

Weather data analysis project

Week 5





Table of content

1. Objectives
2. Data collection and workflow
3. Results
4. Hypothesis
5. Conclusion

Objectives

- Obtain weather data from 4 cities of relatively different geo locations
- Run data through our python code and observe the data visually
- Check weather changes over the decades to verify climate change.
- Apply hypothesis testing using the paired method to forecast the trends.
- Check if COVID-19 had any direct or indirect impact on the current weather patterns





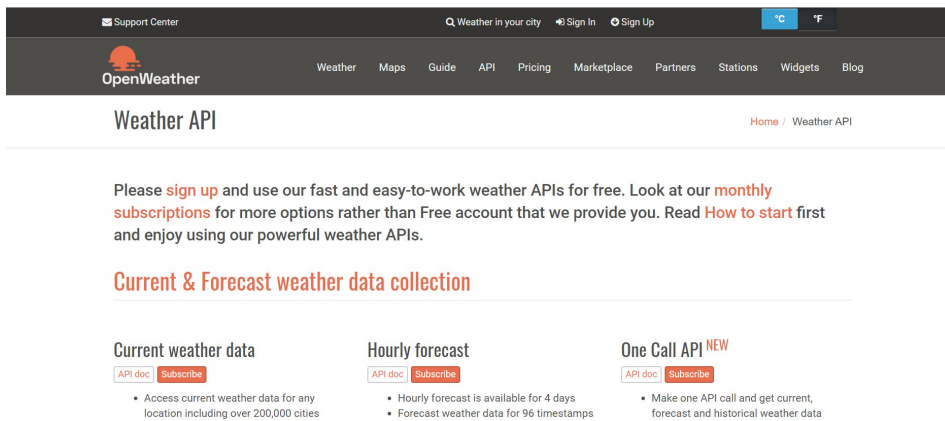


Data collection

Data scraped from : [Open weather API](#)

5 cities scraped in total from 1979 to 2020 (data per hour)

- Berlin, Milan and Hotan in a first csv
- Then Dhaka and Szczecin in a second csv.



The screenshot shows the OpenWeather API website. The header includes a search bar, login/signup links, and temperature units. The main navigation menu lists various services. The 'Weather API' section is highlighted, featuring a call to action to sign up for free APIs. Below this, three service cards are displayed: 'Current weather data', 'Hourly forecast', and 'One Call API NEW'. Each card includes a brief description and a 'Subscribe' button.

Support Center Weather in your city Sign In Sign Up °C °F

OpenWeather Weather Maps Guide API Pricing Marketplace Partners Stations Widgets Blog

Weather API

Home / Weather API

Please [sign up](#) and use our fast and easy-to-work weather APIs for free. Look at our [monthly subscriptions](#) for more options rather than Free account that we provide you. Read [How to start](#) first and enjoy using our powerful weather APIs.

Current & Forecast weather data collection

Current weather data

[API doc](#) [Subscribe](#)

- Access current weather data for any location including over 200,000 cities

Hourly forecast

[API doc](#) [Subscribe](#)

- Hourly forecast is available for 4 days
- Forecast weather data for 96 timestamps

One Call API NEW

[API doc](#) [Subscribe](#)

- Make one API call and get current, forecast and historical weather data



Data collection

	dt	dt_iso	timezone	city_name	lat	lon	temp	feels_like	temp_min	temp_max	...	wind_deg	rain_1h	rain_3h	snow_1h	snow_3h	clouds_all	weather_id	weather_main	weather_description	weather_icon
0	283996800	1979-01-01 00:00:00+0000 UTC	3600	Berlin	52.520007	13.404954	-18.69	-25.24	-20.63	-18.0	...	20	NaN	NaN	NaN	NaN	20	801	Clouds	few clouds	02n
1	284000400	1979-01-01 01:00:00+0000 UTC	3600	Berlin	52.520007	13.404954	-18.43	-24.20	-20.37	-18.0	...	20	NaN	NaN	NaN	NaN	100	804	Clouds	overcast clouds	04n
2	284004000	1979-01-01 02:00:00+0000 UTC	3600	Berlin	52.520007	13.404954	-18.38	-24.89	-20.10	-18.0	...	360	NaN	NaN	NaN	NaN	100	804	Clouds	overcast clouds	04n
3	284007600	1979-01-01 03:00:00+0000 UTC	3600	Berlin	52.520007	13.404954	-18.10	-24.60	-18.62	-18.0	...	10	NaN	NaN	NaN	NaN	75	803	Clouds	broken clouds	04n
4	284011200	1979-01-01 04:00:00+0000 UTC	3600	Berlin	52.520007	13.404954	-18.02	-24.86	-18.17	-18.0	...	330	NaN	NaN	NaN	NaN	68	803	Clouds	broken clouds	04n

```
df.shape
```

```
(1829805, 31)
```

```
df.columns
```

```
Index(['dt', 'dt_iso', 'timezone', 'city_name', 'lat', 'lon', 'temp',  
      'feels_like', 'temp_min', 'temp_max', 'pressure', 'sea_level',  
      'grnd_level', 'humidity', 'wind_speed', 'wind_deg', 'rain_1h',  
      'rain_3h', 'snow_1h', 'snow_3h', 'clouds_all', 'weather_id',  
      'weather_main', 'weather_description', 'weather_icon', 'Date', 'day',  
      'month', 'year', 'delta_T', 'decade'],  
      dtype='object')
```



Workflow

Data cleaning

- Dropping useless columns
- Creating a “date” column, followed by “month” and “year”
- Creating a column “decade” based on the years
- Creating a “delta” column (max-min temperature difference)



Data aggregation

- Splitting the cleaned dataframe per city
- Grouping and creating multiple dataframes per month, decade for our analysis



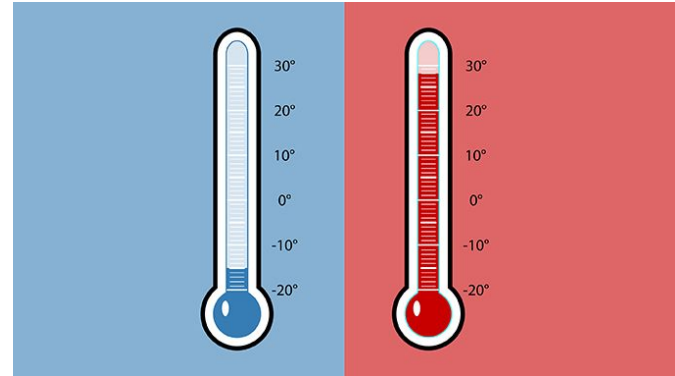
Results & data visualisation

- Creation of 6 line plots and a correlation matrix
- Saving the clean dataframe and graphs obtained in an output folder



Results

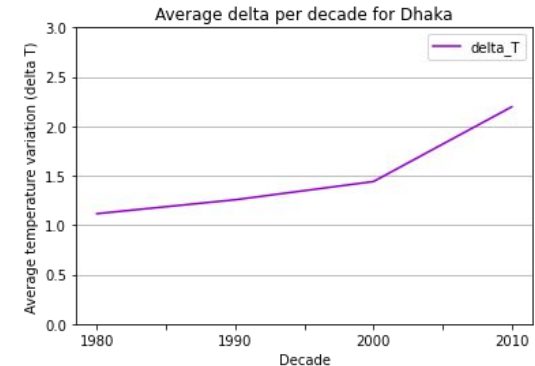
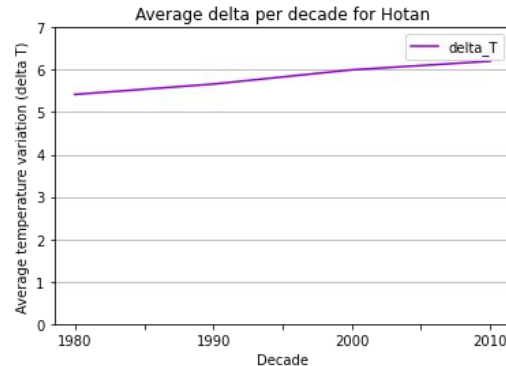
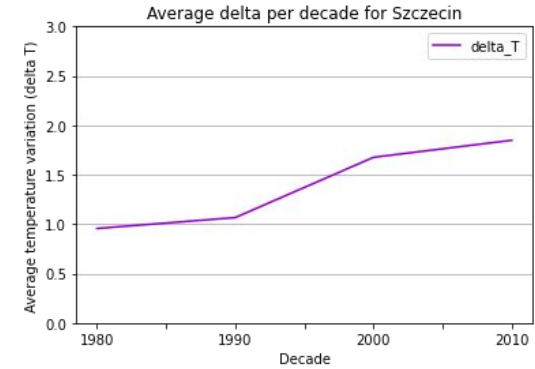
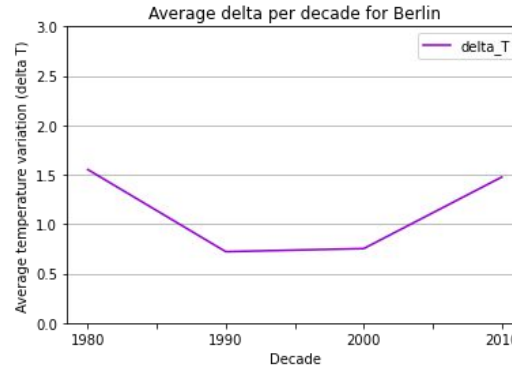
1. Average variation temperature per decade
2. Average minimum and maximum temperature per decade
3. Average temperature per decade
4. Average temperature per month per decade
5. Average minimum and maximum temperature per month per decade
6. Correlation between temperature parameters
7. Checking recent temperatures





1. Average temperature variation per city

1. Impact of the variation of temperature depends on the geological conditions
2. Increase in the variation over the decades
3. Change in temperature in Berlin



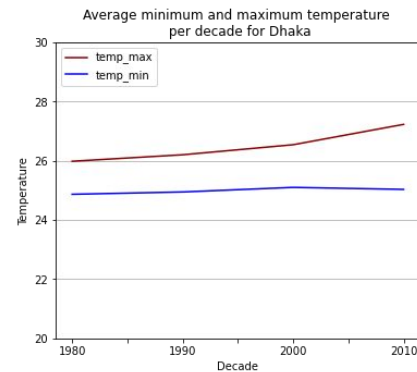
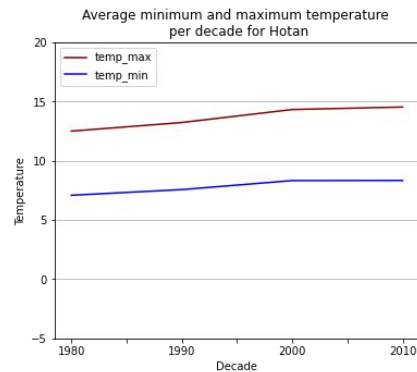
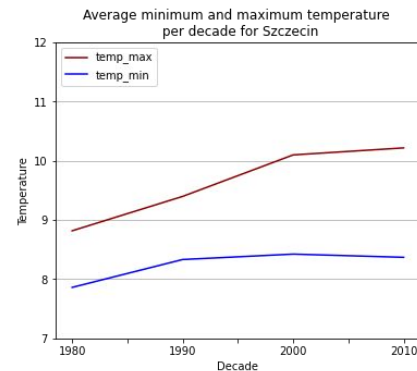
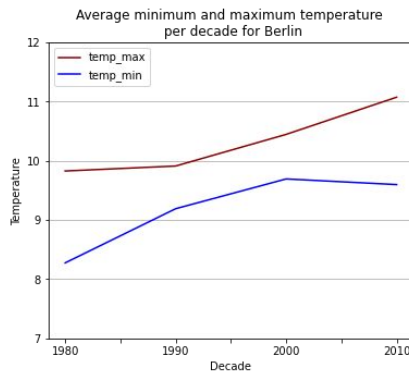


2. Average min and max temperatures per decade

Temperature increase don't respect the same shape
Exemple Berlin :

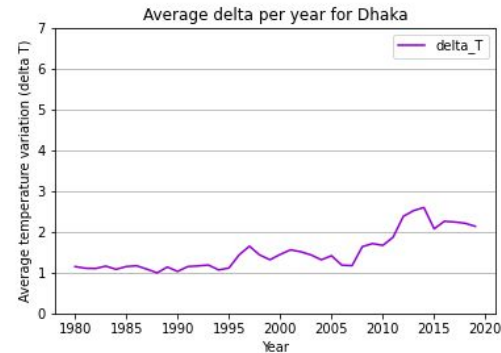
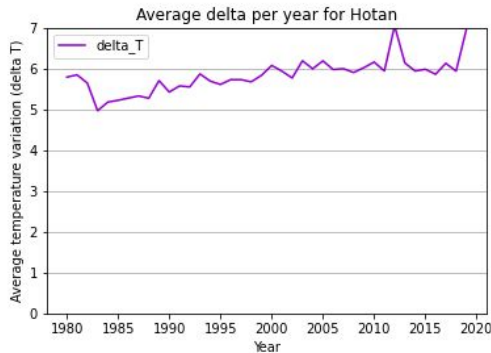
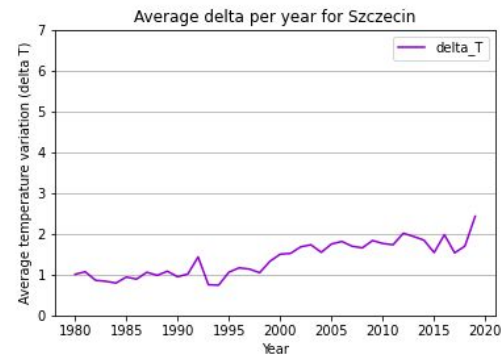
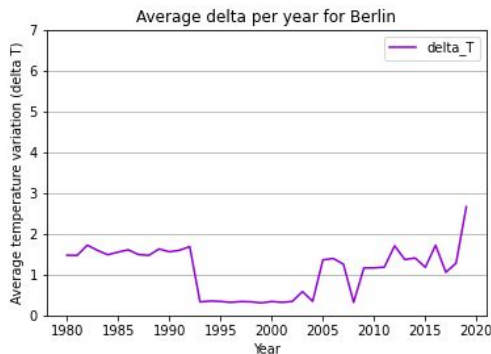
The min temperature in the last decade tends to stabilize against the max temperature continues to increase.

The increase in the minimum of the temperature can explain the appearance of the previous graph (shape of ΔT)



1. Average temperature variation per year

1. Impact of the variation of temperature depends on the geological conditions
2. Increase in the variation over the decades
3. Change in temperature in Berlin





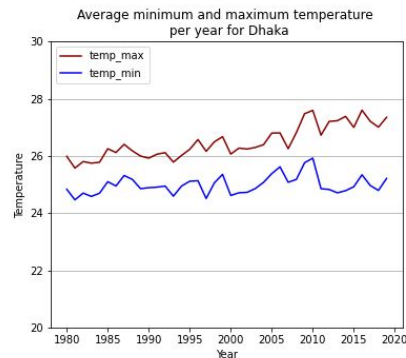
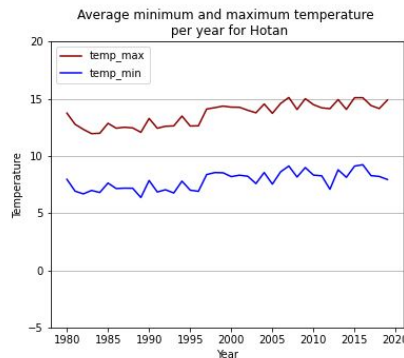
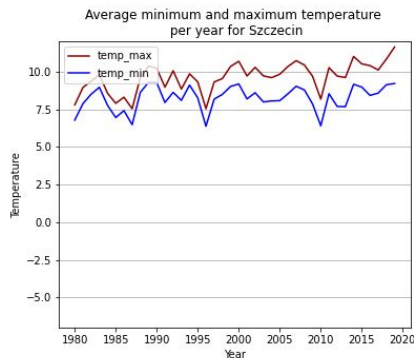
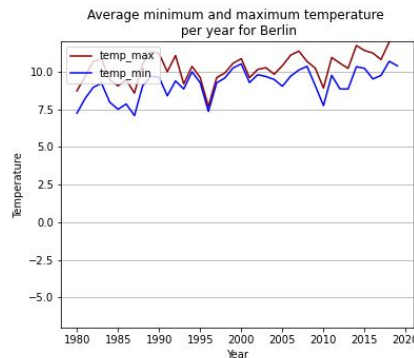
2. Average min and max temperatures per year

Temperature increase don't respect the same shape

Exemple Berlin :

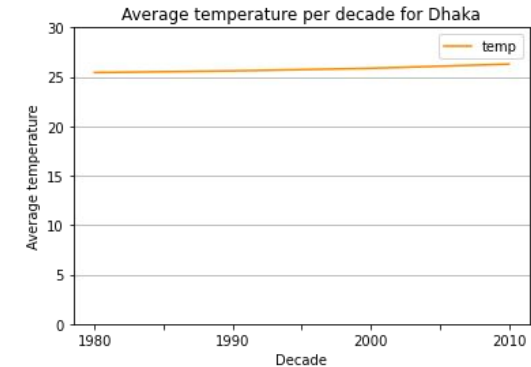
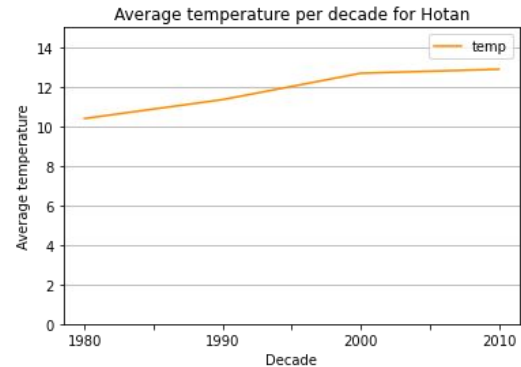
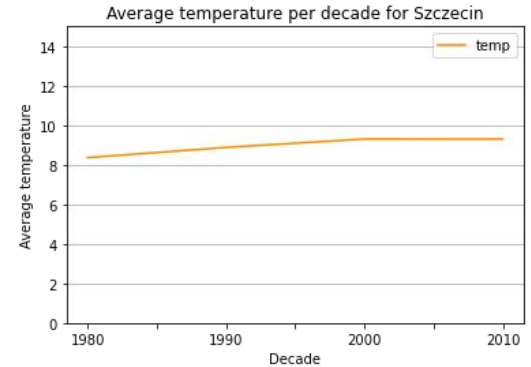
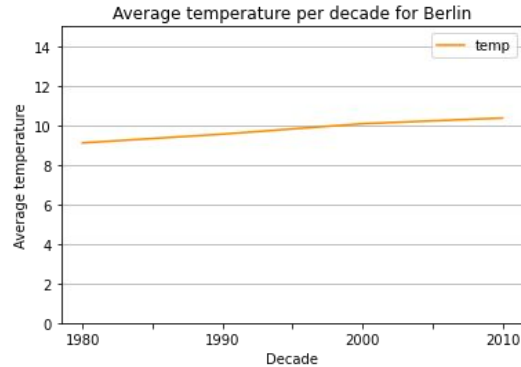
The min temperature in the last decade tends to stabilize against the max temperature continues to increase.

The increase in the minimum of the temperature can explain the appearance of the previous graph (shape of ΔT)



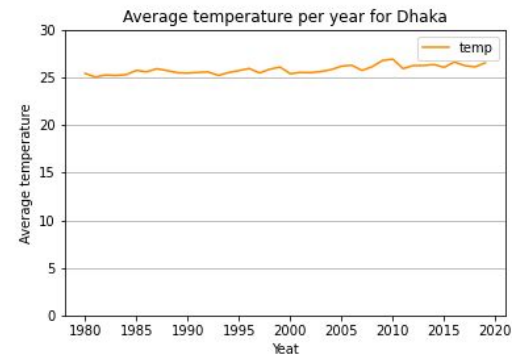
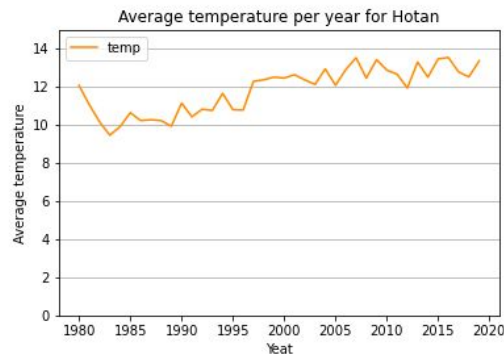
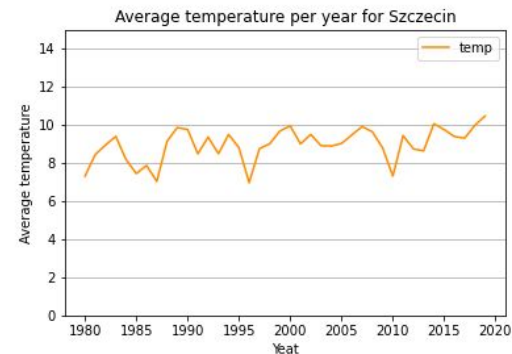
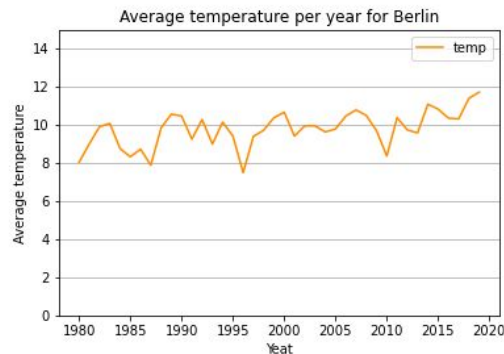
3. Average temperatures per decade

- The average temperature per decade increased in all 4 cities since 1980.



3. Average temperatures per year

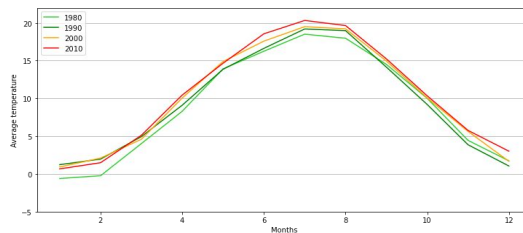
- The average temperature per year increased in all 4 cities since 1980.



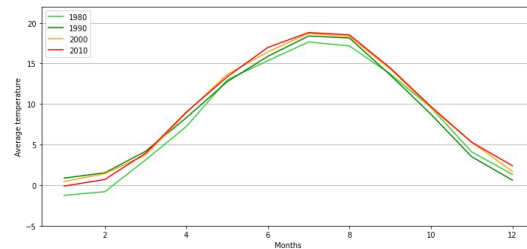
4. Average temperatures per months per decade

- The increase affects all month and season of the year
- Difference between Dhaka and Berlin

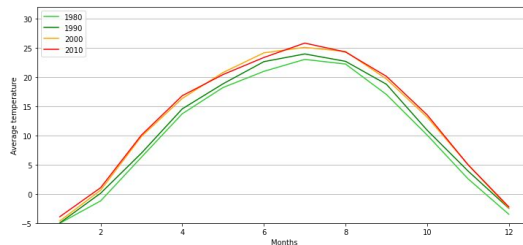
Average temperature per month and per decade for Berlin



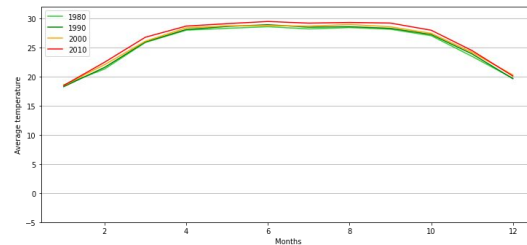
Average temperature per month and per decade for Szczecin



Average temperature per month and per decade for Hotan



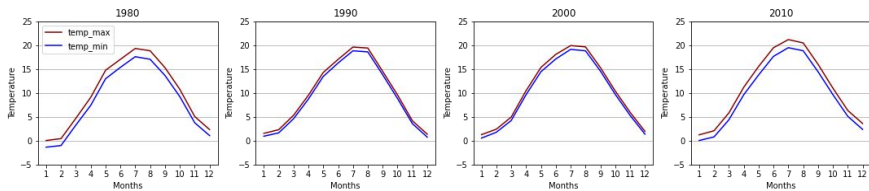
Average temperature per month and per decade for Dhaka



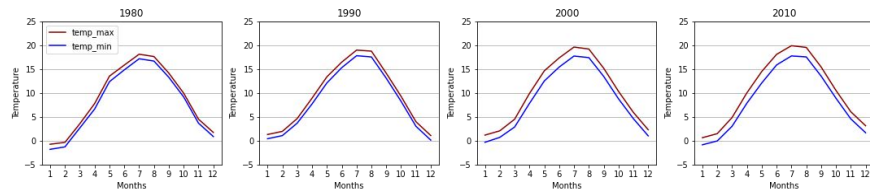


5. Average min and max of the temperature over the 4 last decades

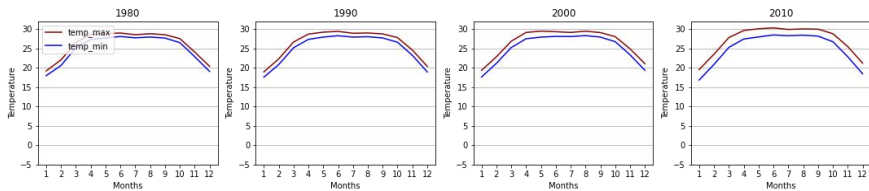
Average minimum and maximum temperature per month and per decade for Berlin



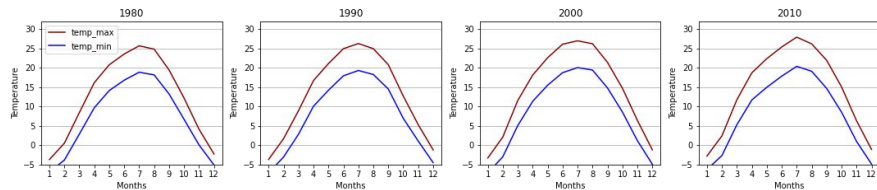
Average minimum and maximum temperature per month and per decade for Szczecin



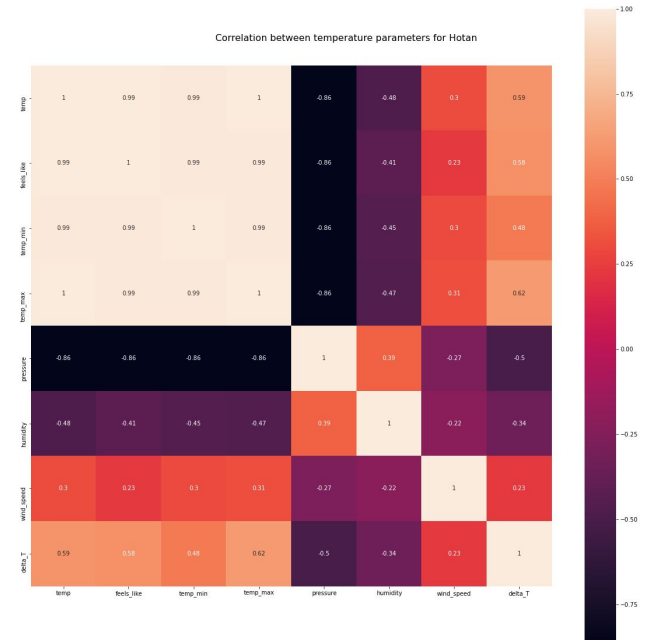
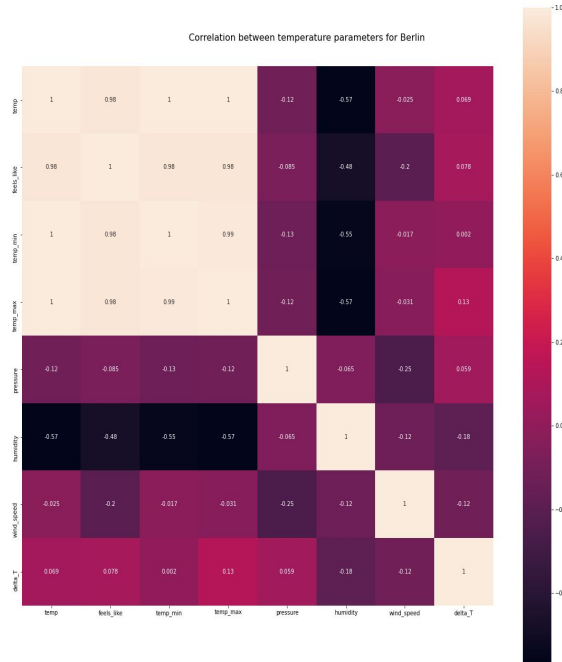
Average minimum and maximum temperature per month and per decade for Dhaka



Average minimum and maximum temperature per month and per decade for Hotan

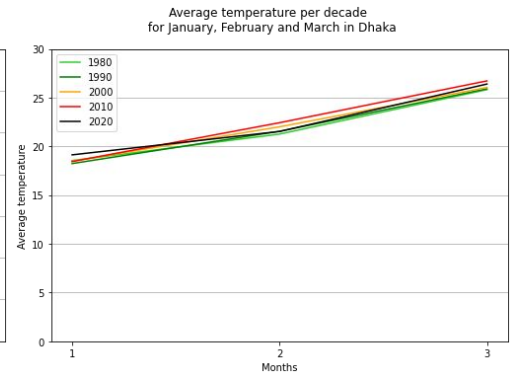
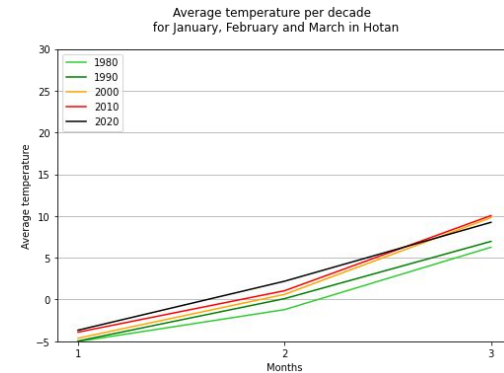
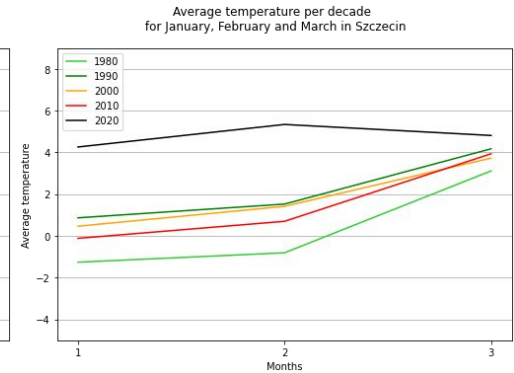
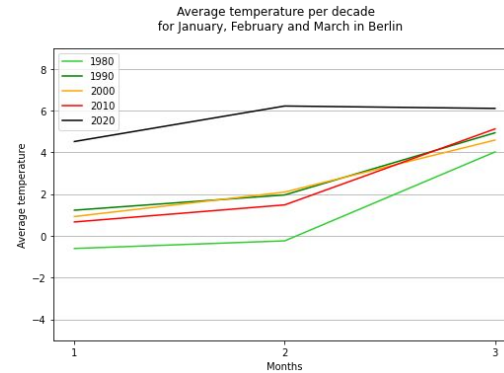


6. Correlations between temperatures parameters



7. 2020 temperatures vs other decades

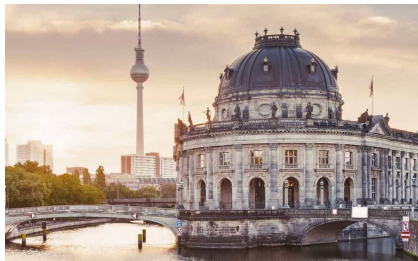
- The graph shows that the temperature have stopped increasing compared with the average temperatures of other decades.



Hypothesis testing

Hypothesis 1: Are Berlin and Szczecin average temperatures comparable ?

BERLIN



Average temperature: 9.76
Standard deviation: 7.81

VS

SZCZECIN



Average temperature :8.98
Standard deviation: 7.56

Results: `Ttest_relResult(statistic=90.53136037419236, pvalue=0.0)`

Conclusion: $pvalue < 0.05$, so the hypothesis that temperatures are comparable is rejected. Both cities average temperatures are significantly different. Another test decreasing Berlin temperatures by 1 degree had the same result.



Hypothesis testing

Hypothesis 2: Is climate change real ?



BERLIN

Average temperature: 9.55

Standard deviation: 7.71

SZCZECIN

Average temperature: 8.91

Standard deviation: 7.48

VS



BERLIN

Average temperature :10.49

Standard deviation: 7.93

SZCZECIN

Average temperature :9.47

Standard deviation: 7.62

Results:

- Berlin: `Ttest_relResult(statistic=-10.013286240252075, pvalue=2.6553851055415264e-23)`
- Szczecin: `TTtest_relResult(statistic=-6.317321778803096, pvalue=2.981668322092181e-10)`

Conclusion: In both cases, $pvalue < 0.05$, so the hypothesis that temperatures are comparable is rejected. Temperature in the 2010s is significantly different from the 1990s.



Conclusion

- Weather data was compiled and visualized
- Results indicate climate change is a real ongoing event
- Data needed cleaning and visual editing in order to read and understand results.
- To further explore COVID impact, pollution level data would need to be included.