In [1]:
```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
```

In [2]:
```
dt=pd.read_csv("Algerian_forest_fires_dataset_UPDATE.csv",header=1)
```

In [3]:
```
dt.head()
```

Out[3]:

	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]:
```
dt.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 246 entries, 0 to 245
Data columns (total 14 columns):
 #   Column       Non-Null Count  Dtype 
---  ------       --------------  ----- 
 0   day          246 non-null    object
 1   month        245 non-null    object
 2   year         245 non-null    object
 3   Temperature  245 non-null    object
 4    RH          245 non-null    object
 5    Ws          245 non-null    object
 6   Rain         245 non-null    object
 7   FFMC         245 non-null    object
 8   DMC          245 non-null    object
 9   DC           245 non-null    object
 10  ISI          245 non-null    object
 11  BUI          245 non-null    object
 12  FWI          245 non-null    object
 13  Classes      244 non-null    object
dtypes: object(14)
memory usage: 27.0+ KB
```

#data cleaning

In [5]:
```
dt[dt.isnull().any(axis=1)]
```

Out[5]:

	day	month	year	Temperature	RH	Ws	Rain	FFMC	DMC	DC	ISI	BUI	FWI
122	Sidi-Bel Abbes 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	14.69	12.5	10.4	fire

In [6]:
```
dt.loc[:122,"Region"]=0
dt.loc[122:,"Region"]=1
df=dt
```

In [7]:
```
df.head()
```

Out[7]:

	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 [8]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 246 entries, 0 to 245
Data columns (total 15 columns):
 #   Column       Non-Null Count  Dtype  
---  ------       --------------  -----  
 0   day          246 non-null    object 
 1   month        245 non-null    object 
 2   year         245 non-null    object 
 3   Temperature  245 non-null    object 
 4    RH          245 non-null    object 
 5    Ws          245 non-null    object 
 6   Rain         245 non-null    object 
 7   FFMC         245 non-null    object 
 8   DMC          245 non-null    object 
 9   DC           245 non-null    object 
 10  ISI          245 non-null    object 
 11  BUI          245 non-null    object 
 12  FWI          245 non-null    object 
 13  Classes      244 non-null    object 
 14  Region       246 non-null    float64
dtypes: float64(1), object(14)
memory usage: 29.0+ KB
```

In [9]:
```
df[["Region"]]=df[["Region"]].astype(int)
```

In [10]:
```
df.head()
```

Out[10]:

	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 [11]:
```
df.tail()
```

Out[11]:

	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 [12]:
```
df.isnull().sum()
```

Out[12]:
```
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 [13]:
```
df=df.dropna().reset_index(drop=True)
```

In [14]:
```
df.isnull().sum()
```

Out[14]:
```
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 [15]:
```
df.iloc[[122]]
```

Out[15]:

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

In [16]:
```
#remove the 122 row
df=df.drop(122).reset_index(drop=True)
```

In [17]:
```
df.iloc[[122]]
```

Out[17]:

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

In [18]:
```
df.columns
```

Out[18]:
```
Index(['day', 'month', 'year', 'Temperature', ' RH', ' Ws', 'Rain ', 'FFMC',
       'DMC', 'DC', 'ISI', 'BUI', 'FWI', 'Classes  ', 'Region'],
      dtype='object')
```

In [19]:
```
#fix spaces in col rows
df.columns=df.columns.str.strip()
df.columns
```

Out[19]:
```
Index(['day', 'month', 'year', 'Temperature', 'RH', 'Ws', 'Rain', 'FFMC',
       'DMC', 'DC', 'ISI', 'BUI', 'FWI', 'Classes', 'Region'],
      dtype='object')
```

In [20]:
```
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
```

In [21]:
```
df[['month','day','year','Temperature','RH','Ws']]=df[['month','day','year','T
```

In [22]:
```
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
```

In [23]:
```
df.head()
```

Out[23]:

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

In [24]:
```
#change the other columns to float datatype
obj=[features for features in df.columns if df[features].dtypes=='O']
obj
```

Out[24]:
```
['Rain', 'FFMC', 'DMC', 'DC', 'ISI', 'BUI', 'FWI', 'Classes']
```

In [25]:
```
for i in obj:
    if i!='Classes':
        df[i]=df[i].astype(float)
```

In [26]:
```
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 [27]:
```
obj
```

Out[27]:
```
['Rain', 'FFMC', 'DMC', 'DC', 'ISI', 'BUI', 'FWI', 'Classes']
```

In [28]:
```
df.describe()
```

Out[28]:

	day	month	year	Temperature	RH	Ws	Rain	F
count	243.000000	243.000000	243.0	243.000000	243.000000	243.000000	243.000000	243.00
mean	15.761317	7.502058	2012.0	32.152263	62.041152	15.493827	0.762963	77.84
std	8.842552	1.114793	0.0	3.628039	14.828160	2.811385	2.003207	14.34
min	1.000000	6.000000	2012.0	22.000000	21.000000	6.000000	0.000000	28.60
25%	8.000000	7.000000	2012.0	30.000000	52.500000	14.000000	0.000000	71.85
50%	16.000000	8.000000	2012.0	32.000000	63.000000	15.000000	0.000000	83.30
75%	23.000000	8.000000	2012.0	35.000000	73.500000	17.000000	0.500000	88.30
max	31.000000	9.000000	2012.0	42.000000	90.000000	29.000000	16.800000	96.00

In [29]:
```
#lect save the the cleaned file
df.to_csv("Algerian_forest_fires_cleaned_dataset.csv",index=False)
```

#EDA

In [30]:
```
df_copy=df.drop(['day','year','month'],axis=1)
```

In [31]:
```
df_copy.head()
```

Out[31]:

	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 [32]:
```
df_copy["Classes"].value_counts()
```

Out[32]:
```
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 [33]:
```
df_copy["Classes"]=np.where(df_copy["Classes"].str.contains('not fire'),0,1)
```

In [34]:
```
df_copy.head()
```

Out[34]:

	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 [35]:
```
df_copy.tail()
```

Out[35]:

	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 [36]:
```
df_copy["Classes"].value_counts()
```

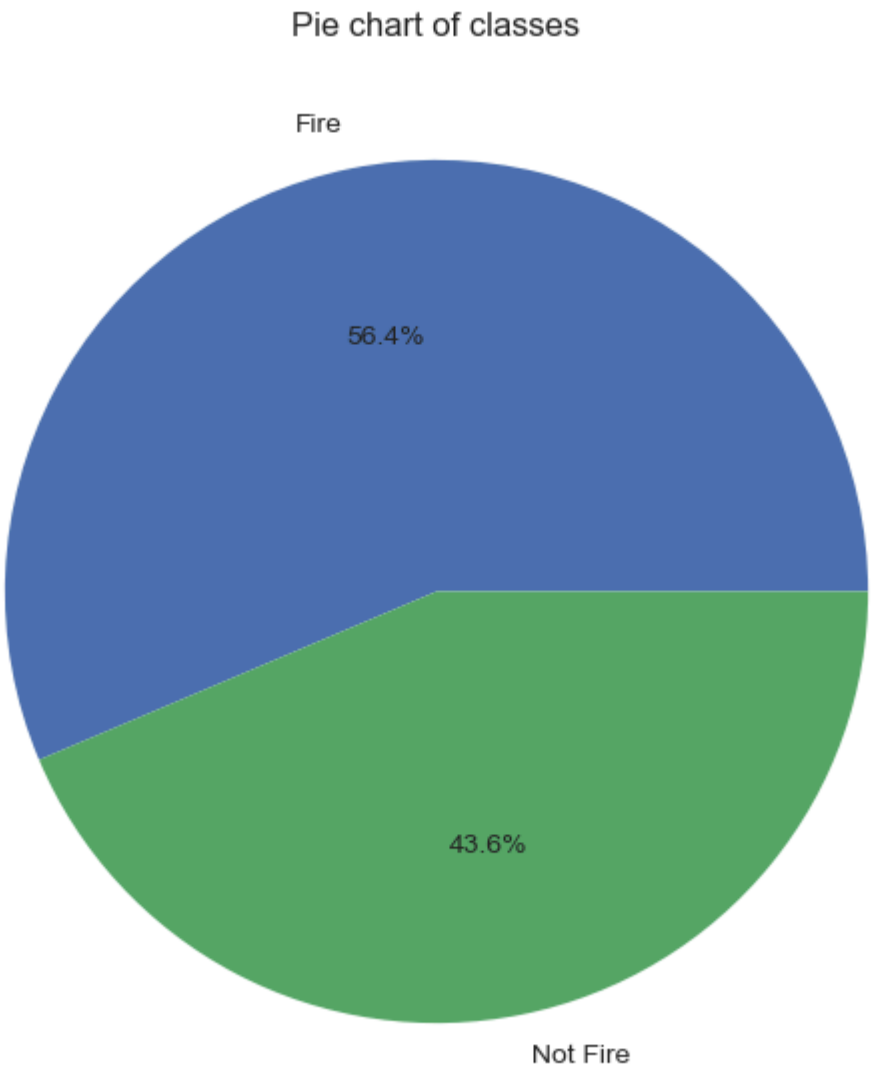
Out[36]:
```
1    137
0    106
Name: Classes, dtype: int64
```

In [37]:
```
#plot density plot for all features
plt.style.use('seaborn')
df_copy.hist(bins=50,figsize=(20,15))
plt.show()
```

```
C:\Users\Hp\AppData\Local\Temp\ipykernel_6568\3743561007.py:2: MatplotlibDep
recationWarning: The seaborn styles shipped by Matplotlib are deprecated sin
ce 3.6, as they no longer correspond to the styles shipped by seaborn. Howev
er, they will remain available as 'seaborn-v0_8-<style>'. Alternatively, dir
ectly use the seaborn API instead.
  plt.style.use('seaborn')
```

In [38]:
```
percentage=df_copy["Classes"].value_counts(normalize=True)*100
```

In [39]:
```
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()
```



In [40]:
```
#correlation
df_copy.corr()
```

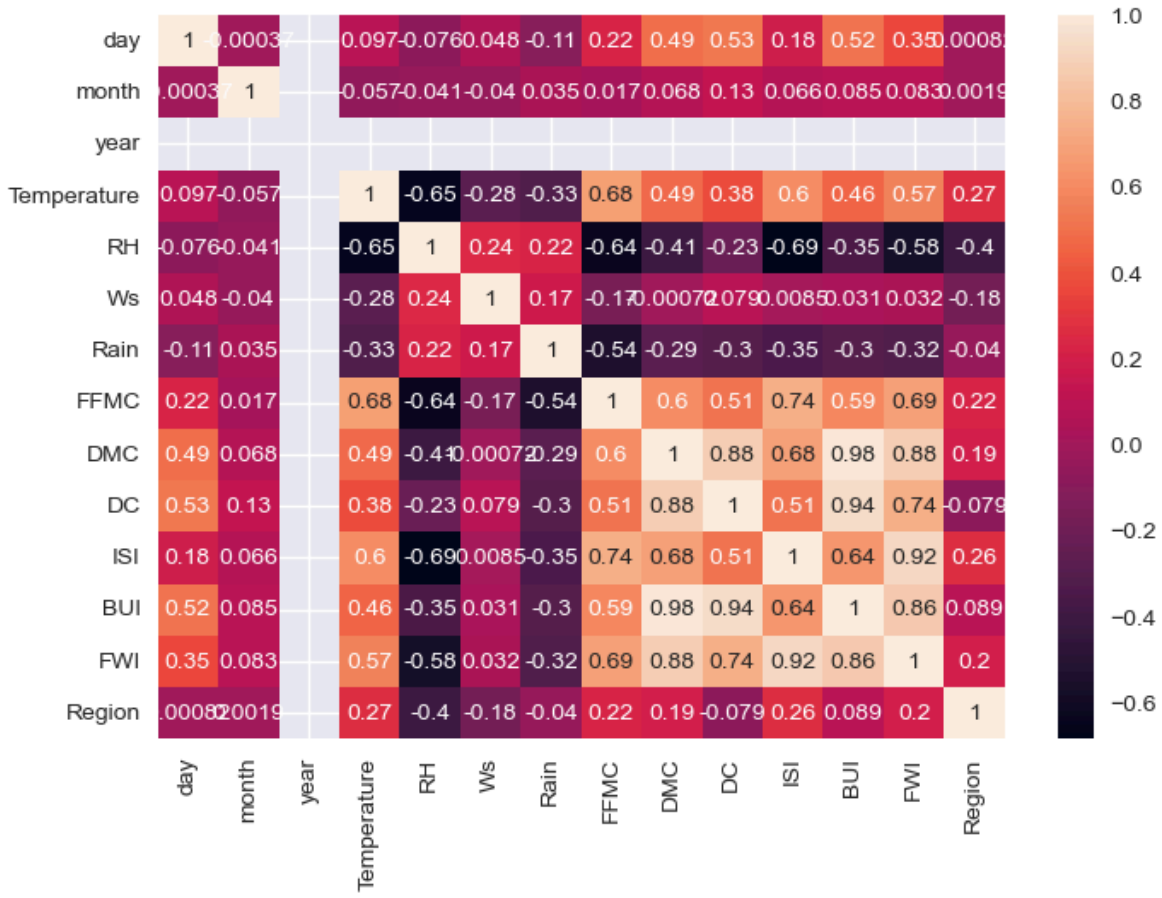
Out[40]:

	Temperature	RH	Ws	Rain	FFMC	DMC	DC	
Temperature	1.000000	-0.651400	-0.284510	-0.326492	0.676568	0.485687	0.376284	0.60
RH	-0.651400	1.000000	0.244048	0.222356	-0.644873	-0.408519	-0.226941	-0.68
Ws	-0.284510	0.244048	1.000000	0.171506	-0.166548	-0.000721	0.079135	0.00
Rain	-0.326492	0.222356	0.171506	1.000000	-0.543906	-0.288773	-0.298023	-0.34
FFMC	0.676568	-0.644873	-0.166548	-0.543906	1.000000	0.603608	0.507397	0.74
DMC	0.485687	-0.408519	-0.000721	-0.288773	0.603608	1.000000	0.875925	0.68
DC	0.376284	-0.226941	0.079135	-0.298023	0.507397	0.875925	1.000000	0.50
ISI	0.603871	-0.686667	0.008532	-0.347484	0.740007	0.680454	0.508643	1.00
BUI	0.459789	-0.353841	0.031438	-0.299852	0.592011	0.982248	0.941988	0.64
FWI	0.566670	-0.580957	0.032368	-0.324422	0.691132	0.875864	0.739521	0.92
Classes	0.516015	-0.432161	-0.069964	-0.379097	0.769492	0.585658	0.511123	0.73
Region	0.269555	-0.402682	-0.181160	-0.040013	0.222241	0.192089	-0.078734	0.26

In [41]:
```
sns.heatmap(df.corr(),annot=True)
```

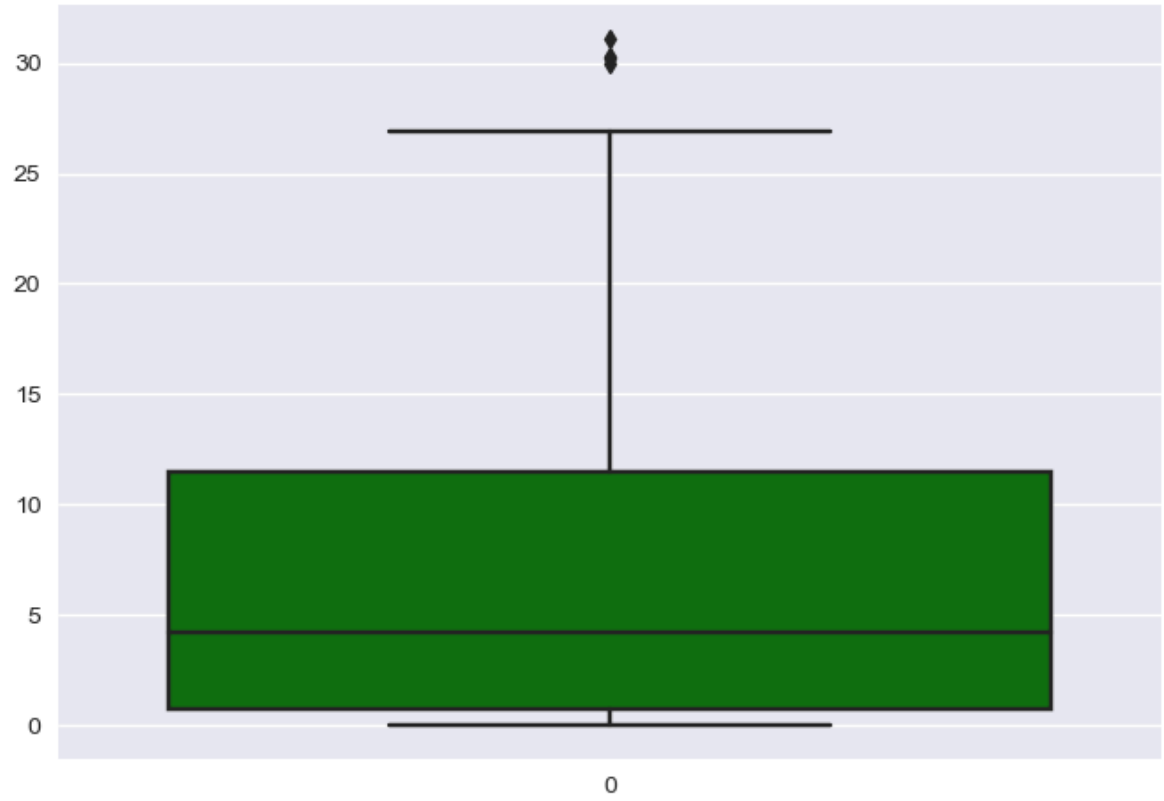
```
C:\Users\Hp\AppData\Local\Temp\ipykernel_6568\4277794465.py:1: FutureWarnin
g: The default value of numeric_only in DataFrame.corr is deprecated. In a f
uture version, it will default to False. Select only valid columns or specif
y the value of numeric_only to silence this warning.
  sns.heatmap(df.corr(),annot=True)
```

Out[41]: <Axes: >



In [42]:
```
sns.boxplot(df["FWI"],color="green")
```

Out[42]: <Axes: >



In [43]:
```
df.head()
```

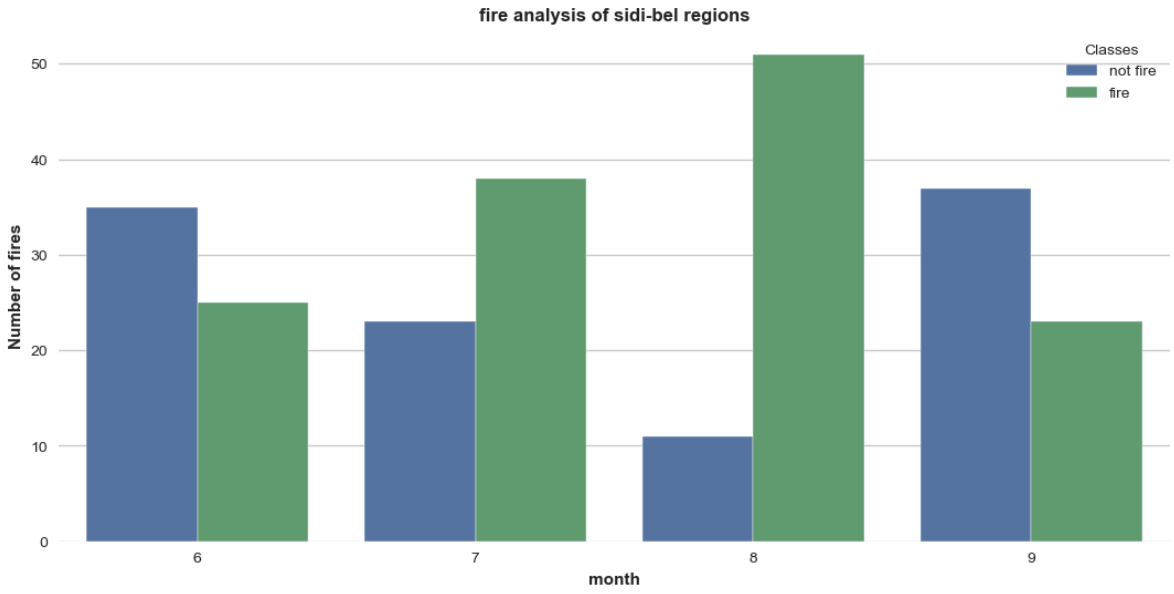
Out[43]:

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

In [44]:
```
df["Classes"]=np.where(df["Classes"].str.contains('not fire'),'not fire','fire
```

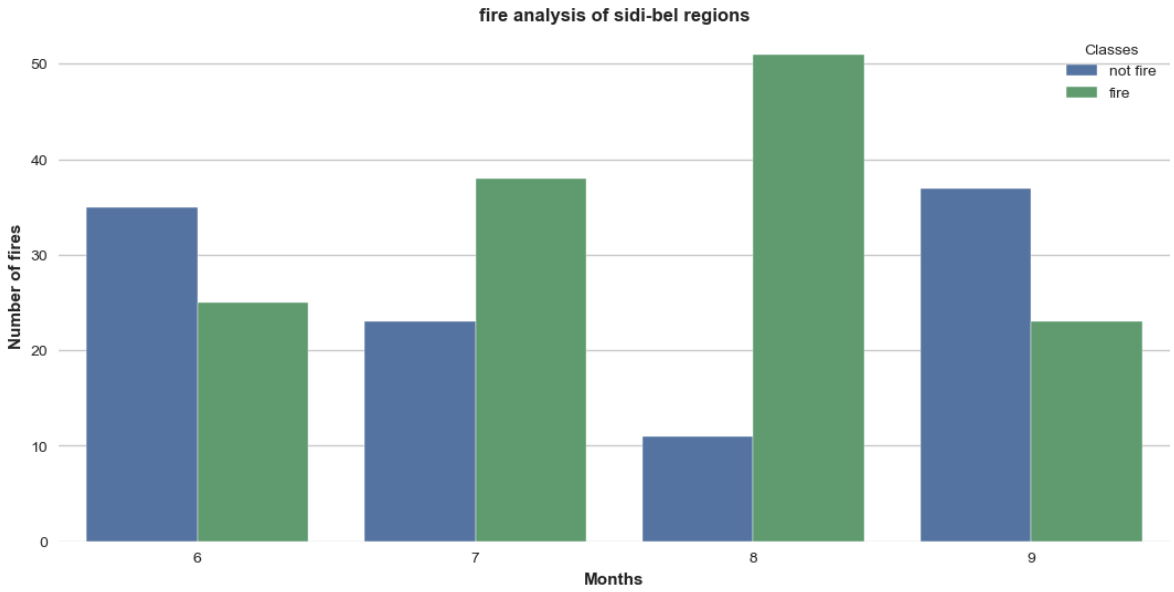
In [47]:
```
#monthly fire analysis
dftemp=df.loc[df["Region"]==1]
plt.subplots(figsize=(13,6))
sns.set_style("whitegrid")
sns.countplot(x="month",hue="Classes",data=df)
plt.ylabel("Number of fires",weight="bold")
plt.xlabel("month",weight="bold")
plt.title("fire analysis of sidi-bel regions",weight="bold")
```

Out[47]:
```
Text(0.5, 1.0, 'fire analysis of sidi-bel regions')
```



In [48]:
```
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 sidi-bel regions",weight="bold")
```

Out[48]:
```
Text(0.5, 1.0, 'fire analysis of sidi-bel regions')
```



it is observed that august and sptember had the most number of forest fires for both regions.and from the above plot of the moths,we can undestand the things. most of the fires happened in august and very high fires happend in 3 mohs-june,july nad augaust

less fires was on september

In [51]:
```
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
Cell In[51], line 1
----> 1 x_train.corr()

NameError: name 'x_train' is not defined
```

In [ ]:
```
```