

Algerian Forest Fires Dataset

Data Set Information:

The dataset includes 244 instances that regroup a data of two regions of Algeria,namely the Bejaia region located in the northeast of Algeria and the Sidi Bel-abbes region located in the northwest of Algeria.

122 instances for each region.

The period from June 2012 to September 2012. The dataset includes 11 attribues and 1 output attribue (class) The 244 instances have been classified into fire(138 classes) and not fire (106 classes) classes.

Attribute Information:

Date : (DD/MM/YYYY) Day, month ('june' to 'september'), year (2012) Weather data observations

Temp : temperature noon (temperature max) in Celsius degrees: 22 to 42

RH : Relative Humidity in %: 21 to 90

Ws :Wind speed in km/h: 6 to 29

Rain: total day in mm: 0 to 16.8 FWI Components

Fine Fuel Moisture Code (FFMC) index from the FWI system: 28.6 to 92.5

Duff Moisture Code (DMC) index from the FWI system: 1.1 to 65.9

Drought Code (DC) index from the FWI system: 7 to 220.4

Initial Spread Index (ISI) index from the FWI system: 0 to 18.5

Buildup Index (BUI) index from the FWI system: 1.1 to 68

Fire Weather Index (FWI) Index: 0 to 31.1

Classes: two classes, namely Fire and not Fire

In [1]:

```
import pandas as pd
import seaborn as sns
import numpy as np
import matplotlib.pyplot as plt
```

In [2]:

```
df=pd.read_csv("-----/Algerian_forest_fires_dataset_UPDATE.csv",header=1)
df.head()
```

Out[2]:

	day	month	year	Temperature	RH	Ws	Rain	FFMC	DMC	DC	ISI	BUI	FWI	Classes
0	01	06	2012	29	57	18	0	65.7	3.4	7.6	1.3	3.4	0.5	not fire
1	02	06	2012	29	61	13	1.3	64.4	4.1	7.6	1	3.9	0.4	not fire
2	03	06	2012	26	82	22	13.1	47.1	2.5	7.1	0.3	2.7	0.1	not fire
3	04	06	2012	25	89	13	2.5	28.6	1.3	6.9	0	1.7	0	not fire
4	05	06	2012	27	77	16	0	64.8	3	14.2	1.2	3.9	0.5	not fire

In [4]:

```
df.tail()
```

Out[4]:

	day	month	year	Temperature	RH	Ws	Rain	FFMC	DMC	DC	ISI	BUI	FWI	Classes
241	26	09	2012	30	65	14	0	85.4	16	44.5	4.5	16.9	6.5	fire
242	27	09	2012	28	87	15	4.4	41.1	6.5	8	0.1	6.2	0	not fire
243	28	09	2012	27	87	29	0.5	45.9	3.5	7.9	0.4	3.4	0.2	not fire
244	29	09	2012	24	54	18	0.1	79.7	4.3	15.2	1.7	5.1	0.7	not fire
245	30	09	2012	24	64	15	0.2	67.3	3.8	16.5	1.2	4.8	0.5	not fire

In [186]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 243 entries, 0 to 242
Data columns (total 12 columns):
 #   Column          Non-Null Count  Dtype  
---  -
 0   Temperature     243 non-null   int64  
 1   RH              243 non-null   int64  
 2   Ws              243 non-null   int64  
 3   Rain           243 non-null   float64 
 4   FFMC           243 non-null   float64 
 5   DMC            243 non-null   float64 
 6   DC             243 non-null   float64 
 7   ISI            243 non-null   float64 
 8   BUI            243 non-null   float64 
 9   FWI            243 non-null   float64 
10  Classes        243 non-null   int32  
11  Region         243 non-null   int64  
dtypes: float64(7), int32(1), int64(4)
memory usage: 22.0 KB
```

In [21]:

```
df.isnull().sum()
```

Out[21]:

```
day          0
month        1
year         1
Temperature  1
RH           1
Ws           1
Rain         1
FFMC         1
DMC          1
DC           1
ISI          1
BUI          1
FWI          1
Classes      2
dtype: int64
```

Observation

- There is null value in every coloumn except day coloumn

Data Cleaning

In [32]:

```
df[df.isnull().any(axis=1)]
```

Out[32]:

		day	month	year	Temperature	RH	Ws	Rain	FFMC	DMC	DC	ISI	BUI	FWI	Classes	
122	Sidi-Bel Abbès Region Dataset		NaN	NaN		NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
167		14	07	2012		37	37	18	0.2	88.9	12.9	14.6 ₉	12.5	10.4	fire	NaN

Observation

- .There is null columns in 122 row and 167row

Add new column with region

In [7]:

```
df.loc[:122,"Region"]=0
df.loc[122:,"Region"]=1
```

In [8]:

```
df.head()
```

Out[8]:

	day	month	year	Temperature	RH	Ws	Rain	FFMC	DMC	DC	ISI	BUI	FWI	Classes	Region	
0	01	06	2012		29	57	18	0	65.7	3.4	7.6	1.3	3.4	0.5	not fire	0.0
1	02	06	2012		29	61	13	1.3	64.4	4.1	7.6	1	3.9	0.4	not fire	0.0
2	03	06	2012		26	82	22	13.1	47.1	2.5	7.1	0.3	2.7	0.1	not fire	0.0
3	04	06	2012		25	89	13	2.5	28.6	1.3	6.9	0	1.7	0	not fire	0.0
4	05	06	2012		27	77	16	0	64.8	3	14.2	1.2	3.9	0.5	not fire	0.0

In [12]:

```
df[['Region']]=df[['Region']].astype(int)
```

In [16]:

```
df.isnull().sum()
```

Out[16]:

```
day          0
month        1
year         1
Temperature  1
RH           1
Ws           1
Rain         1
FFMC         1
DMC          1
DC           1
ISI          1
BUI          1
FWI          1
Classes      2
Region       0
dtype: int64
```

In [17]:

```
df=df.dropna().reset_index(drop=True)
```

In [18]:

```
df.shape
```

Out[18]:

```
(244, 15)
```

In [19]:

```
df.isnull().sum()
```

Out[19]:

```
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          0
Classes      0
Region       0
dtype: int64
```

In [25]:

```
df.iloc[[122]]
```

Out[25]:

	day	month	year	Temperature	RH	Ws	Rain	FFMC	DMC	DC	ISI	BUI	FWI	Classes	Region
122	day	month	year	Temperature	RH	Ws	Rain	FFMC	DMC	DC	ISI	BUI	FWI	Classes	1

In [26]:

```
## removing the122nd row
df=df.drop(122).reset_index(drop=True)
```

In [27]:

```
df.iloc[[122]]
```

Out[27]:

	day	month	year	Temperature	RH	Ws	Rain	FFMC	DMC	DC	ISI	BUI	FWI	Classes	Region
122	01	06	2012	32	71	12	0.7	57.1	2.5	8.2	0.6	2.8	0.2	not fire	1

In [30]:

```
df.duplicated().sum()
```

Out[30]:

```
0
```

Observation

- .There is no duplicate value in dataset

In [34]:

```
df.columns
```

Out[34]:

```
Index(['day', 'month', 'year', 'Temperature', ' RH', ' Ws', 'Rain ', 'FFMC',  
      'DMC', 'DC', 'ISI', 'BUI', 'FWI', 'Classes ', 'Region'],  
      dtype='object')
```

In [35]:

```
#### Fix spaces in columns names  
df.columns=df.columns.str.strip()  
df.columns
```

Out[35]:

```
Index(['day', 'month', 'year', 'Temperature', 'RH', 'Ws', 'Rain', 'FFMC',  
      'DMC', 'DC', 'ISI', 'BUI', 'FWI', 'Classes', 'Region'],  
      dtype='object')
```

In [36]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 243 entries, 0 to 242  
Data columns (total 15 columns):  
#   Column          Non-Null Count  Dtype  
---  ---  
0   day              243 non-null   object  
1   month            243 non-null   object  
2   year             243 non-null   object  
3   Temperature      243 non-null   object  
4   RH               243 non-null   object  
5   Ws               243 non-null   object  
6   Rain            243 non-null   object  
7   FFMC            243 non-null   object  
8   DMC             243 non-null   object  
9   DC              243 non-null   object  
10  ISI             243 non-null   object  
11  BUI             243 non-null   object  
12  FWI             243 non-null   object  
13  Classes         243 non-null   object  
14  Region          243 non-null   int32  
dtypes: int32(1), object(14)  
memory usage: 27.7+ KB
```

Change the required columns as integer data type

In [37]:

```
df[['month', 'day', 'year', 'Temperature', 'RH', 'Ws']] = df[['month', 'day', 'year', 'Temperature', 'RH', 'Ws']].astype(int)
```

In [38]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 243 entries, 0 to 242
Data columns (total 15 columns):
#   Column          Non-Null Count  Dtype
---  -
0   day              243 non-null   int32
1   month            243 non-null   int32
2   year             243 non-null   int32
3   Temperature      243 non-null   int32
4   RH               243 non-null   int32
5   Ws               243 non-null   int32
6   Rain             243 non-null   object
7   FFMC             243 non-null   object
8   DMC              243 non-null   object
9   DC              243 non-null   object
10  ISI              243 non-null   object
11  BUI              243 non-null   object
12  FWI              243 non-null   object
13  Classes          243 non-null   object
14  Region           243 non-null   int32
dtypes: int32(7), object(8)
memory usage: 22.0+ KB
```

Changing the other data type to float data type

In [39]:

```
objects=[features for features in df.columns if df[features].dtype!='O']
```

In [41]:

```
objects
```

Out[41]:

```
['Rain', 'FFMC', 'DMC', 'DC', 'ISI', 'BUI', 'FWI', 'Classes']
```

In [42]:

```
for i in objects:
    if i!='Classes':
        df[i]=df[i].astype(float)
```

In [43]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 243 entries, 0 to 242
Data columns (total 15 columns):
#   Column          Non-Null Count  Dtype
---  -
0   day              243 non-null   int32
1   month            243 non-null   int32
2   year             243 non-null   int32
3   Temperature      243 non-null   int32
4   RH               243 non-null   int32
5   Ws               243 non-null   int32
6   Rain             243 non-null   float64
7   FFMC             243 non-null   float64
8   DMC              243 non-null   float64
9   DC              243 non-null   float64
10  ISI              243 non-null   float64
11  BUI              243 non-null   float64
12  FWI              243 non-null   float64
13  Classes          243 non-null   object
14  Region           243 non-null   int32
dtypes: float64(7), int32(7), object(1)
memory usage: 22.0+ KB
```

In [44]:

```
df.describe()
```

Out[44]:

	day	month	year	Temperature	RH	Ws	Rain	FFMC	DMC	DC
count	243.000000	243.000000	243.0	243.000000	243.000000	243.000000	243.000000	243.000000	243.000000	243.000000
mean	15.761317	7.502058	2012.0	32.152263	62.041152	15.493827	0.762963	77.842387	14.680658	49.43081
std	8.842552	1.114793	0.0	3.628039	14.828160	2.811385	2.003207	14.349641	12.393040	47.66561
min	1.000000	6.000000	2012.0	22.000000	21.000000	6.000000	0.000000	28.600000	0.700000	6.90000
25%	8.000000	7.000000	2012.0	30.000000	52.500000	14.000000	0.000000	71.850000	5.800000	12.35000
50%	16.000000	8.000000	2012.0	32.000000	63.000000	15.000000	0.000000	83.300000	11.300000	33.10000
75%	23.000000	8.000000	2012.0	35.000000	73.500000	17.000000	0.500000	88.300000	20.800000	69.10000
max	31.000000	9.000000	2012.0	42.000000	90.000000	29.000000	16.800000	96.000000	65.900000	220.40000

In [45]:

```
df.head()
```

Out[45]:

	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	0
1	2	6	2012	29	61	13	1.3	64.4	4.1	7.6	1.0	3.9	0.4	not fire	0
2	3	6	2012	26	82	22	13.1	47.1	2.5	7.1	0.3	2.7	0.1	not fire	0
3	4	6	2012	25	89	13	2.5	28.6	1.3	6.9	0.0	1.7	0.0	not fire	0
4	5	6	2012	27	77	16	0.0	64.8	3.0	14.2	1.2	3.9	0.5	not fire	0

In [46]:

```
df.to_csv('Algerian_forest_fires_cleaned_dataset.csv',index=False)
```

EDA

In [47]:

```
df_copy=df.drop(['day','month','year'],axis=1)
```

In [48]:

```
df_copy.head()
```

Out[48]:

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

In [49]:

```
df_copy['Classes'].value_counts()
```

Out[49]:

```
fire          131
not fire      101
fire           4
fire           2
not fire       2
not fire       1
not fire       1
not fire       1
Name: Classes, dtype: int64
```

In [50]:

```
## Encoding of the categories in classes
df_copy['Classes']=np.where(df_copy['Classes'].str.contains('not fire'),0,1)
```

In [51]:

```
df_copy['Classes'].value_counts()
```

Out[51]:

```
1    137
0    106
Name: Classes, dtype: int64
```

In [52]:

```
df_copy.head()
```

Out[52]:

	Temperature	RH	Ws	Rain	FFMC	DMC	DC	ISI	BUI	FWI	Classes	Region
0	29	57	18	0.0	65.7	3.4	7.6	1.3	3.4	0.5	0	0
1	29	61	13	1.3	64.4	4.1	7.6	1.0	3.9	0.4	0	0
2	26	82	22	13.1	47.1	2.5	7.1	0.3	2.7	0.1	0	0
3	25	89	13	2.5	28.6	1.3	6.9	0.0	1.7	0.0	0	0
4	27	77	16	0.0	64.8	3.0	14.2	1.2	3.9	0.5	0	0

In [53]:

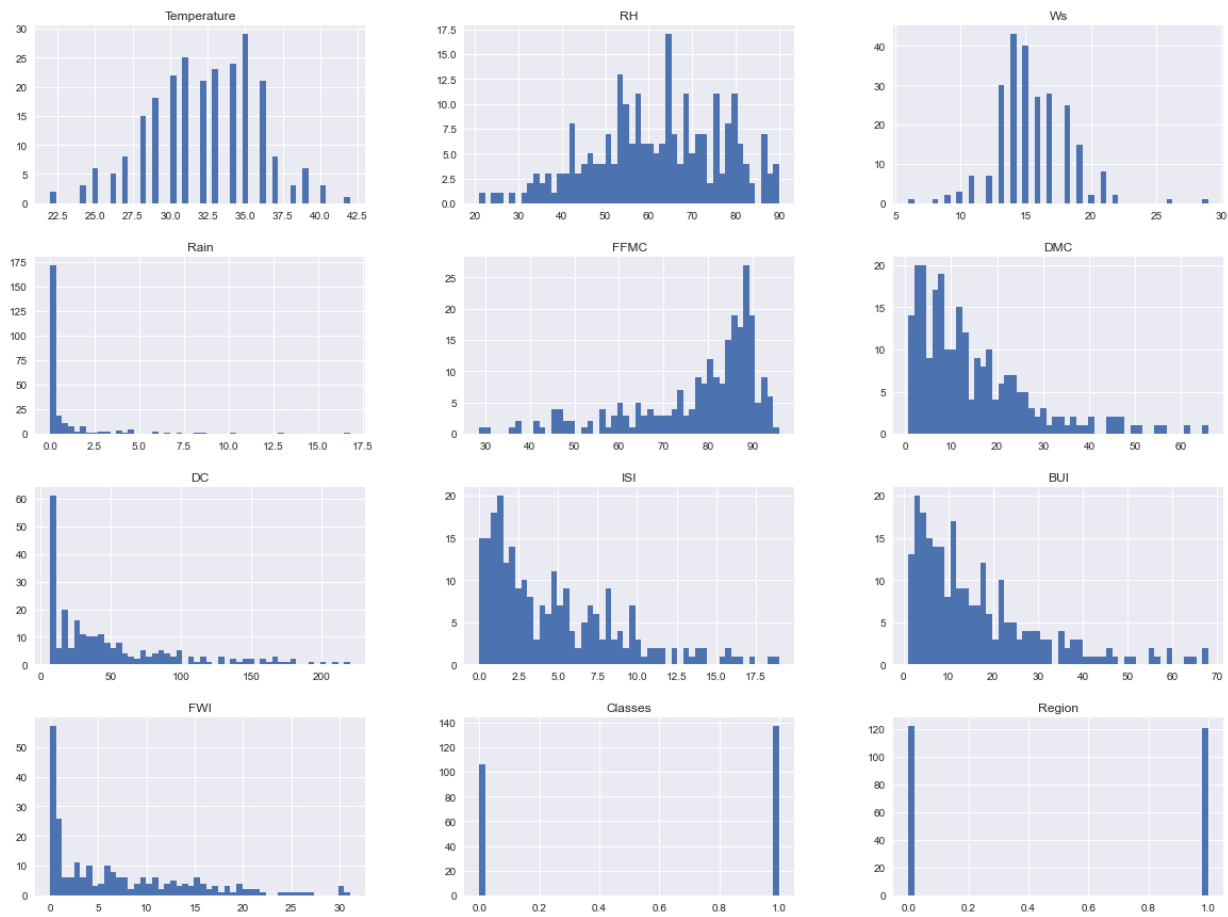
```
df_copy.tail()
```

Out[53]:

	Temperature	RH	Ws	Rain	FFMC	DMC	DC	ISI	BUI	FWI	Classes	Region
238	30	65	14	0.0	85.4	16.0	44.5	4.5	16.9	6.5	1	1
239	28	87	15	4.4	41.1	6.5	8.0	0.1	6.2	0.0	0	1
240	27	87	29	0.5	45.9	3.5	7.9	0.4	3.4	0.2	0	1
241	24	54	18	0.1	79.7	4.3	15.2	1.7	5.1	0.7	0	1
242	24	64	15	0.2	67.3	3.8	16.5	1.2	4.8	0.5	0	1

In [54]:

```
## Plot density plot for all features
plt.style.use('seaborn')
df_copy.hist(bins=50,figsize=(20,15))
plt.show()
```

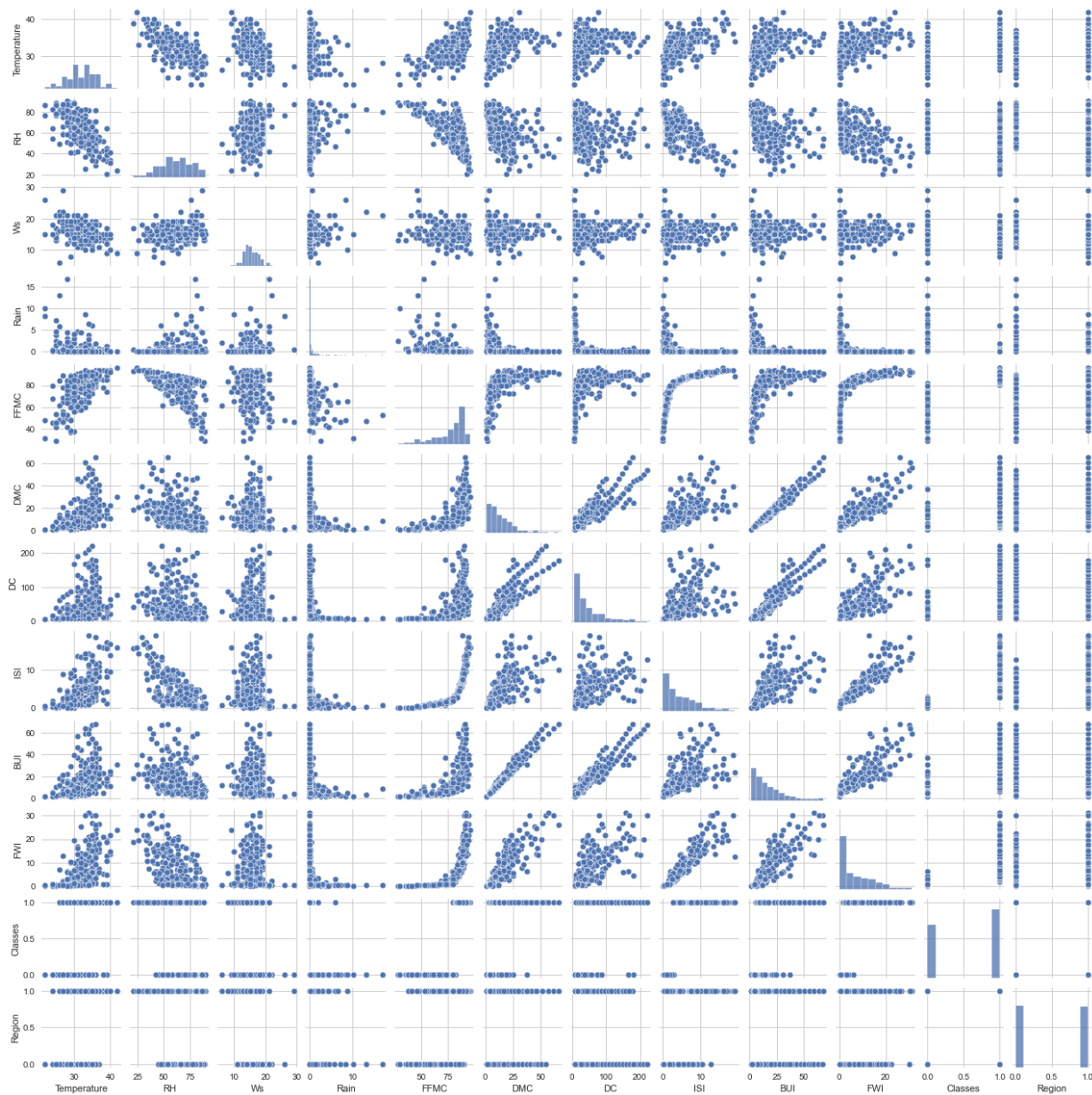


In [187]:

```
sns.pairplot(df,height=1.5,  
aspect=1,)
```

Out[187]:

<seaborn.axisgrid.PairGrid at 0x21a909778e0>

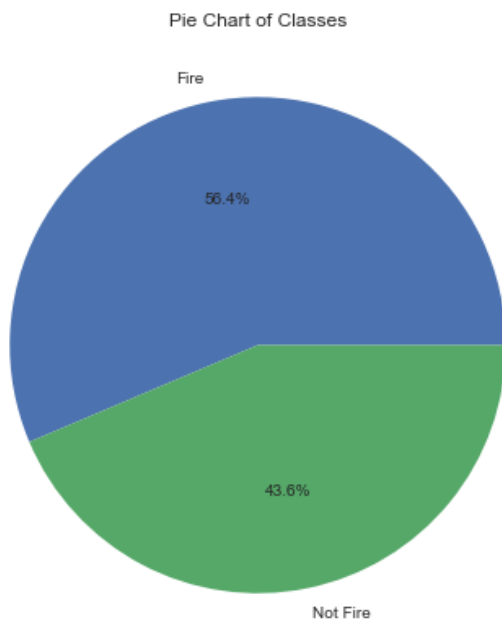


In [55]:

```
## Percentages for pie chart  
percentage=df_copy['Classes'].value_counts(normalize=True)*100
```

In [67]:

```
## Plotting Piechart
classlabels=["Fire","Not Fire"]
plt.figure(figsize=(12,7))
plt.pie(percentage,labels=classlabels,autopct='%1.1f%%')
plt.title("Pie Chart of Classes")
plt.show()
```



Observation

- 56.4% are the chances that forests catch fires

In [68]:

```
## Correlation
df_copy.corr()
```

Out[68]:

	Temperature	RH	Ws	Rain	FFMC	DMC	DC	ISI	BUI	FWI
Temperature	1.000000	-0.651400	-0.284510	-0.326492	0.676568	0.485687	0.376284	0.603871	0.459789	0.566670
RH	-0.651400	1.000000	0.244048	0.222356	-0.644873	-0.408519	-0.226941	-0.686667	-0.353841	-0.580957
Ws	-0.284510	0.244048	1.000000	0.171506	-0.166548	-0.000721	0.079135	0.008532	0.031438	0.032368
Rain	-0.326492	0.222356	0.171506	1.000000	-0.543906	-0.288773	-0.298023	-0.347484	-0.299852	-0.324422
FFMC	0.676568	-0.644873	-0.166548	-0.543906	1.000000	0.603608	0.507397	0.740007	0.592011	0.691132
DMC	0.485687	-0.408519	-0.000721	-0.288773	0.603608	1.000000	0.875925	0.680454	0.982248	0.875864
DC	0.376284	-0.226941	0.079135	-0.298023	0.507397	0.875925	1.000000	0.508643	0.941988	0.739521
ISI	0.603871	-0.686667	0.008532	-0.347484	0.740007	0.680454	0.508643	1.000000	0.644093	0.922895
BUI	0.459789	-0.353841	0.031438	-0.299852	0.592011	0.982248	0.941988	0.644093	1.000000	0.857973
FWI	0.566670	-0.580957	0.032368	-0.324422	0.691132	0.875864	0.739521	0.922895	0.857973	1.000000
Classes	0.516015	-0.432161	-0.069964	-0.379097	0.769492	0.585658	0.511123	0.735197	0.586639	0.719216
Region	0.269555	-0.402682	-0.181160	-0.040013	0.222241	0.192089	-0.078734	0.263197	0.089408	0.197102

In [76]:

```
sns.heatmap(df.corr(),annot=True)
```

Out[76]:

<AxesSubplot:>

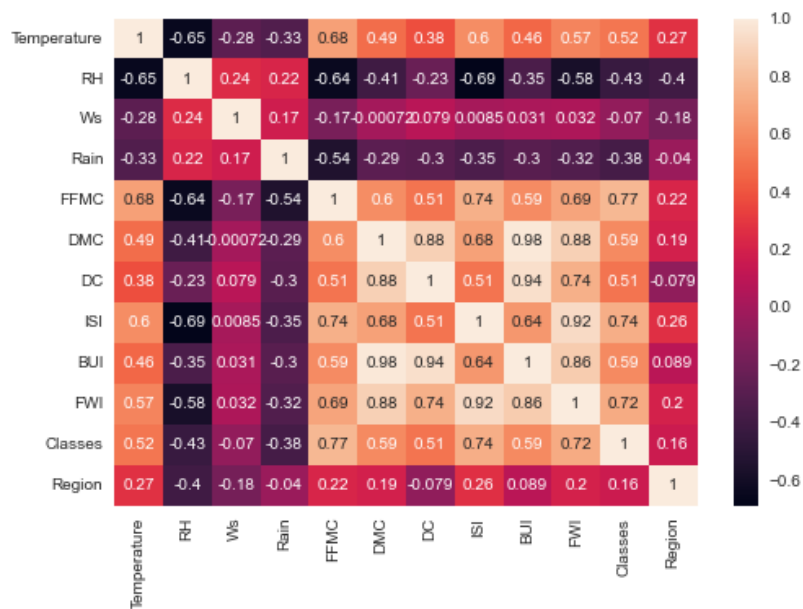


In [75]:

```
sns.heatmap(df_copy.corr(),annot=True)
```

Out[75]:

<AxesSubplot:>

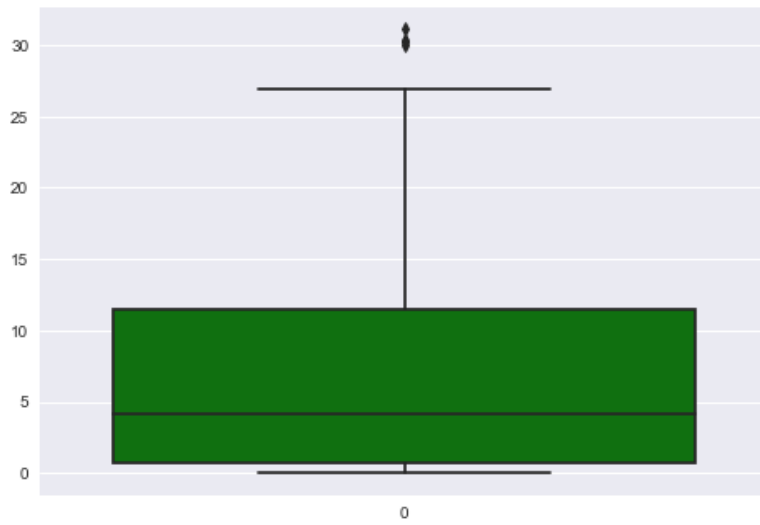


In [71]:

```
## Box Plots
sns.boxplot(df['FWI'],color='green')
```

Out[71]:

<AxesSubplot:>

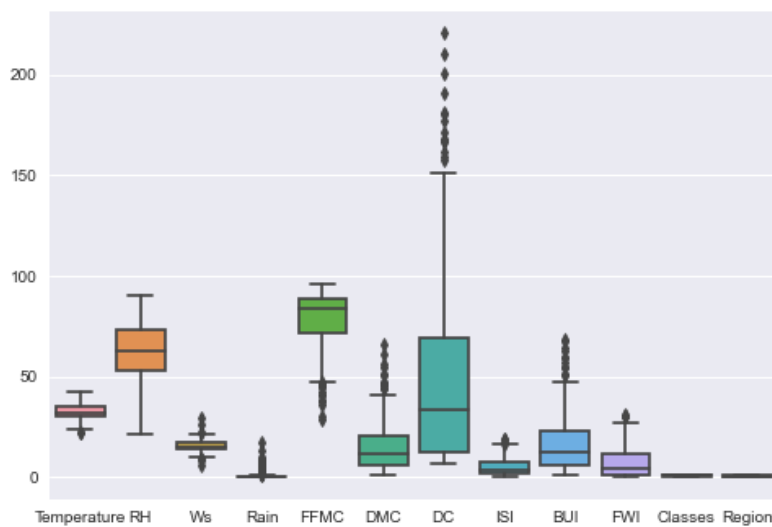


In [73]:

```
sns.boxplot(df_copy)
```

Out[73]:

<AxesSubplot:>



In [77]:

```
df.head()
```

Out[77]:

	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	0
1	2	6	2012	29	61	13	1.3	64.4	4.1	7.6	1.0	3.9	0.4	not fire	0
2	3	6	2012	26	82	22	13.1	47.1	2.5	7.1	0.3	2.7	0.1	not fire	0
3	4	6	2012	25	89	13	2.5	28.6	1.3	6.9	0.0	1.7	0.0	not fire	0
4	5	6	2012	27	77	16	0.0	64.8	3.0	14.2	1.2	3.9	0.5	not fire	0

In [82]:

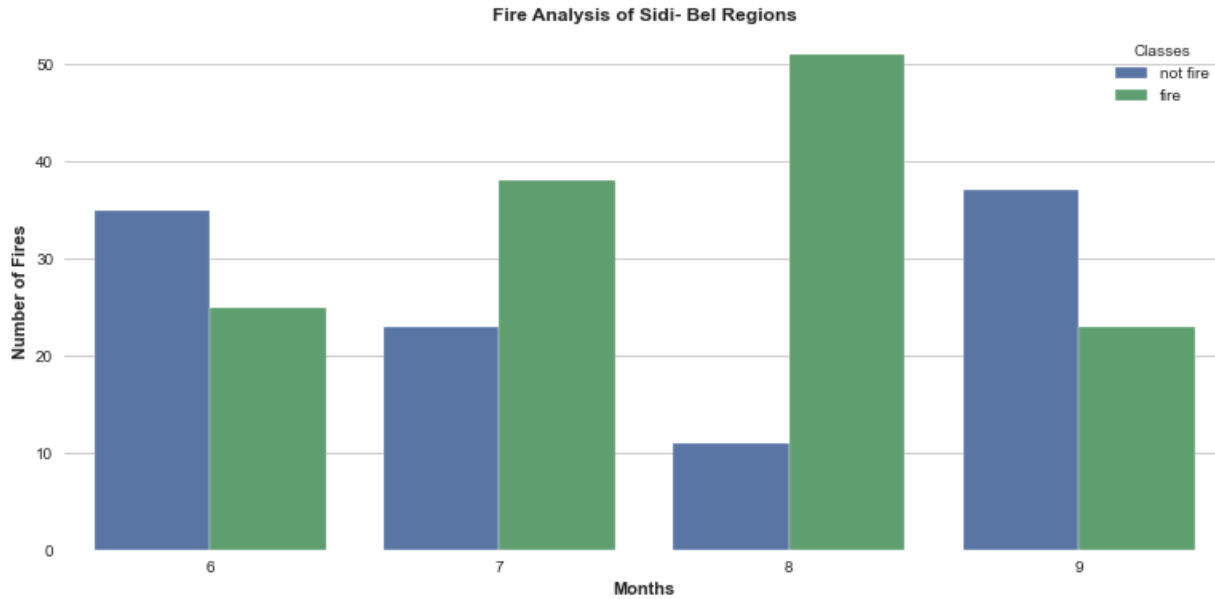
```
df['Classes']=np.where(df['Classes'].str.contains('not fire'),'not fire','fire')
```

In [83]:

```
## Monthly Fire Analysis
dftemp=df.loc[df['Region']==1]
plt.subplots(figsize=(13,6))
sns.set_style('whitegrid')
sns.countplot(x='month',hue='Classes',data=df)
plt.ylabel('Number of Fires',weight='bold')
plt.xlabel('Months',weight='bold')
plt.title("Fire Analysis of Sidi- Bel Regions",weight='bold')
```

Out[83]:

Text(0.5, 1.0, 'Fire Analysis of Sidi- Bel Regions')

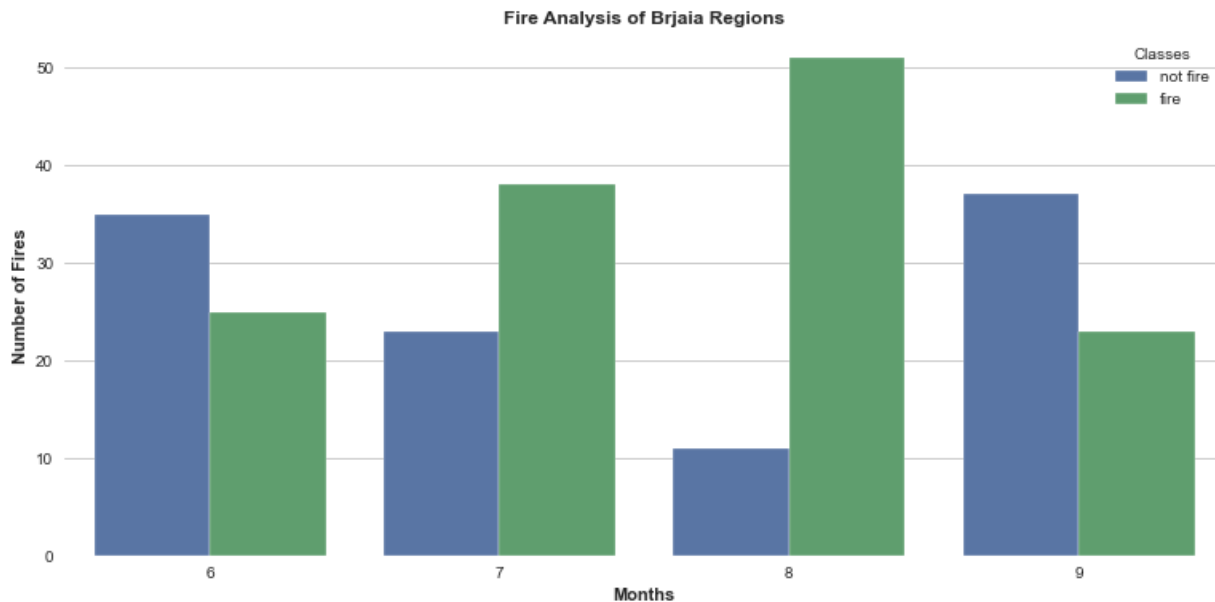


In [84]:

```
## Monthly Fire Analysis
dftemp=df.loc[df['Region']==0]
plt.subplots(figsize=(13,6))
sns.set_style('whitegrid')
sns.countplot(x='month',hue='Classes',data=df)
plt.ylabel('Number of Fires',weight='bold')
plt.xlabel('Months',weight='bold')
plt.title("Fire Analysis of Brjaia Regions",weight='bold')
```

Out[84]:

Text(0.5, 1.0, 'Fire Analysis of Brjaia Regions')



Observation

- Its observed that July and August had the most number of forest fires for both regions
- Most of the fires happend in August and very high fires happend in only 3 months - Junes , July and August
- There is a very less fires on September