# Forest Fires Dataset – 3D Visualization & Predictive Analysis

**Table of Content**


# 1. Introduction:……..…… …….3

# 2. Dataset Overview:……..3

# 3. Data Preprocessing:..... 3

# 4. Exploratory Data Analysis (EDA): ..... 3

a) **3D Scatter Plot of Burned Area:** ..... 3

b) **3D Surface Plot of Temperature:** ..... 4

# 5. Clustering Analysis:..... 5

# 6. Meteorological Feature Visualization: ..... 5

# 7. Predictive Modeling: ..... 6


**Source**: UCI Machine Learning Repository – Forest Fires Dataset

**Prepared By:** Hani Bint E Rauf – master's in data science

**Tools Used**: Python 3.x, Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn, Jupyter Notebook

## 1. Introduction:

This analysis explores the **Forest Fires Dataset** from the UCI Machine Learning Repository. The aim is to use **3D visualization** and **predictive modeling** to study spatial, meteorological, and fire-related factors to better understand patterns in forest fire occurrences.

## 2. Dataset Overview:

- **Instances:** 517
- **Features:** 13 (spatial coordinates, meteorological data, FWI indexes, burned area)
- **Objective:** Identify relationships between environmental conditions and forest fire severity through visualization and modeling.

## 3. Data Preprocessing:

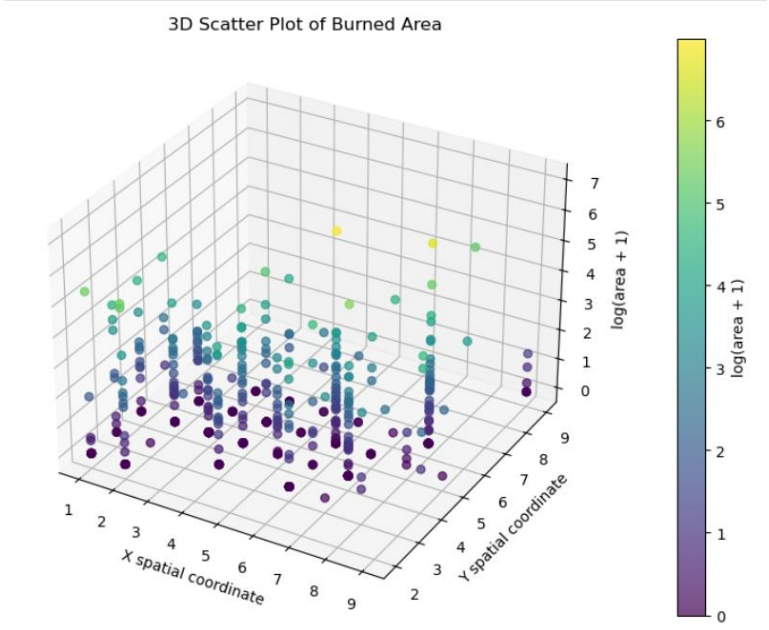
Steps performed:

- Converted categorical features (month, day) into numeric codes for analysis.
- Applied log-transformation on area to handle skewness in burned area distribution.
- Verified and standardized data types.

## 4. Exploratory Data Analysis (EDA):

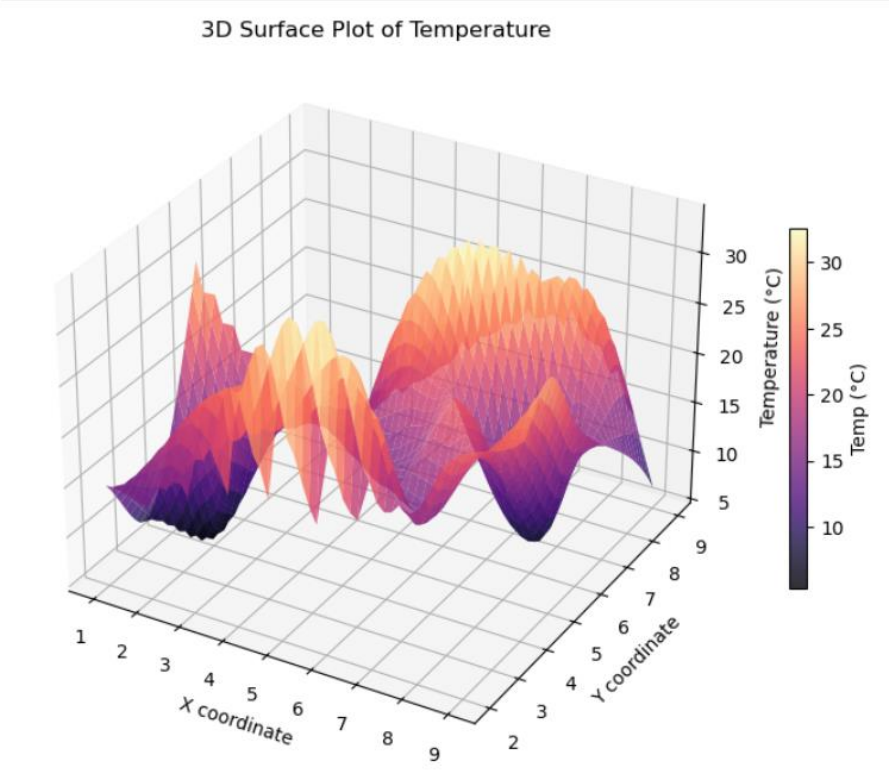
### a) 3D Scatter Plot of Burned Area:

- **Purpose:** To examine how burned area varies with spatial coordinates (X, Y).
- **Notes:** Log-scaling was applied to area for better visibility of variations.



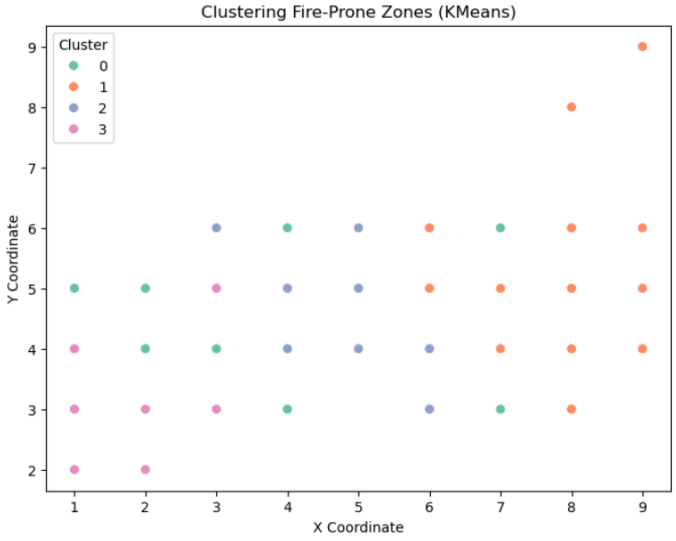

**b) 3D Surface Plot of Temperature:**

a) **Purpose:** To visualize spatial variation in temperature across the forest region.



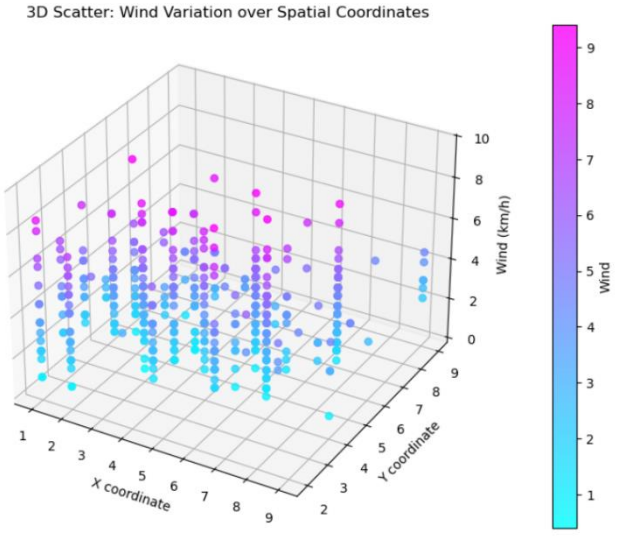
## 5. Clustering Analysis:

I applied **KMeans clustering** on spatial coordinates and burned area to detect fire-prone zones. I identified distinct regions with higher fire occurrence.




## 6. Meteorological Feature Visualization:

Explored **wind speed** and **humidity** in 3D space to study their spatial variations.



## 7. Predictive Modeling:

Two models were applied to predict the burned forest area from weather data: Linear Regression and Random Forest Regression, using key features (FFMC, DMC, DC, ISI, temp, RH, wind, rain). The target was log-transformed to reduce skewness.

**Linear Regression:** Showed low $R^2$ and patterned residuals, indicating it could not capture the dataset's non-linear nature. Predictions underestimated larger fires.

**Random Forest Regression:** Better captured non-linear relationships, achieved higher $R^2$, and identified temp, RH, wind, and ISI as most important. Residuals were more evenly distributed, indicating improved accuracy.

**Conclusion:** Random Forest provided more reliable predictions than Linear Regression.