1. Research objectives

Research questions:

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

2. Data sources

2.1 Daily weather data

Description: This dataset offers weather data for Greece, including variables like temperature, wind speed. It was filtered to cover from week 41 of 2018 to week 41 of 2022.

• Source: <u>Kaggle - Historical Weather Data of All Country Capitals</u>

• Data type: CSV

• License: CC BY-SA 3.0

- Data structure and quality: The dataset includes columns for date, country, city, latitude, longitude, avg temp., min temp., max.temp., wind direction, wind speed. The data is structured in tabular format and cleaned to removee any missing entries.
- **Licensing obligations**: The dataset is under the Creative Commons BY-SA 3.0 license, allowing for sharing and adaptation with proper attribution. The obligations will be fulfilled by crediting the original source in all publications and derived works.

2.2 Wildfire occurrences

Description: This dataset contains records of wildfire occurrences, specifically the number of fire alerts per week. It was filtered to align with the time frame of the weather data.

• Source: Global Forest Watch - VIIRS Fire Alerts

• Data type: CSV

• License: Not specified

- Data structure and quality: The dataset includes columns for year, week, count of alerts, and confidence category. The data is structured in a tabular format and cleaned to focus on high-confidence alerts for the specified time frame.
- **Licensing obligations**: The dataset does not specify a license. Usage will comply with data usage terms stated on the website, ensuring proper attribution.

3. Data pipeline

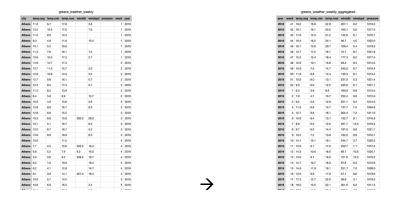
3.1 Overview

The data pipeline is designed to automate the extraction, transformation, and loading of data from the two sources to prepare it for analysis. The pipeline was implemented using Python, leveraging libraries such as Pandas for data manipulation and SQL for data storage.

3.2 Transformation and cleaning steps

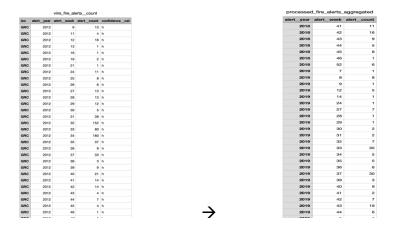
Weather data:

- Filtered to include only records from Greece
- Converted date column to datetime format
- Extracted week number and year from the date
- Aggregated data to provide weekly averages for temperature and wind speed
- Dropped unnecessary columns such as latitude, longitude and city
- Renamed columns for clarity



Wildfire data:

- Filtered to include records from 2018 to 2022 as well
- Dropped columns not relevant to the analysis (e.g., ISO code, confidence category)
- Aggregated data to provide weekly counts of fire alerts
- Renamed columns for clarity



3.3 Merging datasets

- Merged the weather data and wildfire data on the common columns 'year' and 'week'
- Ensured the merged dataset covered the same time period from week 41 of 2018 to week 41 of 2022
- The merged dataset allows for correlation analysis between weather variables and wildfire occurrences

| year | week | temp.avg | temp.min | temp.max | winddir | windspd | pressure | alert_count |
|------|------|----------|----------|----------|---------|---------|----------|-------------|
| 2018 | 41 | 19.2 | 16.6 | 22.8 | 207.1 | 9.2 | 1019.5 | - 11 |
| 2018 | 42 | 19.1 | 16.1 | 23.5 | 163.1 | 5.0 | 1017.5 | 16 |
| 2018 | 43 | 17.8 | 14.3 | 21.2 | 142.6 | 6.1 | 1016.1 | 9 |
| 2018 | 44 | 19.4 | 16.5 | 24.1 | 46.7 | 4.5 | 1022.0 | 5 |
| 2018 | 45 | 16.1 | 12.6 | 20.7 | 108.4 | 5.4 | 1019.5 | 8 |
| 2018 | 46 | 13.7 | 11.5 | 16.1 | 15.7 | 8.7 | 1021.6 | 1 |
| 2018 | 52 | 9.0 | 6.0 | 12.2 | 235.6 | 6.1 | 1021.1 | 6 |
| 2019 | 7 | 8.8 | 6.5 | 12.0 | 291.7 | 13.0 | 1019.3 | 1 |
| 2019 | 8 | 9.7 | 6.0 | 14.4 | 187.0 | 6.8 | 1021.1 | 8 |
| 2019 | 9 | 10.5 | 7.2 | 13.8 | 192.0 | 9.9 | 1016.1 | 1 |
| 2019 | 12 | 14.3 | 10.6 | 18.0 | 69.7 | 10.6 | 1020.1 | 5 |
| 2019 | 14 | 14.7 | 10.7 | 18.3 | 67.6 | 9.3 | 1012.6 | 1 |
| 2019 | 24 | 27.1 | 22.5 | 32.2 | 151.3 | 5.9 | 1012.3 | 1 |
| 2019 | 27 | 28.5 | 23.0 | 34.0 | 99.7 | 7.6 | 1012.1 | 7 |
| 2019 | 28 | 27.4 | 23.2 | 32.8 | 132.4 | 6.4 | 1008.7 | 1 |
| 2019 | 29 | 26.3 | 21.8 | 30.4 | 123.6 | 9.1 | 1010.9 | 1 |
| 2019 | 30 | 28.8 | 24.4 | 33.0 | 51.7 | 9.1 | 1011.3 | 2 |
| 2019 | 31 | 29.7 | 25.0 | 34.4 | 218.9 | 6.7 | 1007.1 | 2 |
| 2019 | 32 | 29.4 | 24.9 | 33.9 | 58.9 | 10.2 | 1011.9 | 7 |
| 2019 | 33 | 28.4 | 24.2 | 33.7 | 83.7 | 7.9 | 1009.9 | 35 |
| 2019 | 34 | 29.2 | 25.1 | 32.9 | 7.4 | 12.2 | 1014.8 | 5 |
| 2019 | 35 | 28.4 | 24.1 | 32.9 | 2.7 | 11.6 | 1013.7 | 5 |
| 2019 | 36 | 27.0 | 22.8 | 31.6 | 155.1 | 8.0 | 1012.3 | 6 |
| 2019 | 37 | 25.6 | 22.0 | 29.8 | 105.6 | 10.9 | 1016.8 | 30 |
| 2019 | 39 | 23.0 | 19.0 | 28.1 | 231.1 | 4.3 | 1014.2 | 3 |
| 2019 | 40 | 23.1 | 19.5 | 28.2 | 213.6 | 5.6 | 1010.2 | 9 |
| 2019 | 41 | 20.9 | 17.7 | 25.2 | 69.9 | 6.2 | 1018.1 | 2 |
| 2019 | 42 | 21.2 | 17.4 | 26.5 | 11.3 | 2.9 | 1018.3 | 7 |
| 2019 | 43 | 21.3 | 17.4 | 26.5 | 256.4 | 8.3 | 1018.8 | 19 |
| 2019 | 44 | 18.1 | 15.0 | 22.7 | 131.9 | 4.8 | 1016.8 | 6 |
| | | | | | | | | |

3.4 Problems and solutions

- Data gaps: Managed by using available data and documenting the missing weeks
- **Dataset disparity**: Acknowledged the difference in geographical scope and adjusted the analysis accordingly
- Sample size: Limited number of observations addressed by aggregating data

3.5 Errorhandling

The pipeline includes error handling mechanisms to manage missing data and inconsistencies. It is designed to handle changes in the input data structure by dynamically adjusting the data cleaning steps based on the presence of expected columns.

4. Result and limitations

4.1 Output data

The final output of the data pipeline is a merged dataset that combines the weather and wildfire data on a weekly basis from week 41 of 2018 to week 41 of 2022. This dataset is saved in CSV due to its compatibility and ease of analysis.

4.2 Data structure and format choice

The merged dataset includes columns for year, week, avg. temp., min. temp., max. temp., wind direction, wind speed, pressure, and count of fire alerts. The data is structured in a tabular format and cleaned to ensure consistency.

4.4 Potential issues

- **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.
- 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.

References

Daily Weather Data: <u>Kaggle Dataset</u>
Global Forest Watch: <u>VIIRS Fire Alerts</u>