In [1]:
```
import pandas as pd
import numpy as np
```

In [2]:
```
pwd
```

Out[2]:
```
'C:\\Users\\Nancy'
```

In [3]:
```
df=pd.read_csv("C:\\Users\\Nancy\\Desktop\\Climate.csv")
print(df)
```

```
         day  temperature  windspeed   event  duration
0  01-01-2017         32.0        6.0    Rain         5
1  01-04-2017          NaN        9.0   Sunny         4
2  01-05-2017         28.0        NaN    Snow         3
3  01-06-2017          NaN        7.0     NaN         2
4  01-07-2017         32.0        NaN    Rain         1
5  01-08-2017          NaN        NaN   Sunny         5
6  01-09-2017          NaN        NaN     NaN         3
7  01-10-2017         34.0        8.0  Cloudy         2
8  01-11-2017         40.0       12.0   Sunny         1
```

In [4]:
```
df.head()
```

Out[4]:

| | day | temperature | windspeed | event | duration |
|---|---|---|---|---|---|
| **0** | 01-01-2017 | 32.0 | 6.0 | Rain | 5 |
| **1** | 01-04-2017 | NaN | 9.0 | Sunny | 4 |
| **2** | 01-05-2017 | 28.0 | NaN | Snow | 3 |
| **3** | 01-06-2017 | NaN | 7.0 | NaN | 2 |
| **4** | 01-07-2017 | 32.0 | NaN | Rain | 1 |

In [5]:
```
df.tail()
```

Out[5]:

| | day | temperature | windspeed | event | duration |
|---|---|---|---|---|---|
| **4** | 01-07-2017 | 32.0 | NaN | Rain | 1 |
| **5** | 01-08-2017 | NaN | NaN | Sunny | 5 |
| **6** | 01-09-2017 | NaN | NaN | NaN | 3 |
| **7** | 01-10-2017 | 34.0 | 8.0 | Cloudy | 2 |
| **8** | 01-11-2017 | 40.0 | 12.0 | Sunny | 1 |

In [6]:
```
df.shape
```

Out[6]:
```
(9, 5)
```

In [8]:
```
df.replace(" ",np.nan, inplace=True)
df.head(9)
```

Out[8]:

| | day | temperature | windspeed | event | duration |
|---|---|---|---|---|---|
| **0** | 01-01-2017 | 32.0 | 6.0 | Rain | 5 |
| **1** | 01-04-2017 | NaN | 9.0 | Sunny | 4 |
| **2** | 01-05-2017 | 28.0 | NaN | Snow | 3 |

| | day | temperature | windspeed | event | duration |
|---|---|---|---|---|---|
| **3** | 01-06-2017 | NaN | 7.0 | NaN | 2 |
| **4** | 01-07-2017 | 32.0 | NaN | Rain | 1 |
| **5** | 01-08-2017 | NaN | NaN | Sunny | 5 |
| **6** | 01-09-2017 | NaN | NaN | NaN | 3 |
| **7** | 01-10-2017 | 34.0 | 8.0 | Cloudy | 2 |
| **8** | 01-11-2017 | 40.0 | 12.0 | Sunny | 1 |

In [9]:
```
df.head()
```

Out[9]:

| | day | temperature | windspeed | event | duration |
|---|---|---|---|---|---|
| **0** | 01-01-2017 | 32.0 | 6.0 | Rain | 5 |
| **1** | 01-04-2017 | NaN | 9.0 | Sunny | 4 |
| **2** | 01-05-2017 | 28.0 | NaN | Snow | 3 |
| **3** | 01-06-2017 | NaN | 7.0 | NaN | 2 |
| **4** | 01-07-2017 | 32.0 | NaN | Rain | 1 |

In [10]:
```
missingdata = df.isnull()
print(missingdata)
```

```
     day  temperature  windspeed  event  duration
0  False        False      False  False     False
1  False         True      False  False     False
2  False        False       True  False     False
3  False         True      False   True     False
4  False        False       True  False     False
5  False         True       True  False     False
6  False         True       True   True     False
7  False        False      False  False     False
8  False        False      False  False     False
```

In [11]:
```
df.isnull().sum()
```

Out[11]:
```
day            0
temperature    4
windspeed      4
event          2
duration       0
dtype: int64
```

In [12]:
```
df.dropna()
```

Out[12]:

| | day | temperature | windspeed | event | duration |
|---|---|---|---|---|---|
| **0** | 01-01-2017 | 32.0 | 6.0 | Rain | 5 |
| **7** | 01-10-2017 | 34.0 | 8.0 | Cloudy | 2 |
| **8** | 01-11-2017 | 40.0 | 12.0 | Sunny | 1 |

In [13]:
```
df
```

Out[13]:

| | day | temperature | windspeed | event | duration |
|---|---|---|---|---|---|

| | day | temperature | windspeed | event | duration |
|---|---|---|---|---|---|
| **0** | 01-01-2017 | 32.0 | 6.0 | Rain | 5 |
| **1** | 01-04-2017 | NaN | 9.0 | Sunny | 4 |
| **2** | 01-05-2017 | 28.0 | NaN | Snow | 3 |
| **3** | 01-06-2017 | NaN | 7.0 | NaN | 2 |
| **4** | 01-07-2017 | 32.0 | NaN | Rain | 1 |
| **5** | 01-08-2017 | NaN | NaN | Sunny | 5 |
| **6** | 01-09-2017 | NaN | NaN | NaN | 3 |
| **7** | 01-10-2017 | 34.0 | 8.0 | Cloudy | 2 |
| **8** | 01-11-2017 | 40.0 | 12.0 | Sunny | 1 |

In [14]:
```
df.dropna(axis=1)
```

Out[14]:

| | day | duration |
|---|---|---|
| **0** | 01-01-2017 | 5 |
| **1** | 01-04-2017 | 4 |
| **2** | 01-05-2017 | 3 |
| **3** | 01-06-2017 | 2 |
| **4** | 01-07-2017 | 1 |
| **5** | 01-08-2017 | 5 |
| **6** | 01-09-2017 | 3 |
| **7** | 01-10-2017 | 2 |
| **8** | 01-11-2017 | 1 |

In [15]:
```
df.dropna(how='all')
```

Out[15]:

| | day | temperature | windspeed | event | duration |
|---|---|---|---|---|---|
| **0** | 01-01-2017 | 32.0 | 6.0 | Rain | 5 |
| **1** | 01-04-2017 | NaN | 9.0 | Sunny | 4 |
| **2** | 01-05-2017 | 28.0 | NaN | Snow | 3 |
| **3** | 01-06-2017 | NaN | 7.0 | NaN | 2 |
| **4** | 01-07-2017 | 32.0 | NaN | Rain | 1 |
| **5** | 01-08-2017 | NaN | NaN | Sunny | 5 |
| **6** | 01-09-2017 | NaN | NaN | NaN | 3 |
| **7** | 01-10-2017 | 34.0 | 8.0 | Cloudy | 2 |
| **8** | 01-11-2017 | 40.0 | 12.0 | Sunny | 1 |

In [16]:
```
df.dropna(subset=["event"])
```

Out[16]:

| | day | temperature | windspeed | event | duration |
|---|---|---|---|---|---|

| | day | temperature | windspeed | event | duration |
|---|---|---|---|---|---|
| **0** | 01-01-2017 | 32.0 | 6.0 | Rain | 5 |
| **1** | 01-04-2017 | NaN | 9.0 | Sunny | 4 |
| **2** | 01-05-2017 | 28.0 | NaN | Snow | 3 |
| **4** | 01-07-2017 | 32.0 | NaN | Rain | 1 |
| **5** | 01-08-2017 | NaN | NaN | Sunny | 5 |
| **7** | 01-10-2017 | 34.0 | 8.0 | Cloudy | 2 |
| **8** | 01-11-2017 | 40.0 | 12.0 | Sunny | 1 |

In [18]:
```
mean_value=df["temperature"].mean()
print(mean_value)
```

```
33.2
```

In [19]:
```
mean_value1=df["windspeed"].mean()
print(mean_value1)
```

```
8.4
```

In [20]:
```
df["temperature"] = df["temperature"].fillna(mean_value)
print(df)
```

```
          day  temperature  windspeed    event  duration
0  01-01-2017         32.0        6.0     Rain         5
1  01-04-2017         33.2        9.0    Sunny         4
2  01-05-2017         28.0        NaN     Snow         3
3  01-06-2017         33.2        7.0      NaN         2
4  01-07-2017         32.0        NaN     Rain         1
5  01-08-2017         33.2        NaN    Sunny         5
6  01-09-2017         33.2        NaN      NaN         3
7  01-10-2017         34.0        8.0   Cloudy         2
8  01-11-2017         40.0       12.0    Sunny         1
```

In [21]:
```
df["windspeed"] = df["windspeed"].fillna(mean_value1)
print(df)
```

```
          day  temperature  windspeed    event  duration
0  01-01-2017         32.0        6.0     Rain         5
1  01-04-2017         33.2        9.0    Sunny         4
2  01-05-2017         28.0        8.4     Snow         3
3  01-06-2017         33.2        7.0      NaN         2
4  01-07-2017         32.0        8.4     Rain         1
5  01-08-2017         33.2        8.4    Sunny         5
6  01-09-2017         33.2        8.4      NaN         3
7  01-10-2017         34.0        8.0   Cloudy         2
8  01-11-2017         40.0       12.0    Sunny         1
```

In [22]:
```
new_df = df.fillna({'temperature' : 0, 'windspeed' : 0, 'event' : 'no_event'})
print(new_df)
```

```
          day  temperature  windspeed      event  duration
0  01-01-2017         32.0        6.0       Rain         5
1  01-04-2017         33.2        9.0      Sunny         4
2  01-05-2017         28.0        8.4       Snow         3
3  01-06-2017         33.2        7.0  no_event         2
4  01-07-2017         32.0        8.4       Rain         1
5  01-08-2017         33.2        8.4      Sunny         5
6  01-09-2017         33.2        8.4  no_event         3
7  01-10-2017         34.0        8.0     Cloudy         2
8  01-11-2017         40.0       12.0      Sunny         1
```

In [23]:
```
df.dtypes
```

Out[23]:
```
day             object
temperature    float64
windspeed      float64
event           object
duration         int64
dtype: object
```

In [25]:
```
df[["duration"]]=df[["duration"]].astype("float")
```

In [26]:
```
df.dtypes
```

Out[26]:
```
day             object
temperature    float64
windspeed      float64
event           object
duration       float64
dtype: object
```

In [27]:
```
df
```

Out[27]:

| | day | temperature | windspeed | event | duration |
|---|---|---|---|---|---|
| **0** | 01-01-2017 | 32.0 | 6.0 | Rain | 5.0 |
| **1** | 01-04-2017 | 33.2 | 9.0 | Sunny | 4.0 |
| **2** | 01-05-2017 | 28.0 | 8.4 | Snow | 3.0 |
| **3** | 01-06-2017 | 33.2 | 7.0 | NaN | 2.0 |
| **4** | 01-07-2017 | 32.0 | 8.4 | Rain | 1.0 |
| **5** | 01-08-2017 | 33.2 | 8.4 | Sunny | 5.0 |
| **6** | 01-09-2017 | 33.2 | 8.4 | NaN | 3.0 |
| **7** | 01-10-2017 | 34.0 | 8.0 | Cloudy | 2.0 |
| **8** | 01-11-2017 | 40.0 | 12.0 | Sunny | 1.0 |

In [28]:
```
df3 = df.copy()
```

In [30]:
```
df3 = pd.get_dummies(df3, columns = ["event"])
display(df3)
```

| | day | temperature | windspeed | duration | event_Cloudy | event_Rain | event_Snow | event_Sunny |
|---|---|---|---|---|---|---|---|---|
| **0** | 01-01-2017 | 32.0 | 6.0 | 5.0 | 0 | 1 | 0 | 0 |
| **1** | 01-04-2017 | 33.2 | 9.0 | 4.0 | 0 | 0 | 0 | 1 |
| **2** | 01-05-2017 | 28.0 | 8.4 | 3.0 | 0 | 0 | 1 | 0 |
| **3** | 01-06-2017 | 33.2 | 7.0 | 2.0 | 0 | 0 | 0 | 0 |

| | day | temperature | windspeed | duration | event_Cloudy | event_Rain | event_Snow | event_Sunny |
|---|---|---|---|---|---|---|---|---|
| **4** | 01-07-2017 | 32.0 | 8.4 | 1.0 | 0 | 1 | 0 | 0 |
| **5** | 01-08-2017 | 33.2 | 8.4 | 5.0 | 0 | 0 | 0 | 1 |
| **6** | 01-09-2017 | 33.2 | 8.4 | 3.0 | 0 | 0 | 0 | 0 |
| **7** | 01-10-2017 | 34.0 | 8.0 | 2.0 | 1 | 0 | 0 | 0 |
| **8** | 01-11-2017 | 40.0 | 12.0 | 1.0 | 0 | 0 | 0 | 1 |

In [ ]: