Analysis Report

1 Question

How did the average temperature in Alabama during the year 2020 correlate with the number of COVID-19 death cases in the state?

2 Brief Explanation on Why I Had to Change the Previous Dataset

The dataset lacked information about the location where the data was collected, which was crucial for solar radiation and weather data as these parameters can vary significantly by geographic region. Secondly, The dataset provided did not share common columns or keys with the other dataset, which was necessary for a merge. Without shared columns, finding a basis for merging was problematic.

3 Introduction

The objective of this analysis is to explore the correlation between average temperature and COVID-19 death cases in Alabama during the year 2020. The motivation behind this study stems from understanding how environmental factors, such as temperature, might influence the spread and severity of infectious diseases. In this case, the focus is on COVID-19.

4 Used Data

The data used in this analysis includes:

- **Weather Data:** Daily weather data for Alabama in 2020, sourced from Meteostat.

  Data Source: Meteostat
  URL: https://bulk.meteostat.net/v2/daily/KGVQ0.csv.gz
  License: Open data license from Meteostat
  Data Quality: The dataset is structured with a clear format and includes daily records with minimal missing values.

  Description: This dataset includes COVID-19 cases and deaths data for counties in the USA.
  Data Source: Kaggle (imdevskp/corona-virus-report)
  URL: https://www.kaggle.com/imdevskp/corona-virus-report
  License: Kaggle Open Data license
  Data Quality: The dataset is comprehensive and contains daily records of COVID-19 cases and deaths. Some counties may have inconsistencies in reporting, which could affect data quality.

5 Data Pipeline

5.1 High-Level Description

The data pipeline involves several steps to prepare the data for analysis. These steps include extracting, transforming, and loading the data, as well as merging the datasets for a comprehensive analysis.

5.2 Overview of the Extract

ccc

Dataset	Task	Description
Weather Data	Extract	Downloaded dataset from Meteostat and loaded into a DataFrame.
	Load	Stored cleaned data into a CSV flie
	Transform	Converted date to datetime, aggregated daily averages, and handled missing values.
COVID-19 Data	Extract	Downloaded dataset from Kaggle and loaded into a DataFrame.
	Load	Stored data into a CSV file.
	Transform	Filtered data for Alabama and for the year 2020, summarized daily death cases, handled missing values.
Merged Data	Transform	Merged weather and COVID-19 datasets on date.
	Load	Stored merged data into a CSV file for analysis.

Table 1: Extract, Transform, and Load Tasks for Each Dataset

6 Analysis

The analysis involves the following steps:

6.1 Method

The merged dataset, containing daily records of average temperature and COVID-19 death cases in Alabama for 2020, was analyzed to determine whether temperature and the number of death cases correlate. The method included:

- Plotting the daily average temperature and COVID-19 death cases.

- Calculating the rolling average to smoothen the trends.

- Using scatter plots to visualize the correlation between temperature and death cases.

- Computing the correlation coefficient to quantify the relationship.

6.2 Results

The analysis shows that there is no strong visual correlation between daily average temperature and COVID-19 death cases when plotted on a daily basis. The scatter plot and correlation coefficient calculation indicated a weak correlation, suggesting that other factors might have a more significant impact on COVID-19 death cases.
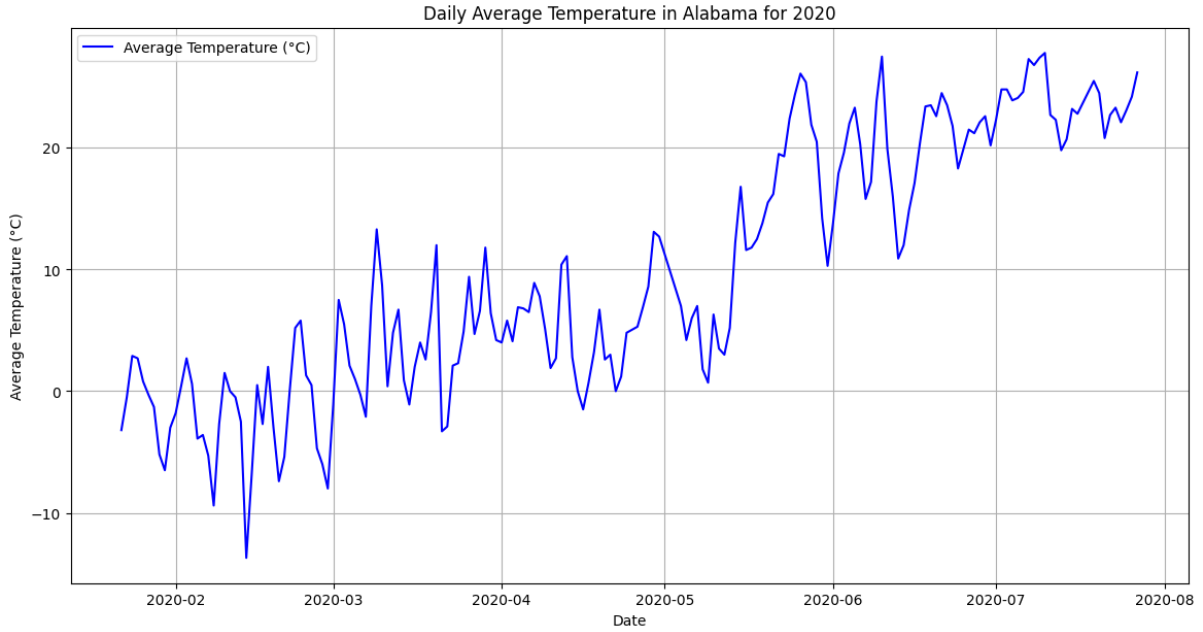
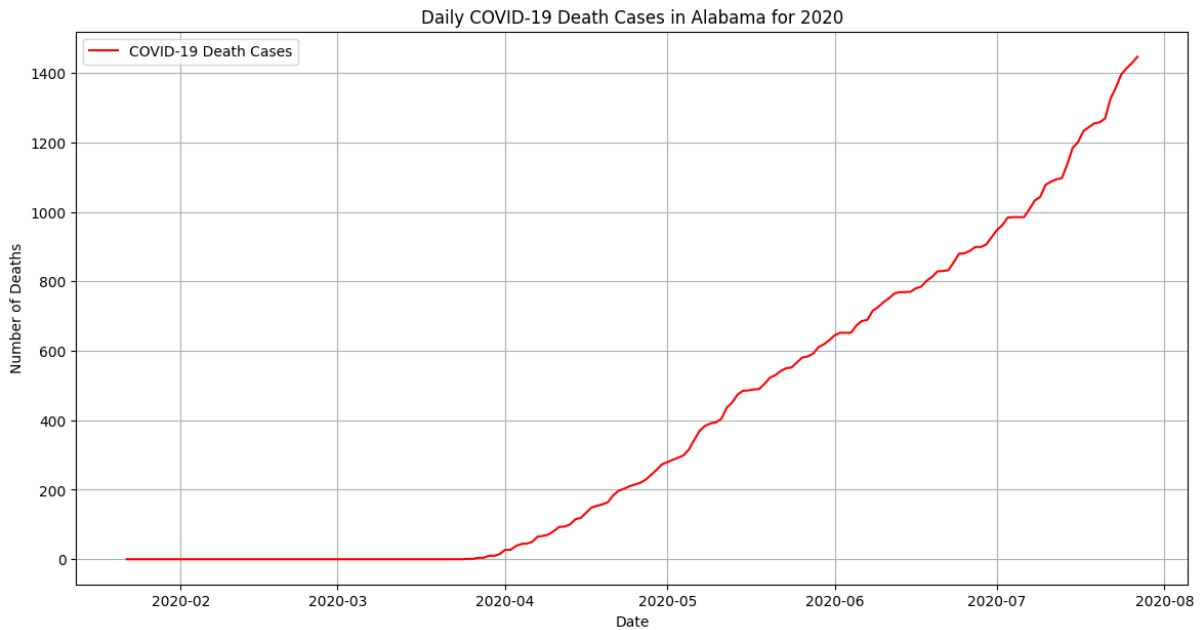Figure 1: Daily Average Temperature in Alabama for 2020

Figure 2: Daily COVID-19 Death Cases in Alabama for 2020

7 Conclusions

The analysis aimed to determine the correlation between average temperature and COVID-19 death cases in Alabama during the year 2020. The findings suggest that there is no strong correlation between the two variables. This implies that temperature alone is not a significant factor influencing the number of COVID-19 death cases.

However, it is important to consider that other factors, such as public health measures, population density, and healthcare infrastructure, may play a more critical role in determining the number of COVID-19 death cases. Further studies incorporating these factors would provide a more comprehensive understanding of the dynamics influencing COVID-19 mortality.
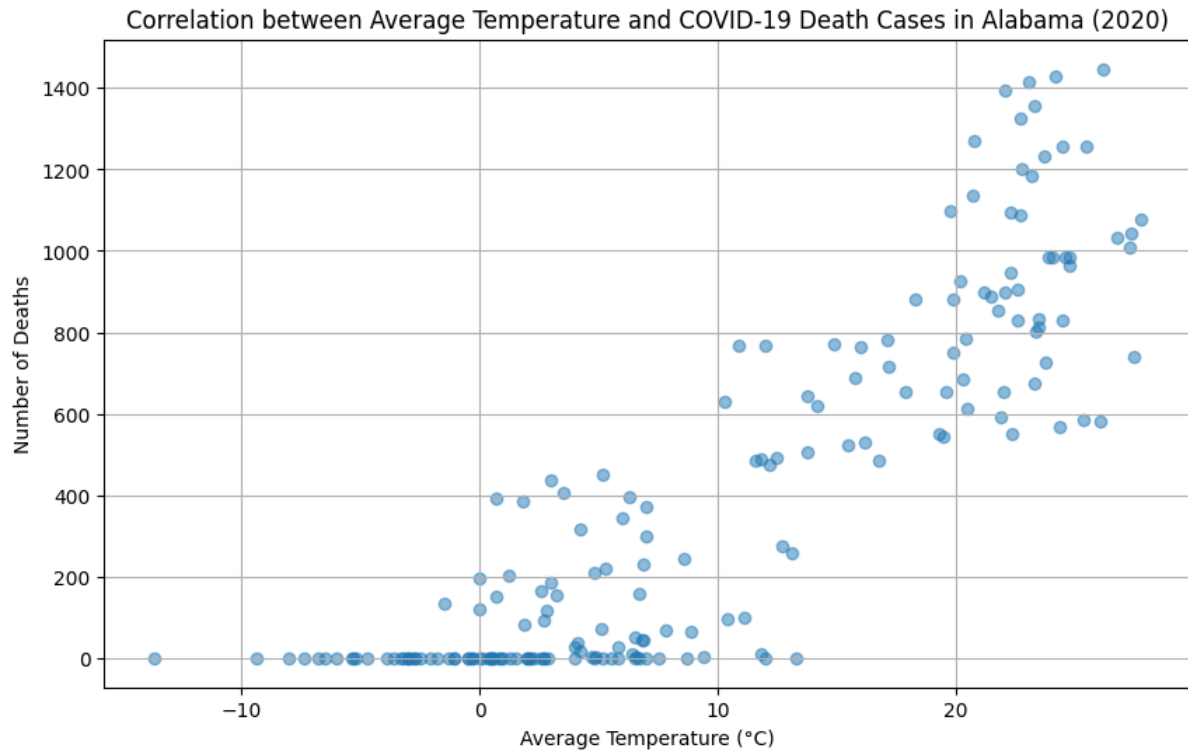
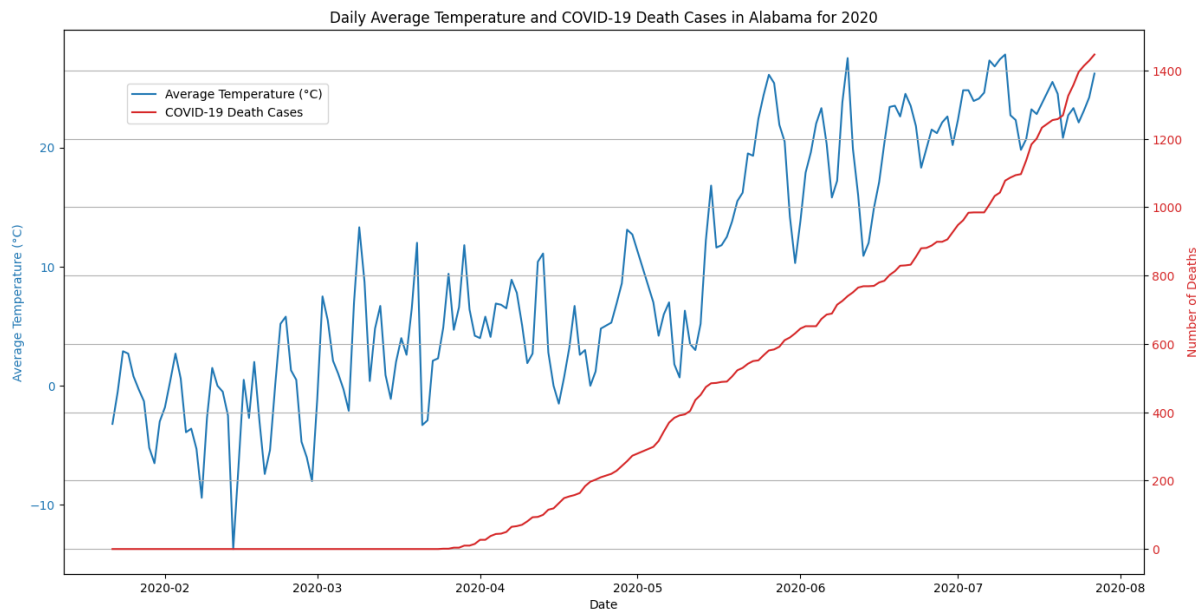Figure 3: Correlation between Average Temperature and COVID-19 Death Cases in Alabama (2020)

Figure 4: Daily Average Temperature and COVID-19 Death Cases in Alabama for 2020

7.1 Limitations

- The analysis is limited to data from Alabama for the year 2020. Results might vary with a larger dataset or a different geographic location.

- The quality and completeness of the data might affect the results. Any missing or inaccurate records could influence the analysis.

- Only temperature was considered as an environmental factor.