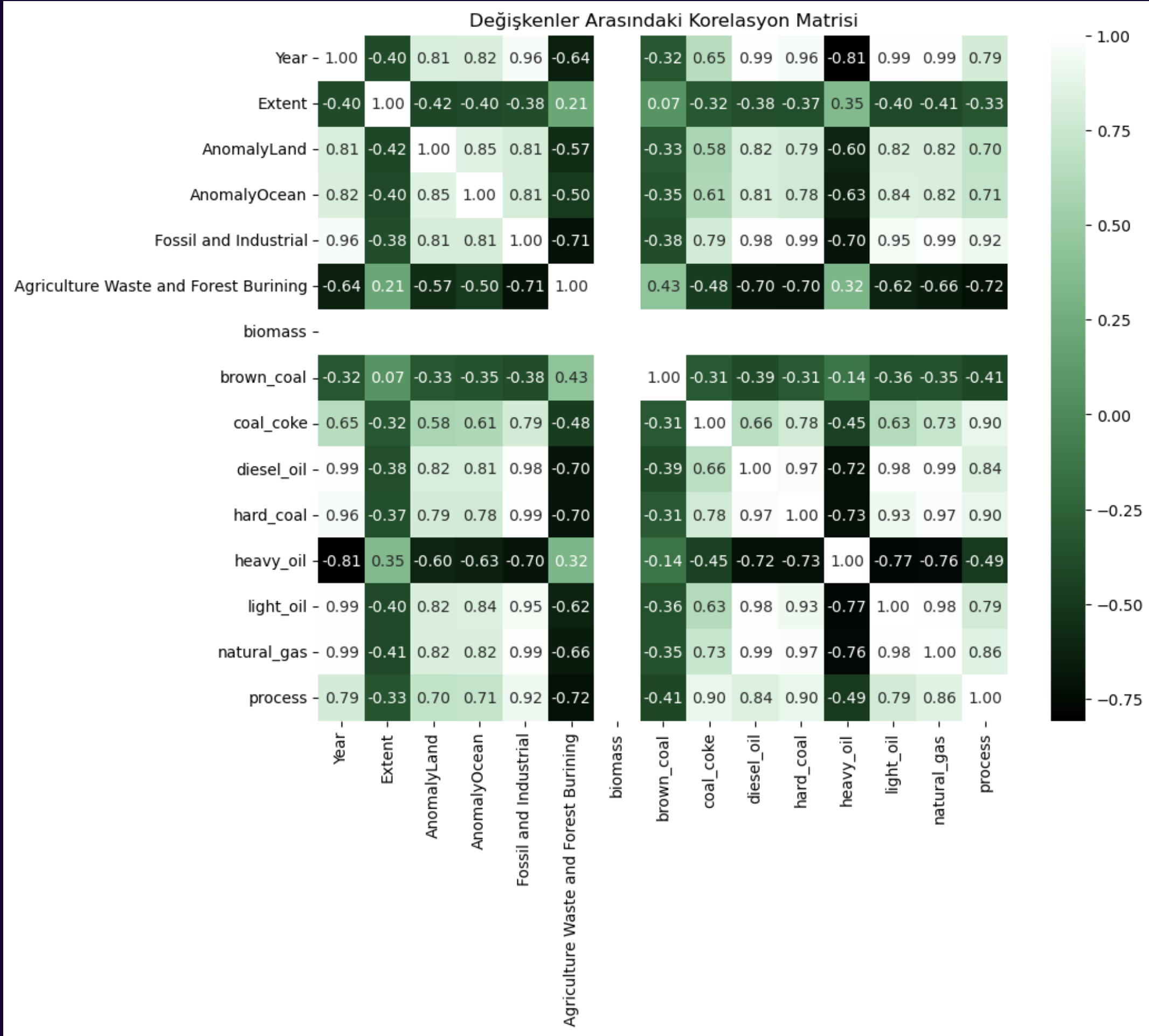
-matplotlib

-scatterplot

-seaborn

-pairplot



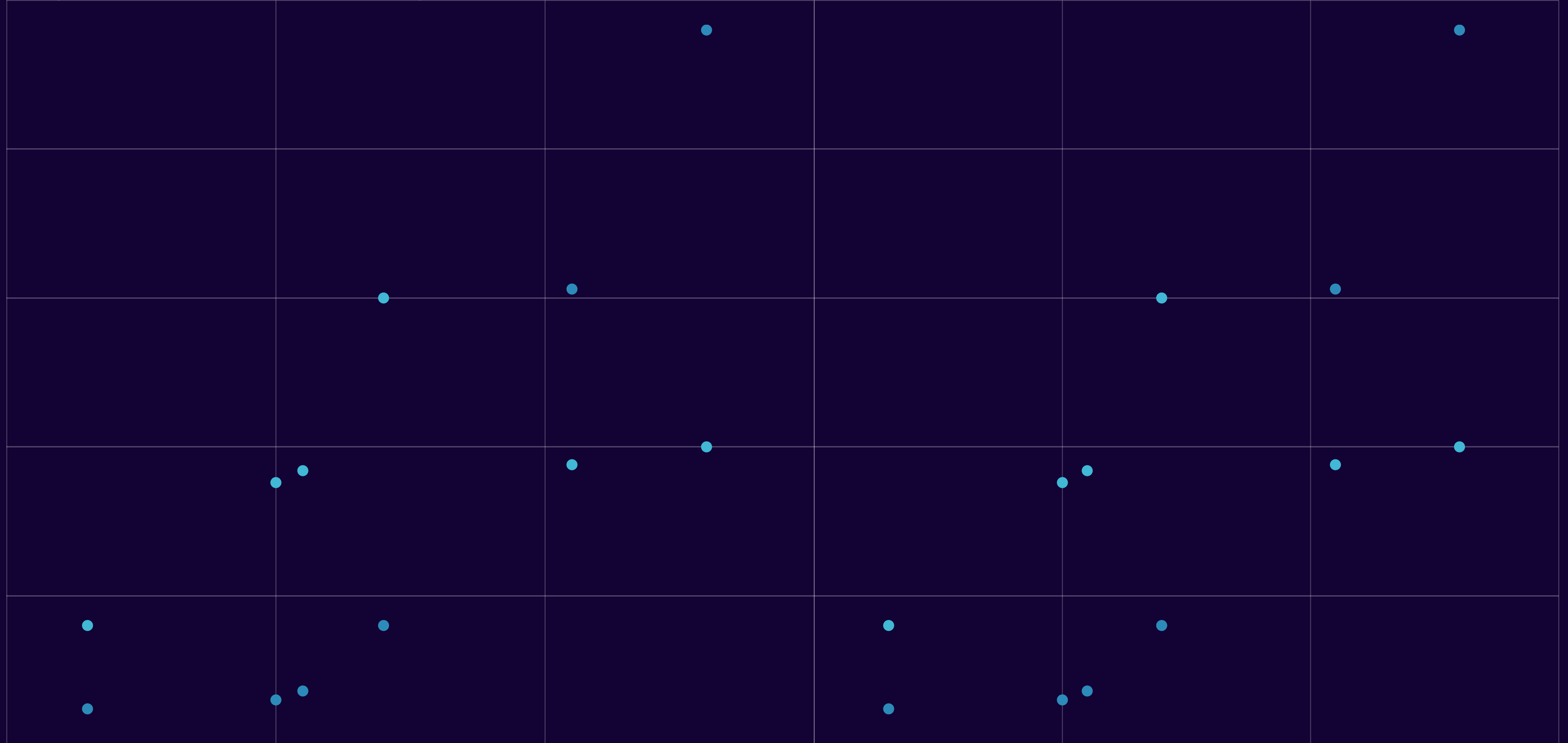


Değerler Arasındaki Korelasyon Matrisi

Güçlü ve zayıf korelasyonarı inceleyelim.(sunumda aktarılacak)

- Year<>Fossil and Industrial
- Year <>Diesel_oil
- Diesel_oil<>hard_coal
- Extent><AnomalyLand

Korelasyon matrisinde yakaladığımız korelasyonları pairplot üzerinde de görebilir miyiz?(görsel çok büyük ekstra sunulacak)



Neden Random Forest Regressor ?

- **Feature Importance:**

Random Forest, her bir öznitelik (feature) için önem değerini ölçerek, hangi özelliklerin tahminleme sürecinde daha etkili olduğunu belirlemenize yardımcı olacak.

- **Outlier'lara Direnç:**

Random Forest, outlier'ları (aykırı değerleri) etkili bir şekilde ele alabilir. Çünkü birden fazla ağacın bir araya getirilmesi, tek bir ağacın outlier'ları etkileyebileceği şekilde etkileme olasılığını düşürür. Mevsim anomalilerinin etkisinden kurtulmak için bu özelliğinden faydalanacağız.

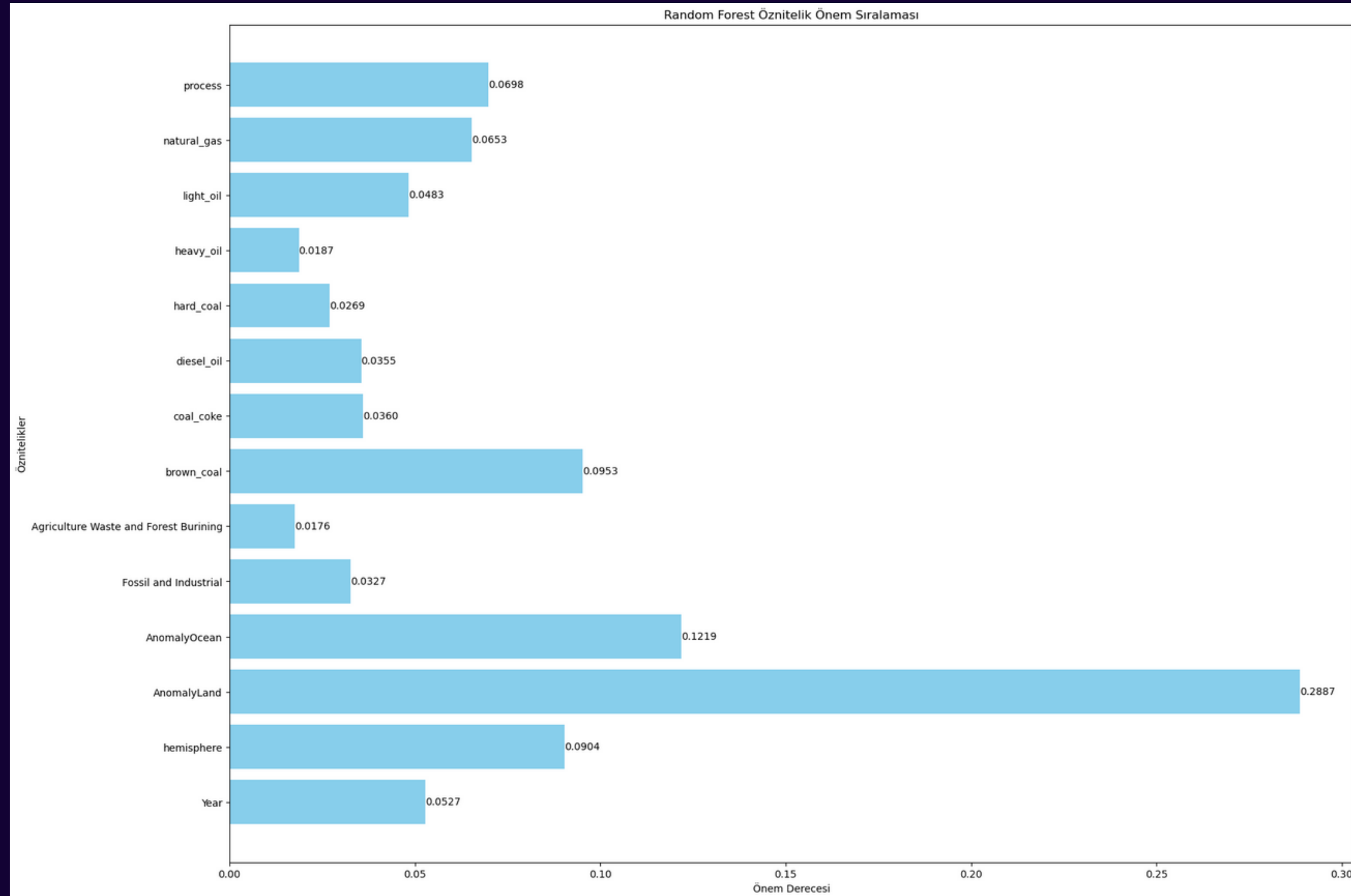
- **Parametre Ayarı Kolaylığı:**

Bazı karmaşık modellerde olduğu gibi çok fazla hiperparametre ayarı yapmaya gerek yoktur. Genellikle, ağaç sayısı ve öznitelik alt kümesinin büyüklüğü gibi birkaç temel parametre ile iyi sonuçlar alınabilir.

- **Veri Setinin Büyüklüğüne Uyum Sağlama:**

Hem küçük hem de büyük veri setleri üzerinde etkili bir şekilde çalışabilir. Veri setimiz başlangıçta büyük bir setti. Fakat gerekli düzenlemelerle 84 satıra düştü. Burada un

Öznitelik Önem Derecesi (Feature Importance)



Öznitelik Önem Derecesi (Feature Importance) /Yorum

1. **Anomaly Land**
2. **Anomaly Ocean**
3. **brown_coal**
4. **hemisphere**
5. **process**
6. **natural_gas**
7. **year**

Test Veri Seti Üzerinde Model Tahminlerinin Sonuç Metrikleri/ Yorum

	Metric	Value
0	MSE	0.035440
1	MAE	0.131430
2	RMSE	0.188256
3	R2 Score	0.895171

Neden bu metrikler seçildi?
Model performansı nasıl?



Yorum:

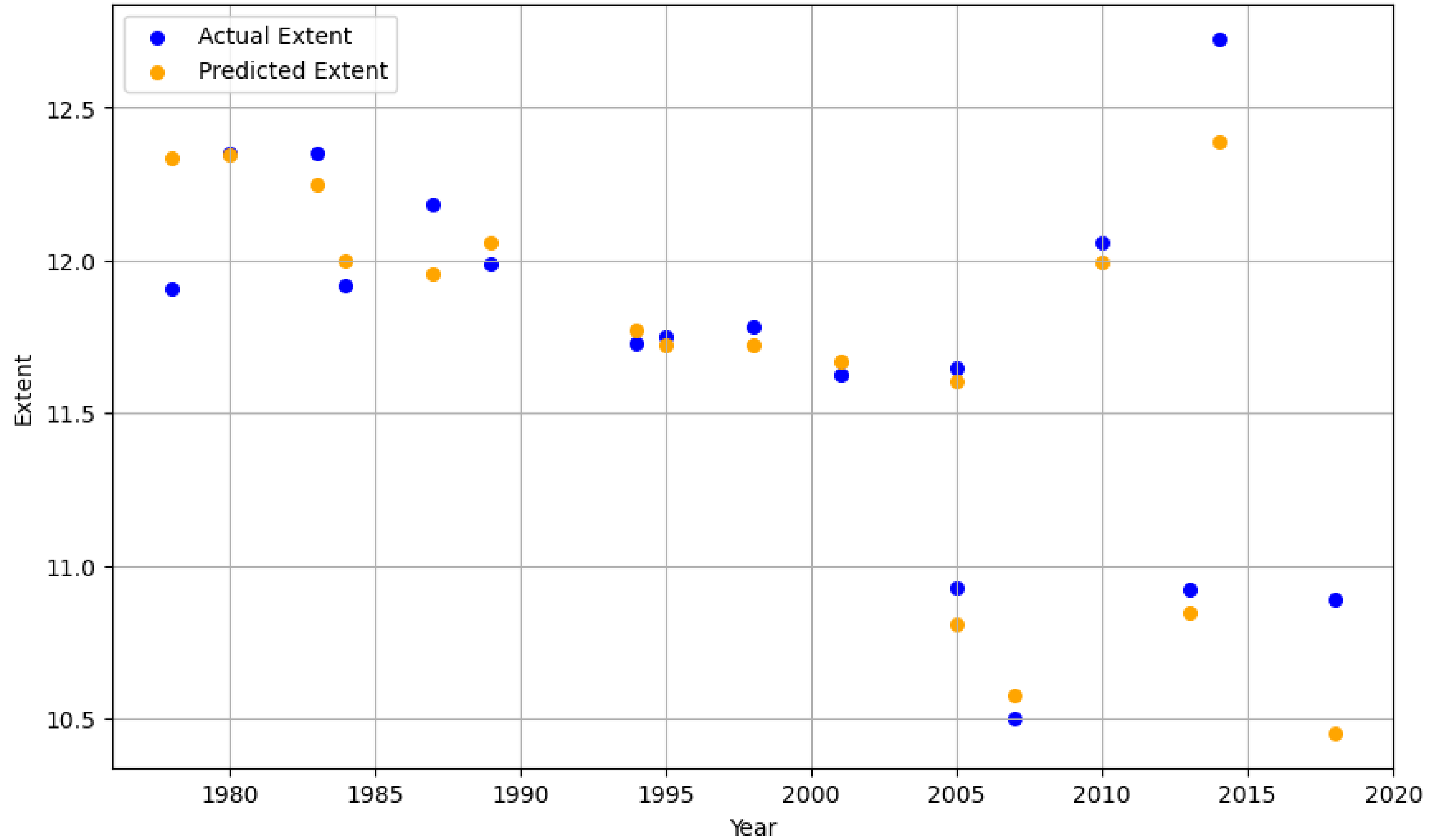
MSE (Mean Squared Error): Bu metrik, modelin tahminlerinin gerçek değerlerden ne kadar uzak olduğunu ölçer. MSE'nin düşük olması, modelin iyi bir tahmin yaptığını gösterir. Örneğinizde, MSE değeri 0.035440 olduğu için modelinizin ortalama karesel hatası oldukça düşüktür.

MAE (Mean Absolute Error): MAE, modelin tahminlerinin gerçek değerlere olan mutlak farklarının ortalamasını ifade eder. MAE'nin düşük olması, modelin genel olarak iyi tahminler yaptığını gösterir. Örneğinizde, MAE değeri 0.131430 olduğu için modelinizin ortalama mutlak hatası kabul edilebilir bir seviyededir.

RMSE (Root Mean Squared Error): RMSE, MSE'nin karekökünü alarak büyük hataları daha fazla vurgular ve birimleri orijinal veri birimleriyle aynı hale getirir. RMSE'nin düşük olması, modelin iyi tahminler yaptığını gösterir. Örneğinizde, RMSE değeri 0.188256 olduğu için modelinizin karesel hatası oldukça düşüktür.

R2 Score (Coefficient of Determination): R2, modelin bağımsız değişkenlerce açıklanan varyansın toplam varyansa oranını ölçer. R2'nin 1'e yaklaşması, modelin verilerdeki varyansın büyük bir kısmını açıkladığını gösterir. Örneğinizde, R2 Score değeri 0.895171 olduğu için modelinizin veri setinizdeki varyansın büyük bir kısmını başarılı bir şekilde açıkladığı söylenebilir.

Actual vs Predicted Extent Over the Years



Güney Yarımküre Deniz Buzu Kapsamı-Penguen Popölasyonları





ASYA

AVRUPA

KUZEY
AMERİKA

AFRİKA

GÜNEY
AMERİKA

GÜNEY
AMERİKA

OCEANIA

ANTARCTICA

Atlas
Okyanusu

Atlas
Okyanusu

Hint
Okyanusu

Büyük
Okyanus



Yeni bir araştırma, Antarktika'da iklim değişikliğinin tetiklediği deniz buzunu erimesinin imparator penguenlere ağır bir darbe vurduğunu ve 2100 yılına kadar tüm popülasyonları yok edebileceğini öne sürüyor.

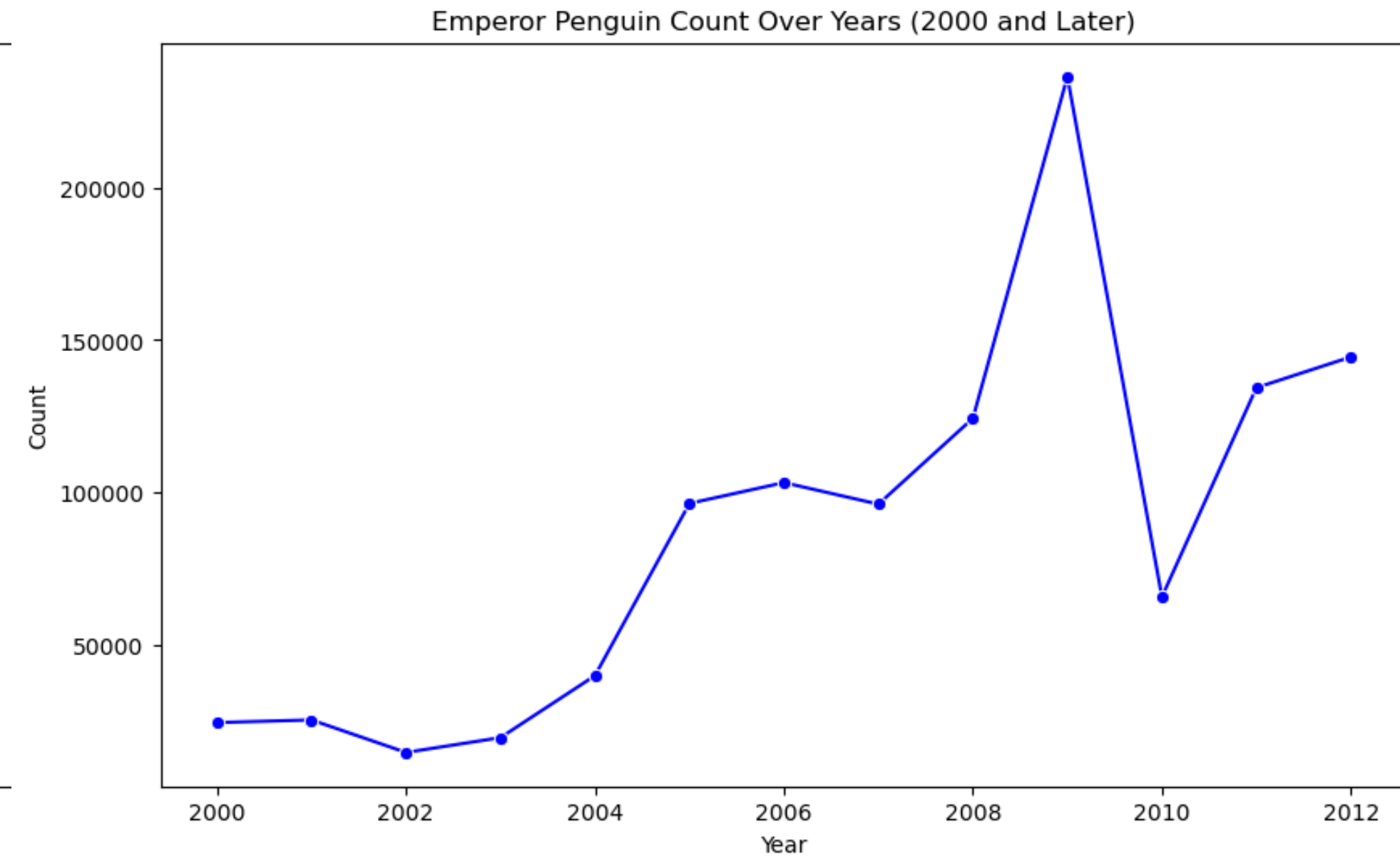
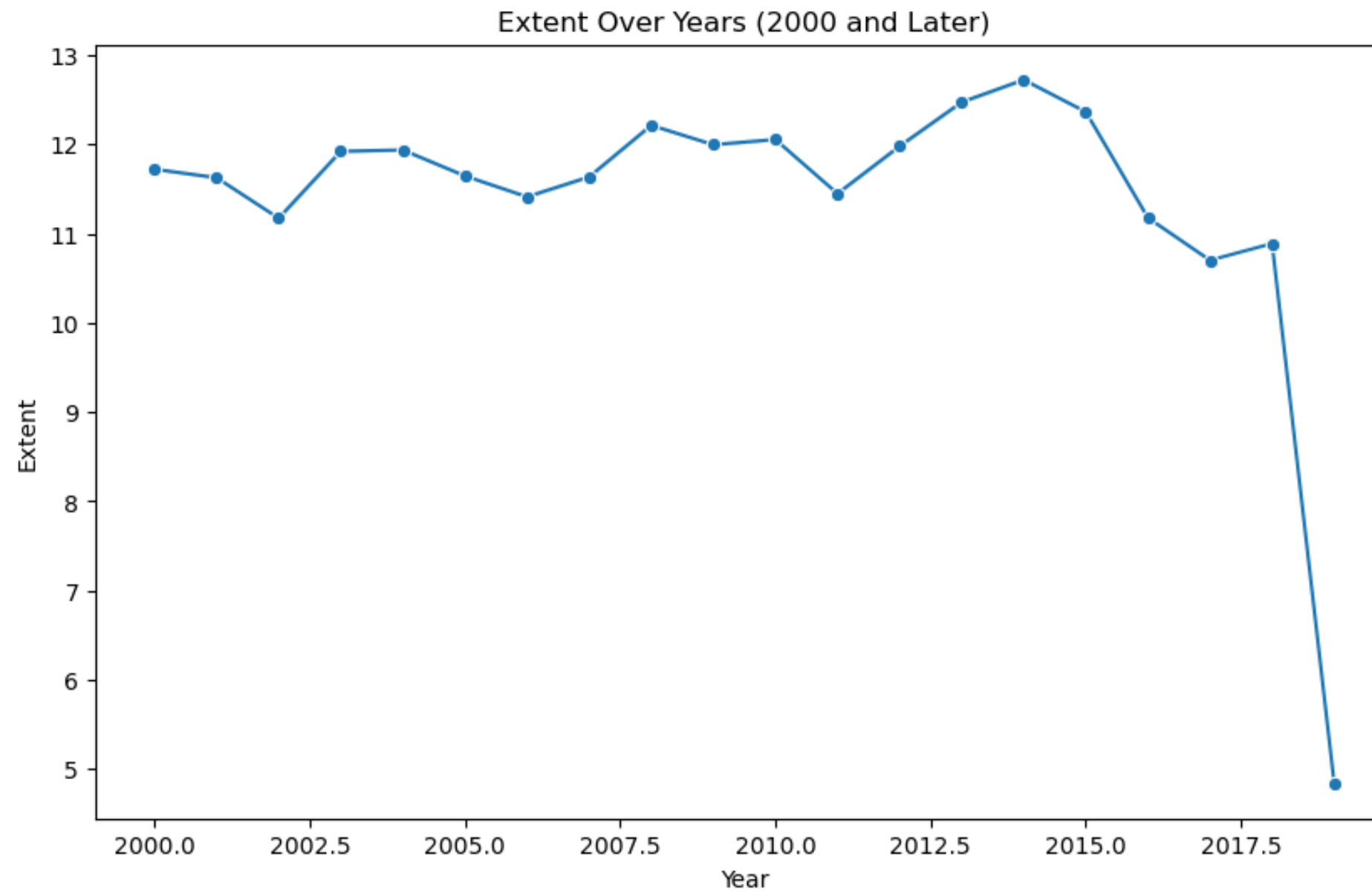
Antarktika'nın en hassas türlerinden biri olarak sınıflandırılan imparator penguenler üremek için Nisan ve Ocak ayları arasında karaya bağlı deniz buzuna ihtiyaç duyuyor. Bu da onları, 2022 bahar mevsiminde rekor seviyelere ulaşan bölgedeki hızlı deniz buzunu kaybına karşı son derece savunmasız hale getiriyor.

Communications Earth & Environment dergisinde yayınlanan çalışmaya göre, 2022'deki rekor düşük deniz buzunu, imparator penguenlerin ilk kez kaydedilen büyük ölçekli üreme başarısızlığı ile aynı zamana denk geldi. Araştırmacılar, Antarktika'daki beş üreme alanından "biri hariç hepsinin deniz buzunun kırılmasından sonra tamamen üreme başarısızlığı yaşadığını" tespit etti. Araştırmacılar, mevcut ısınma oranlarının devam etmesi halinde 2100 yılına kadar tür kolonilerinin %90'ından fazlasının "yarı yok" olacağını ileri sürüyor.

Araştırmanın başyazarı Peter Fretwell bulguları yorumlarken, araştırmacıların "imparator penguenlerin tek bir sezon boyunca bu ölçekte üreyemediğini hiç görmediklerini" söyledi.



2000 Sonrası Deniz Buzu ve Penguen Popülasyonu Veri Seti Görselleştirmeleri



Aşağıda penguen popülasyonu ile ilgili veri setini görüyoruz. Yıl bazlı hale getirilen sea_ice_extent veri setimiz ile bunu birleştirdim. Burada ön işleme olarak;

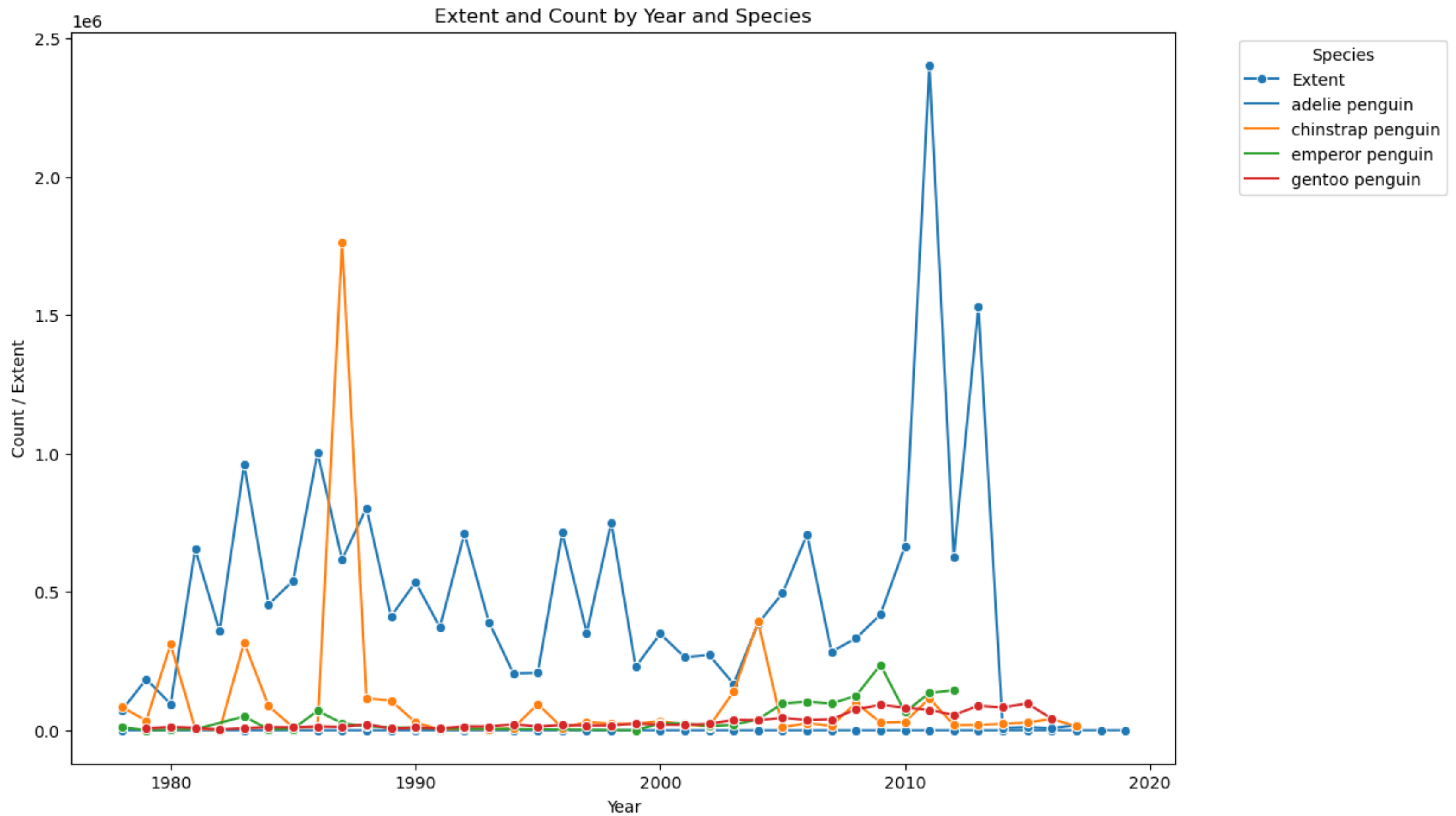
- sea_ice_extent in sadece hemishepere=south satırlarının alınması
- penguen popülasyonunu 1895-2017 aralığında verilere sahipti. Fakat aşağıda görüldüğü üzere gün, ay, yıl sütunları var. Burada yıl ve common.name'e göre bir grupta yapıp toplam count değerini yazdırdık.
- daha sonra bu iki seti merge ettik.

site.name	site.id	ccamlr.region	Longitude.EPSG.4326	Latitude.EPSG.4326	common.name	day	month	year	season.starting	count	accuracy	count.type
Adams Island	ADAM	58.4.1	92.5488	-88.5459	adelie penguin	12	11	2009	2009	76	5	nests
Adams Island	ADAM	58.4.1	92.5488	-88.5459	adelie penguin	21	1	2011	2010	425	1	adults
Cape Adare	ADAR	88.1	170.1999	-71.3083	adelie penguin	0	0	1982	1982	220900	2	chicks
Cape Adare	ADAR	88.1	170.1999	-71.3083	adelie penguin	0	12	1983	1983	258806	2	nests
Cape Adare	ADAR	88.1	170.1999	-71.3083	adelie penguin	0	12	1986	1986	282307	2	nests

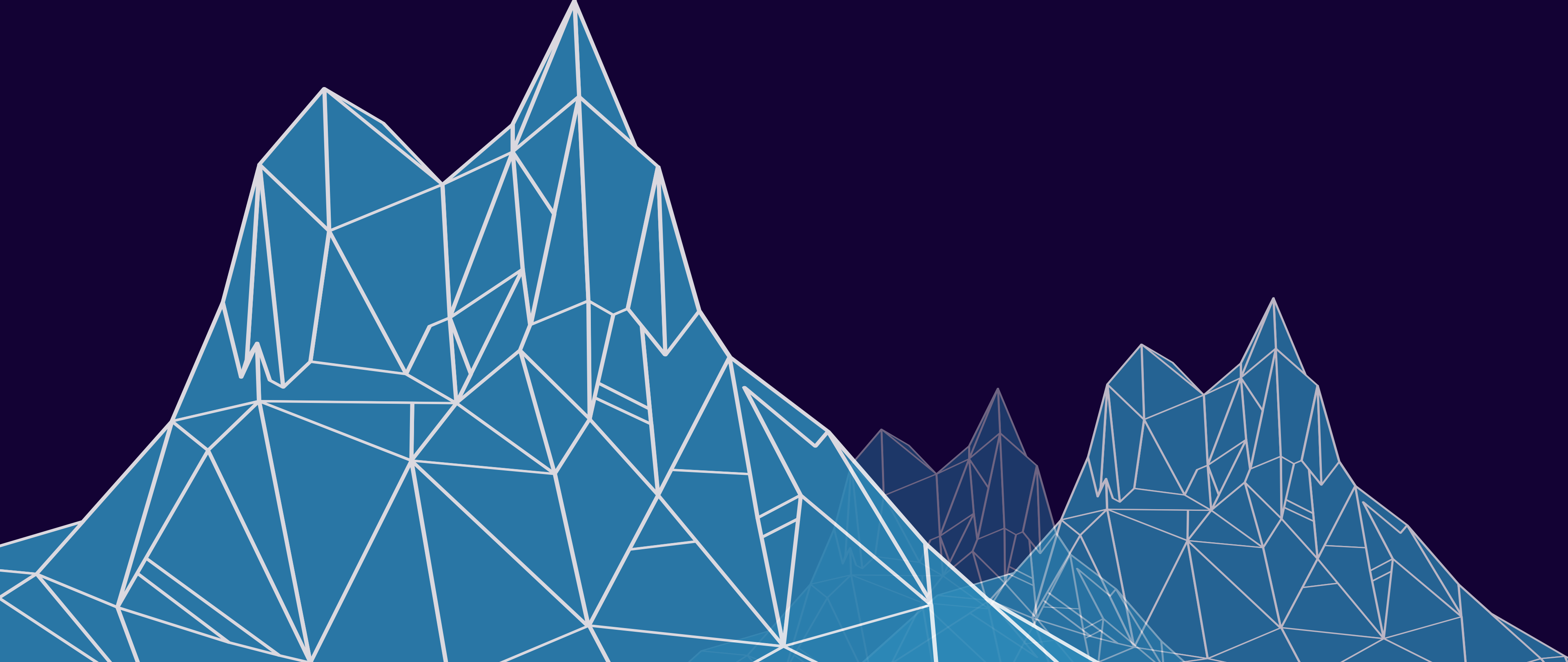
Elde edilen set ve describe() ile detayları:

	Year	hemisphere	Extent	common.name	count
0	1978	south	14.663550	adelie penguin	70828.0
1	1978	south	14.663550	chinstrap penguin	84425.0
2	1978	south	14.663550	emperor penguin	11165.0
3	1979	south	11.653819	adelie penguin	184855.0
4	1979	south	11.653819	chinstrap penguin	34257.0

	Year	Extent	count
count	145.000000	145.000000	1.430000e+02
mean	1997.517241	11.641430	1.833857e+05
std	11.629168	0.805225	3.389143e+05
min	1978.000000	4.839170	3.100000e+01
25%	1987.000000	11.379159	1.370350e+04
50%	1998.000000	11.628244	3.729900e+04
75%	2007.000000	11.748206	2.064355e+05
max	2019.000000	14.663550	2.403432e+06

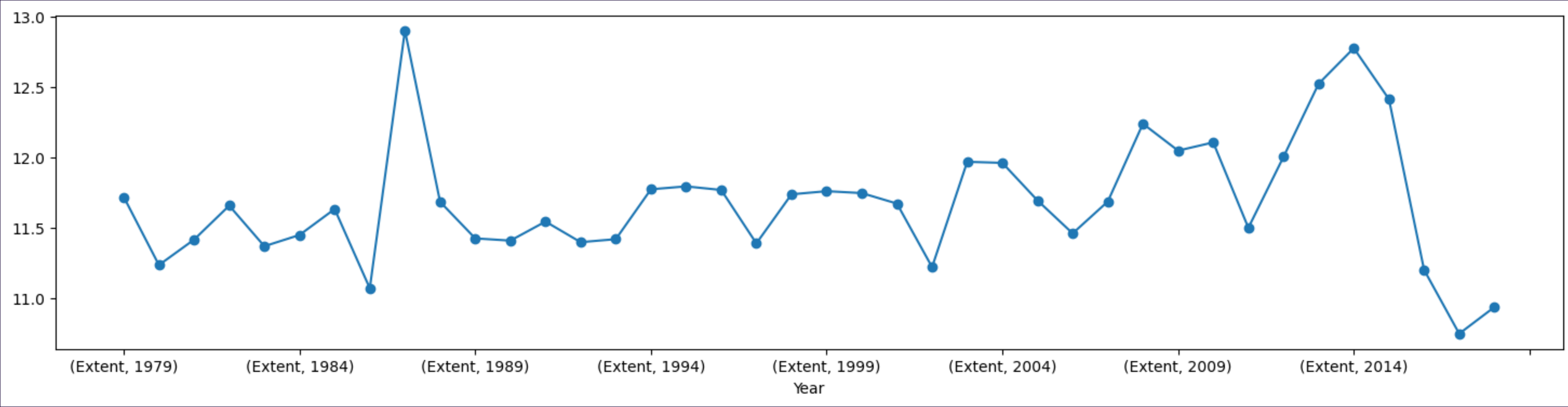


Deniz Buzu Kapsamı ve Penguen Popölasyonu Arasındaki İlişkinin İncelenmesi ve Forecast

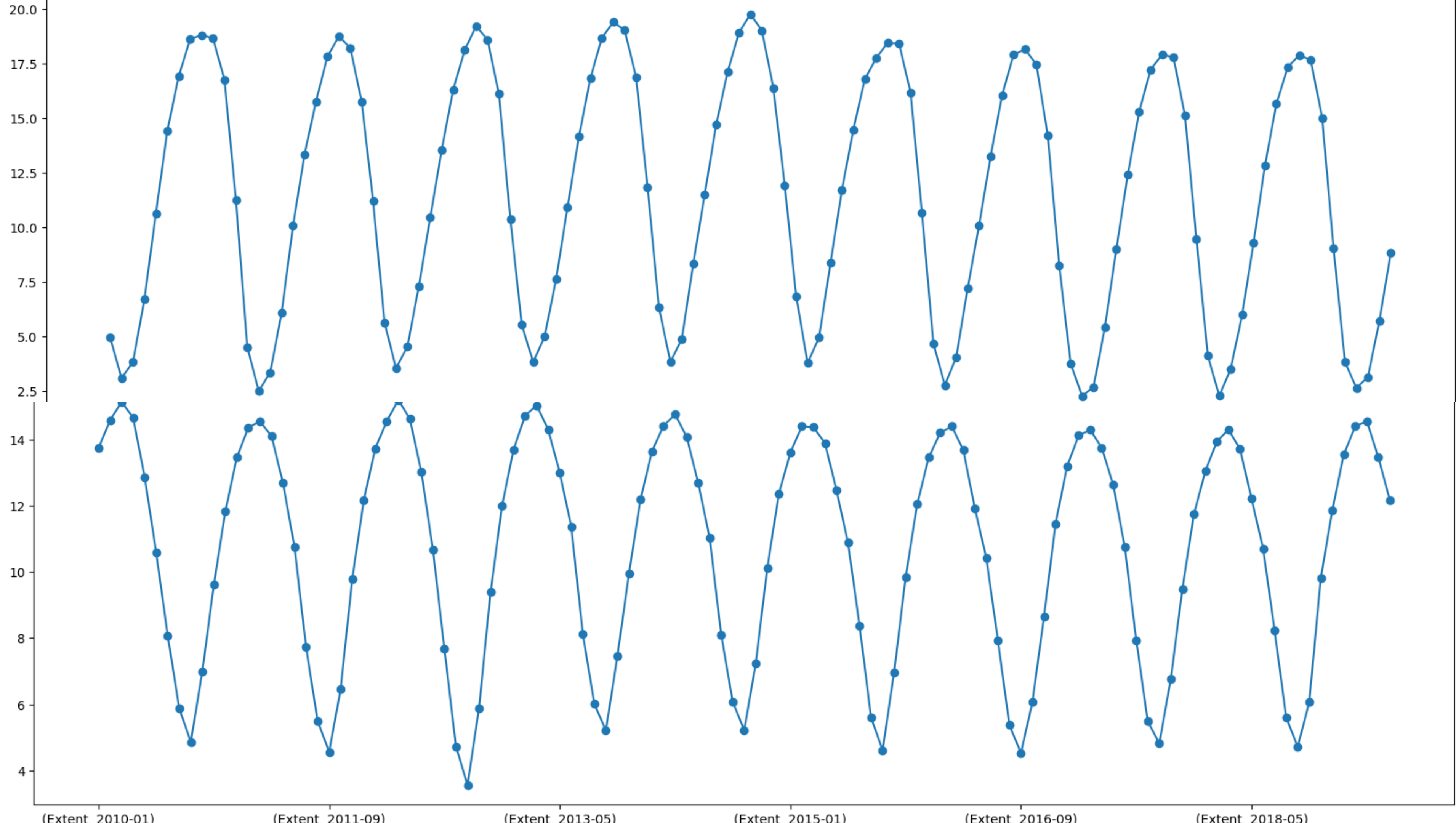


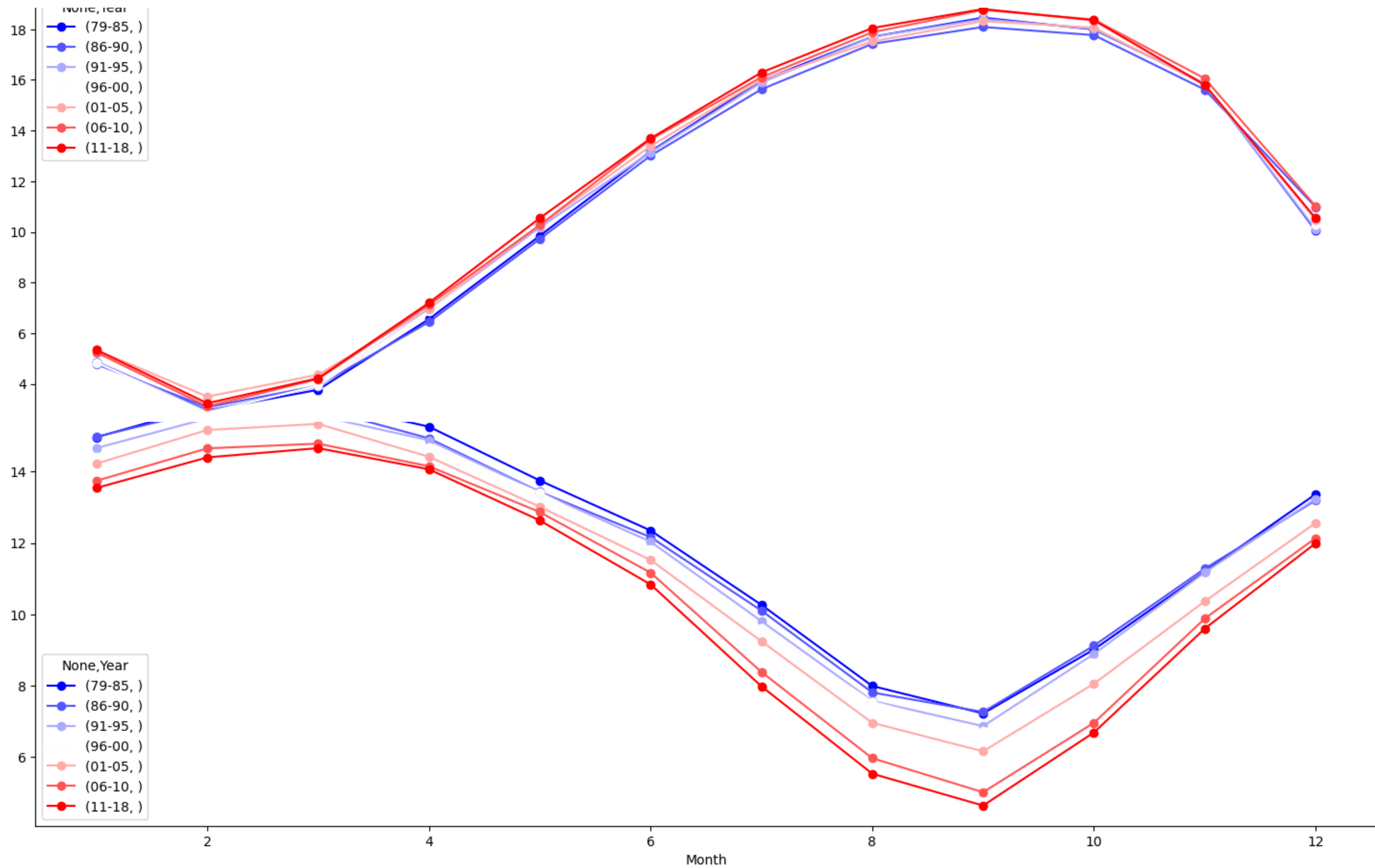
GYK

1978-2019 arası



KYK





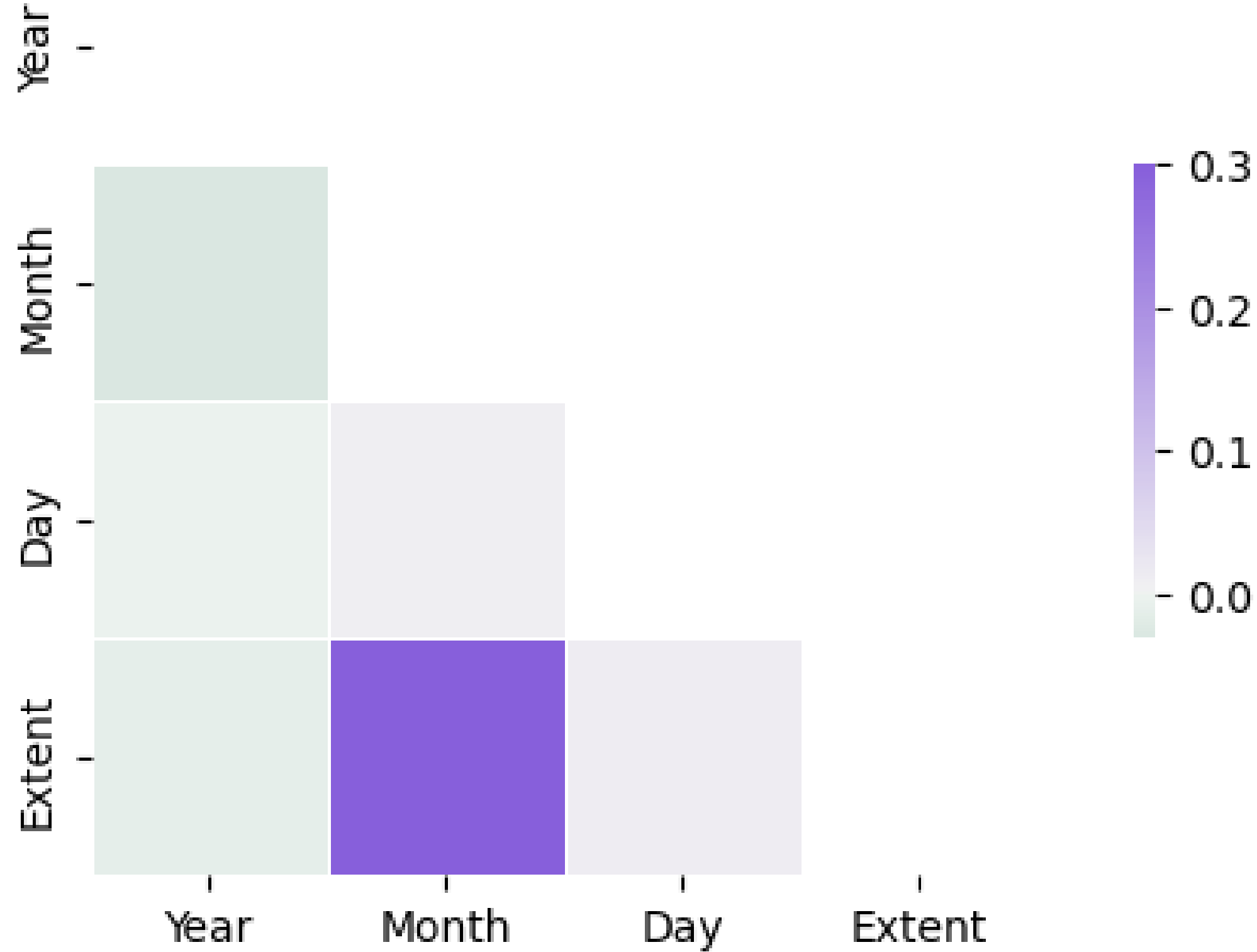
Yorum:

Aylık rakamları özetleyen bir grafik oluşturduk. Bu grafiğe göre kuzey için en büyük deniz buzu miktarı Mart ayında, en küçük ise Eylül ayında görülürken güney için tam tersi görülüyor. Bu da aslında yarım kürelerin mevsimselliğine uygun bir çıktıdır.

Ayrıca renkler maviden kırmızıya dönüştükçe daha yakın tarihli veri olacak şekilde ayarlandı.

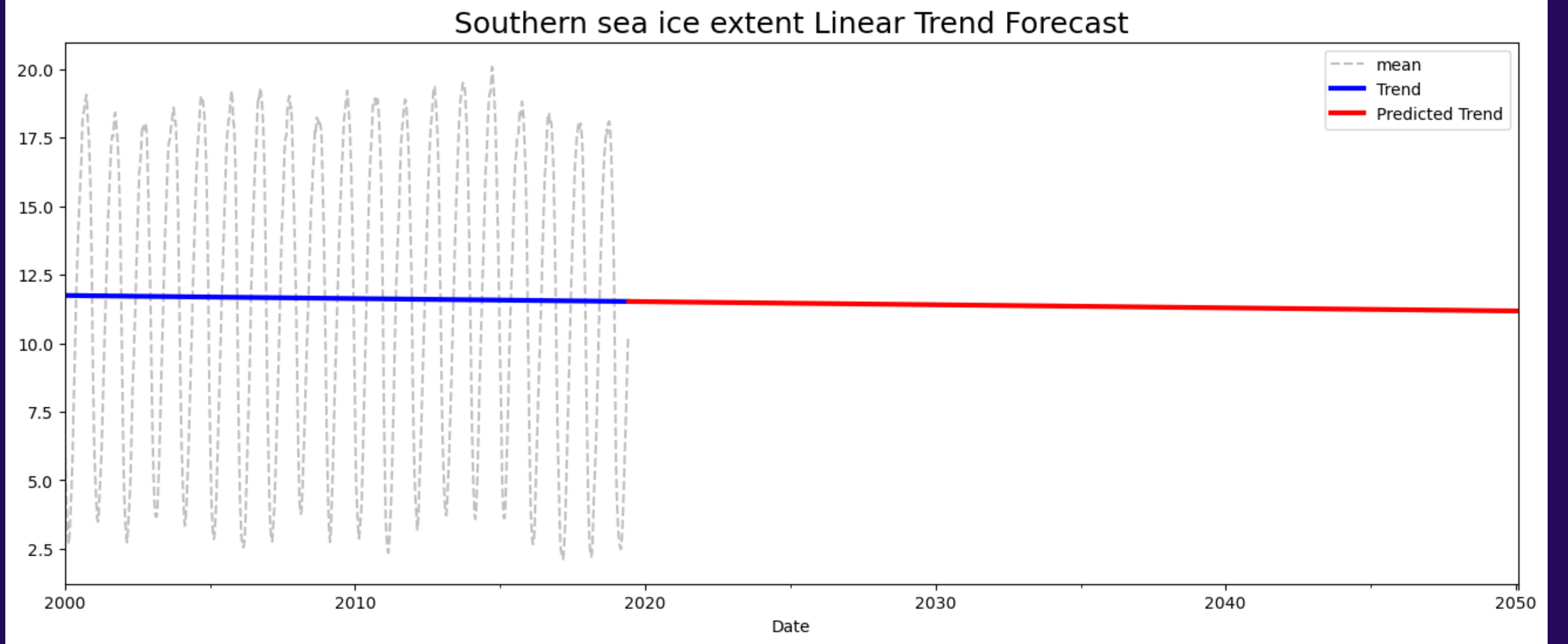
En koyu mavi 1979'dan 1985'e kadardır. En koyu kırmızı 2011'den 2018'e kadardır. Maviden kırmızıya geçiş, deniz buzu miktarının önceki yıllardan günümüze kadar azaldığını göstermektedir.

Kuzeyde deniz buzu kapsamında 2011-2018 ortalamaları; Mart için 14,6 milyon kilometrekare, Eylül için 4,6 milyon kilometrekaredir. Aradaki fark 10 milyon kilometrekare, yani Kanada'nın yüzölçümü kadar.



Mevsimsel
değişikliklerden için
genellikle ay bazlı
analizler yapılmaktadır.
'**seasonality**' denileni
yakalamak için. Burada
ise deniz buzu kapsamı
için de en anlamlı
değerin güçlü bir
korelasyon ile 'Month'
olduğunu görebiliriz.

Lineer Regresyon ile Güneydeki Deniz Buzu Kapsamı İçin Forecast-Sonucun Penguen Popölasyonu ile İlişkilendirilmesi

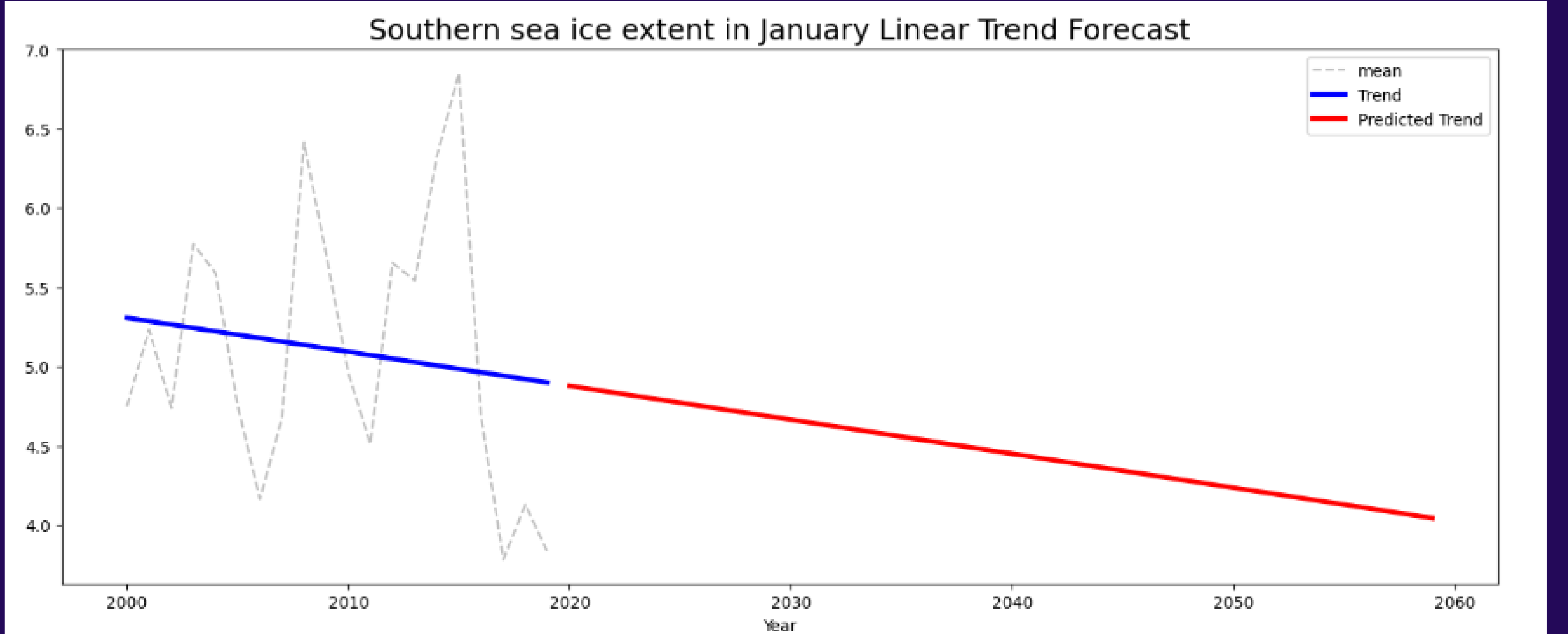


Yorum

Aslında güney için durum kuzeydeki kadar vahim durmamaktadır. Bunun sebebi için aşağıdaki alıntıya bakalım:

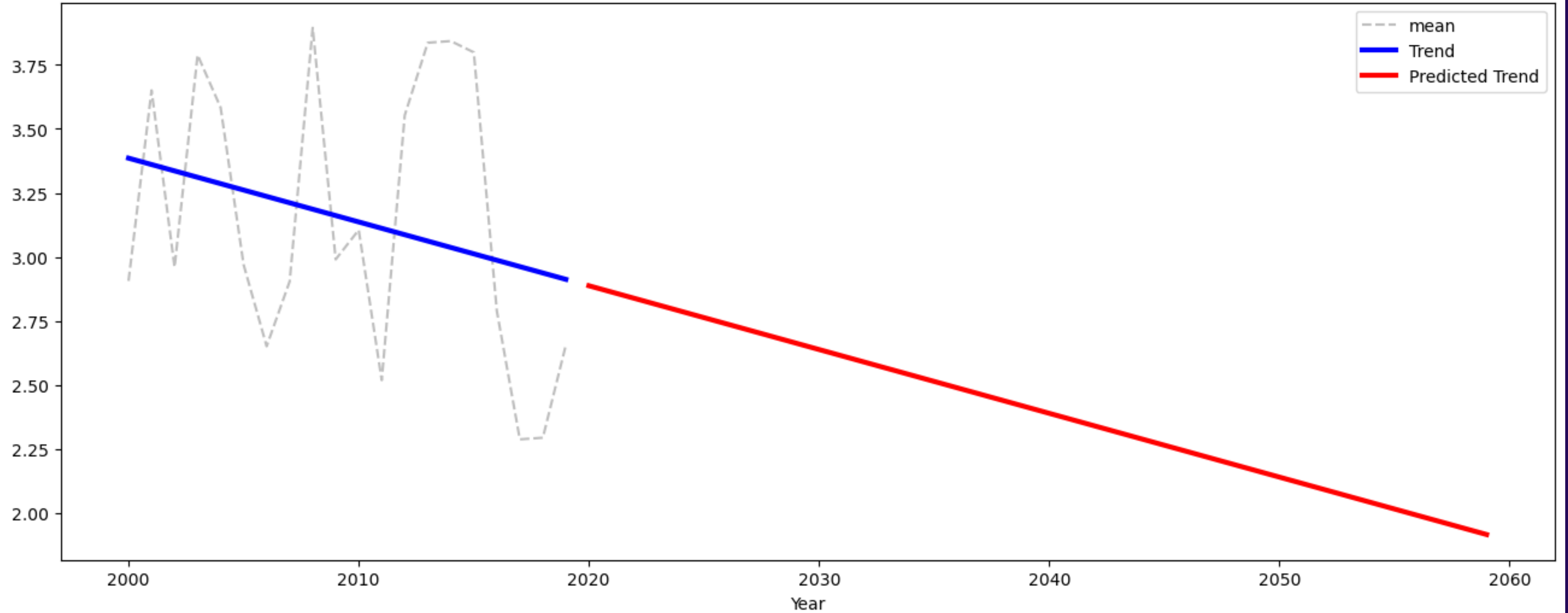
“NSIDC'nin NASA tarafından finanse edilen Arktik Deniz Buzu Haberleri ve Analizi (ASINA) projesi, Arktik deniz buzunun nasıl değiştiğine ve buz davranışında hangi koşulların rol oynayabileceğine dair neredeyse gerçek zamanlı veriler ve aylık bilgiler, analizler sağlıyor. NSIDC'ye göre iklim değişikliği Kuzey Kutbu'nu dünyadaki diğer yerlerden daha fazla etkiliyor. Bu bölge dünyanın geri kalanından iki kat daha hızlı ısınıyor ve bunun ciddi sonuçlarından biri de önemli miktarda deniz buzunun kaybı. Deniz buzu kaybı hem Arktik ekosistemleri hem de bir bütün olarak Dünya'yı etkiliyor. Arktik deniz buzu açık renkli olduğundan, deniz buzu yüzeyine çarpan güneş ışığının çoğunu uzaya geri yansıtır. Bu, okyanusa çok fazla ısının emilmesini önler ve bölgenin serin kalmasına yardımcı olur. Ancak daha fazla deniz buzu kaybedildikçe daha fazla ısı emilir ve bu da daha fazla erimeye neden olur. Bu ısınmayı artırır ve döngü devam eder. Sonuç olarak deniz buzu Arktik ortamın en hızlı değişen alanlarından biridir. Ki deniz buzu, Dünya ikliminin düzenlenmesinde kritik bir rol oynuyor ve küresel hava durumunu ve okyanus sirkülasyonlarını etkiliyor.

<https://nsidc.org/our-research/featured-projects/arctic-sea-ice-news-analysis>

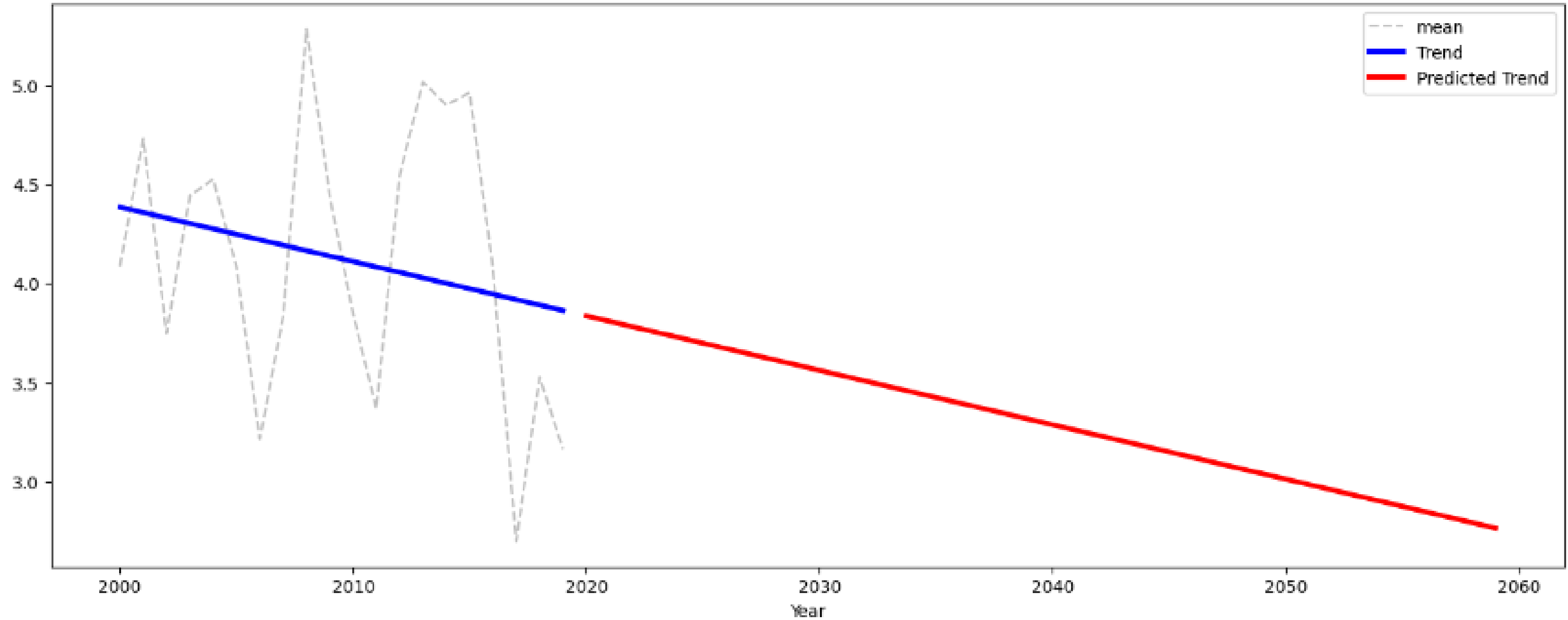


Güney için Şubat ayına odaklanılmak önemlidir. Çünkü grafikten de görüldüğü üzere 2060 ve özellikle sonrasında dönemsel olarak deniz buzu kapsamının sıfıra ulaşacaktır. Nisan ve Ocak ayları arasında üreme döneminde olan penguenler için bu büyük bir dezavantaj oluşturmaktadır.

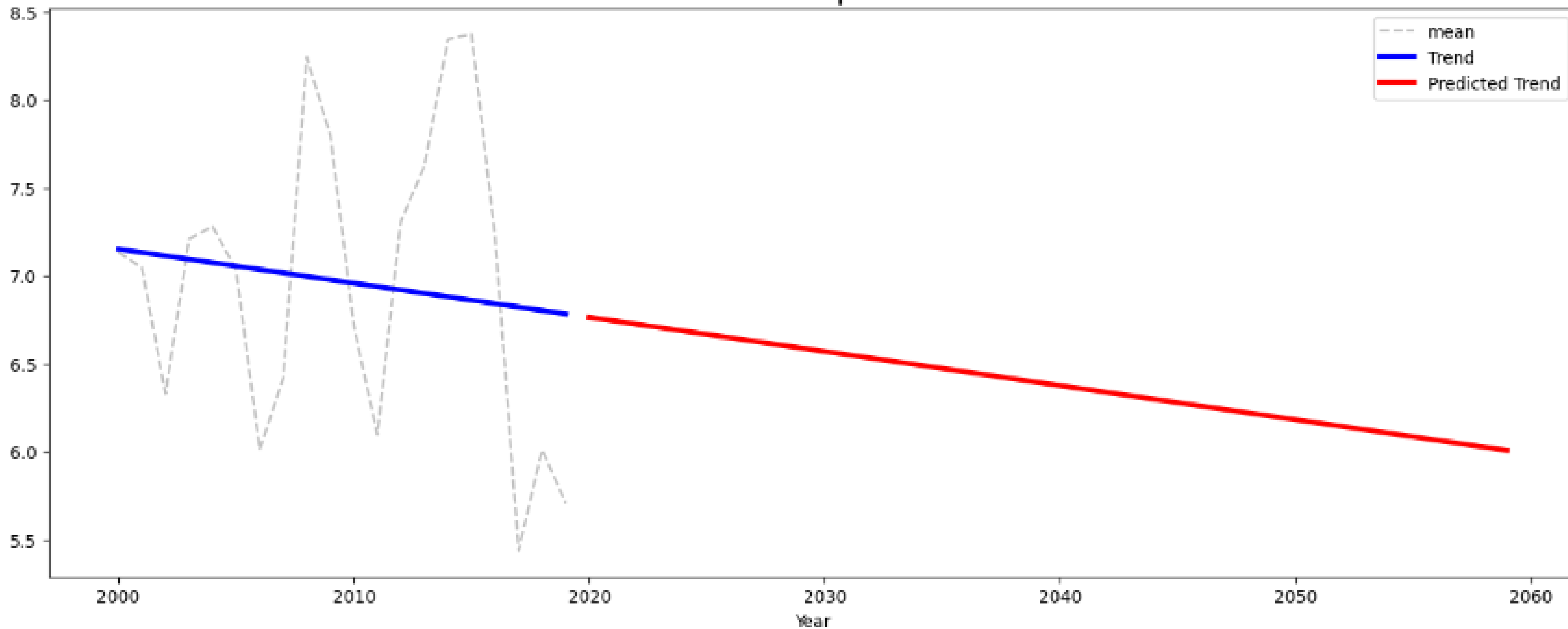
Southern sea ice extent in February Linear Trend Forecast



Southern sea ice extent in March Linear Trend Forecast



Southern sea ice extent in April Linear Trend Forecast



GR_DAILY_SEA_ICE3

HÜSNA NUR KAYA,19052080

YTÜ,VERİ MADENCİLİĞİNE GİRİŞ,MTM4641

husna.kaya@std.yildiz.edu.tr

