In [3]:
```
import pandas as pd
```

In [4]:
```
df=pd.read_csv('g:/dataset/analysis/weather_data.txt')
```

In [5]:
```
df
```

Out[5]:

	day	temperature	windspeed	event
0	1/1/2017	32.0	6.0	Rain
1	1/4/2017	NaN	9.0	Sunny
2	1/5/2017	28.0	NaN	Snow
3	1/6/2017	NaN	7.0	NaN
4	1/7/2017	32.0	NaN	Rain
5	1/8/2017	NaN	NaN	Sunny
6	1/9/2017	NaN	NaN	NaN
7	1/10/2017	34.0	8.0	Cloudy
8	1/11/2017	40.0	12.0	Sunny
9	1/12/2017	30.0	10.0	Sunny

In [7]:
```
 df.isnull().sum()
```

Out[7]:
```
day            0
temperature    4
windspeed      4
event          2
dtype: int64
```

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

Out[10]:

	day	temperature	windspeed	event
0	1/1/2017	32.0	6.0	Rain
1	1/4/2017	NaN	9.0	Sunny
2	1/5/2017	28.0	NaN	Snow
3	1/6/2017	NaN	7.0	NaN
4	1/7/2017	32.0	NaN	Rain
5	1/8/2017	NaN	NaN	Sunny
6	1/9/2017	NaN	NaN	NaN
7	1/10/2017	34.0	8.0	Cloudy
8	1/11/2017	40.0	12.0	Sunny
9	1/12/2017	30.0	10.0	Sunny

In [13]:
```
df.temperature.mean()
```

Out[13]:
```
32.666666666666664
```

In [14]:
```
df.windspeed.mean()
```

Out[14]:
```
8.666666666666666
```

In [16]:
```
df.event.mode()
```

Out[16]:
```
0    Sunny
Name: event, dtype: object
```

In [17]:
```
df.fillna({'temperature':32,'windspeed':9,'event':'Sunny'})
```

Out[17]:

	day	temperature	windspeed	event
0	1/1/2017	32.0	6.0	Rain
1	1/4/2017	32.0	9.0	Sunny
2	1/5/2017	28.0	9.0	Snow
3	1/6/2017	32.0	7.0	Sunny
4	1/7/2017	32.0	9.0	Rain
5	1/8/2017	32.0	9.0	Sunny
6	1/9/2017	32.0	9.0	Sunny
7	1/10/2017	34.0	8.0	Cloudy
8	1/11/2017	40.0	12.0	Sunny
9	1/12/2017	30.0	10.0	Sunny

In [18]:
```
df.ffill()
```

Out[18]:

	day	temperature	windspeed	event
0	1/1/2017	32.0	6.0	Rain
1	1/4/2017	32.0	9.0	Sunny
2	1/5/2017	28.0	9.0	Snow
3	1/6/2017	28.0	7.0	Snow
4	1/7/2017	32.0	7.0	Rain
5	1/8/2017	32.0	7.0	Sunny
6	1/9/2017	32.0	7.0	Sunny
7	1/10/2017	34.0	8.0	Cloudy
8	1/11/2017	40.0	12.0	Sunny
9	1/12/2017	30.0	10.0	Sunny

In [21]:
```
df['temperature']=df.temperature.ffill()
df['windspeed']=df.windspeed.bfill()
```

In [22]:
```
df
```

Out[22]:

	day	temperature	windspeed	event
0	1/1/2017	32.0	6.0	Rain
1	1/4/2017	32.0	9.0	Sunny
2	1/5/2017	28.0	7.0	Snow
3	1/6/2017	28.0	7.0	NaN
4	1/7/2017	32.0	8.0	Rain
5	1/8/2017	32.0	8.0	Sunny
6	1/9/2017	32.0	8.0	NaN
7	1/10/2017	34.0	8.0	Cloudy
8	1/11/2017	40.0	12.0	Sunny
9	1/12/2017	30.0	10.0	Sunny

In [23]:
```
df.at[3,'event']='Sunny'
```

In [24]:
```
df
```

Out[24]:

	day	temperature	windspeed	event
0	1/1/2017	32.0	6.0	Rain
1	1/4/2017	32.0	9.0	Sunny
2	1/5/2017	28.0	7.0	Snow
3	1/6/2017	28.0	7.0	Sunny
4	1/7/2017	32.0	8.0	Rain
5	1/8/2017	32.0	8.0	Sunny
6	1/9/2017	32.0	8.0	NaN
7	1/10/2017	34.0	8.0	Cloudy
8	1/11/2017	40.0	12.0	Sunny
9	1/12/2017	30.0	10.0	Sunny

In [25]:
```
df
```

Out[25]:

	day	temperature	windspeed	event
0	1/1/2017	32.0	6.0	Rain
1	1/4/2017	32.0	9.0	Sunny
2	1/5/2017	28.0	7.0	Snow
3	1/6/2017	28.0	7.0	Sunny
4	1/7/2017	32.0	8.0	Rain
5	1/8/2017	32.0	8.0	Sunny
6	1/9/2017	32.0	8.0	NaN
7	1/10/2017	34.0	8.0	Cloudy
8	1/11/2017	40.0	12.0	Sunny
9	1/12/2017	30.0	10.0	Sunny

In [26]:
```
df=pd.read_csv('g:/dataset/analysis/weather_data.txt')
```

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

Out[27]:

	day	temperature	windspeed	event
0	1/1/2017	32.0	6.0	Rain
1	1/4/2017	NaN	9.0	Sunny
2	1/5/2017	28.0	NaN	Snow
3	1/6/2017	NaN	7.0	NaN
4	1/7/2017	32.0	NaN	Rain
5	1/8/2017	NaN	NaN	Sunny
6	1/9/2017	NaN	NaN	NaN
7	1/10/2017	34.0	8.0	Cloudy
8	1/11/2017	40.0	12.0	Sunny
9	1/12/2017	30.0	10.0	Sunny

In [28]:
```
df.dropna(axis=0)
```

Out[28]:

	day	temperature	windspeed	event
0	1/1/2017	32.0	6.0	Rain
7	1/10/2017	34.0	8.0	Cloudy
8	1/11/2017	40.0	12.0	Sunny
9	1/12/2017	30.0	10.0	Sunny

In [35]:
```
df.dropna(axis=0,thresh=3)
```

Out[35]:

	day	temperature	windspeed	event
0	1/1/2017	32.0	6.0	Rain
1	1/4/2017	NaN	9.0	Sunny
2	1/5/2017	28.0	NaN	Snow
4	1/7/2017	32.0	NaN	Rain
7	1/10/2017	34.0	8.0	Cloudy
8	1/11/2017	40.0	12.0	Sunny
9	1/12/2017	30.0	10.0	Sunny

In [36]:
```
df.dropna(axis=0,thresh=3).shape
```

Out[36]:
```
(7, 4)
```

In [37]:
```
df.dropna(axis=0,thresh=2)
```

Out[37]:

	day	temperature	windspeed	event
0	1/1/2017	32.0	6.0	Rain
1	1/4/2017	NaN	9.0	Sunny
2	1/5/2017	28.0	NaN	Snow
3	1/6/2017	NaN	7.0	NaN
4	1/7/2017	32.0	NaN	Rain
5	1/8/2017	NaN	NaN	Sunny
7	1/10/2017	34.0	8.0	Cloudy
8	1/11/2017	40.0	12.0	Sunny
9	1/12/2017	30.0	10.0	Sunny

In [38]:
```
df.dropna(axis=0,thresh=2).shape
```

Out[38]:
```
(9, 4)
```

In [39]:
```
df.dropna(axis=0,subset=['temperature'])
```

Out[39]:

	day	temperature	windspeed	event
0	1/1/2017	32.0	6.0	Rain
2	1/5/2017	28.0	NaN	Snow
4	1/7/2017	32.0	NaN	Rain
7	1/10/2017	34.0	8.0	Cloudy
8	1/11/2017	40.0	12.0	Sunny
9	1/12/2017	30.0	10.0	Sunny

In [43]:
```
df.dropna(axis=0,subset=['temperature','windspeed'])
```

Out[43]:

	day	temperature	windspeed	event
0	1/1/2017	32.0	6.0	Rain
7	1/10/2017	34.0	8.0	Cloudy
8	1/11/2017	40.0	12.0	Sunny
9	1/12/2017	30.0	10.0	Sunny

In [44]:
```
df
```

Out[44]:

	day	temperature	windspeed	event
0	1/1/2017	32.0	6.0	Rain
1	1/4/2017	NaN	9.0	Sunny
2	1/5/2017	28.0	NaN	Snow
3	1/6/2017	NaN	7.0	NaN
4	1/7/2017	32.0	NaN	Rain
5	1/8/2017	NaN	NaN	Sunny
6	1/9/2017	NaN	NaN	NaN
7	1/10/2017	34.0	8.0	Cloudy
8	1/11/2017	40.0	12.0	Sunny
9	1/12/2017	30.0	10.0	Sunny

In [45]:
```
df.dropna(axis=1)
```

Out[45]:

	day
0	1/1/2017
1	1/4/2017
2	1/5/2017
3	1/6/2017
4	1/7/2017
5	1/8/2017
6	1/9/2017
7	1/10/2017
8	1/11/2017
9	1/12/2017

In [46]:
```
df
```

Out[46]:

	day	temperature	windspeed	event
0	1/1/2017	32.0	6.0	Rain
1	1/4/2017	NaN	9.0	Sunny
2	1/5/2017	28.0	NaN	Snow
3	1/6/2017	NaN	7.0	NaN
4	1/7/2017	32.0	NaN	Rain
5	1/8/2017	NaN	NaN	Sunny
6	1/9/2017	NaN	NaN	NaN
7	1/10/2017	34.0	8.0	Cloudy
8	1/11/2017	40.0	12.0	Sunny
9	1/12/2017	30.0	10.0	Sunny

In [47]:
```
df.drop('event',axis=1)
```

Out[47]:

	day	temperature	windspeed
0	1/1/2017	32.0	6.0
1	1/4/2017	NaN	9.0
2	1/5/2017	28.0	NaN
3	1/6/2017	NaN	7.0
4	1/7/2017	32.0	NaN
5	1/8/2017	NaN	NaN
6	1/9/2017	NaN	NaN
7	1/10/2017	34.0	8.0
8	1/11/2017	40.0	12.0
9	1/12/2017	30.0	10.0

In [48]:
```
df.drop(['event','day'],axis=1)
```

Out[48]:

	temperature	windspeed
0	32.0	6.0
1	NaN	9.0
2	28.0	NaN
3	NaN	7.0
4	32.0	NaN
5	NaN	NaN
6	NaN	NaN
7	34.0	8.0
8	40.0	12.0
9	30.0	10.0

In [49]:
```
df.drop([0,5,6],axis=0)
```

Out[49]:

	day	temperature	windspeed	event
1	1/4/2017	NaN	9.0	Sunny
2	1/5/2017	28.0	NaN	Snow
3	1/6/2017	NaN	7.0	NaN
4	1/7/2017	32.0	NaN	Rain
7	1/10/2017	34.0	8.0	Cloudy
8	1/11/2017	40.0	12.0	Sunny
9	1/12/2017	30.0	10.0	Sunny

In [50]:
```
df=pd.read_csv('g:/dataset/analysis/titanic.csv')
```

In [51]:
```
df
```

Out[51]:

	PassengerId	Survived	Pclass	Name	gender	Age	SibSp	Parch	Ticket	Fare	Ca
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	I
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	I
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	I
...	...	...	...	...	...	...	...	...	...	...	
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0000	I
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000	
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4500	I
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0000	C
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.7500	I

891 rows × 12 columns

In [52]:
```
df.dropna(axis=0,subset=['Age','gender','Pclass','Survived'],inplace=True)
```

In [53]:
```
df
```

Out[53]:

	PassengerId	Survived	Pclass	Name	gender	Age	SibSp	Parch	Ticket	Fare	Ca
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	N
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	N
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	N
...	...	...	...	...	...	...	...	...	...	...	
885	886	0	3	Rice, Mrs. William (Margaret Norton)	female	39.0	0	5	382652	29.1250	N
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0000	N
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000	
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0000	C
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.7500	N

714 rows × 12 columns

In [54]:
```
df[df.Age<=10].shape
```

Out[54]:
```
(64, 12)
```

In [57]:
```
df[df.Age>=60].shape
```

Out[57]:
```
(26, 12)
```

In [55]:
```
df[(df.Age<=10)&(df.Survived==1)].shape
```

Out[55]: (38, 12)

In [56]:
```
df[(df.Age>=60)&(df.Survived==1)].shape
```

Out[56]: (7, 12)

In [58]:
```
df=pd.read_csv('g:/dataset/analysis/weather_data2.txt')
df
```

Out[58]:

	day	temperature	windspeed	event
0	1/1/2017	32	6	Rain
1	1/4/2017	-99999	9	Sunny
2	1/5/2017	28	-99999	Snow
3	1/6/2017	-99999	7	No event
4	1/7/2017	32	-88888	Rain
5	1/11/2017	40	12	No event
6	1/1/2017	32	6	Rain
7	1/4/2017	-99999	9	Sunny
8	1/5/2017	28	-99999	Snow
9	1/6/2017	-99999	7	No event
10	1/7/2017	32	-88888	Rain
11	1/11/2017	40	12	No event
12	1/1/2017	32	6	Rain
13	1/4/2017	-99999	9	Sunny
14	1/5/2017	28	-99999	Snow
15	1/6/2017	-99999	7	No event
16	1/7/2017	32	-88888	Rain
17	1/11/2017	40	12	No event

In [59]:
```
df.drop_duplicates()
```

Out[59]:

	day	temperature	windspeed	event
0	1/1/2017	32	6	Rain
1	1/4/2017	-99999	9	Sunny
2	1/5/2017	28	-99999	Snow
3	1/6/2017	-99999	7	No event
4	1/7/2017	32	-88888	Rain
5	1/11/2017	40	12	No event

In [60]:
```
df.drop_duplicates(subset=['temperature'])
```

Out[60]:

	day	temperature	windspeed	event
0	1/1/2017	32	6	Rain
1	1/4/2017	-99999	9	Sunny
2	1/5/2017	28	-99999	Snow
5	1/11/2017	40	12	No event

In [61]:
```
df.drop_duplicates(inplace=True)
```

In [62]:
```
df
```

Out[62]:

	day	temperature	windspeed	event
0	1/1/2017	32	6	Rain
1	1/4/2017	-99999	9	Sunny
2	1/5/2017	28	-99999	Snow
3	1/6/2017	-99999	7	No event
4	1/7/2017	32	-88888	Rain
5	1/11/2017	40	12	No event

In [63]:
```
df.replace({'temperature':-99999},30,inplace=True)
df.replace({'windspeed':[-99999,-88888]},10,inplace=True)
```

In [64]:
```
df
```

Out[64]:

	day	temperature	windspeed	event
0	1/1/2017	32	6	Rain
1	1/4/2017	30	9	Sunny
2	1/5/2017	28	10	Snow
3	1/6/2017	30	7	No event
4	1/7/2017	32	10	Rain
5	1/11/2017	40	12	No event

In [65]:
```
df=pd.read_csv('g:/dataset/analysis/weather_data3.txt')
```

In [66]:
```
df
```

Out[66]:

	day	temperature	windspeed	event
0	1/1/2017	32F	6mph	Rain
1	1/4/2017	NaN	9mph	Su@nny
2	1/5/2017	2_8	NaN	Snow
3	1/6/2017	NaN	7	NaN
4	1/7/2017	32C	NaN	Rain
5	1/8/2017	NaN	NaN	Sunny
6	1/9/2017	NaN	NaN	NaN
7	1/10/2017	34	8	Cloudy
8	1/11/2017	40	12	Sunny

In [67]:
```
df.replace({'temperature':'[^0-9]'},'',regex=True)
```

Out[67]:

	day	temperature	windspeed	event
0	1/1/2017	32	6mph	Rain
1	1/4/2017	NaN	9mph	Su@nny
2	1/5/2017	28	NaN	Snow
3	1/6/2017	NaN	7	NaN
4	1/7/2017	32	NaN	Rain
5	1/8/2017	NaN	NaN	Sunny
6	1/9/2017	NaN	NaN	NaN
7	1/10/2017	34	8	Cloudy
8	1/11/2017	40	12	Sunny

In [69]:
```
df
```

Out[69]:

	day	temperature	windspeed	event
0	1/1/2017	32F	6mph	Rain
1	1/4/2017	NaN	9mph	Su@nny
2	1/5/2017	2_8	NaN	Snow
3	1/6/2017	NaN	7	NaN
4	1/7/2017	32C	NaN	Rain
5	1/8/2017	NaN	NaN	Sunny
6	1/9/2017	NaN	NaN	NaN
7	1/10/2017	34	8	Cloudy
8	1/11/2017	40	12	Sunny

In [76]:
```
df2=df.replace({'temperature':'[^0-9]'},'',regex=True)
```

In [77]:
```
df2
```

Out[77]:

	day	temperature	windspeed	event
0	1/1/2017	32	6mph	Rain
1	1/4/2017	NaN	9mph	Su@nny
2	1/5/2017	28	NaN	Snow
3	1/6/2017	NaN	7	NaN
4	1/7/2017	32	NaN	Rain
5	1/8/2017	NaN	NaN	Sunny
6	1/9/2017	NaN	NaN	NaN
7	1/10/2017	34	8	Cloudy
8	1/11/2017	40	12	Sunny

In [82]:
```
df2.replace({'event':'[^A-Z,a-z]'},'',regex=True)
```

Out[82]:

	day	temperature	windspeed	event
0	1/1/2017	32	6mph	Rain
1	1/4/2017	NaN	9mph	Sunny
2	1/5/2017	28	NaN	Snow
3	1/6/2017	NaN	7	NaN
4	1/7/2017	32	NaN	Rain
5	1/8/2017	NaN	NaN	Sunny
6	1/9/2017	NaN	NaN	NaN
7	1/10/2017	34	8	Cloudy
8	1/11/2017	40	12	Sunny

In [ ]:
```

```