CLIMATE CHANGE

NIDORAN

İçimi DEMİRAĞ 191180033

Eylül DALKIRAN 191180030

Abdussamet DOĞAN 191180036

Muhammet İzzet DEMİR 191180032

ABSTRACT

In this paper, we describe our project topic, methodology, datasets, experiments, results, conclusions. In this project, we purpose to see climate change how effect on the natural environment. Therefore, we analyze about effects of climate change to make deduction about global temperature, gas emission, sea level and bird species.

Keywords

Climate change; Gas emissions; Sea level; Temperature; Bird species;

1. INTRODUCTION

Climate change, one of the biggest problems of humanity. It is one of the most important issues of the century we live in. Our primary motivation is to take care of our future by understanding this change and its effects, which will shape most of our lives. Another source of motivation is to raise people's awareness by explaining the consequences of climate change, for example, sea level, temperature and CO2 emissions. We aim to show the analysis about that relationship between temperature change and climate change and change most over the years since temperature changes.

2. METHODOLOGY

2.1 Temperature

We used px.choropleth to visualize earth temperatures on the map. "px.choropleth" is a Plotly Express function for creating choropleth maps. A choropleth map is a type of map that displays divided regions that are coloured or patterned in relation to a data given. The px.choropleth function makes it easy to create choropleth maps by providing options for customizing the map colors, labels, and markers.

We used plotly.graph_objs to visualize the average land temperature in Turkey and in World as a line graph. The "plotly.graph_objs" library is a module in the Plotly library that provides classes for creating and customizing various types of plots and charts using Plotly. It allows you to create a wide range of static, interactive, and animated visualizations using Python.

We used the LinearRegression() and polyfit() functions to make regression models. We also tested the accuracy of our model with the r²_score() function.

2.2 CO2 Emission

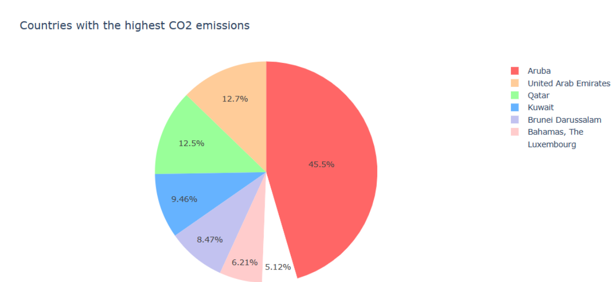
To summarize the data in the DataFrame, you can use the describe() method, which calculates various statistical measures

Trinidad and Tobago	Tunisia	Turkey	Tanzania	Uganda	Upper middle income	Uruguay	United States	St. Vincent and the Grenadines	Venezuela, RB	Vietnam
3.044399	0.413370	0.612271	0.082444	0.062317	2.573291	1.701585	15.999779	0.135865	7.009414	0.181947
5.321474	0.417045	0.616879	0.068047	0.058283	2.408432	1.602728	15.681256	0.133884	6.153191	0.183099
8.231023	0.417336	0.750243	0.071950	0.059458	2.370116	1.540660	16.013937	0.132162	6.188716	0.217694
1.467708	0.444553	0.767638	0.073218	0.057991	2.435563	1.639287	16.482762	0.174204	6.208593	0.196997
4.204809	0.618138	0.870790	0.091548	0.063657	2.523331	1.710104	16.968119	0.215409	6.041541	0.209870
5.699133	0.541833	0.884281	0.099910	0.078519	2.630170	2.049518	17.451725	0.170540	6.271781	0.217934
4.412122	0.621834	0.994631	0.114427	0.088752	2.741215	1.985095	18.121073	0.210973	5.690063	0.225434

for each numeric column as you can see the figure..

You can then use Seaborn to create a line plot showing the relationship between two continuous variables as a line. To do this, you can use the lineplot function from the Seaborn library.

```
top_countries = df2.sort_values(by='co2emission', ascending=False).head(150)
top_countries = top_countries.drop_duplicates('Country Name')
# Create the pie chart
fig = px.pie(top_countries, values='co2emission', names='Country Name', title='Countries with the highest CO2 emissions',
             color_discrete_sequence=['#ff6666', '#ffcc99', '#99ff99', '#66b3ff', '#c2c2f0'])
fig.show()
```



as you can see To create a pie chart visualizing the top countries with the highest and lowest CO2 emissions, you can sort the data in descending order based on the 'co2emission' column, select the top and bottom 9 countries, and use the pie chart function from a library like Matplotlib to create the chart.

2.3 Bird Species

Performed Exploratory Data Analysis by making some simple maps, examining column and data structure. Using Pandas and GeoPandas created maps for the CBC and BBS bird species separately for each predicted year. Using CRS(Coordinate Reference System). It is an attribute of the geometry for each species.

Created layer-separated GeoDataFrames, by iterating first over keep_layers, then again over all the bird species data using os.listdir(), then finally each layer using fiona.listlayers() and [layer string] in [filename.geometry.layer] logic to separate the layers. Used geoDataFrame.bounds.sort_values() to limit the target area bounds, and used np.linspace to create equally-sized squares.Read map of North America, created at this website and stored locally here.Used GeoDataFrame.intersects() and iterated over every species in sum_00. Used sum() to add the boolean values into the count column. Mapped the "pixels" over the map of North America and the layered polygons of sum_00, colored using GeoPandas' built in chloropleth argument in their version of .plot(). NorthAmerica using plt.subplots(), I created and then saved my output maps as side-by-side, "Summer" and "Winter" maps for each year.

2.4 Sea Level

The datasets type is time series. To analyze and predict values The VAR function from the statsmodels library is used to fit a vector autoregressive (VAR) model to time series data. A VAR model is a statistical model used to describe the evolution of a multivariate time series. It specifies a relationship between multiple time series variables such that each variable is a function of its own past values and the past values of the other variables.

In a VAR model, the variable being modeled is treated as a linear combination of its own lagged values, the lagged values of the other variables, and an error term. The coefficients of the lagged values are the parameters of the model and are estimated from the data.

In the dataset there are dates, GMSL and GMSL Uncertainty attributes.

```
RangeIndex: 1608 entries, 0 to 1607
Data columns (total 3 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   Time              1608 non-null   object
 1   GMSL              1608 non-null   float64
 2   GMSL uncertainty  1608 non-null   float64
```

There are not any missing values in the 1608 records. This dataset contains one record for each month since 1880 and 2014. We used LinearRegression from sklearn.linear_model to make predictions. Actually there are two different models used; Linear regression and Polynomial regression. For the training we split the data as 75 percent will be training set and 25 percent will be test set.

```
X= sea_level_data2015.dt_ordinal.to_numpy().reshape(-1,1) # linear
Y= sea_level_data2015.GMSL      #values
x_train,x_test,y_train,y_test=train_test_split(X,Y,test_size=0.25)
```

Statsmodels is a Python module that provides classes and functions for the estimation of many different statistical models, as well as for conducting statistical tests, and statistical data exploration. Statsmodels also includes tools for performing decompositions of time series data, such as trend and seasonality decomposition, as well as tools for conducting time series cross-validation and model selection.

This includes tools for modeling time series using statistical models such as SARIMAX (Seasonal ARIMA), and Vector Autoregression (VAR). It also includes tools for estimating the parameters of these models and for conducting statistical tests to assess the goodness of fit of the models.

3. DATASETS

3.1 Temperature

The GlobalLandTemperaturesByCountry dataset contains the average temperature values of countries from 1743 to 2013. With this dataset, a visualization of average temperature values on 2D and 3D maps, as well as a graph showing the temperature trend in Turkey, is also made. In this direction, you can see the null values of the data set deleted and the year data required for visualization extracted into a separate column below.

	Year	dt	AverageTemperature	AverageTemperatureUncertainty	Country
0	1743	1743-11-01	4.384	2.294	Åland
1	1744	1744-04-01	1.530	4.680	Åland
2	1744	1744-05-01	6.702	1.789	Åland
3	1744	1744-06-01	11.609	1.577	Åland
4	1744	1744-07-01	15.342	1.410	Åland
...	...	...	...	...	...
544806	2013	2013-04-01	21.142	0.495	Zimbabwe
544807	2013	2013-05-01	19.059	1.022	Zimbabwe
544808	2013	2013-06-01	17.613	0.473	Zimbabwe
544809	2013	2013-07-01	17.000	0.453	Zimbabwe
544810	2013	2013-08-01	19.759	0.717	Zimbabwe

The other dataset is GlobalTemperatures and includes the world's average land temperature values from 1750 to 2015. With this data set, a graph showing the temperature trend in the world is made. The last data set, city_temperature, was used to examine the temperature change according to the continents, as it also includes region information. For this, first of all, null and erroneous data were deleted and then the temperature values from Fahrenheit in the data set were converted to Celcius.

3.2 CO2 Emission

First dataset we have a table with data on different countries' GDP per capita in different years. The table has data for 1960 through 1964. The values are given in terms of a percentage of the world's GDP per capita in that year. The rows represent different countries or regions, and the columns represent different years.

Country Name	Africa Eastern and Southern	Afghanistan	Africa Western and Central	Angola	Albania	Arab World	United Arab Emirates
1960	0.906060	0.046057	0.090880	0.100835	1.258195	0.609268	0.119037
1961	0.922474	0.053589	0.095283	0.082204	1.374186	0.662618	0.109136
1962	0.930816	0.073721	0.096612	0.210533	1.439956	0.727117	0.163542
1963	0.940570	0.074161	0.112376	0.202739	1.181681	0.853116	0.175833
1964	0.996033	0.086174	0.133258	0.213562	1.111742	0.972381	0.132815

first data

second data is a table of data on CO2 emissions (in metric tons per capita) for different countries in different years. The table has data for the years 1990 through 2019. The rows represent different countries, and the columns represent different years. The "country_code" column gives a code for each country, and the "Region" column specifies the region in which the country is located.

	Country Name	country_code	Region	Indicator Name	1990	1991	1992	1993	1994	1995	...
0	Aruba	ABW	Latin America & Caribbean	CO2 emissions (metric tons per capita)	NaN	NaN	NaN	NaN	NaN	NaN	...
1	Afghanistan	AFG	South Asia	CO2 emissions (metric tons per capita)	0.191745	0.167682	0.095958	0.084721	0.075546	0.068468	...
2	Angola	AGO	Sub-Saharan Africa	CO2 emissions (metric tons per capita)	0.553662	0.544539	0.543557	0.708984	0.836804	0.912141	...

second data

3.3 Bird Species

For mapping, 616 different bird species in North America using data from the Audubon Society's Christmas Bird Counts and Breeding Bird Surveys. The data includes historic observations as well as predictions for the future based on the Intergovernmental Panel on Climate Change's (IPCC) Special Report on Emissions Scenarios (SRES).The predictions cover the years 2020, 2050, and 2080 and are based on projected levels of carbon dioxide and other greenhouse gases.

	value	Shape_Leng	Shape_Area	geometry
0	1.0	5500000.0	7.410000e+10	MULTIPOLYGON (((-1844911.231 -655519.546, -184...
1	1.0	121380000.0	2.459500e+12	MULTIPOLYGON (((-834911.231 -705519.546, -8249...
2	1.0	103120000.0	2.774100e+12	MULTIPOLYGON (((-1854911.231 2274480.454, -184...
3	1.0	15100000.0	1.159300e+12	MULTIPOLYGON (((335088.769 -1165519.546, 34508...
4	1.0	1640000.0	1.120000e+10	MULTIPOLYGON (((1755088.769 -1555519.546, 1765...

For mass wing ratio, We used a dataset containing data on bird species living in the Amazon rainforest between 1980 and 2020 from a prestigious university, LSU.

species	spMeanMass	spMeanWing	spMeanMW	year	mass	wing	mw	temp_lag0_dry	temp_lag2_dry	precip_lag0_dry	precip_lag2_dry
Deconychura_stictolaema	16.7182	78.3903	0.2141	1980	20.0	85.0	0.235294	25.919	26.432	960.083	941.847
Myiobius_barbatus	10.4659	61.8618	0.1691	1980	11.5	62.0	0.185484	25.919	26.432	960.083	941.847
Thalurania_furcata	4.0815	50.4655	0.0804	1980	3.5	49.0	0.071429	25.919	26.432	960.083	941.847
Thalurania_furcata	4.0815	50.4655	0.0804	1980	4.0	51.0	0.078431	25.919	26.432	960.083	941.847
Mionectes_macconnelli	12.2338	62.9850	0.1947	1980	11.5	58.0	0.198276	25.919	26.432	960.083	941.847
Formicarius_analis	61.8459	90.3300	0.6846	1980	60.0	89.0	0.674157	25.919	26.432	960.083	941.847

3.3 Sea Level

We have two datasets to analyze. We found this datasets from different sources so they have some differences about records, attributes, and record counts. Firstly, we get information about datasets and extract the first 5 records to see.

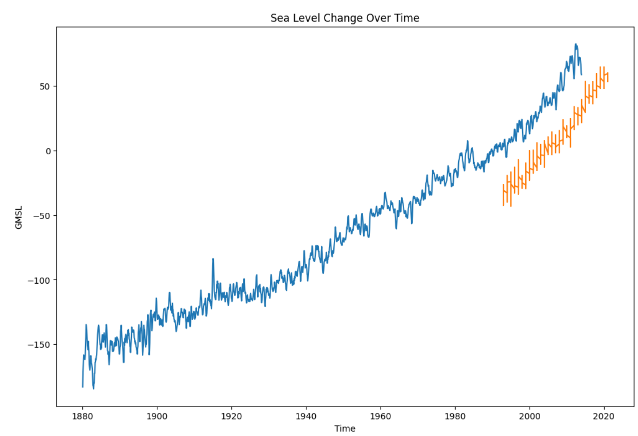
First dataset about sea level in 1993-2021:

```
 1993-2021 datas:  (1048, 9)
   Year  TotalWeightedObservations  GMSL_noGIA  StdDevGMSL_noGIA
0  1993                  327401.31      -38.59             89.86
1  1993                  324498.41      -41.97             90.86
2  1993                  333018.19      -41.93             87.27
3  1993                  297483.19      -42.67             90.75
...
 8   SmoothedGSML_GIA_sigremoved  1048 non-null   float64
dtypes: float64(8), int64(1)
```

Second datasets:

```
RangeIndex: 1608 entries, 0 to 1607
Data columns (total 3 columns):
 #   Column           Non-Null Count  Dtype
---  ------           --------------  -----
 0   Time             1608 non-null   object
 1   GMSL             1608 non-null   float64
 2   GMSL uncertainty 1608 non-null   float64
dtypes: float64(2), object(1)
```

We generally use the second dataset because there is more data. Secondly we compare the datasets to see the difference between them. And we visualize the result to see what the difference is. Here is the line graphic:



In the graphic, there is a difference but they look like they have the same increasing ratio.

4. EXPERIMENTS

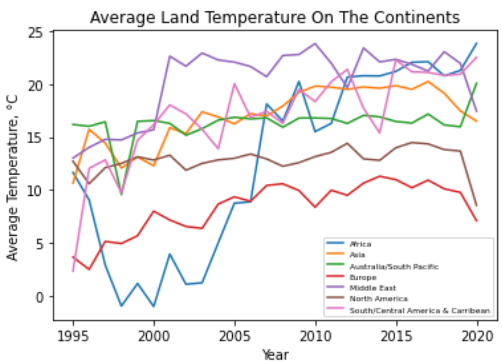
In this project, firstly we defined a subject. After that, we did some research about theoretical knowledge. Another thing is, which we did secondly, researching to collect datasets. We looked at different sources like kaggle, nasa. After we found enough datasets to answer our questions we shared topics to do analyzing.

Everyone purified these datasets from missing values and unnecessary attributes. In addition to that, we visualized these datasets and used several different regression models to prediction the future. Therefore, we learnt making predictions with trained datas and visualizing the results.

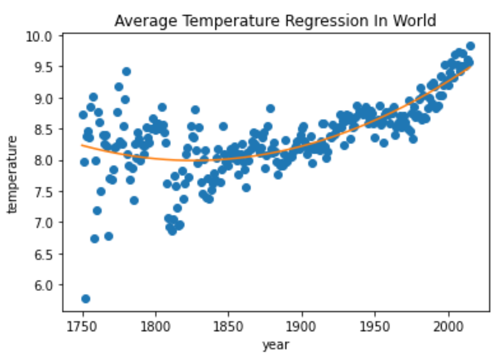
5. RESULTS

5.1 Temperature

When we look at all the outputs and the results of the trained models, we can say that there is a general temperature increase in the world. For example, when we look at the image showing the continents below, we see that there is a general increase in temperature.

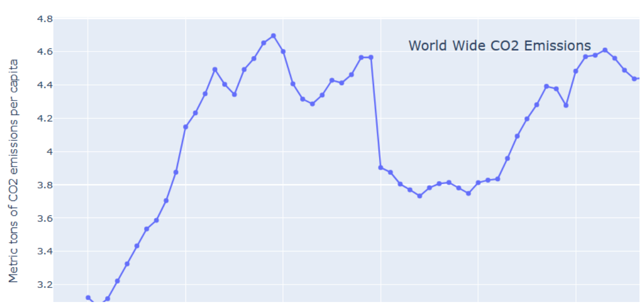


Or when we look at the quadratic regression model, as seen below, we arrive at the same conclusion.



5.2 CO2 Emission

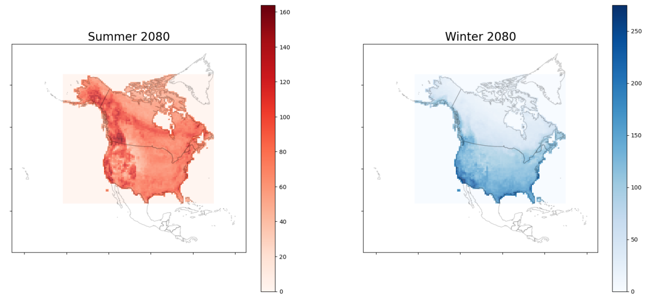
Carbon dioxide (CO2) emissions are a major impact of climate change and are increasing over time. As we can see from the values in the pie charts, there is a very serious difference between the country with the highest CO2 emissions and the country with the least CO2 emissions.



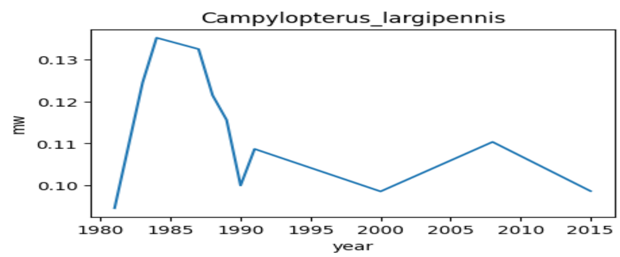
We expect emissions to increase over the years, we can say that the reason for this is the increasing industrial activities and the increasing number of vehicles, we can say that this increase has decreased in recent years. We can say that the reason for this abnormal decrease is the decreased human activities during the pandemic process.

5.3 Bird Species

Large numbers of bird species near the mountain range continue Northward into the Coast Mountains and Alaska Range in the future.

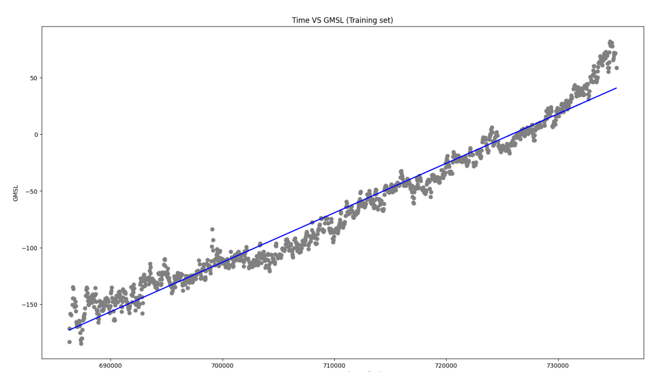


In order to adapt to changing climate, the wing lengths increase while the masses of birds decrease.



5.4 Sea Level

There was a question about the sea level change in the beginning. We wanted to reach the answer of the effects of climate change on the sea level. Therefore, firstly we visualized the datasets. After that, we used different types of regression models to train data. And we extracted some predictions from the trained model by using these regression models.



The regression's training set ratio is 75 percent. We used prediction line with test data. Global Mean Sea Level is increasing because of climate change over time. We can see that sea level will be more increased after coming years owing to these graphics of regressions. We expected this result in the beginning while detect questions. After these visualizations and predictions we reached the result which we expected in the beginning.

6. CONCLUSIONS

When we look at all the outputs and the results of the trained models, we can say that there is a general temperature increase in the world. Then, we say the Carbon dioxide (CO_2) emissions are a major impact of climate change.then we came to the following conclusion about the birds "Many bird species near the mountain range will continue to migrate northward into the Coast Mountains and Alaska Range in the future. In order to adapt to changing climate, the wing lengths of these birds will increase while their masses will decrease."

Global Mean Sea Level is increasing because of climate change over time. We can see that sea level will be more increased after coming years owing to these graphics of regressions.

After performing data analysis, the experience gained allows you to learn to discover meaningful relationships between data and draw conclusions based on the data.

7. REFERENCES

[1] https://www.statsmodels.org/0.9.0/generated/statsmodels.tsa.statespace.sarimax.SARIMAX.html

[2] https://www.statsmodels.org/stable/vector_ar.html

[3] https://geopandas.org/en/stable/

[4] https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html

[5] https://www.ipcc.ch/report/emission-scenarios

[6] https://matplotlib.org/stable/plot_types/index.html

[7] https://www.sciencebase.gov/catalog/item/55897deae4b0b6d21dd61c9d

[8] https://plotly.com/python/plotly-express/

[9] https://plotly.com/python-api-reference/generated/plotly.graph_objects.Figure.html