## Data Collection:

| Requirement | Verification Output | Status |
|---|---|---|
| **Collected a structured dataset...** | Loaded 244 entries and 15 columns. | Passed |
| **Ensured the dataset includes...** | All required columns (Temperature, RH, Ws, Rain, FFMC, DMC, ISI, Region, FWI) are present. | Passed |
| **Verified data types, consistency, and formatting** | Column names were cleaned (spaces removed). Most columns are int64 or float64. | Passed |
| **Loaded the dataset into a Pandas DataFrame** | Displaying the first 5 rows confirms successful loading. | Passed |
| **Conducted initial inspection...** | DC and FWI were initially object types and successfully converted to float64 for better consistency (though this revealed 1 missing value in each, which is handled in the next module). | Passed |

First 5 rows of the dataset:

```
  day  month  year  Temperature   RH   Ws  Rain   FFMC  DMC   DC  ISI  BUI  FWI   Classes
Region
0   1     6  2012          29   57   18   0.0  65.7  3.4   7.6  1.3  3.4  0.5  not fire    Bejaia
1   2     6  2012          29   61   13   1.3  64.4  4.1   7.6  1.0  3.9  0.4  not fire    Bejaia
2   3     6  2012          26   82   22  13.1  47.1  2.5   7.1  0.3  2.7  0.1  not fire    Bejaia
3   4     6  2012          25   89   13   2.5  28.6  1.3   6.9  0.0  1.7    0  not fire    Bejaia
4   5     6  2012          27   77   16   0.0  64.8  3.0  14.2  1.2  3.9  0.5  not fire    Bejaia
```

## Data types of each column:

day int64

month int64

year int64

Temperature int64

RH int64

Ws int64

Rain float64

FFMC float64

DMC float64

DC object

ISI float64

BUI float64

FWI object

Classes object

Region object

dtype: object

Shape of the dataset (rows, columns):

(244, 15)

Column names:

Index(['day', 'month', 'year', 'Temperature', ' RH', ' Ws', 'Rain ', 'FFMC',

'DMC', 'DC', 'ISI', 'BUI', 'FWI', 'Classes  ', 'Region'], dtype='object')

## Dataset Information:

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 244 entries, 0 to 243

Data columns (total 15 columns):

# Column Non-Null Count Dtype

--- ------ -------------- ----- 0 day

244 non-null int64

| | | | |
|---|---|---|---|
| 1 | month | 244 non-null | int64 |
| 2 | year | 244 non-null | int64 |
| 3 | Temperature | 244 non-null | int64 |
| 4 | RH | 244 non-null | int64 |
| 5 | Ws | 244 non-null | int64 |
| 6 | Rain | 244 non-null | float64 |
| 7 | FFMC | 244 non-null | float64 |
| 8 | DMC | 244 non-null | float64 |
| 9 | DC | 244 non-null | object |
| 10 | ISI | 244 non-null | float64 |
| 11 | BUI | 244 non-null | float64 |
| 12 | FWI | 244 non-null | object |
| 13 | Classes | 243 non-null | object |
| 14 | Region | 244 non-null | object |

dtypes: float64(5), int64(6), object(4)

## Data Exploration and Data Preprocessing:

### Missing values in each column:

| | |
|---|---|
| day | 0 |
| month | 0 |
| year | 0 |
| Temperature | 0 |
| RH | 0 |
| Ws | 0 |
| Rain | 0 |
| FFMC | 0 |
| DMC | 0 |
| DC | 0 |

| | |
|---|---|
| ISI | 0 |
| BUI | 0 |
| FWI | 1 |
| Classes | 244 |
| Region | 244 |
| dtype: | int64 |

Number of duplicate rows: 0

Outliers in day: 0

Outliers in month: 0

Outliers in year: 0

Outliers in Temperature: 2

Outliers in  RH: 0

Outliers in  Ws: 8

Outliers in Rain : 35

Outliers in FFMC: 16

Outliers in DMC: 12

Outliers in DC: 15

Outliers in ISI: 4

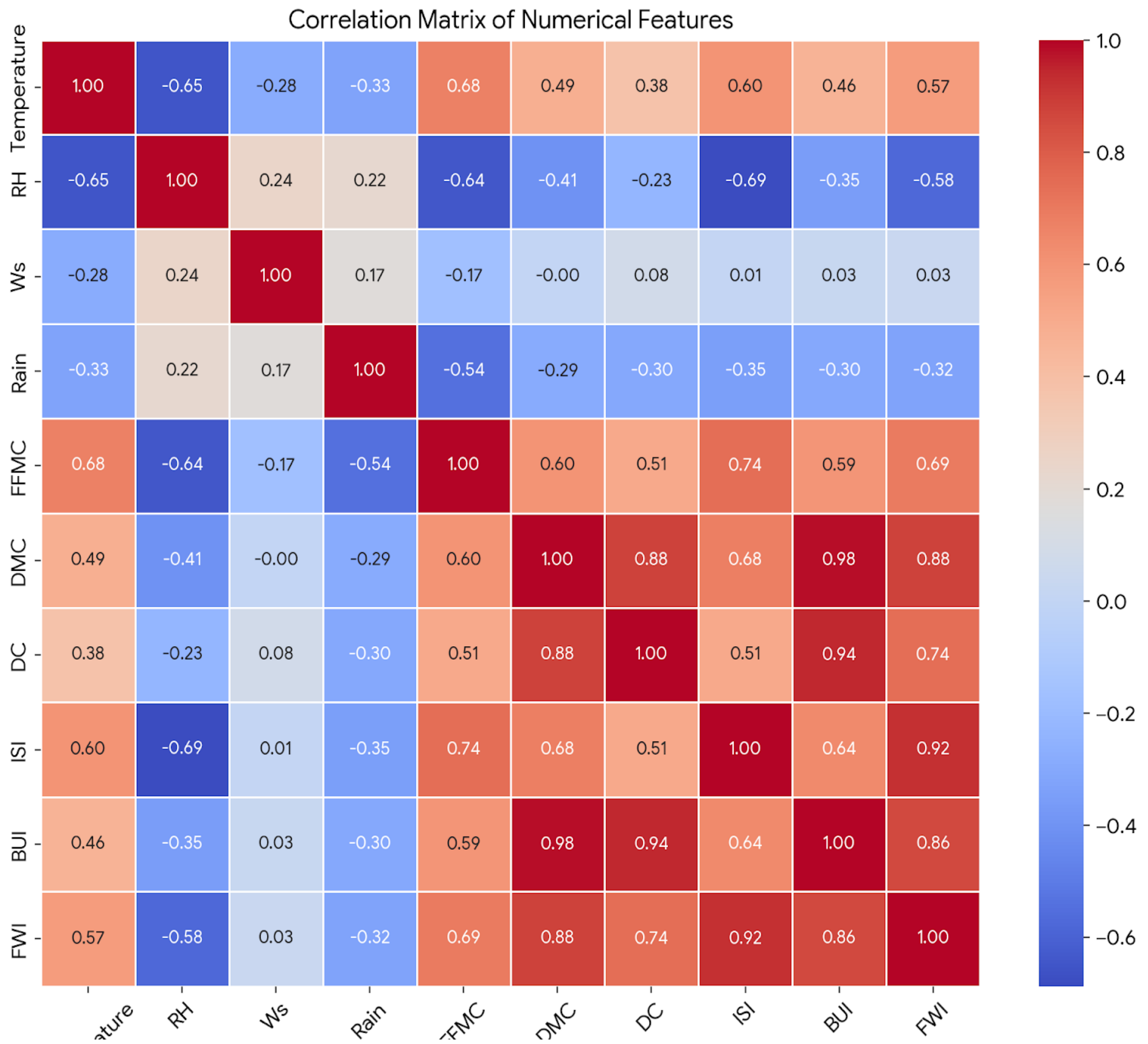
Outliers in BUI: 12

Outliers in FWI: 4

Outliers in Classes  : 0

Outliers in Region: 0

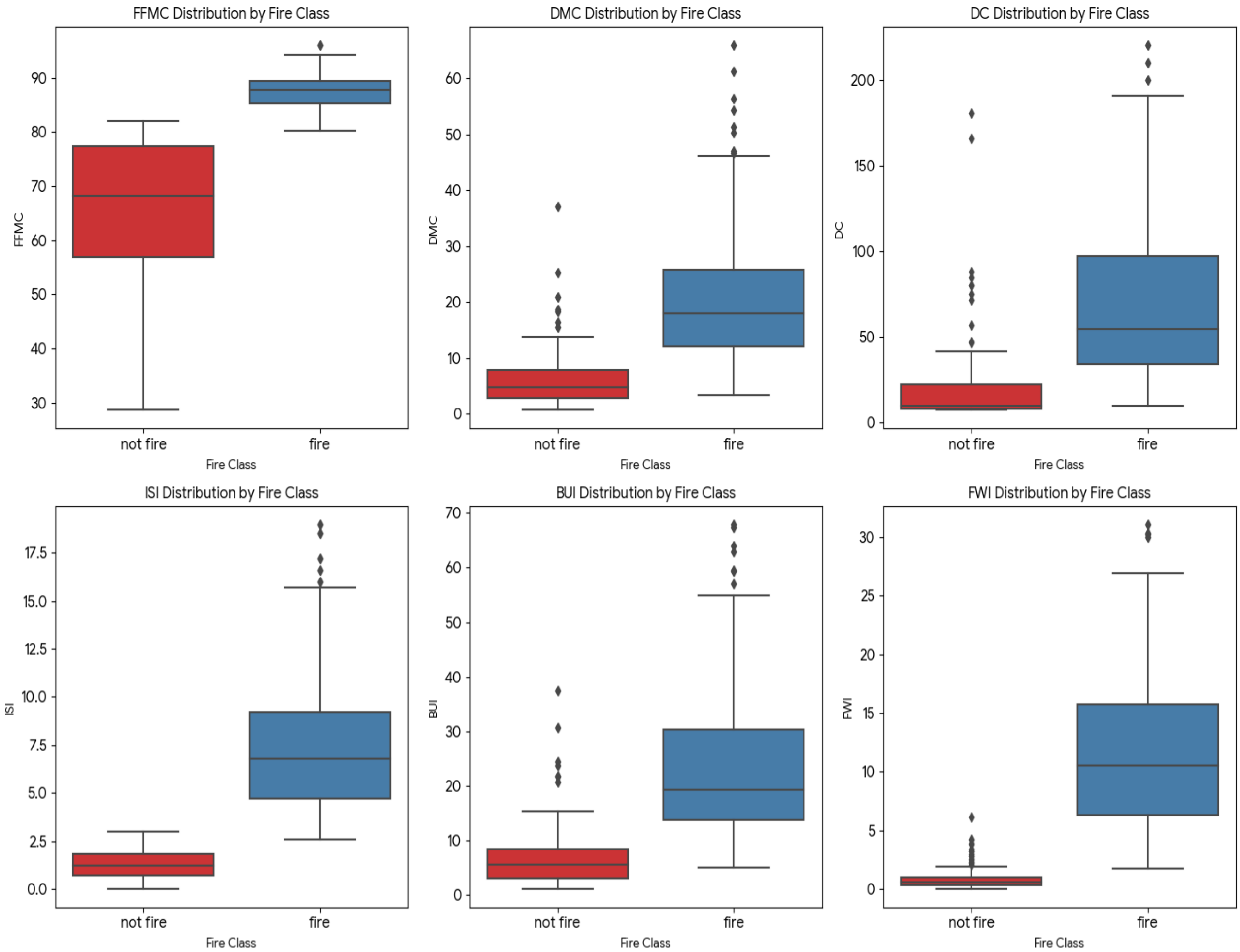
## Correlation Matrix:




## Correlation Matrix Heatmap:

This figure shows the correlation between all numerical features. It confirms that the FWI components (FFMC, DMC, DC, ISI, BUI) are highly correlated with each other and, most importantly, with the target variable, FWI (Fire Weather Index).

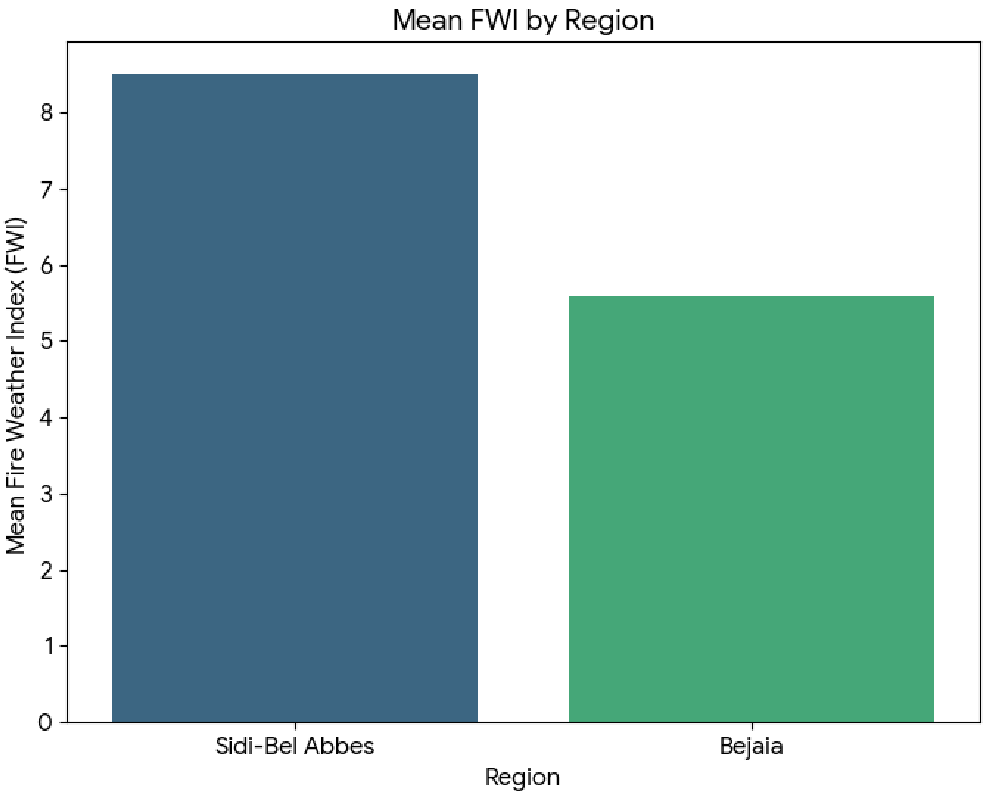
## FWI Components Distribution by Fire Class (Boxplots):




This figure compares the distributions of the FWI and its components (FFMC, DMC, DC, ISI, BUI, FWI) across the two categorical classes: **'fire'** and **'not fire'**.

- **Observation:** The boxplots clearly show that all six indices are significantly higher on days classified as **'fire'** compared to days classified as **'not fire'**, which validates the relationship between these features and the fire occurrence.

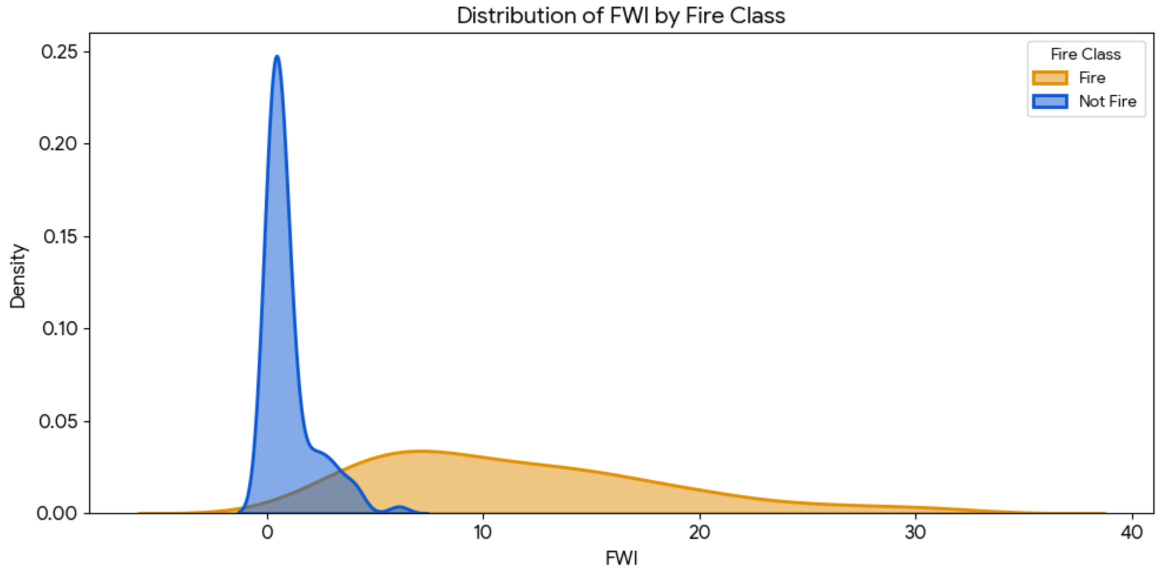
## Mean FWI by Region (Bar Chart):

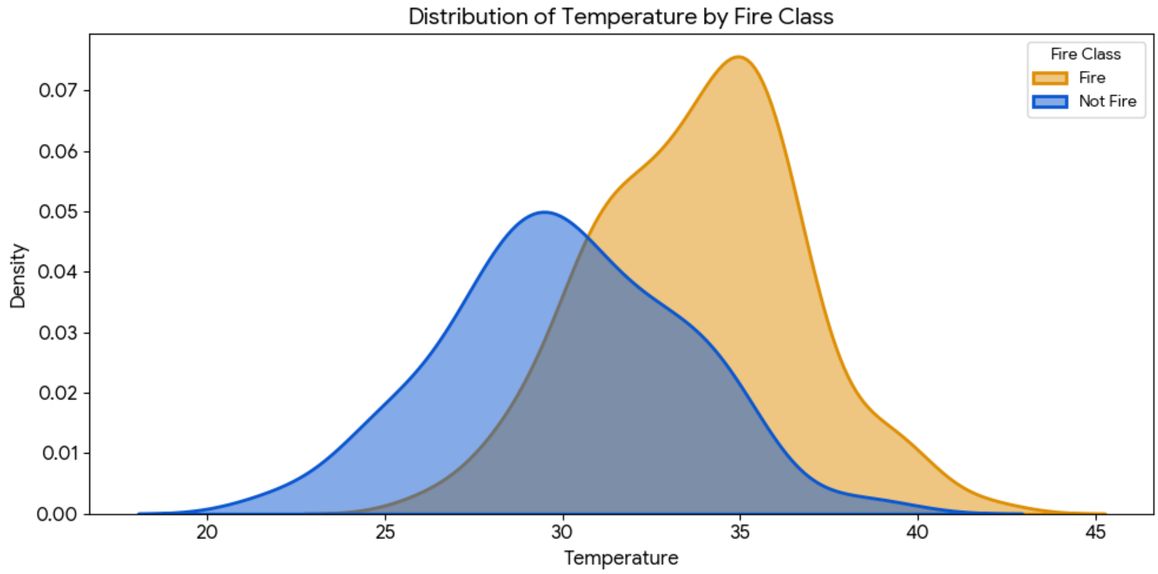



This bar chart compares the average Fire Weather Index (FWI) between the two recorded regions: Bejaia and Sidi-Bel Abbes.

- Observation: The chart visually represents the difference in average fire potential between the regions, allowing for a quick comparison of regional risk profiles.

## Comparison Histograms by Fire Class:




### Distribution of FWI by Fire Class:

- Observation: The distribution of the Fire Weather Index (FWI) for the 'fire' class is clearly shifted to the right (higher values) compared to the 'not fire' class, demonstrating that FWI is an effective predictor of fire occurrence.

### Distribution of Temperature by Fire Class:

- Observation: The temperature distribution for the 'fire' class is also shifted slightly to the right (higher temperatures) compared to the 'not fire' class, indicating that higher temperatures are associated with fire days, as expected.

### Python libraries were used for the data exploration, preprocessing, and generating the comparison graphs:

1. **pandas** (as pd): Used for data manipulation, loading the CSV file, cleaning the data (e.g., stripping spaces, converting types), handling missing values, and encoding the Region column.
2. **numpy** (as np): Used for general numerical operations, although its specific use was minor in the final code output, it is standard practice to import it during data cleaning and manipulation.
3. **matplotlib.pyplot** (as plt): Used for creating and customizing the plots, including the histograms and structuring the boxplots.
4. **seaborn** (as sns): Used for generating the statistical visualizations, specifically the boxplots, histograms, and the correlation heatmap.