Climate Change Project

Introduction

- Overview: The goal of our project is to investigate patterns across regional and global temperatures and forecast future temperatures using statistical techniques and machine learning algorithms.

- Dataset: “Climate Change: Earth Surface Temperature Data”
  - Link to data: https://www.kaggle.com/datasets/berkeleyearth/climate-change-earth-surface-temperature-data

- Dataset background: 7 variables - date, average temperature, average temperature uncertainty, city, country, latitude longitude

Overview of dataset

- Cleaning the dataset: we removed the average temperature uncertainty column since it wasn't relevant to our analysis.

- The dataset was large so for some parts of our analysis, we decided to subset and focus on specific geographical locations or factors of interest.
    - 8235082 rows × 6 columns

	dt	AverageTemperature	City	Country	Latitude	Longitude
0	1743-11-01	6.068	Århus	Denmark	57.05N	10.33E
5	1744-04-01	5.788	Århus	Denmark	57.05N	10.33E
6	1744-05-01	10.644	Århus	Denmark	57.05N	10.33E
7	1744-06-01	14.051	Århus	Denmark	57.05N	10.33E
8	1744-07-01	16.082	Århus	Denmark	57.05N	10.33E
...	...	...	...	...	...	...
8599206	2013-04-01	7.710	Zwolle	Netherlands	52.24N	5.26E
8599207	2013-05-01	11.464	Zwolle	Netherlands	52.24N	5.26E
8599208	2013-06-01	15.043	Zwolle	Netherlands	52.24N	5.26E
8599209	2013-07-01	18.775	Zwolle	Netherlands	52.24N	5.26E
8599210	2013-08-01	18.025	Zwolle	Netherlands	52.24N	5.26E

General statistics - EDA

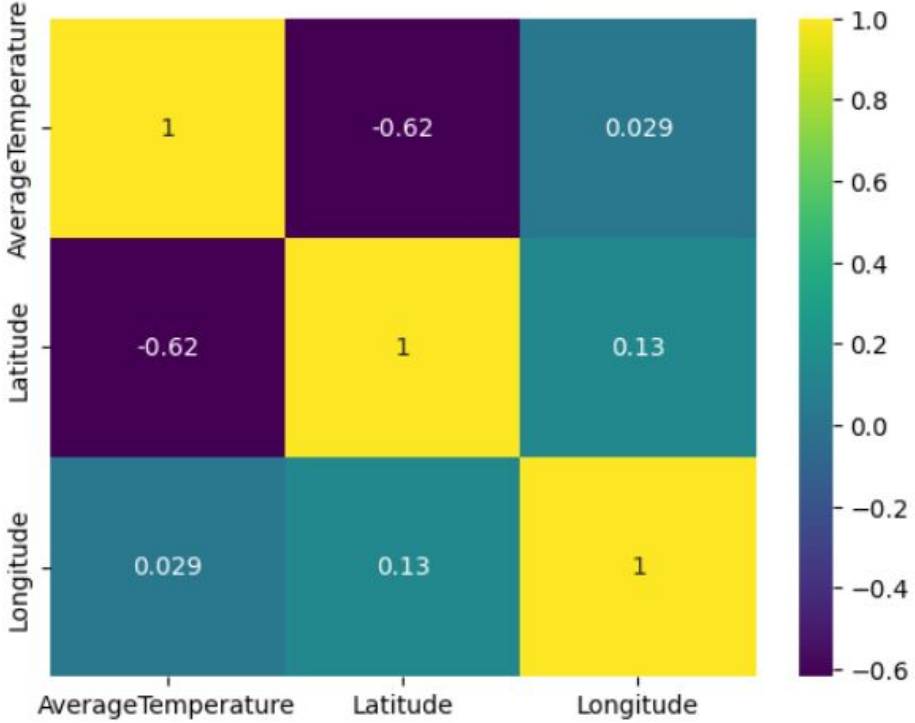
- Average temperature of all data points:
  - 16.727432636247972 degrees Celsius

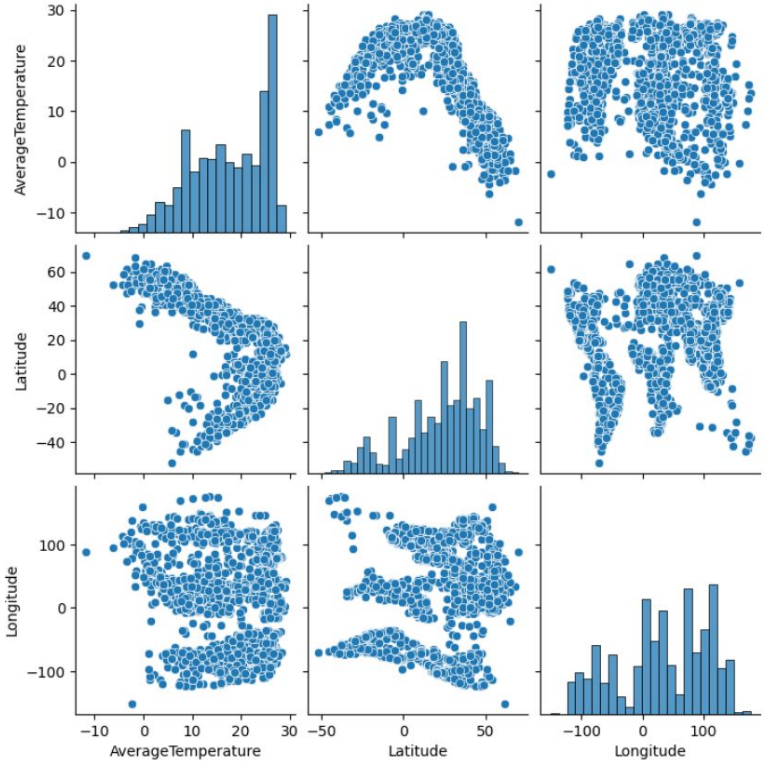
- Average temperature by country and city

Country	AverageTemperature
Mongolia	-3.365485
Iceland	1.500089
Russia	3.347268
Norway	3.612553
Finland	3.711645
...	...
Mali	27.590491
Burkina Faso	27.815295
Sudan	28.072831
Niger	28.145552

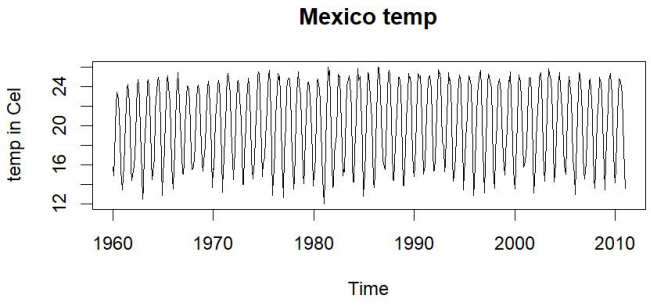
City	AverageTemperature
Norilsk	-11.854750
Kyzyl	-6.222452
Chita	-4.360300
Ust Ilimsk	-3.996800
Surgut	-3.538281
...	...
Kassala	28.938776
Niamey	29.062560
Umm Durman	29.081291
Khartoum	29.081291
Jibuti	29.152790

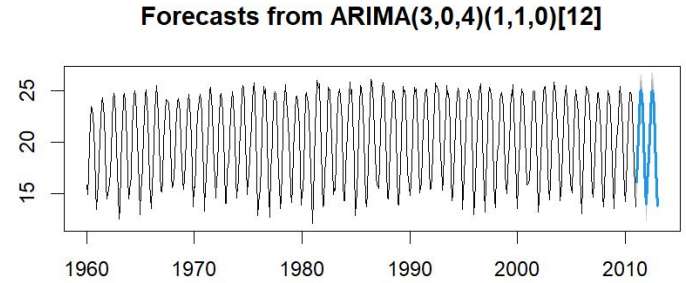
Data Visualization





Forecasting (time series)





The model shows good performance on the training set, with low error measures.

However, when tested on unseen data, the model's performance declines, with larger errors and a negative bias in predictions.

The average absolute percentage difference (MAPE) between predicted and actual values is around 6.3%.

Considering alternative modeling techniques or exploring different time series models can lead to more robust and accurate predictions.

Future extensions to project

- Due to time constraints and time taken to load the data, we didn't have a fully functioning dashboard.
- However, we were able to play around with an existing dataset in Tableau and explore different ways to visualize data.
- Given more time, in the future we would create a dashboard with our dataset and create more visualizations.