**MDSC-102**

**Final Assignment Report**

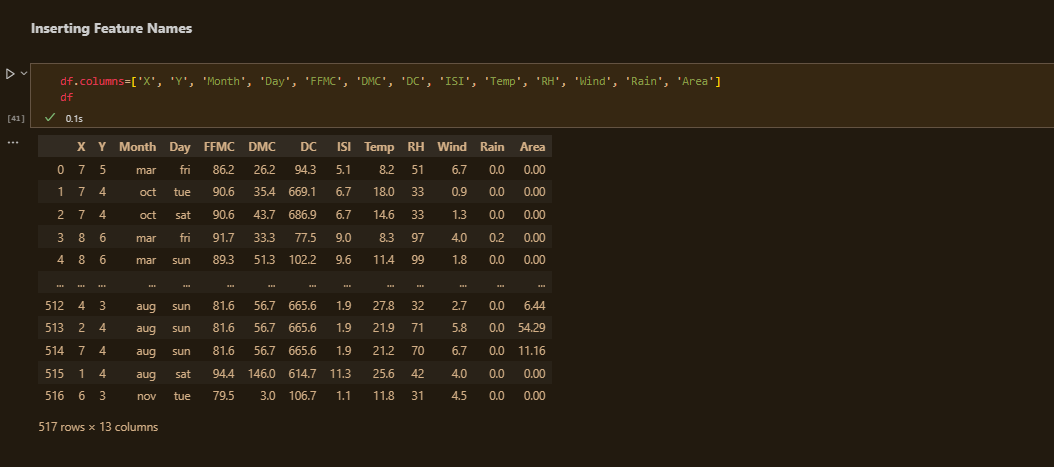
* **Sumukha Sreenidhi (23912)**

**Dattase**: Forest Fire Dataset

**INTRODUCTION:**

**Features**:

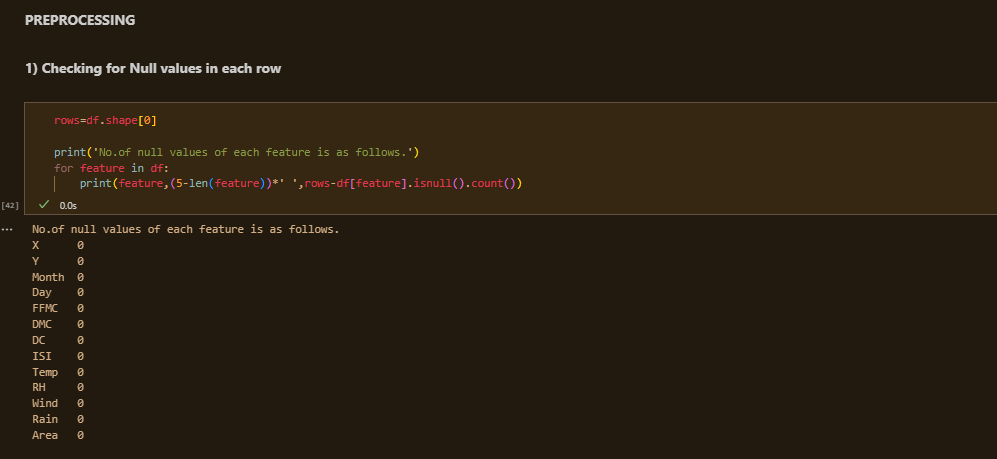
1. **X, Y**: The spatial coordinates of the fire's location within a grid or on a map.
2. **Month**: The month in which the fire occurred.
3. **Day**: The day of the week when the fire occurred.
4. **FFMC (Fine Fuel Moisture Code)**: A numerical rating of the moisture content of litter and other cured fine fuels.
5. **DMC (Duff Moisture Code)**: A numerical rating of the moisture content of decomposed organic material.
6. **DC (Drought Code)**: A numerical rating of the drought conditions in the area.
7. **ISI (Initial Spread Index)**: A numerical rating of the fire's potential rate of spread.
8. **Temperature**: The temperature on the day of the fire.
9. **Relative Humidity (RH)**: The relative humidity on the day of the fire.
10. **Wind**: The wind speed on the day of the fire.
11. **Rain**: The amount of rain on the day of the fire.
12. **Area**: The burned area in hectares caused by the forest fire.



Our aim is to study the features and their role through plotting various plots, analyzing and testing hypotheses.

**PREPROCESSING:**

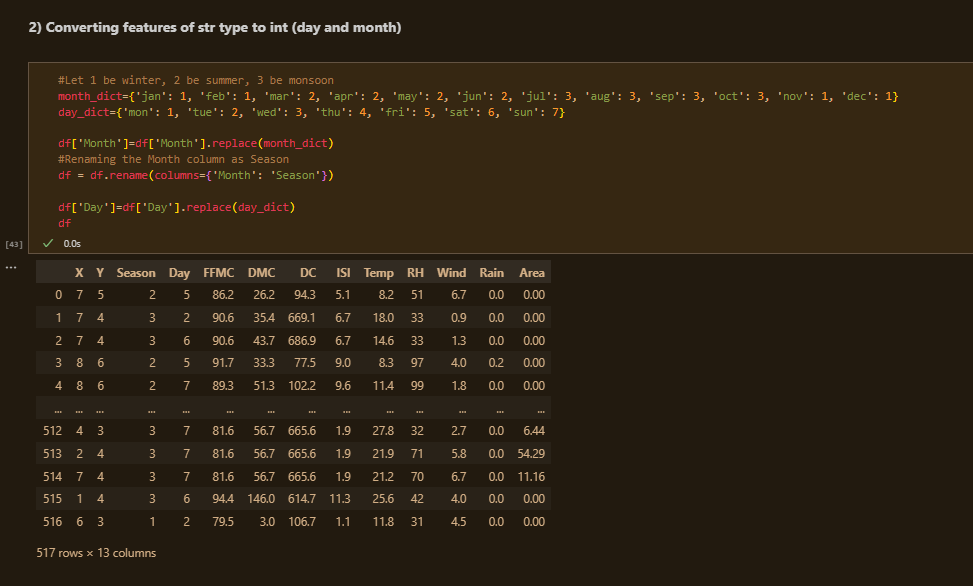
i) Checked for Null Values in the dataframe

****

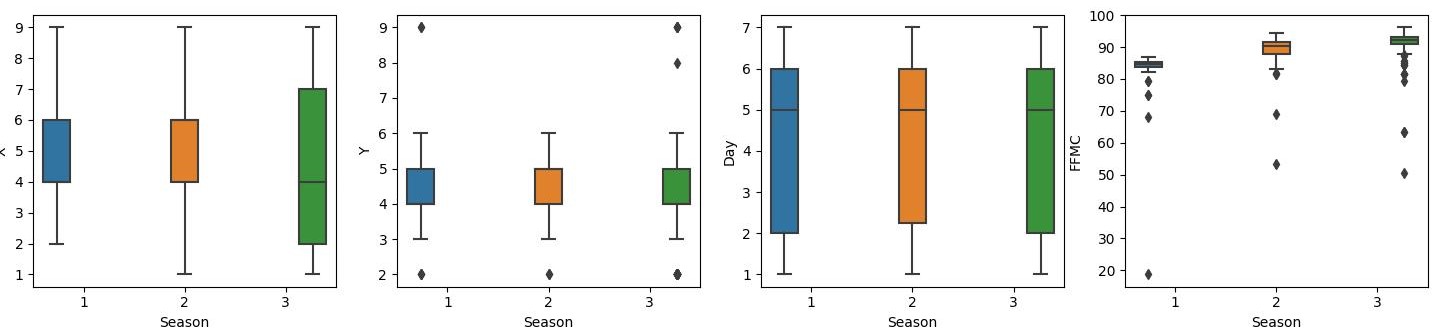
ii) Converted string type values to integer.

Also converted ‘Month’ feature into ‘Season’ so that the season column can be used as a hue variable for better understanding and analysis

(1 is winter, 2 is summer, 3 is monsoon)

****

**BOXPLOTS:**



Inferences:

* X coordinates-

Season 1: Mostly around 5, i.e., 4-6

Season 2: Mostly around 5, i.e., 4-6

Season 3: Mostly around 2-7, little right-skewed

* Y coordinates-

Season 1: Mostly around 4-5

Season 2: Mostly around 4-5

Season 3: Mostly around 4-5

* Day-

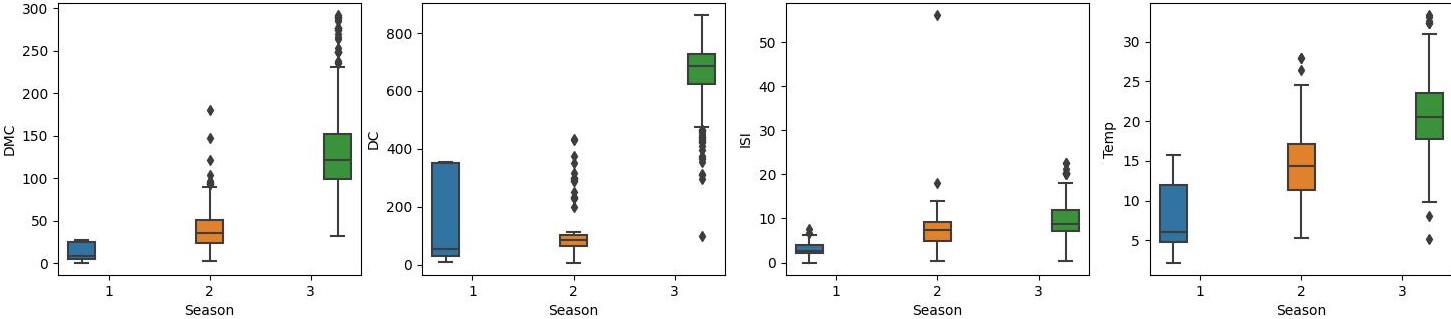
Season 1: Left skewed

Season 2: Left skewed

Season 3: Left skewed

* FFMC

Can’t make significant inference



* DMC-

Season 1: Right skewed

Season 2: Right skewed

Season 3: Slightly right skewed

* DC-

Season 1: Highly Right skewed

Season 2: Left skewed

Season 3: Slightly left skewed

* ISI-

Season 1: Right skewed

Season 2: Close to normal distribution

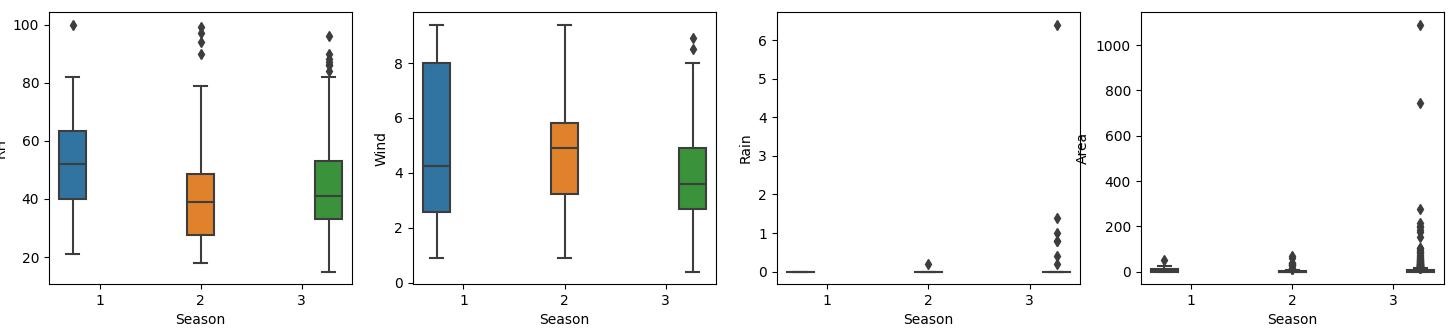
Season 3: Close to normal distribution

* Temp

Season 1: Right skewed

Season 2: Close to normal distribution

Season 3: Close to normal distribution



* RH-

Season 1: Mostly around 5, i.e., 4-6

Season 2: Mostly around 5, i.e., 4-6

Season 3: Mostly around 2-7, little right-skewed

* Wind-

Season 1: Mostly around 4-5

Season 2: Mostly around 4-5

Season 3: Mostly around 4-5

* Rain-

Season 1: Left skewed

Season 2: Left skewed

Season 3: Left skewed

* Season-

Can’t make significant inference