

Final Report

Assessing the impact of weather conditions on wildfires in Greece

Methods of Advanced Data Engineering

2024/07/03

Arvanitidis Alexis
22558160



Faculty of Engineering
Methods of Advanced Data Engineering
Friedrich-Alexander-Universität Erlangen-Nürnberg

1 Introduction

This project aims to assess the impact of weather conditions on wildfire occurrences in Greece. By analyzing historical data, we seek to identify trends in weather patterns and their correlation with the frequency and severity of wildfires. Our study focuses on key weather variables, such as temperature and wind speed, to determine their influence on wildfire events. This investigation is crucial for enhancing our understanding of how climatic factors contribute to wildfire risk, therefore supporting the development of more effective wildfire management and prevention strategies.

2 Research Questions

- “What are the trends in weather conditions in Greece and how do they correlate with wildfire occurrences?”
- “How do specific weather variables, such as temperature and wind speed, influence the frequency and severity of wildfires?”

3 Data Pipeline

To address our research question, we constructed an ETL (Extract, Transform, Load) data pipeline that generates the final dataset for analysis. This pipeline automates the extraction, processing, and storage of data from diverse sources, ensuring the data is thoroughly cleaned, properly structured, and ready for analysis. The following figure presents an overview of the Data Pipeline.

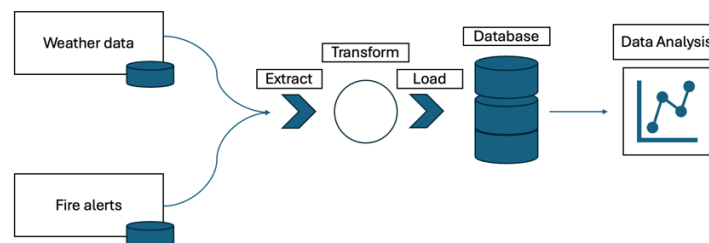


Figure 1: Data Pipeline

4 Used data

This report utilizes a merged dataset produced by an ETL data pipeline, derived from two primary sources: Kaggle and Global Forest Watch. The resulting dataset offers a comprehensive overview of weather conditions, including temperature and wind speed, alongside wildfire occurrences in Greece.

Description of Dataset


The dataset is structured as a merged CSV file and includes the following columns:

- **year & week:** The year and week in which the data was recorded.
- **avg_temp:** The average temperature for each week, measured in degrees Celsius.
- **min_temp:** The minimum temperature for each week, measured in degrees Celsius.
- **wind_speed:** The average wind speed for each week, measured in meters per second.
- **fire_alerts:** The number of wildfire alerts recorded each week.

Weather Data: This portion of the dataset includes weekly averages of various weather conditions, such as temperature and wind speed, collected for Greece.

Wildfire Occurrences: This section contains weekly counts of wildfire alerts, indicating the frequency and intensity of wildfires over the specified period.

The merged dataset spans from week 41 of 2018 to week 41 of 2022.



greece_weather_weekly						
city	temp_avg	temp_min	temp_max	wind_avg	precip_avg	week_start
Athens	11.6	6.7	17.6	5.8		1 2018
Athens	13.4	10.5	17.0	7.8		1 2018
Athens	11.2	6.9	14.3			1 2018
Athens	8.5	4.8	11.8	10.4		1 2018
Athens	10.1	5.3	16.6			1 2018
Athens	11.5	7.8	16.1	7.5		1 2018
Athens	13.6	10.5	17.5	2.7		1 2018
Athens	14.8	12.7	17.3			2 2018
Athens	13.7	11.5	16.7	2.9		2 2018
Athens	13.8	10.9	16.5	5.5		2 2018
Athens	12.7	8.9	16.1	9.7		2 2018
Athens	12.5	8.5	17.3	5.1		2 2018
Athens	11.2	6.2	15.9			2 2018
Athens	8.4	5.8	8.9	10.7		2 2018
Athens	10.5	4.9	15.8	2.8		3 2018
Athens	10.8	6.9	16.1	8.3		3 2018
Athens	12.8	8.9	16.5			3 2018
Athens	10.3	6.8	15.8	20.0	28.9	3 2018
Athens	10.1	5.1	16.7	6.3		3 2018
Athens	13.5	8.7	16.7	5.3		3 2018
Athens	13.9	8.8	16.8	8.3		3 2018
Athens	10.9		11.0			4 2018
Athens	7.7	4.5	10.8	20.0	18.4	4 2018
Athens	8.8	5.2	7.8	6.0	16.0	4 2018
Athens	5.8	3.8	8.1	10.0	16.7	4 2018
Athens	6.5	1.9	10.8	16.4		4 2018
Athens	8.2	4.1	12.8	14.7		4 2018
Athens	8.1	3.9	14.1	107.0	16.4	4 2018
Athens	10.0	5.7	14.5			5 2018
Athens	10.8	6.9	16.0	2.4		5 2018

greece_weather_weekly_aggregated						
year	week	temp_avg	temp_min	temp_max	wind_avg	precip_avg
2018	41	10.2	5.8	22.8	207.1	5.0
2018	42	10.1	6.1	23.5	165.1	5.0
2018	43	11.9	14.3	21.2	142.8	6.1
2018	44	10.4	10.5	24.1	46.7	4.5
2018	45	10.7	10.8	20.7	128.4	8.4
2018	46	12.7	11.5	16.1	16.7	8.7
2018	47	10.5	10.4	18.4	177.3	8.2
2018	48	10.9	10.1	16.6	83.3	8.4
2018	49	10.9	7.4	14.7	245.0	8.7
2018	50	11.8	8.8	14.4	133.3	8.1
2018	51	10.6	8.0	15.1	237.0	5.3
2018	52	9.9	8.0	12.2	238.6	8.1
2019	1	6.3	3.8	8.8	188.0	8.8
2019	2	7.8	4.7	13.7	238.4	6.8
2019	3	9.2	5.8	12.9	251.1	8.4
2019	4	11.8	8.8	14.7	137.7	7.5
2019	5	12.7	8.8	16.1	254.4	7.2
2019	6	10.8	8.4	12.7	132.7	8.1
2019	7	8.8	6.0	12.0	291.7	10.0
2019	8	8.7	6.0	14.4	187.2	6.8
2019	9	10.5	7.2	13.8	182.0	8.9
2019	10	14.1	10.7	18.1	244.7	5.7
2019	11	12.8	8.1	17.0	253.7	7.1
2019	12	14.3	10.8	16.0	88.7	10.8
2019	13	10.0	8.7	14.5	151.0	12.2
2019	14	14.7	10.7	18.3	67.8	8.3
2019	15	14.8	11.8	16.1	251.7	7.8
2019	16	10.5	8.8	17.9	87.1	6.8
2019	17	17.3	12.7	20.0	38.8	5.1
2019	18	14.8	12.8	22.1	251.6	8.3
2020	5	10.9	8.1	17.6	248.4	5.8

Figure 2: Data Preparation

Data Preparation Process

To prepare the dataset for analysis (Figure 2), the following steps were undertaken:

- Filtered for greece:** Included only relevant data.
- Date conversion:** Standardized date columns to datetime format.
- Weekly sggregation:** Aggregated weather data to weekly averages.
- Dropped unnecessary columns:** Removed latitude, longitude, and city columns.
- Renamed columns:** Standardized column names for clarity and consistency.

merged_weather_fire_alerts							
year	week	temp_avg	temp_min	temp_max	wind_avg	precip_avg	alert_count
2018	42	10.1	5.8	23.5	165.1	5.0	140
2018	43	11.9	14.3	21.2	142.8	6.1	1016.1
2018	44	10.4	10.5	24.1	46.7	4.5	1020.0
2018	45	10.7	10.8	20.7	128.4	8.4	1019.5
2018	46	12.7	11.5	16.1	16.7	8.7	1017.8
2018	47	10.5	10.4	18.4	177.3	8.2	1017.4
2018	48	10.9	10.1	16.6	83.3	8.4	1014.5
2018	49	10.9	7.4	14.7	245.0	8.7	1016.4
2018	50	11.8	8.8	14.4	133.3	8.1	1014.4
2018	51	10.6	8.0	15.1	237.0	5.3	1017.4
2018	52	9.9	8.0	12.2	238.6	8.1	1015.1
2019	1	6.3	3.8	8.8	188.0	8.8	1013.5
2019	2	7.8	4.7	13.7	238.4	6.8	1012.5
2019	3	9.2	5.8	12.9	251.1	8.4	1014.3
2019	4	11.8	8.8	14.7	137.7	7.5	1004.8
2019	5	12.7	8.8	16.1	254.4	7.2	1011.9
2019	6	10.8	8.4	12.7	132.7	8.1	1016.9
2019	7	8.8	6.0	12.0	291.7	10.0	1019.0
2019	8	8.7	6.0	14.4	187.2	6.8	1025.1
2019	9	10.5	7.2	13.8	182.0	8.9	1016.1
2019	10	14.1	10.7	18.1	244.7	5.7	1020.5
2019	11	12.8	8.1	17.0	253.7	7.1	1010.0
2019	12	14.3	10.8	16.0	88.7	10.8	1020.1
2019	13	10.0	8.7	14.5	151.0	12.2	1016.2
2019	14	14.7	10.7	18.3	67.8	8.3	1018.8
2019	15	14.8	11.8	16.1	251.7	7.8	1008.0
2019	16	10.5	8.8	17.9	87.1	6.8	1019.8
2019	17	17.3	12.7	20.0	38.8	5.1	1019.5
2019	18	14.8	12.8	22.1	251.6	8.3	1011.3
2020	5	10.9	8.1	17.6	248.4	5.8	1017.8
2020	6	10.9	8.1	17.6	248.4	5.8	1017.8
2020	7	10.9	8.1	17.6	248.4	5.8	1017.8
2020	8	10.9	8.1	17.6	248.4	5.8	1017.8
2020	9	10.9	8.1	17.6	248.4	5.8	1017.8
2020	10	10.9	8.1	17.6	248.4	5.8	1017.8
2020	11	10.9	8.1	17.6	248.4	5.8	1017.8
2020	12	10.9	8.1	17.6	248.4	5.8	1017.8
2020	13	10.9	8.1	17.6	248.4	5.8	1017.8
2020	14	10.9	8.1	17.6	248.4	5.8	1017.8
2020	15	10.9	8.1	17.6	248.4	5.8	1017.8
2020	16	10.9	8.1	17.6	248.4	5.8	1017.8
2020	17	10.9	8.1	17.6	248.4	5.8	1017.8
2020	18	10.9	8.1	17.6	248.4	5.8	1017.8
2020	19	10.9	8.1	17.6	248.4	5.8	1017.8
2020	20	10.9	8.1	17.6	248.4	5.8	1017.8
2020	21	10.9	8.1	17.6	248.4	5.8	1017.8
2020	22	10.9	8.1	17.6	248.4	5.8	1017.8
2020	23	10.9	8.1	17.6	248.4	5.8	1017.8
2020	24	10.9	8.1	17.6	248.4	5.8	1017.8
2020	25	10.9	8.1	17.6	248.4	5.8	1017.8
2020	26	10.9	8.1	17.6	248.4	5.8	1017.8
2020	27	10.9	8.1	17.6	248.4	5.8	1017.8
2020	28	10.9	8.1	17.6	248.4	5.8	1017.8
2020	29	10.9	8.1	17.6	248.4	5.8	1017.8
2020	30	10.9	8.1	17.6	248.4	5.8	1017.8
2020	31	10.9	8.1	17.6	248.4	5.8	1017.8
2020	32	10.9	8.1	17.6	248.4	5.8	1017.8
2020	33	10.9	8.1	17.6	248.4	5.8	1017.8
2020	34	10.9	8.1	17.6	248.4	5.8	1017.8
2020	35	10.9	8.1	17.6	248.4	5.8	1017.8
2020	36	10.9	8.1	17.6	248.4	5.8	1017.8
2020	37	10.9	8.1	17.6	248.4	5.8	1017.8
2020	38	10.9	8.1	17.6	248.4	5.8	1017.8
2020	39	10.9	8.1	17.6	248.4	5.8	1017.8
2020	40	10.9	8.1	17.6	248.4	5.8	1017.8
2020	41	10.9	8.1	17.6	248.4	5.8	1017.8
2020	42	10.9	8.1	17.6	248.4	5.8	1017.8
2020	43	10.9	8.1	17.6	248.4	5.8	1017.8
2020	44	10.9	8.1	17.6	248.4	5.8	1017.8
2020	45	10.9	8.1	17.6	248.4	5.8	1017.8
2020	46	10.9	8.1	17.6	248.4	5.8	1017.8
2020	47	10.9	8.1	17.6	248.4	5.8	1017.8
2020	48	10.9	8.1	17.6	248.4	5.8	1017.8
2020	49	10.9	8.1	17.6	248.4	5.8	1017.8
2020	50	10.9	8.1	17.6	248.4	5.8	1017.8
2020	51	10.9	8.1	17.6	248.4	5.8	1017.8
2020	52	10.9	8.1	17.6	248.4	5.8	1017.8

Figure 3: Aggregated data

These data preparation steps were necessary in transforming the raw data into a structured and clean format.

5 Data Analysis

Method Used

Addressing the research questions regarding the correlation between weather and wildfire:

1. Data Preparation:

- A scatter plot was created to visualize the relationship between average temperature and the number of wildfire alerts. Similarly, another scatter plot was generated for wind speed and wildfire alerts.
- Pearson Correlation Coefficients (PCC, r-values) were calculated to assess the strength of the correlation between the variables.

2. Visualization of Relationships:

- A 3D scatter plot was created to visualize the combined effect of temperature and wind speed on wildfire occurrences. This plot includes data points for all weeks in the dataset and helps in understanding the multi-dimensional relationship between these variables.

3. Interaction Effects Analysis:

- Interaction terms between temperature and wind speed were included in a regression model to explore how their combination affects wildfire occurrences. This model helps understanding whether the combined effect of high temperatures, wind speeds leads to more significant risks.

Results

1. Temperature and Wildfire Alerts:

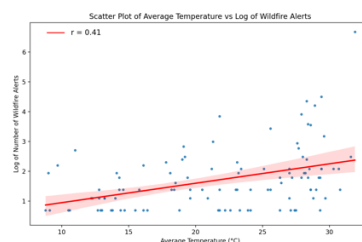


Figure 4: Correlation of average Temp & number of wildfire alerts

- The scatter plot by the regression line shows a moderate positive correlation between average temperature and the log of wildfire alerts ($r = 0.41$), indicating that higher temperatures are associated with an increase in wildfire occurrences.

○

2. Wind Speed and Wildfire Alerts:

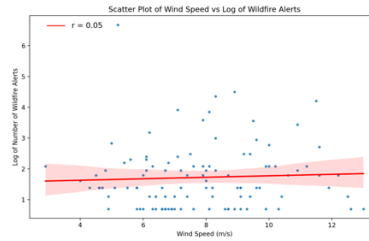


Figure 5: Correlation of wind speed & number of wildfire alerts

- The scatter plot for wind speed and log of wildfire alerts reveals a very weak positive correlation ($r = 0.05$), suggesting that wind speed has a small impact on frequency of wildfires.

3. Combined Effect of Temperature and Wind Speed:

3D Scatter Plot of Temperature, Wind Speed, and Wildfire Alerts

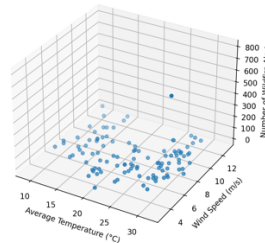


Figure 6: Interaction effects of temp & wind speed

- The 3D scatter plot indicates that regions with both higher temperatures and higher wind speeds tend to have more wildfire alerts. However, the effect of temperature is more dominant compared to wind speed.
- The regression model including the interaction term supports this observation, showing that the combined effect of temperature and wind speed significantly influences wildfire occurrences.

6 Conclusion

Q1.: “What are the trends in weather conditions in Greece and how do they correlate with wildfire occurrences?”:

By analyzing the weekly aggregated data of temperature and wind speed alongside wildfire occurrences, one could identify potential correlations. For instance, higher temperatures and increased wind speeds might correspond to higher frequencies of wildfire alerts.

Q.2: “How do specific weather variables, such as temperature and wind speed, influence the frequency and severity of wildfires?”

Increased average, minimum, and maximum temperatures likely contribute to drier conditions, which can enhance the likelihood of wildfires. Higher wind speeds can exacerbate wildfire conditions by spreading flames more rapidly and over larger areas.

7 Limitations

- a. **Geographical scope:** The weather data pertains to the capital city of Greece, while the wildfire data covers the entire country. This discrepancy can limit the comparability of the datasets.
- b. **Sample size:** The number of observations (N) is relatively low, especially after filtering and merging the datasets, which could impact the robustness of the findings.

Therefore, the research question cannot be fully answered with the current dataset. To achieve a more comprehensive and accurate answer, additional and better-suited datasets are needed.