

pr1

May 3, 2024

```
[1]: import pandas as pd
```

```
[4]: data=pd.read_csv("C:\\Users\\nayan\\Downloads\\weather_data.csv")
```

```
[5]: data
```

```
[5]:
```

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

```
[6]: data.head()
```

```
[6]:
```

	day	temperature	windspeed	event	duration
0	01-01-2017	32.0	6.0	Rain	2
1	01-04-2017	NaN	9.0	Sunny	3
2	01-05-2017	28.0	NaN	Snow	4
3	01-06-2017	NaN	7.0	NaN	5
4	01-07-2017	32.0	NaN	Rain	7

```
[7]: data.isnull()
```

```
[7]:
```

	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
```

```
[8]: data.isnull().sum()
```

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

```
[9]: data.describe()
```

```
[9]:
```

	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

```
[10]: data.size
```

```
[10]: 45
```

```
[11]: data.shape
```

```
[11]: (9, 5)
```

```
[12]: data.dtypes
```

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

```
[13]: data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9 entries, 0 to 8
Data columns (total 5 columns):
#   Column          Non-Null Count  Dtype
---  -
0   day              9 non-null     object
```

```

1  temperature  5 non-null    float64
2  windspeed    5 non-null    float64
3  event        7 non-null    object
4  duration     9 non-null    int64
dtypes: float64(2), int64(1), object(2)
memory usage: 492.0+ bytes

```

```
[14]: #delete null value rows
data.dropna()
```

```
[14]:
```

	day	temperature	windspeed	event	duration
0	01-01-2017	32.0	6.0	Rain	2
7	01-10-2017	34.0	8.0	Cloudy	4
8	01-11-2017	40.0	12.0	Sunny	2

```
[16]: #delete event column contains the null value row
data.dropna(subset=['event'])
```

```
[16]:
```

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

```
[20]: # which column contains null value that column delete
data.dropna(axis=1)
```

```
[20]:
```

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

```
[21]: #data fomatting

mean_value = data['temperature'].mean()
```

```
[22]: mean_value
```

[22]: 33.2

```
[25]: data['temperature']=data['temperature'].fillna(mean_value)
```

```
[26]: data
```

```
[26]:
```

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

```
[27]: mena_value=data['windspeed'].mean()
```

```
[28]: mean_value
```

[28]: 33.2

```
[29]: data['windspeed']=data['windspeed'].fillna(mean_value)
```

```
[30]: data
```

```
[30]:
```

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

```
[31]: #normalization
# creating a copy of the original data frame
data2 = data.copy()
```

```
[ ]:
```

```
[ ]:
```

```
[41]: # categories convert into number

data["event"].replace({"Rain":0,"Sunny":1}, inplace=True)
```

```
[42]: data
```

```
[42]:
```

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

```
[ ]:
```