Wildfire Risk Forecasting in the Iberian Peninsula: A Data-Driven Approach Using AI and Climate Data

---

Index

Introduction

Wildfires are an increasingly severe consequence of climate change, with rising global temperatures and extreme weather events exacerbating fire frequency and intensity. This project aims to analyze historical wildfire patterns in the Iberian Peninsula using climate and fire data from 2000 onwards. By studying the three days leading up to each fire, I will explore the relationships between climate conditions, weather variability, and soil properties to identify key drivers of wildfire occurrences.

While the connection between climate change and extreme weather events is well established, this study seeks to analyze localized wildfire trends within a geologically similar region under two distinct governments—Portugal and Spain. This comparison will allow for an assessment of how differences in wildfire prevention policies impact fire frequency and severity.

Through this study, I hope to generate valuable insights that can be used to support sustainability efforts, environmental protection policies, and disaster preparedness strategies. By highlighting the causes and consequences of wildfires, this research aims to contribute to better resource allocation for government agencies, climate organizations, and policymakers. Ultimately, this project seeks to shed light on the broader impacts of global warming on ecosystems and communities.

Scope

- **Geographical**: The dataset includes Portugal, Spain, and their respective archipelagos—the Azores, Madeira, Canary Islands, and Balearic Islands. Latitude and longitude boundaries have been carefully selected to ensure comprehensive coverage of wildfire and climate data across these regions.

- **Temporal:** The study covers wildfire records from 2000 to the present. While some data exists prior to 2000, technological advancements in measurement devices have significantly improved data accuracy in recent years, making earlier records less reliable.

Project Approach

This project follows the standard framework for data science, whose aim is to analyse and predict wildfire trends in the Iberian Peninsula. The approach consists of collecting historical climate and fire data, carrying out Exploratory Data Analysis to identify patterns and trends, and applying Machine Learning Models to predict the likelihood of a fire occurring based on the weather parameters but also to predict the potential intensity of that fire. I will also aim to explore key government standings, procedures and policies within each country in order to pinpoint areas of improvement, supported by the results of this project.

Main Objectives:

- Trend Exposure
- Statistical and Comparison analysis between Countries
- Isolate parameters with the highest correlation with wildfires for targeted interventions.
- ML: Fire occurrence prediction (Yes/No)
- ML: Fire Intensity forecasting (severity estimation)

Methodology Summary

- Source Data
- Transform Copernicus Raw files into dataframes and save them as .parquet files due to size (additional steps needed for .grib files)
- Pre-Process both Datasets
- Enhance FIRMS data with Copernicus Climate Data
- Filter out any records that have the same Latitude and Longitude as Industrial properties.
- Use specific tools to handle the high level and heavy data files in this project. Such as:
    - Make use of Dask Python Library instead of pandas (ideal for data that ranges from Gb to TB)
    - Optimize repetitive tasks with functions
    - Partition processes where possible (mainly done for the API call)
    - Delete variables that store large volumes of data once the script doesn't need them anymore.
    - Make use of gc.collect to free further RAM once done with a specific RAM heavy process.
- Carry out Exploratory Analysis
    - Data Distribution
    - Boxplots to confirm outliers
    - Analyse Fire Count over Time
        - Include Portugal and Spain separate breakdowns
    - Analyse Fire Density per Country Area
    - Average Fire Radiative Power and Brightness Temperature per Country
    - Confidence Levels Distribution
    - Iberian FIRMS Fire Types Distribution over Time
        - Include Portugal and Spain Separate breakdowns
    - Correlation between Fire Radiative Power and Brightness Correlation
    - Top 30 Fire-Prone Areas Per Country (Vegetation Fires & Confidence ≥ 30)
    - Correlation between Climate Variables and Fire Intensity Variables
    - Temporal Climate Patterns Leading to Fire Events

Data Validation Protocol

Structural Validation

- **Schema Consistency:** Verify that all expected columns exist and have the correct data types.
- **Missing Values**: Identify and handle null values in key fields such as acq_date, latitude, longitude, Climate_Date, and climate variables.
- **Duplicate Records:** Remove exact duplicate rows, ensuring each fire event is unique.
- Time Continuity: Ensure Climate_Date properly expands to cover the required 3-day interval per fire event.

Geospatial Validation

- **Latitude & Longitude Range:** Confirm that fire occurrences fall within valid geographical boundaries (e.g., Iberian Peninsula). Will add “Country Column as part of this check.
- **Rounding Consistency:** FIRMS Latitude & Longitude to match Copernicus grid resolution.

Temporal Validation

- **Date Format Consistency:** Ensure acq_date and Climate_Date are properly formatted as YYYY-MM-DD.
- **Correct Time Expansions:** Confirm that each fire event expands correctly to include acq_date and the previous 3 days.

Climate Data Validation

- **Physical Range Checks:**
  - Air temperature (t2m) between -50°C and 60°C.
  - Wind speeds (u10, v10) between 0 m/s and 50 m/s.
  - Precipitation (tp) non-negative.
- **Anomaly Detection:** Identify and flag extreme outliers using statistical thresholds (e.g., Z-score, IQR).
- **NaN Handling**: Determine appropriate handling (drop, interpolate, forward fill) for missing climate values.

Data Collection

Data Sources

This study uses two primary data sources for fire occurrence and environmental conditions:

- **Fire Information for Resource Management System (FIRMS)**
    - Source: NASA Earth Observing System Data and Information System (EOSDIS).
    - Description: Provides near real-time and archived active fire locations detected by MODIS and VIIRS satellites.
    - Access: https://firms.modaps.eosdis.nasa.gov/download/
    - Citation: *NASA Earth Observing System Data and Information System (EOSDIS), 2024.*
- **ERA5-Land Reanalysis Dataset (Copernicus Climate Data Store)**
    - Source: Copernicus Climate Change Service (C3S).
    - Description: High-resolution meteorological data, including temperature, precipitation, wind speed, and soil moisture.
    - Documentation: https://cds.climate.copernicus.eu/datasets/reanalysis-era5-land?tab=overview
    - API Documentation:
        - https://cds.climate.copernicus.eu/how-to-api/
    - Citation: *Muñoz-Sabater, J. (2019). ERA5-Land hourly data from 1950 to present. DOI: 10.24381/cds.e2161bac.*
- **Overpass Turbo (OpenStreetMap Data Query Tool)**
    - Source: OpenStreetMap (OSM)
    - Description: A web-based data extraction tool that allows users to query and visualize geospatial data from OpenStreetMap using Overpass API. It enables retrieval of features such as roads, buildings, waterways, and points of interest.
    - Documentation: https://overpass-turbo.eu/
    - Citation: *Overpass API, OpenStreetMap contributors (2024). Retrieved from https://wiki.openstreetmap.org/wiki/Overpass_API.*

Datasets & Variables

Fire Information for Resource Management System (FIRMS)

Source Method: Portal Download, free

Format: .JSON

Within the FIRMS database there are multiple datasets available through direct download from the website. The main differences between these datasets are the satellite that captured the data, and the time span available. In order to ensure the wildfire data was as comprehensive as possible, I sourced three datasets:

- **fire_archive_M-C61_571821.json:** Historical fire data for Portugal (2000–2024)
- **fire_archive_M-C61_571826.json:** Historical fire data for Spain (2000–2024)
- **fire_archive_SV-C2_571830.json:** Additional historical fire detections (Portugal & Spain)
- **fire_nrt_M-C61_571826.json:** Near real-time fire data for recent fire detections (validation)

The data in these files is either captured by MODIS (Moderate Resolution Imaging Spectroradiometer - M-C61) or VIIRS (Visible Infrared Imaging Radiometer Suite - SV-C2).

Resolutions

- MODIS (Moderate Resolution Imaging Spectroradiometer): Lower spatial resolution (1 km).
- VIIRS (Visible Infrared Imaging Radiometer Suite): Higher spatial resolution (375 m) and includes additional thermal and reflectance bands.

Example Download process:

Id	Source	Area of Interest	Request Date	Status	Delete
571830	SUOMI VIIRS C2	2000-11-01 : 2025-01-28 Spain	2025-01-28 20:59:17	**Processed** on 2025-01-28 21:08:44 **Download**	
571829	J2 VIIRS C1	2000-11-01 : 2025-01-28 Spain	2025-01-28 20:59:17	**Processed** on 2025-01-28 21:08:11 **Download**	
571828	J1 VIIRS C1	2000-11-01 : 2025-01-28 Spain	2025-01-28 20:59:17	**Processed** on 2025-01-28 21:08:29 **Download**	

Key Variables

Dataset Variable	Summary
Latitude	Center of 1 km fire pixel, but not necessarily the actual location of the fire as one or more fires can be detected within the 1 km pixel.
Longitude	Center of 1 km fire pixel, but not necessarily the actual location of the fire as one or more fires can be detected within the 1 km pixel.
brightness	Brightness temperature in Kelvin (K) detected at the fire location (varies by satellite).
scan	The size of the area scanned in the X-direction (in km).
track	The size of the area scanned in the Y-direction (in km).
acq_date	Date of the fire detection (YYYY-MM-DD).
acq_time	Time of the fire detection in UTC (HHMM).
satellite	The satellite that detected the fire (e.g., "A" for Aqua, "T" for Terra, "N" for NOAA-20).
confidence	Confidence level in the fire detection (Low, Nominal, High for MODIS; 0-100% for VIIRS).
version	The FIRMS dataset version.
bright_t31	Brightness temperature at 31 µm (Kelvin), used for atmospheric correction.
frp (Fire Radiative Power)	The energy released by the fire in megawatts (MW), indicating fire intensity.
Type	Inferred hot spot type: 0 = presumed vegetation fire 1 = active volcano 2 = other static land source 3 = offshore
daynight	Whether the detection occurred during the day or night. D= Daytime fire, N= Nighttime fire

ERA5-Land Reanalysis Dataset (Copernicus Climate Data Store)

Source Method: Creation of Account, API with Token, free

Format: .GRIB file zipped

The ERA5-Land dataset is extremely large and contains vast amounts of data.

The data is captured via satellite and is available to the public via Portal download or API call. In the webpage for the dataset, you set your parameters (variables, location).

For this project, I have selected all available variables aside from Lake specific data and Snow specific data as they are not pertinent for the geographical location I am analysing.

Accessing API

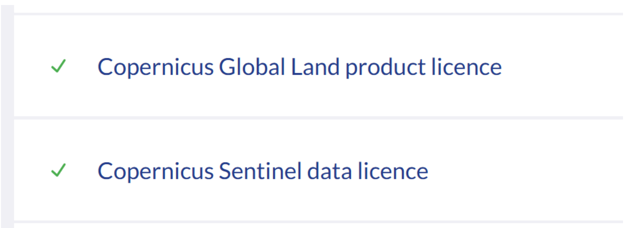
In order to access the API, you create a free account, which will automatically assign an API token. The API access is free, and the limit is 10,000 requests per month, however a single month is considered 1 request, and up to 4 concurrent calls are allowed.

Once the account is setup, follow the next link for a detailed how-to guide to setup the rest of the requirements.

Note: When using the API, before you try to make any requests, make sure to:

- Go to your profile
- Licenses Tab
- Agree to any relevant Licenses for ERA5-Land Reanalysis Dataset.

When getting the script with the parameters directly from the portal, the user is prompted to agree to licenses, however, I found that ticking that box alone doesn't cover everything, and the user still has to manually find the licenses and agree to them under the tab in their profile:



Key Variables

Variable	Description	Explanation
skt	Skin Temperature	Temperature of the land surface, measured in Kelvin (K)

stl1	Soil Temperature Layer 1	Temperature of the top soil layer (~0-7 cm depth)
stl2	Soil Temperature Layer 2	Temperature of the second soil layer (~7-28 cm depth)
stl3	Soil Temperature Layer 3	Temperature of the third soil layer (~28-100 cm depth)
stl4	Soil Temperature Layer 4	Temperature of the deepest soil layer (~100-289 cm depth)
swvl1	Soil Moisture Layer 1	Water content in the top soil layer (~0-7 cm), in m^3/m^3
swvl2	Soil Moisture Layer 2	Water content in the second soil layer (~7-28 cm), in m^3/m^3
swvl3	Soil Moisture Layer 3	Water content in the third soil layer (~28-100 cm), in m^3/m^3
swvl4	Soil Moisture Layer 4	Water content in the deepest soil layer (~100-289 cm), in m^3/m^3
u10	Wind U Component 10m	Eastward (horizontal) wind speed at 10 meters above ground, in m/s
v10	Wind V Component 10m	Northward (vertical) wind speed at 10 meters above ground, in m/s
d2m	Dew Point Temperature 2m	Temperature at which air reaches saturation at 2m height, in °C
t2m	Air Temperature 2m	Temperature of the air at 2 meters above ground, in °C
e	Evaporation	Total amount of water evaporated from the surface, in meters
evavt	Evaporation Vegetation	Evaporation due to transpiration from vegetation, in meters
lai_hv	Leaf Area Index High Vegetation	Leaf surface area per unit ground area for high vegetation
lai_lv	Leaf Area Index Low Vegetation	Leaf surface area per unit ground area for low vegetation
pev	Potential Evaporation	Maximum possible evaporation under current conditions, in meters
sp	Surface Pressure	Atmospheric pressure at the surface, in Pascals (Pa)

sro	Surface Runoff	Water runoff from land surface due to excess precipitation, in meters
sshf	Surface Sensible Heat Flux	Heat exchange between land surface and atmosphere, in W/m²
ssr	Surface Solar Radiation	Total solar energy received at the surface, in Joules/m²
ssrd	Surface Solar Radiation Downward	Downward shortwave solar radiation at the surface, in W/m²
tp	Total Precipitation	Total accumulated precipitation (rain & snow), in meters

Data Cleaning

Copernicus ERA5_Land Dataset

- Check for Nulls: none were found
- Verify Data Types: Data types were consistent and no conversions were needed.
- Unnecessary Columns: No Columns were deemed unnecessary at this stage - to be refined once correlations are calculated
- Unnecessary Rows: Upon further analysis, we should only keep the rows with readings for 12h and 17h, as they bring the most relevant insights for the scale of this project.
- Rename columns for clarity - column “time” was renamed to “date” and column “step” was renamed to “time” as this was more reflective of the data.
- Extract time from renamed date column
- Forward fill any Latitude, Longitude and Dates, as some spaces were identified in sequential intervals.
- Round Latitude and Longitude to 1 decimal point.

FIRMS Dataset

- Check for Nulls: none were found
- Check for Duplicates: 1 duplicate was found and removed
- Verify Data Types: Several columns were found to be datatype “object”. I carried out further investigation and found the unique values stored as objects - none showed different data formats other than the “Confidence” column, as it had numeric and string values. I converted all “object” columns to “string” to avoid compatibility issues.
- Later in the process, converted string Confidence Level classifications (“l”, “n”, “h”) to numeric references (0.0, 50.0, 100.0)

Overpass Turbo Dataset (Industrial Locations)

- Extract rows that include keywords related to industrial properties
- Extract coordinates from “geometry” column
- Expand coordinates to store in a long format for reference later.

Data Pre-Processing

Copernicus ERA5_Land Dataset

Convert measures:

- - Kelvin to Celsius
- - Wind Speed from M/s to KM/s
- - Pressure from Pascals to hPa
- - Precipitation & evaporation from meters to mm

FIRMS Dataset

In order to enhance the FIRMS dataset, I need to determine the matching criteria. Upon investigation, merging Copernicus data with FIRMS data will need to be done on the combination of Latitude, Longitude, Date and Time

The reason for time to be included is because the climate data provides us with readings at timed intervals.

It will be easier to merge both datasets, if the FIRMS file has the 3 day period dates and time steps created beforehand.

Pre-Processing FIRMS data for Climate Data Enhancement:

- Latitude_1 and Longitude_1 were created with the Latitude and Longitude values rounded to 1 decimal place as that is the format that matches Copernicus Data.
- Expand the dataset so that an additional column called “Climate_Date” is created, and each fire incident row has 3 additional rows appended, which display the day of the fire, and the prior three days. For example:

  For Fire Date “2024-08-15”:

acq_date (Fire Date)	Climate_Date
2024-08-15	2024-08-12
2024-08-15	2024-08-13

2024-08-15	2024-08-14
2024-08-15	2024-08-15

- Expand the dataset even further to include the time steps from the climate data.

  For Fire Date “2024-08-15”:

acq_date (Fire Date)	Climate_Date	Climate_Time
2024-08-15	2024-08-12	12.0
2024-08-15	2024-08-13	12.0
2024-08-15	2024-08-14	12.0
2024-08-15	2024-08-15	12.0
2024-08-15	2024-08-12	17.0
2024-08-15	2024-08-13	17.0
2024-08-15	2024-08-14	17.0
2024-08-15	2024-08-15	17.0

- As it can be observed by this example, the dataset is exponentially increasing, so I have made sure to utilize tools such as dask (optimal for large data being used in local environments)

Merge FIRMS data with Copernicus (Climate) Data

Once the FIRMS dataset has been prepared, the datasets are merged with a “left” method in order to keep all of the FIRMS records. Matching Criteria:

- Latitude
- Longitude
- Climate Date
- Climate Time

Data integrity checks are carried out after the merge to understand the result:

- Total Number of Rows
- Total Number of rows that are completely null
- A preview of a specific Fire event, to ensure that data has appended correctly, including the prior 3 days and the time intervals.
- Earliest and Latest Dates available in the dataset
- Number of Unique Fires
- Number of rows with any null values

- Percentage of null values within the dataset
- Preview of unique rows with NaN values

Filter Industrial Locations from Dataset

At this stage the code resets Dask to clear the memory.

With the refresh, it's the best time to filter the dataset. Using the file saved down from cleaning the Overpass Turbo dataset, I filter the records, and check the row count before and after.

The filter is only applied to records where the Fire Type is 0 (Presumed Vegetation Fire).

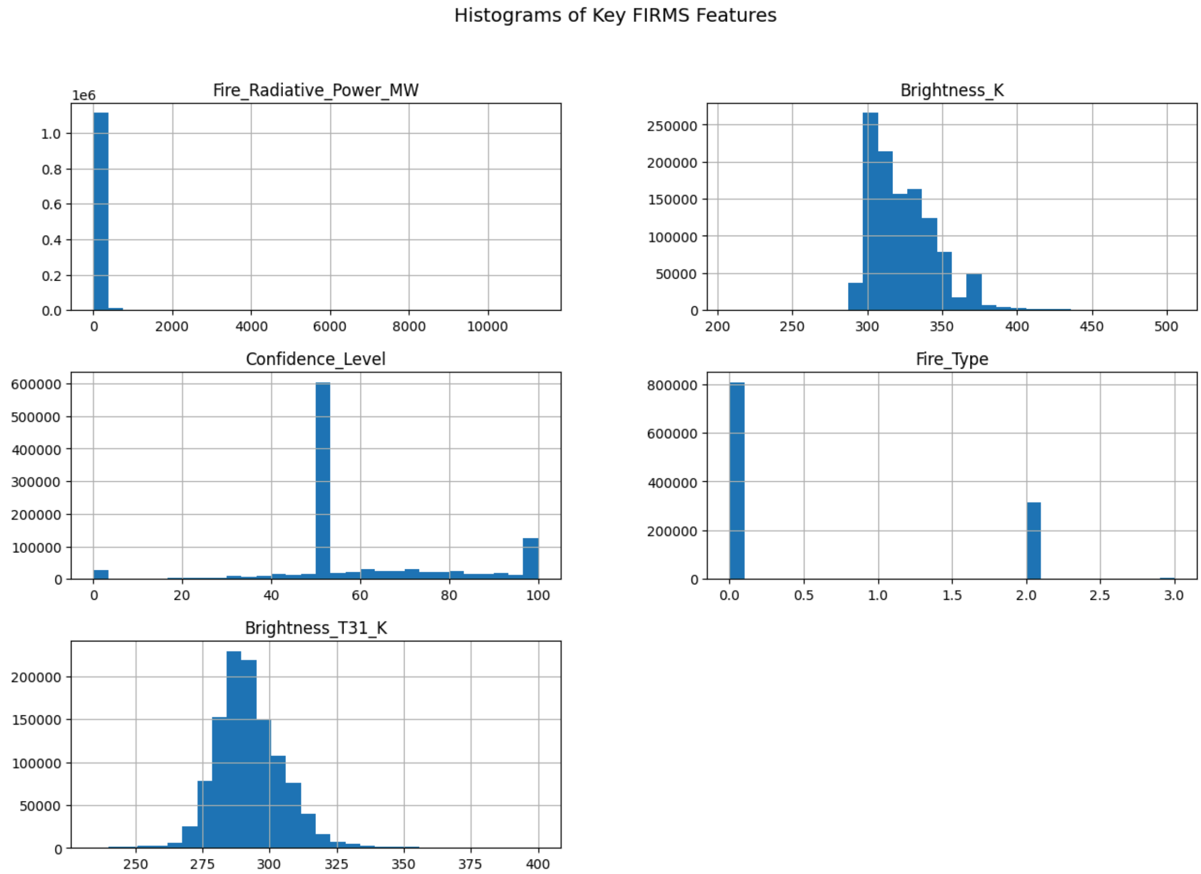
Split the Merged Dataset

Split the Merged Dataset into two variables, one that stores FIRMS data only, and one variable that stores the merged variables, but removes any null rows.

I split these datasets so we can use them for different analyses and reduce the memory required when running.

Exploratory Data Analysis (EDA)

Distribution and Outliers - FIRMS Data Only



Fire Radiative Power (MW)

- Most of the data points are concentrated on the lower end of the scale
- This suggests that the majority of detected fires are of relatively low power, with a small number showing extreme energy levels - something to explore at a later stage.

Brightness (K)

- The brightest data points are mostly between 300K and 350K, with the peak at around 300k and showing a downward slope.
- Only a few fire records have a brightness above 400K, which suggests that extremely intense fires are rare.

Confidence Level Distribution

- There is a large spike around 60% and 100%, which indicates that many fire detections are medium or high confidence level - this indicates that the data appears reliable.

Fire Type Distribution

- The vast majority of recorded fires belong to type 0 (Presumed Vegetation Fires)
- There is another cluster at fire type 2, which is the Other Static Land Source, understandable as covering a wide area - might capture city fires, agricultural burnings, etc.

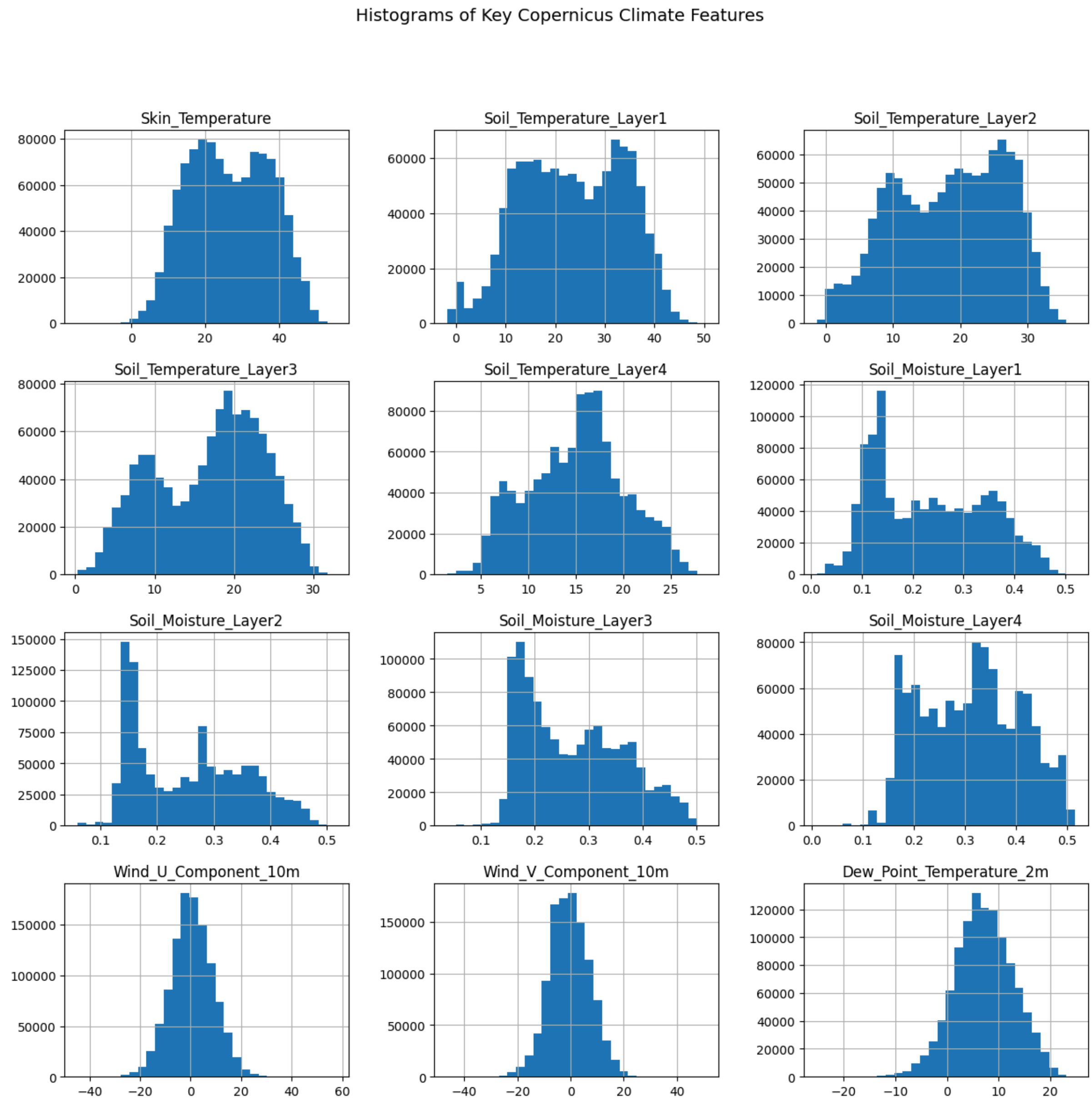
Brightness T31 (K) Distribution

- This follows a normal distribution centered around 300K, similar to the Brightness_K.
- The spread is slightly less skewed compared to Fire Radiative Power, and similar to Brightness_K

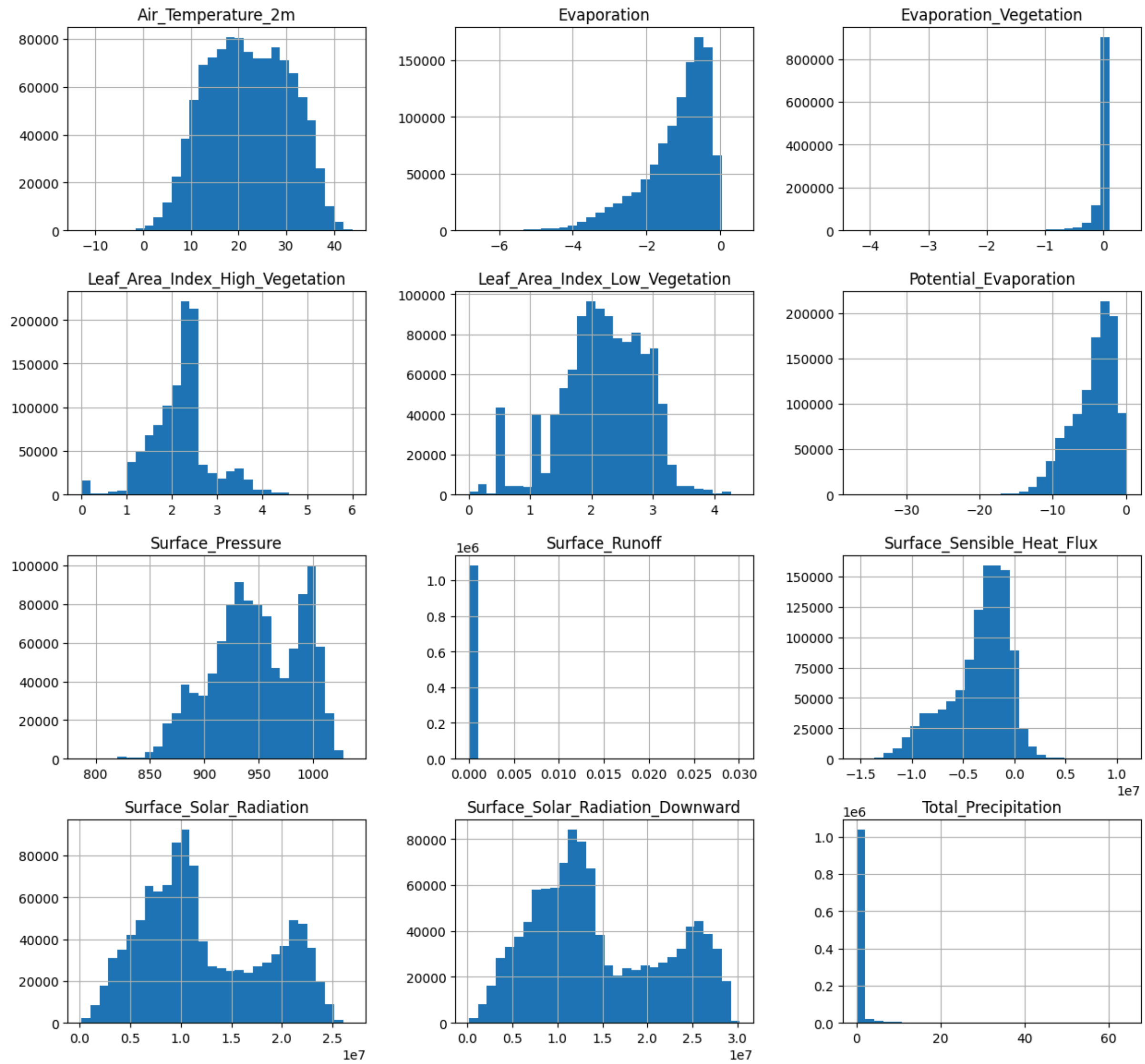
Summary

- Most fires detected are of low radiative power.
- The majority of fires fall within the Brightness range of 300K and 350K.
- Confidence in detection is mostly high or moderate, indicating a potentially trustworthy dataset.
- Fire types appear to be dominated by two categories.
- No outliers detected at this stage.

Distribution and Outliers - Climate Variables



Histograms of Key Copernicus Climate Features



Skin Temperature & Air Temperature (2m)

- Follows a normal distribution, peaking around 20-30°C, suggesting a wide range of temperatures but warm conditions.

Soil Temperature Layers (1-4)

- The deeper soil layers tend to have lower temperature, while the upper layers show broader distributions. This suggests that surface soil temperatures fluctuate more due to atmospheric conditions.

Soil Moisture (Layers 1-4)

- The distributions show high concentrations at lower moisture levels, particularly in layers 1 and 2. This might indicate that the region had very low soil moisture. Exploring regional variations might be something to explore in the future.

Wind U & V Components at 10m

- These follow a normal distribution centered around zero, which is expected, as wind speeds fluctuate a lot and have high variation.

Dew Point Temperature at 2m

- Shows a normal distribution peaking around 5-15°C, aligned with humidity and condensation points in temperate climates.

Evaporation & Potential Evaporation

- These distributions are skewed towards zero, indicating that evaporation was quite low.
- There are some cases of negative values, possibly due to seasonal variation (dry vs. rainy periods).

Surface Runoff

- Shows a large concentration of near-zero values, meaning the covered area was experiencing very low runoff.

Leaf Area Index (High & Low Vegetation)

- The high vegetation index has a strong presence at around 3-4, indicating significant forested or highly vegetated areas, in greater concentration than low vegetation index, indicating that the areas covered have a bigger presence of high level vegetation (trees rather than bushes).

Surface Pressure:

- Shows a bimodal distribution, likely due to seasonal or geographic variations (high-pressure vs. low-pressure zones).

Solar Radiation (Total & Downward)

- These two distributions are very similar, peaking at mid-to-high values, indicating significant solar energy reaching the surface, which is relevant for fire risk analysis, and somewhat expected due to the region we are exploring.

Total Precipitation

- The large spike at near-zero values suggests this region was experiencing low precipitation when fire events were recorded.

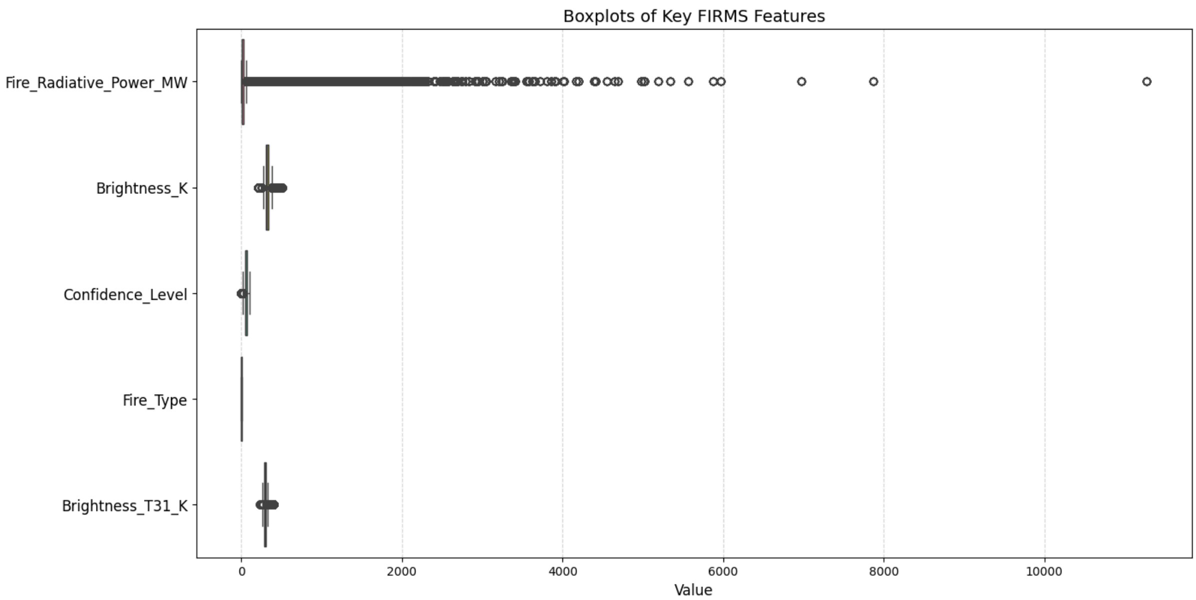
Surface Sensible Heat Flux

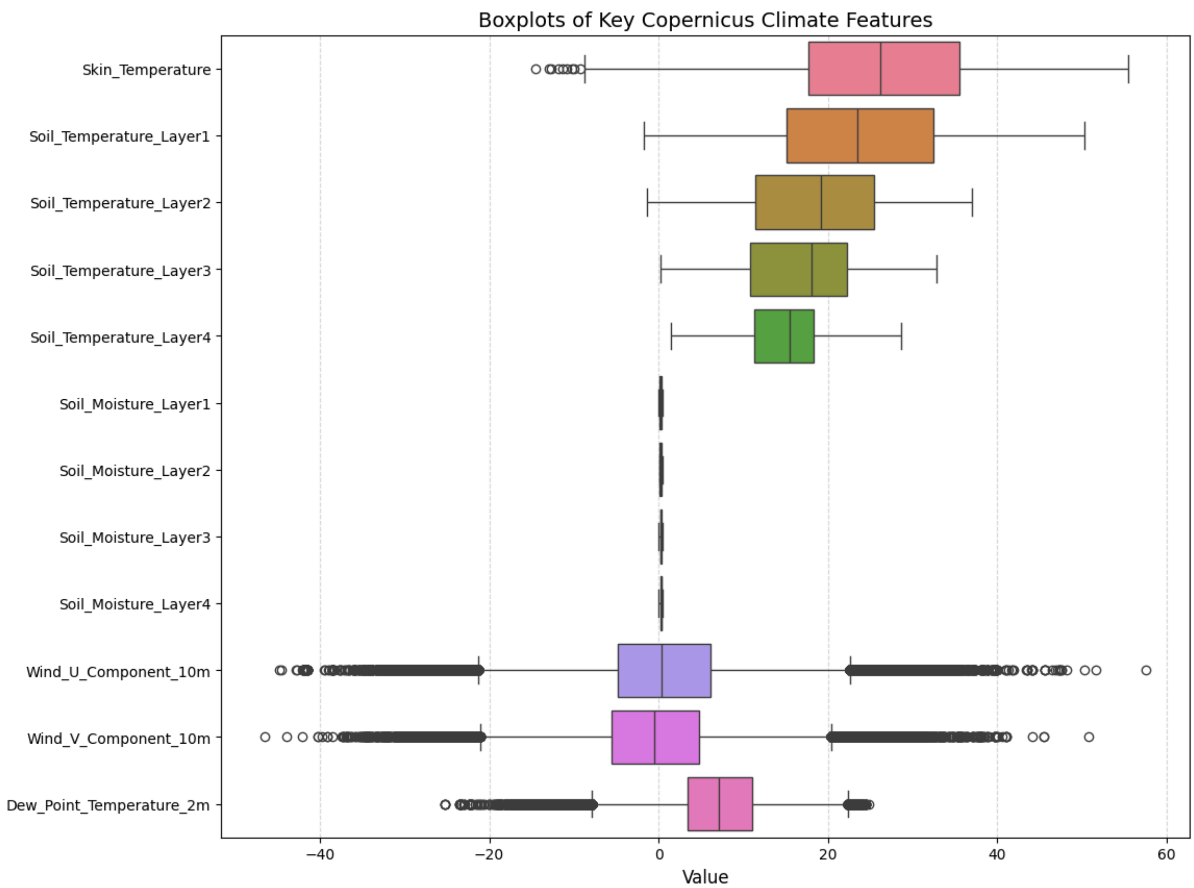
- The majority of the distribution is below zero, indicating that the land surface was losing heat to the atmosphere at the time of most recorded fire events.

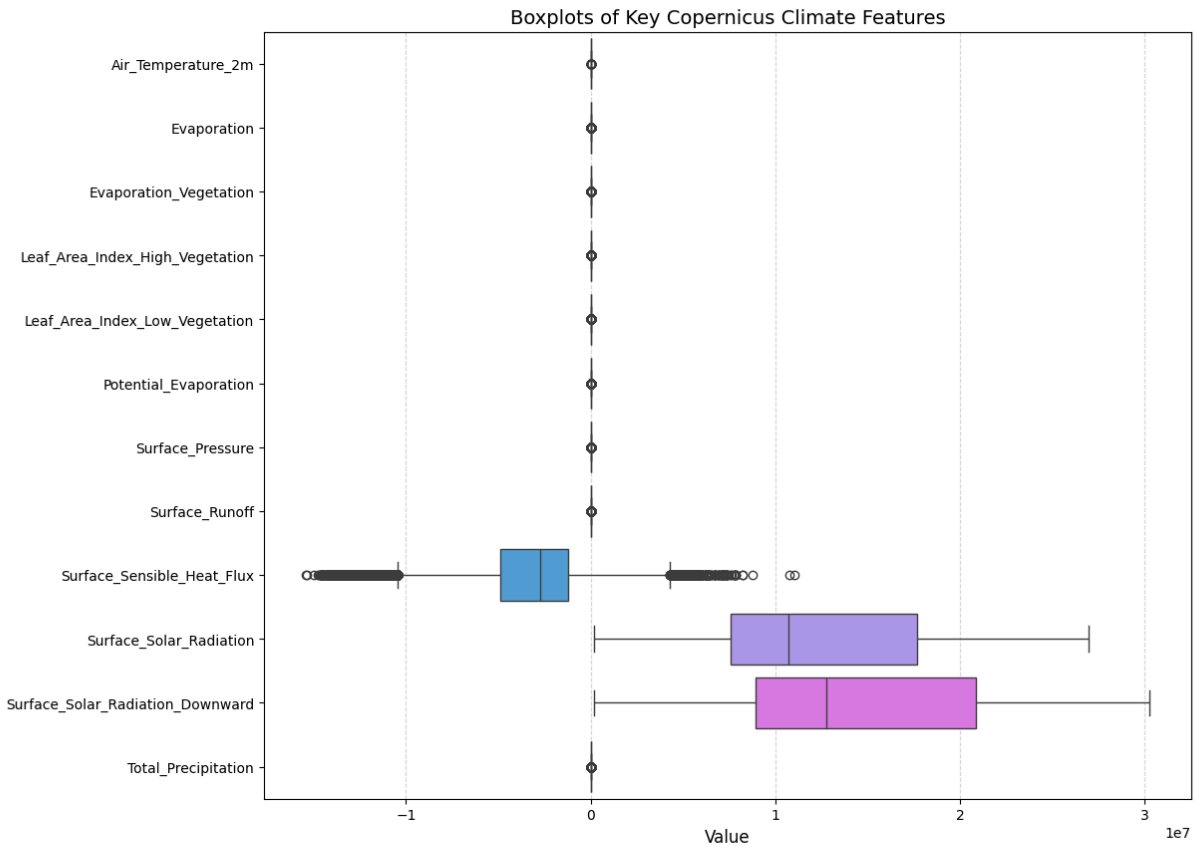
Summary

- Soil moisture and temperature patterns suggest that surface layers fluctuate more, while deeper layers are more stable.
- Wind and pressure behave as expected, with regional variations in pressure.
- Evaporation and runoff data show that water retention was low in most areas at the time of the recorded fire event.
- Solar radiation levels are high, reinforcing the importance of considering fire risks in dry areas.
- Leaf Area Index trends suggest a mix of dense and sparse vegetation zones.
- Precipitation and heat flux variability indicate that these regions are prone to extreme weather events.

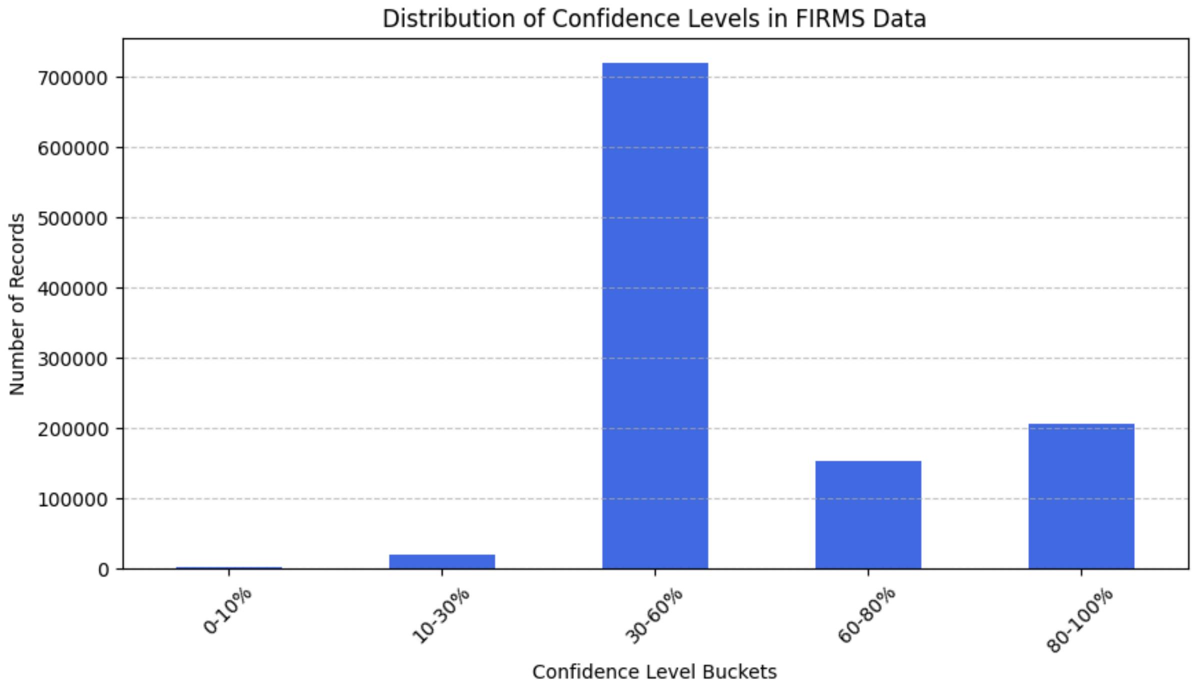
Boxplots to reinforce the findings:




Boxplots of Key Copernicus Climate Features
Skin_Temperature
Soil_Temperature_Layer1
Soil_Temperature_Layer2
Soil_Temperature_Layer3
Soil_Temperature_Layer4
Soil_Moisture_Layer1
Soil_Moisture_Layer2
Soil_Moisture_Layer3
Soil_Moisture_Layer4
Wind_U_Component_10m
Wind_V_Component_10m
Dew_Point_Temperature_2m
−40
−20
0
20
40
60
Value

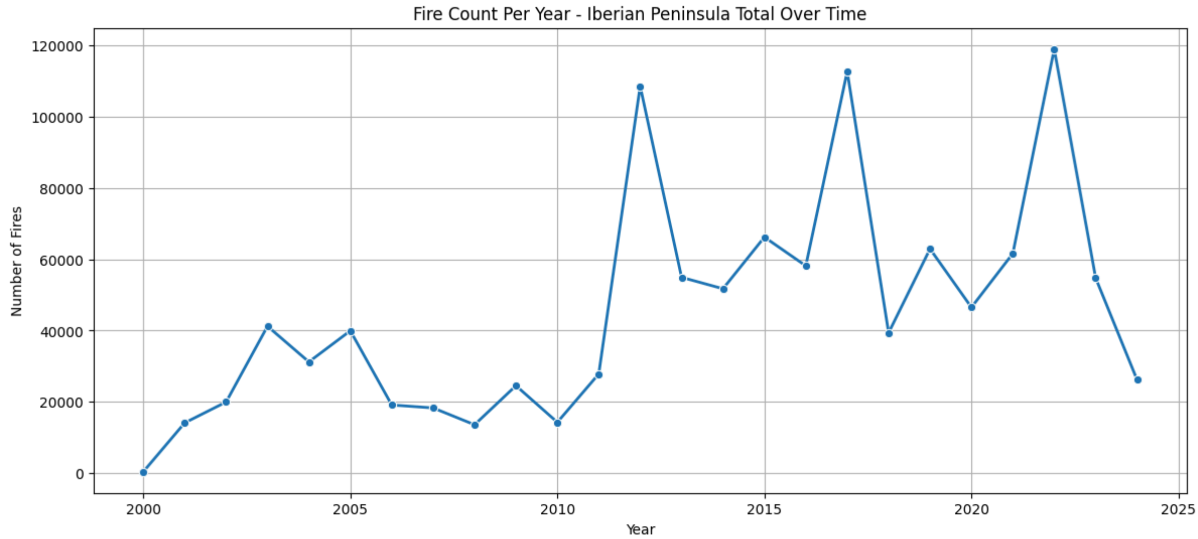

Boxplots of Key Copernicus Climate Features
Air_Temperature_2m
Evaporation
Evaporation_Vegetation
Leaf_Area_Index_High_Vegetation
Leaf_Area_Index_Low_Vegetation
Potential_Evaporation
Surface_Pressure
Surface_Runoff
Surface_Sensible_Heat_Flux
Surface_Solar_Radiation
Surface_Solar_Radiation_Downward
Total_Precipitation
−1
0
1
2
3
1e7
Value

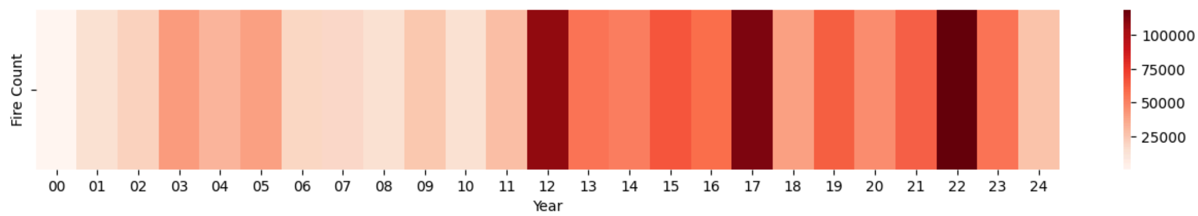
Confidence Levels Distribution - A more detailed view

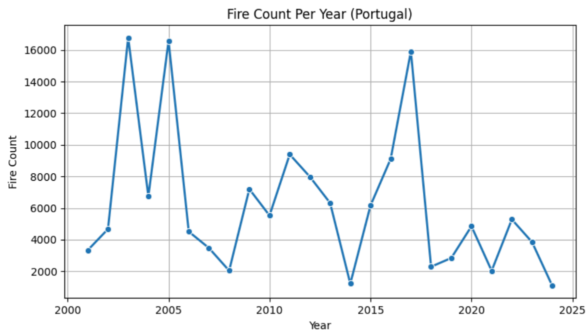


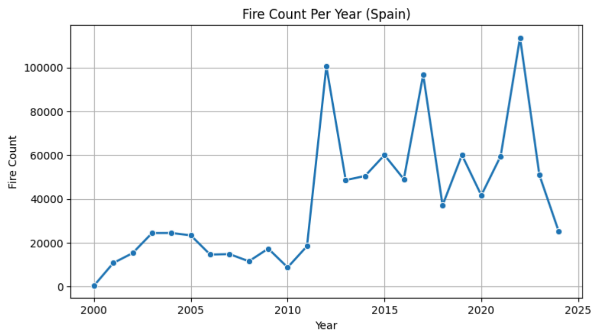
- Most fire detections fall in the 30-60% confidence range, meaning they are moderately reliable but may need additional verification.
- High-confidence (80-100%) fires exist but are fewer, indicating confirmed fire events.
- Low-confidence (0-10%) detections are minimal.
- A significant number of fires (60-80%) are likely real but will need validation

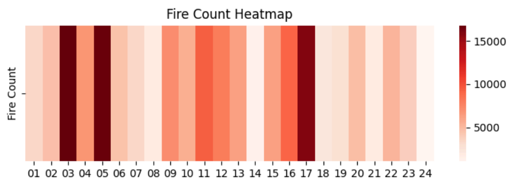
Iberian FIRMS Fire Count over the Years

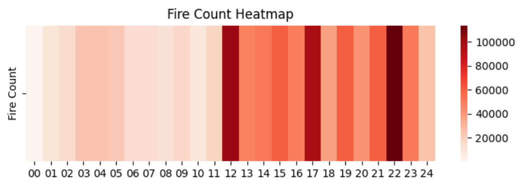












- The fire count fluctuates significantly year by year, with some years experiencing higher peaks.
- The highest peaks in fire count occurred around 2012, 2017, and 2022, indicating particularly bad fire seasons.
- The lowest fire count levels were observed in the early 2000s, suggesting that fire frequency has become more extreme in recent years.
- The heatmap confirms these trends, highlighting periods of more intense wildfire activity.

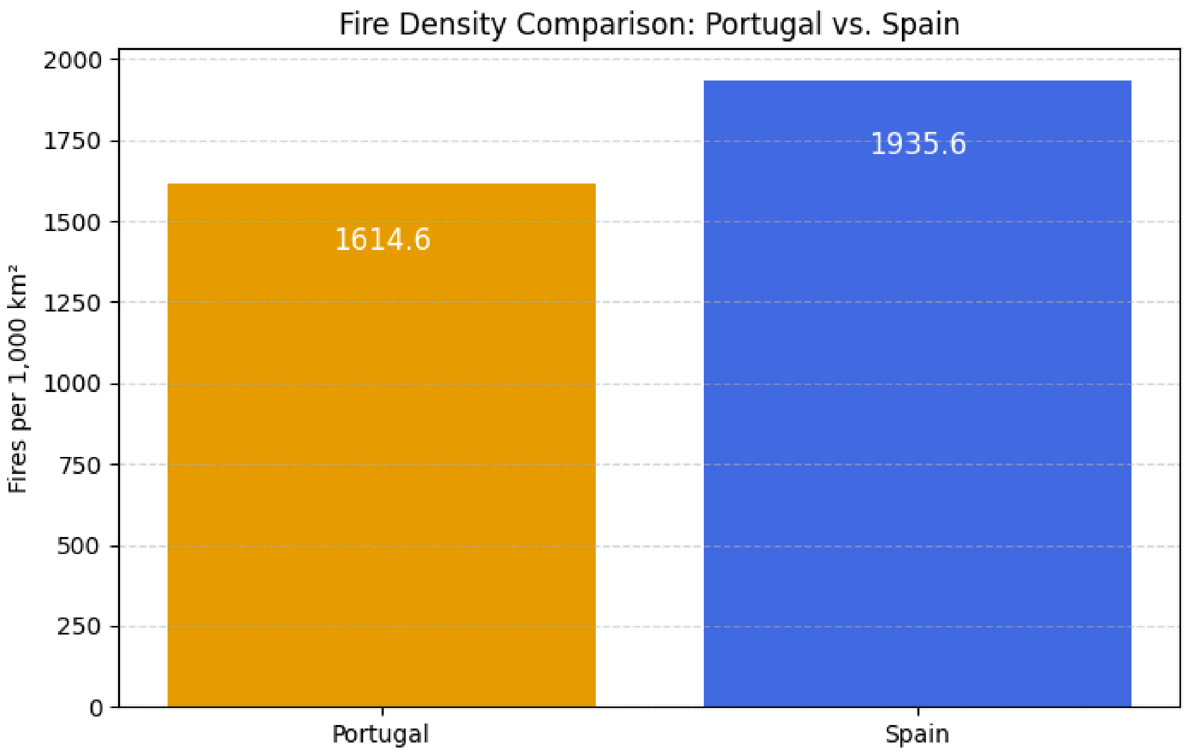
Portugal vs. Spain Wildfire Trends

- Portugal:
  - Wildfire numbers are generally lower than in Spain, possibly due to the smaller country area.
  - Large spikes occurred in the early 2000s, 2013, and 2017, with the most extreme fire season around 2017, when it had one of the worst fire incidents in recent times, which resulted in 66 deaths and 204 injured people
  - After 2017, fire activity decreased but remained sporadic.
  - Portugal seems to have more variability in fire count, with some years having very low numbers and others having extreme peaks - might be related to burned landscape (which would prevent repeated fires in one location until vegetation regrew).
- Spain:
  - Spain experiences a much higher total number of fires than Portugal, sometimes exceeding 100,000 fires per year, but then again, it has a much higher surface area.
  - Extreme peaks occurred around 2012, 2017, and 2022, similar to Portugal but on a larger scale.
  - Unlike Portugal, where some years have very few fires, Spain seems to have consistently high fire activity across most years.
  - The fire heatmap for Spain shows intense wildfire seasons happening frequently, particularly in the past 15 years.

Summary

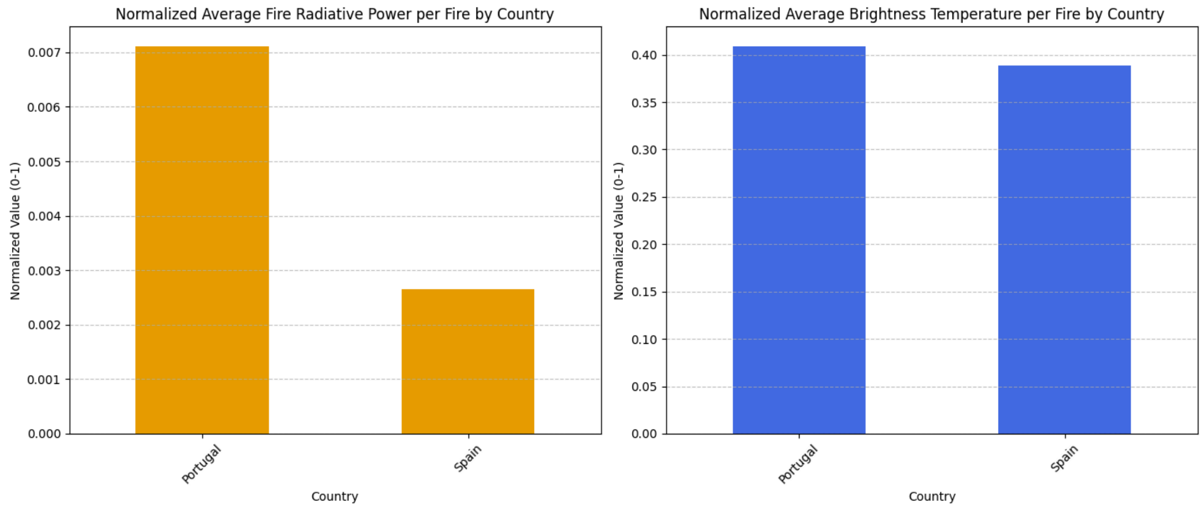
- Fire activity in both countries peaked in similar periods (2012, 2017, 2022), indicating that regional climatic or environmental factors are driving fire trends.
- 2017 stands out as a particularly extreme wildfire season for both Portugal and Spain, aligning with known severe wildfire events in both countries.
- The trend appears cyclical, with fire seasons occurring in waves rather than a steady increase.
- The recent peak in the year 2022 suggests that wildfires are still a major issue.
- Climate factors such as rising temperatures, drought, and extreme weather could be contributing to these trends.

Fire Density Comparison (Portugal vs. Spain)



- Spain has a higher fire density (1,935.6 fires per 1,000 km²) compared to Portugal (1,614.6 fires per 1,000 km²).
- This indicates that Spain experiences more fires relative to its size than Portugal. Even though Spain has a higher density, Portugal's fire density is also very high.
- Spain, with its larger area, may need more extensive fire response infrastructure, while Portugal, being smaller, may be better off to invest in fire prevention strategies.

Average Fire Radiative Power and Brightness Temperature per Country



Fire Radiative Power per Fire (Left Chart)

- Portugal has significantly higher average Fire Radiative Power (FRP) per fire compared to Spain. This suggests that fires in Portugal tend to produce more energy/power. Possible explanations:
    - Drier conditions or more combustible vegetation in Portugal.
    - Higher wind speeds or flammable vegetation leading to more intense fire spread.
    - Differences in fire management policies or suppression strategies.

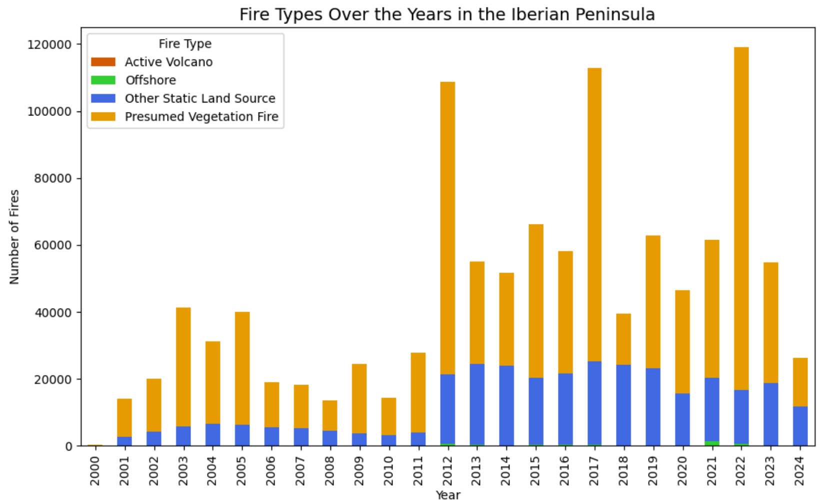
Brightness Temperature per Fire (Right Chart)

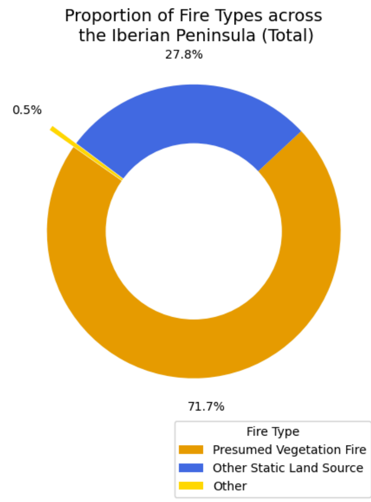
- Brightness temperature measures the thermal intensity of the fire, often linked to flame temperature.
- Portugal and Spain have similar Brightness Temperature values per fire, with Portugal slightly higher. The similarity suggests that the physical temperature of fires is comparable in both countries, even though Portugal’s fires are releasing more energy overall.

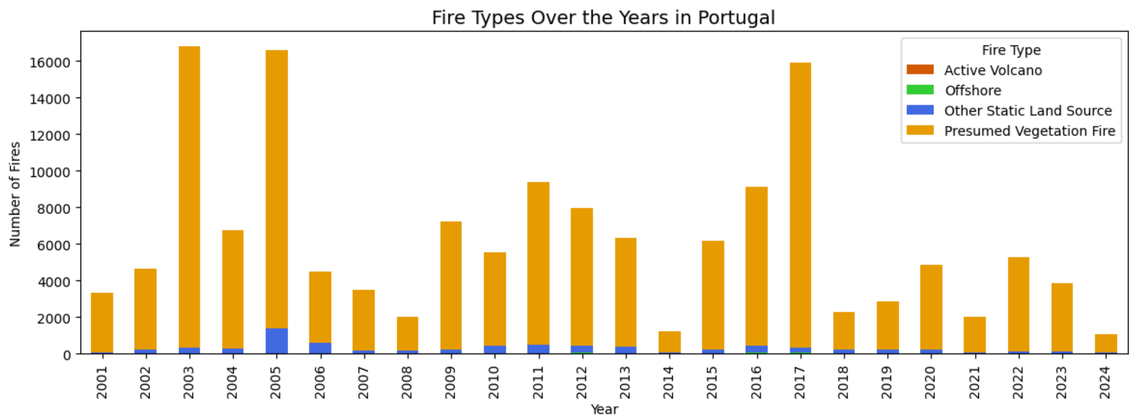
Summary

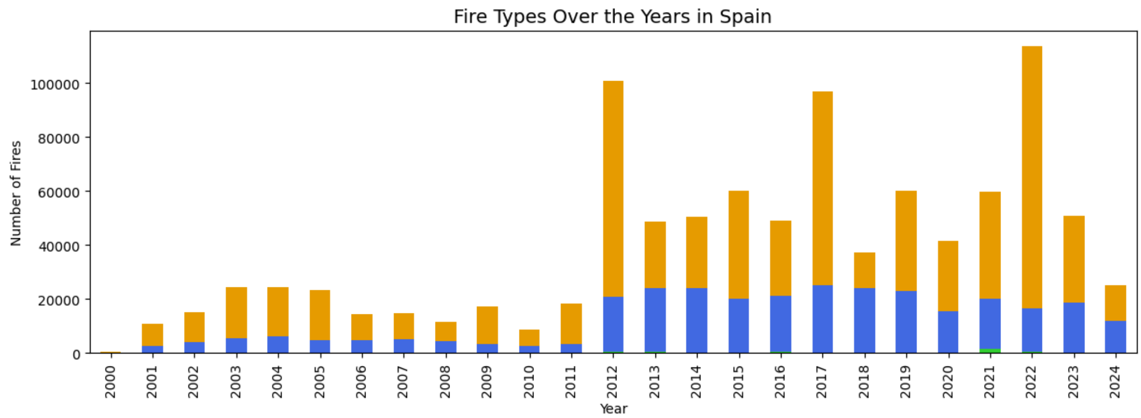
- While the fire activity in Spain is more frequent (as seen in previous graphs), Portugal's fires appear to be more destructive due to the higher energy output.
- Spain may need more fire detection and early suppression due to the sheer number of fires.
- Portugal may require better land management and firebreaks to reduce fire intensity.

Iberian FIRMS Fire Types Distribution over Time









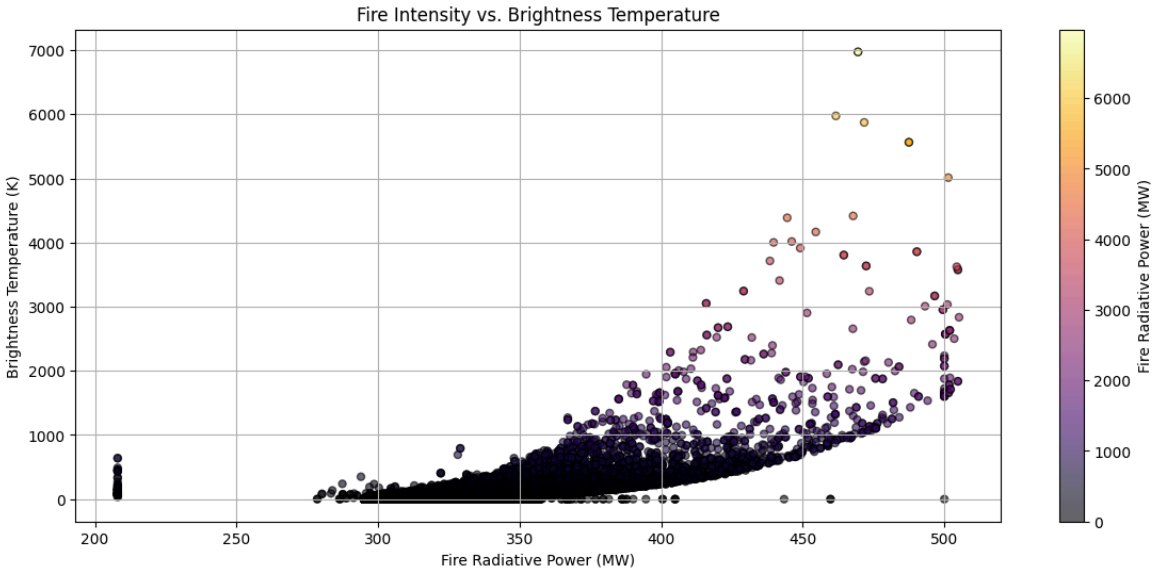
- 71.7% of all fires in the Iberian Peninsula are classified as presumed vegetation fires.
- 27.8% are classified as "Other Static Land Source".
- Offshore and active volcano fires are extremely rare, comprising less than 0.5% of all records, possibly influenced by the island territories.
- Both Spain and Portugal show that most fire activity is related to vegetation fires, reinforcing concerns about climate change, drought, and land management.
- The increase in vegetation fires since 2012 suggests worsening wildfire conditions, potentially due to rising temperatures and drier conditions.

- The blue bars (representing "Other Static Land Source" fires) are more frequent in Spain than in Portugal.
- This suggests that Spain has more complex fire causes, possibly related to industrial, agricultural, or energy-related activities.

Summary

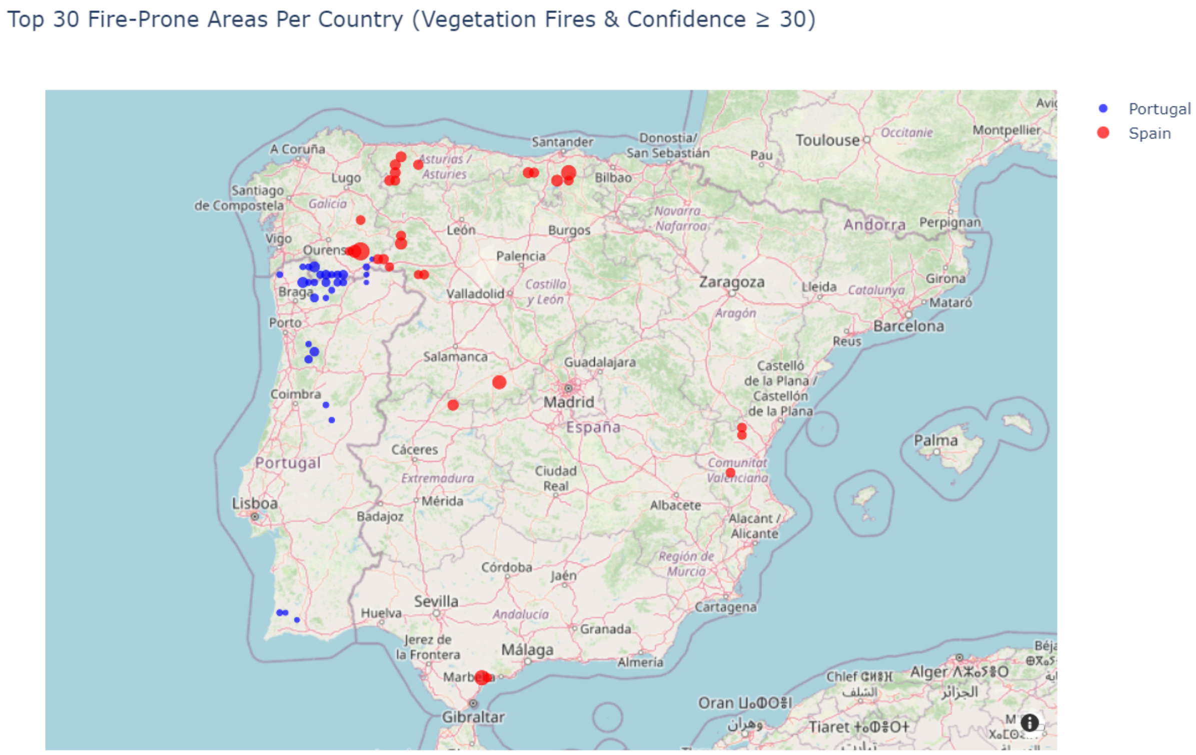
- Vegetation fires dominate the Iberian Peninsula, making them the primary fire risk.
- "Other Static Land Source" fires are more common in Spain, suggesting different fire causes beyond vegetation.

Correlation between Fire Radiative Power and Brightness



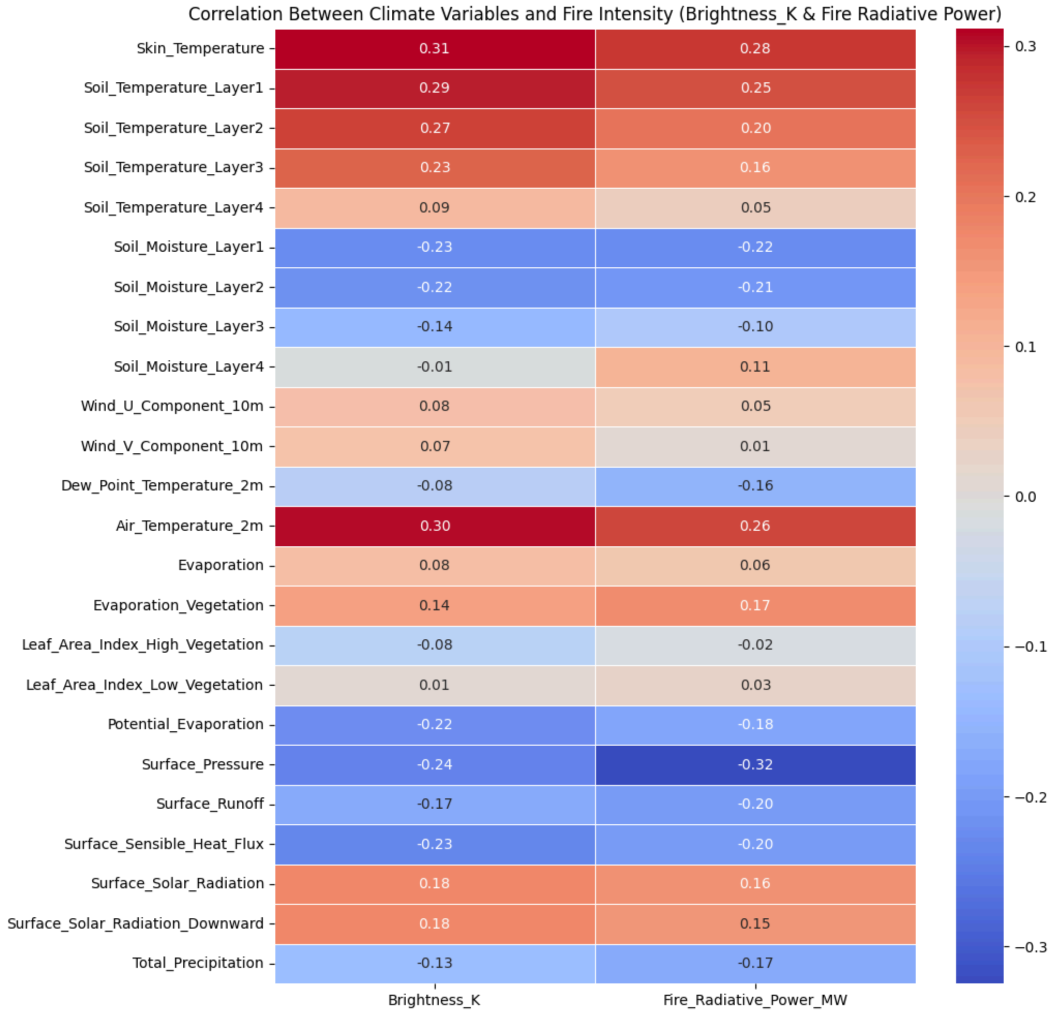
- There is a clear upward trend indicating that as Fire Radiative Power (FRP) increases, brightness temperature also tends to increase, as expected.
- The majority of points are clustered between 300-450 MW FRP and below 1000 K in brightness temperature. This suggests that most wildfires are of moderate intensity and do not reach extreme temperatures.
- Most fires fall within a moderate intensity range, but some extreme cases exist as per the outlier identified in this graph.

Top 30 Fire-Prone Areas Per Country (Vegetation Fires only & Confidence ≥ 30)



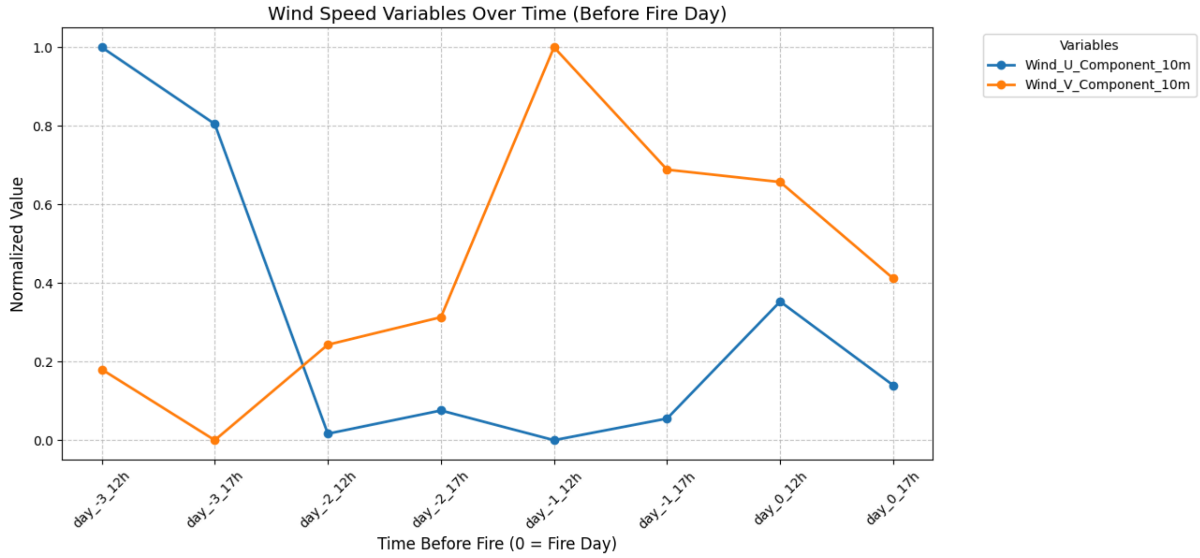
- Most fire-prone areas in Portugal (blue dots) are clustered in the northwest, particularly around Braga, Porto and Coimbra.
- Spain’s fire-prone areas (red dots) are also concentrated in Galicia, Asturias, and León, suggesting similar environmental conditions contribute to fire outbreaks in both countries.
- There are some fire hotspots seen in central Spain (around Salamanca) and southeastern Spain (Valencia and Marbella). This suggests that fires occur in both northern humid regions (Galicia, Asturias) and drier Southern areas.
- Portugal’s fires are more concentrated whilst Spain’s fires are more widely distributed, including northern, central, and coastal regions, indicating a greater geographical spread of fire risk.
- The Galicia-Northern Portugal border is the most active fire-prone region, which might be due to:
    - Dense forests and vegetation
    - Frequent droughts and heatwaves
    - Human activity (agriculture, land clearing, or arson cases)

Correlation between Climate Variables and Fire Intensity Variables

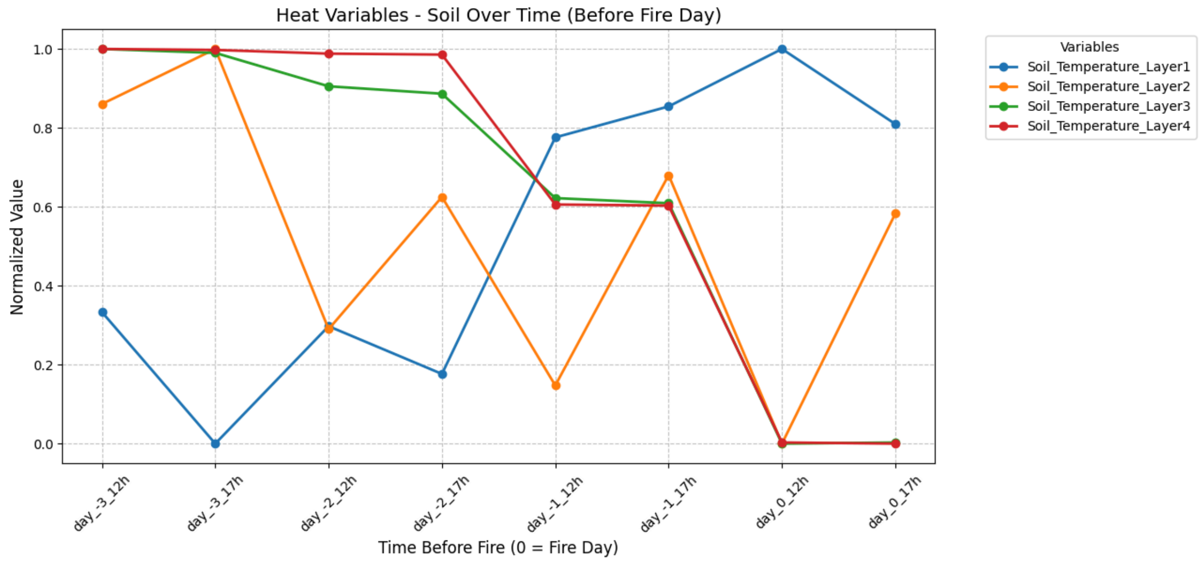


- Correlation between environmental factors and wildfires is very complex, so correlation rates may appear low, but still worth exploring.
- Skin temperature has the highest correlation with both fire variables.
- Air temperature at 2m height also shows a strong link.
- Soil temperature is positively correlated, but the effect weakens with depth as expected.
- Soil moisture (top layers) has a moderate negative correlation—less moisture leads to higher fire intensity, possibly indicating sustained dry periods in the lead up to a fire.
- Surface pressure has one of the strongest negative correlations, possibly due to increased wind and dryness.
- Solar radiation and evaporation slightly increase fire risk by drying out vegetation.
- Precipitation and surface runoff help reduce fire intensity, but their effect is weaker than temperature and moisture levels.
- High temperatures and low moisture are the strongest predictors of fire intensity.
- Low surface pressure and dry conditions create fire-friendly environments.

Temporal Climate Patterns Leading to Fire Events

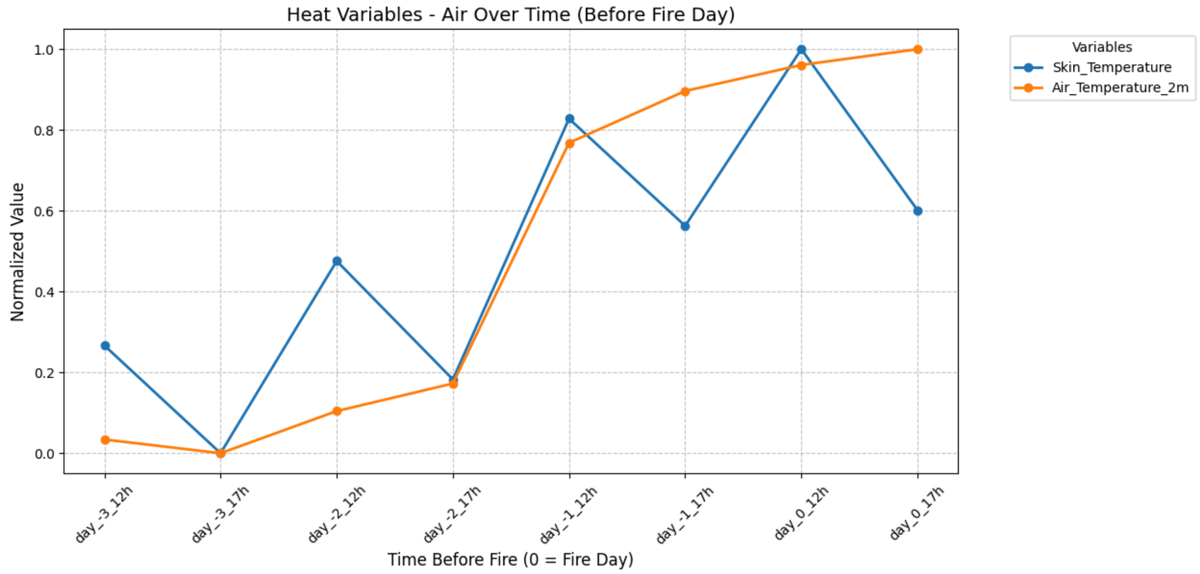


- Wind patterns shift before the fire: The U-component (west-east wind) decreases significantly, while the V-component (north-south wind) peaks a day before the fire.
- Potential wind direction change: The opposing trends suggest a shift in dominant wind direction leading up to the fire. Exploring the wind patterns with the wider geographical journey across the Iberian Peninsula might be worth exploring, since the northwestern part of the Iberian Peninsula is the most affected.
- Increased wind activity may contribute to fire spread: The spike in wind speed just before the fire could have influenced fire behavior and spread.

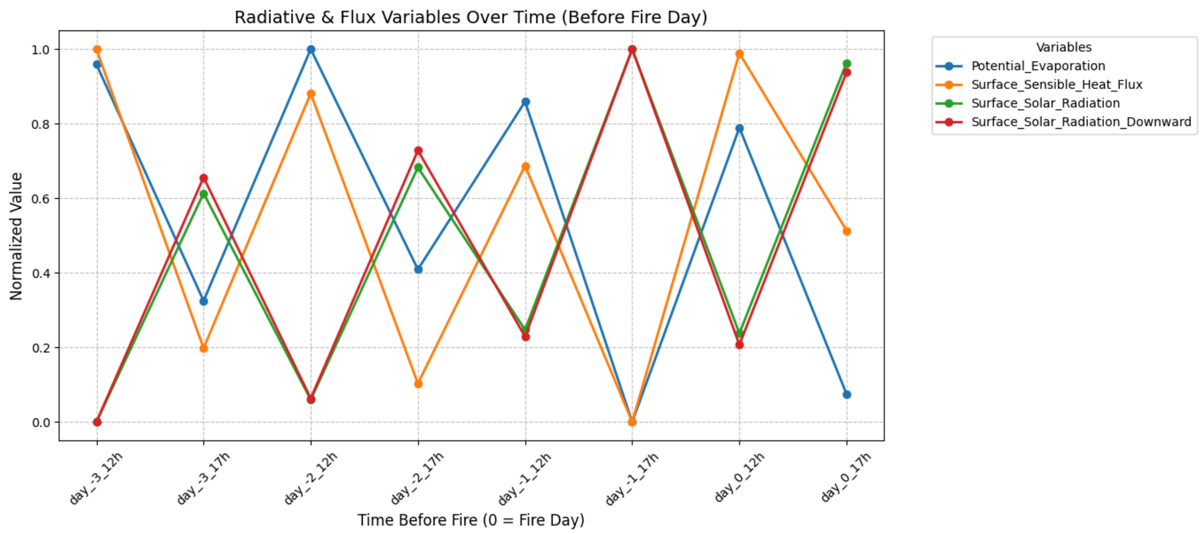


- Surface layers (Layer 1 & 2) show more variability, with steeper drops and spikes.

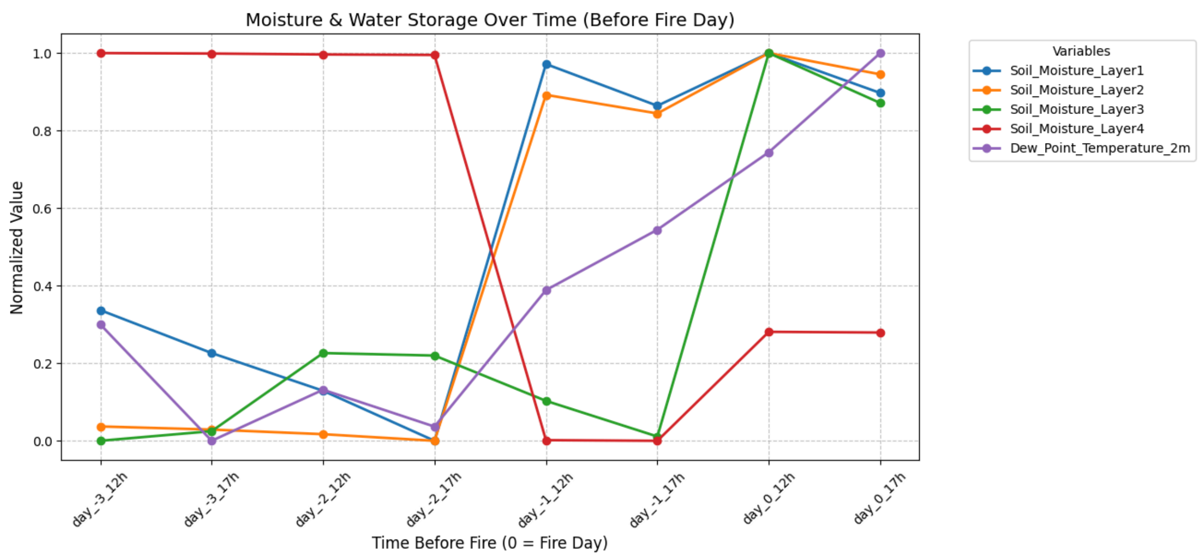
- Deeper layers (Layer 3 & 4) are more stable but decline closer to the fire day.
- Layer 1 (surface) increases sharply just before the fire, peaking at fire day.
- Deeper layers cool down, possibly due to heat transfer or changing weather conditions.
- Rising surface temperatures may contribute to drier conditions, increasing fire risk.
- The cooling of lower layers suggests heat is concentrated at the top, possibly due to solar radiation or drying effects.



- Skin temperature fluctuates but peaks sharply just before fire day.
- Air temperature (2m height) steadily increases, reaching its highest point on fire day.
- Both variables follow a similar upward trend, indicating warming conditions on the lead up to the fire.
- The rising temperatures, especially at the surface, suggest increased dryness, which might contribute to fire ignition and spread.

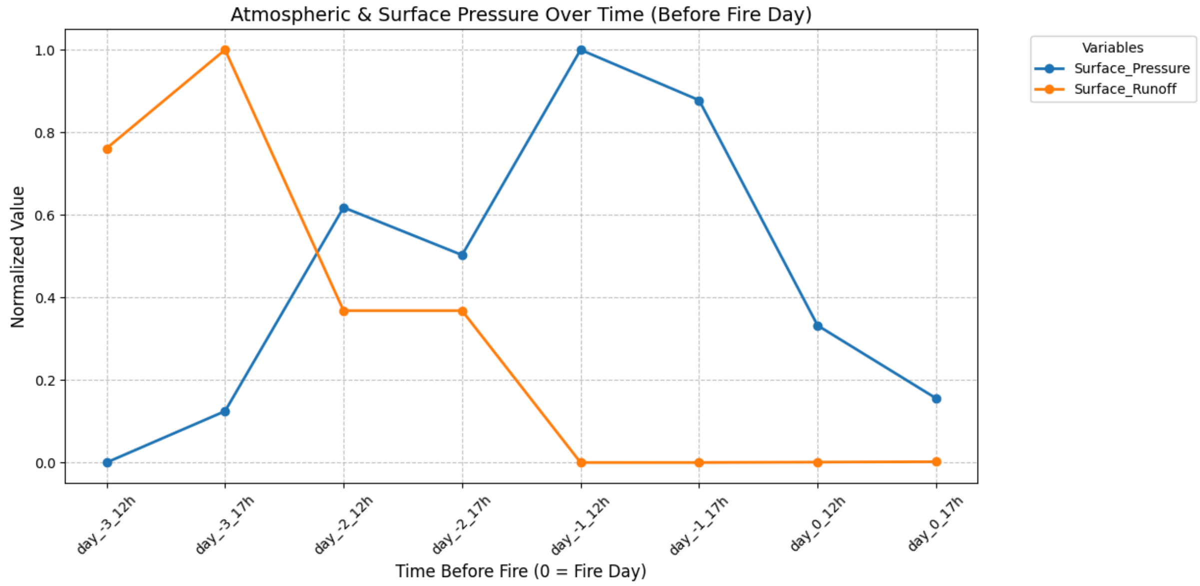


- Surface solar radiation and downward radiation show periodic spikes, indicating the natural fluctuation of the day and night.
- Surface Sensible Heat Flux does have a sharper increase on the evening before a fire, suggesting that the ground is releasing more heat into the atmosphere.
- The afternoon readings on the day before a fire do show the widest distance between Solar Radiation and Surface Sensible Heat Flux as well as Potential Evaporation. This suggests that heat is being absorbed and stored at the surface, while moisture levels drop. This is likely to contribute to drying vegetation, increasing surface temperatures, and setting the stage to start a fire and potentially rapid spreading.

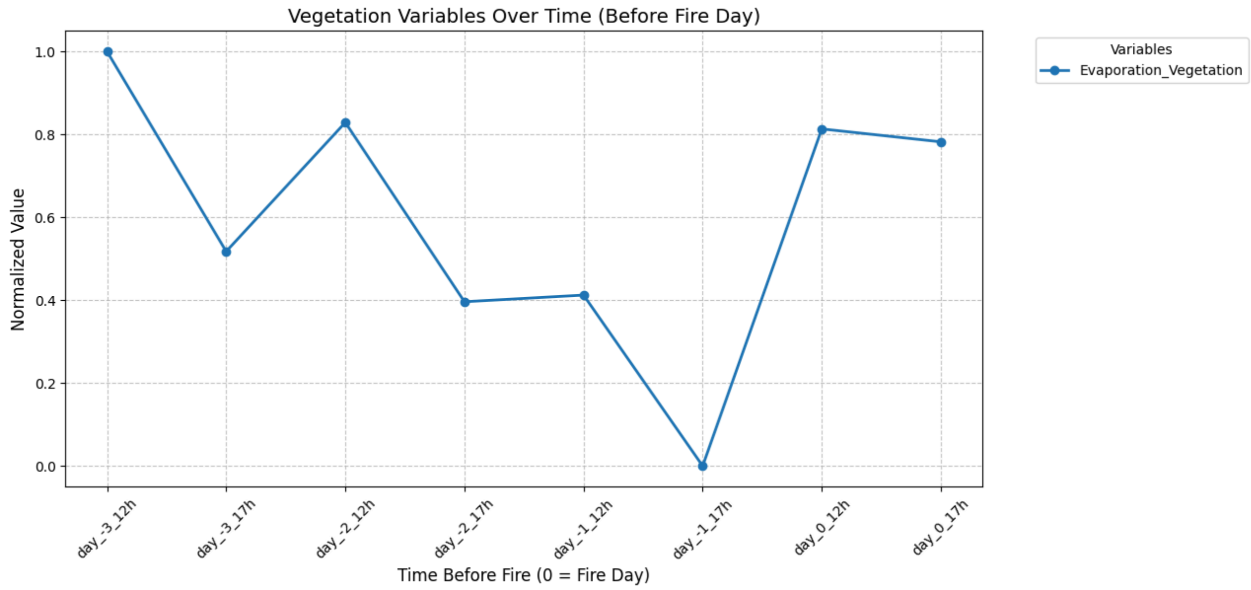


- Soil moisture (Layer 4) drops significantly before the fire:
- Shallow layers (1 & 2) show the biggest incline.
- Deeper layers (3 & 4) remain stable longer but eventually drop too.
- Dew point temperature increases on the lead up to the fire.

- The sudden rise in Soil Moisture before the fire could suggest a reaction to changing weather conditions but is probably too late to prevent ignition.



- Surface pressure increases before the fire, peaking about a day before, then drops sharply.
- Suggests a shift in atmospheric conditions, possibly linked to incoming low-pressure fronts that can bring drier, windier weather. This may also be impacted by seasons.
- Surface runoff is initially high but drops completely just before the fire, indicating drying conditions and lack of recent rainfall.
- Reduced water availability could contribute to drier vegetation and higher fire risk.



- Evaporation from vegetation starts high, decreases, then spikes again just before fire day, suggesting variable moisture loss.
- Fluctuating evaporation suggests changing water availability and potential vegetation drying before the fire.

Summary

By analysing each of these parameters we start to build a picture on what the lead up to a fire event looks like. Most of the observations indicate that temperatures (Air and Soil), moisture retention and heat transfer seem to have trends worth emphasising in the continuation of this project.

Going forward I believe it is important to re-analyse these parameters within months associated with Wildfire Season. This will give a more real-life view of what the lead up to a fire is, accounting for seasonal changes.

Model Development

Fire Likelihood Prediction (Classification)

Fire Intensity Prediction (Regression)

Model Hyperparameter Tuning

Results & Discussion

Model Performance

Feature Importance

Insights & Interpretation

Conclusion & Future Work

Key Findings

Future Improvements