DSBDA Lab Assignment No. 1

Name: Akash Ganesh Padir
Roll No.: TEB04

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

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

Out[2]:
```
'C:\\Users\\hp\\DSBDA Assignment 1'
```

In [3]:
```
df= pd.read_csv("weather_data.csv")
df
```

Out[3]:

	Day	temperature	windspeed	event	duration
0	1/1/2017	32.0	6.0	rain	2
1	1/4/2017	NaN	9.0	sunny	3
2	1/5/2017	28.0	NaN	snow	4
3	1/6/2017	NaN	7.0	NaN	5
4	1/7/2017	32.0	NaN	rain	7
5	1/8/2017	NaN	NaN	sunny	5
6	1/9/2017	NaN	NaN	NaN	6
7	1/10/2017	34.0	8.0	cloud	4
8	1/11/2017	40.0	12.0	sunny	2

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

Out[4]:

	Day	temperature	windspeed	event	duration
0	1/1/2017	32.0	6.0	rain	2
1	1/4/2017	NaN	9.0	sunny	3
2	1/5/2017	28.0	NaN	snow	4
3	1/6/2017	NaN	7.0	NaN	5
4	1/7/2017	32.0	NaN	rain	7

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

Out[5]:

	Day	temperature	windspeed	event	duration
4	1/7/2017	32.0	NaN	rain	7
5	1/8/2017	NaN	NaN	sunny	5
6	1/9/2017	NaN	NaN	NaN	6
7	1/10/2017	34.0	8.0	cloud	4
8	1/11/2017	40.0	12.0	sunny	2

In [6]:
```
df.describe()
```

Out[6]:

	temperature	windspeed	duration
count	5.00000	5.000000	9.000000
mean	33.20000	8.400000	4.222222
std	4.38178	2.302173	1.715938
min	28.00000	6.000000	2.000000
25%	32.00000	7.000000	3.000000
50%	32.00000	8.000000	4.000000
75%	34.00000	9.000000	5.000000
max	40.00000	12.000000	7.000000

Working with missing values

In [7]:
```
df.replace(" ", np.nan, inplace= True)
df.head(9)
```

Out[7]:

	Day	temperature	windspeed	event	duration
0	1/1/2017	32.0	6.0	rain	2
1	1/4/2017	NaN	9.0	sunny	3
2	1/5/2017	28.0	NaN	snow	4
3	1/6/2017	NaN	7.0	NaN	5
4	1/7/2017	32.0	NaN	rain	7
5	1/8/2017	NaN	NaN	sunny	5
6	1/9/2017	NaN	NaN	NaN	6
7	1/10/2017	34.0	8.0	cloud	4
8	1/11/2017	40.0	12.0	sunny	2

In [8]:
```
missingdata = df.isnull()
df.isnull().sum()
```

Out[8]:
```
Day            0
temperature    4
windspeed      4
event          2
duration       0
dtype: int64
```

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

Out[9]:

	Day	temperature	windspeed	event	duration
0	1/1/2017	32.0	6.0	rain	2
7	1/10/2017	34.0	8.0	cloud	4
8	1/11/2017	40.0	12.0	sunny	2

In [10]:
```
df
```

Out[10]:

	Day	temperature	windspeed	event	duration
0	1/1/2017	32.0	6.0	rain	2
1	1/4/2017	NaN	9.0	sunny	3
2	1/5/2017	28.0	NaN	snow	4
3	1/6/2017	NaN	7.0	NaN	5
4	1/7/2017	32.0	NaN	rain	7
5	1/8/2017	NaN	NaN	sunny	5
6	1/9/2017	NaN	NaN	NaN	6
7	1/10/2017	34.0	8.0	cloud	4
8	1/11/2017	40.0	12.0	sunny	2

In [11]:
```
df.dropna(axis=1)
```

Out[11]:

	Day	duration
0	1/1/2017	2
1	1/4/2017	3
2	1/5/2017	4
3	1/6/2017	5
4	1/7/2017	7
5	1/8/2017	5
6	1/9/2017	6
7	1/10/2017	4
8	1/11/2017	2

In [12]:
```
df.dropna(how = 'all')
```

Out[12]:

	Day	temperature	windspeed	event	duration
0	1/1/2017	32.0	6.0	rain	2
1	1/4/2017	NaN	9.0	sunny	3
2	1/5/2017	28.0	NaN	snow	4
3	1/6/2017	NaN	7.0	NaN	5
4	1/7/2017	32.0	NaN	rain	7
5	1/8/2017	NaN	NaN	sunny	5
6	1/9/2017	NaN	NaN	NaN	6
7	1/10/2017	34.0	8.0	cloud	4
8	1/11/2017	40.0	12.0	sunny	2

In [13]:
```
df.dropna(subset=['event'])
```

Out[13]:

	Day	temperature	windspeed	event	duration
0	1/1/2017	32.0	6.0	rain	2
1	1/4/2017	NaN	9.0	sunny	3
2	1/5/2017	28.0	NaN	snow	4
4	1/7/2017	32.0	NaN	rain	7
5	1/8/2017	NaN	NaN	sunny	5
7	1/10/2017	34.0	8.0	cloud	4
8	1/11/2017	40.0	12.0	sunny	2

In [14]:
```
mean_value= df["temperature"].mean()
mean_value
```

Out[14]: 33.2

In [16]:
```
df["temperature"] = df["temperature"].fillna(mean_value)
df
```

Out[16]:

	Day	temperature	windspeed	event	duration
0	1/1/2017	32.0	6.0	rain	2
1	1/4/2017	33.2	9.0	sunny	3
2	1/5/2017	28.0	NaN	snow	4
3	1/6/2017	33.2	7.0	NaN	5
4	1/7/2017	32.0	NaN	rain	7
5	1/8/2017	33.2	NaN	sunny	5
6	1/9/2017	33.2	NaN	NaN	6
7	1/10/2017	34.0	8.0	cloud	4
8	1/11/2017	40.0	12.0	sunny	2

In [17]:
```
mean_v = df['windspeed'].mean()
mean_v
```

Out[17]: 8.4

In [18]:
```
df['windspeed'] = df['windspeed'].fillna(mean_value)
df
```

Out[18]:

	Day	temperature	windspeed	event	duration
0	1/1/2017	32.0	6.0	rain	2
1	1/4/2017	33.2	9.0	sunny	3
2	1/5/2017	28.0	33.2	snow	4
3	1/6/2017	33.2	7.0	NaN	5
4	1/7/2017	32.0	33.2	rain	7
5	1/8/2017	33.2	33.2	sunny	5
6	1/9/2017	33.2	33.2	NaN	6
7	1/10/2017	34.0	8.0	cloud	4
8	1/11/2017	40.0	12.0	sunny	2

In [20]:
```
new_df = df.fillna({'temperature':0, 'windspeed':0, 'event': 'no_event'})
new_df
```

Out[20]:

	Day	temperature	windspeed	event	duration
0	1/1/2017	32.0	6.0	rain	2
1	1/4/2017	33.2	9.0	sunny	3
2	1/5/2017	28.0	33.2	snow	4
3	1/6/2017	33.2	7.0	no_event	5
4	1/7/2017	32.0	33.2	rain	7
5	1/8/2017	33.2	33.2	sunny	5
6	1/9/2017	33.2	33.2	no_event	6
7	1/10/2017	34.0	8.0	cloud	4
8	1/11/2017	40.0	12.0	sunny	2

In [21]:
```
df.dtypes
```

Out[21]:
```
Day             object
temperature    float64
windspeed      float64
event           object
duration         int64
dtype: object
```

In [22]:
```
df[["duration"]]=df[["duration"]].astype('float')
```

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

Out[23]:

	Day	temperature	windspeed	event	duration
0	1/1/2017	32.0	6.0	rain	2.0
1	1/4/2017	33.2	9.0	sunny	3.0
2	1/5/2017	28.0	33.2	snow	4.0
3	1/6/2017	33.2	7.0	NaN	5.0
4	1/7/2017	32.0	33.2	rain	7.0

In [24]:
```
df3= df.copy()
```

In [25]:
```
df3= pd.get_dummies(df3,columns=['event'])
display(df3)
```

	Day	temperature	windspeed	duration	event_cloud	event_rain	event_snow	event_sunny
0	1/1/2017	32.0	6.0	2.0	0	1	0	0
1	1/4/2017	33.2	9.0	3.0	0	0	0	1
2	1/5/2017	28.0	33.2	4.0	0	0	1	0
3	1/6/2017	33.2	7.0	5.0	0	0	0	0
4	1/7/2017	32.0	33.2	7.0	0	1	0	0
5	1/8/2017	33.2	33.2	5.0	0	0	0	1
6	1/9/2017	33.2	33.2	6.0	0	0	0	0
7	1/10/2017	34.0	8.0	4.0	1	0	0	0
8	1/11/2017	40.0	12.0	2.0	0	0	0	1

Importing csv using url

In [26]:
```
import pandas as pd
```

In [29]:
```
csv_url = "https://archive.ics.uci.edu/ml/datasets/iris/iris.csv"
```

In [30]:
```
iris = pd.read_csv(csv_url, header= None)
```

In [35]:
```
col_name=["sepal.length","sepal.width", "petal.length", "petal.width", "variety"
```

In [37]:
```
iris= pd.read_csv(csv_url, names= col_name)
```

In [38]:
```
print(iris)
```

```
     sepal.length  sepal.width  petal.length  petal.width     variety
0    sepal.length  sepal.width  petal.length  petal.width     variety
1             5.1          3.5           1.4           .2      Setosa
2             4.9            3           1.4           .2      Setosa
3             4.7          3.2           1.3           .2      Setosa
4             4.6          3.1           1.5           .2      Setosa
..            ...          ...           ...          ...         ...
146           6.7            3           5.2          2.3   Virginica
147           6.3          2.5             5          1.9   Virginica
148           6.5            3           5.2            2   Virginica
149           6.2          3.4           5.4          2.3   Virginica
150           5.9            3           5.1          1.8   Virginica

[151 rows x 5 columns]
```