practical1

May 4, 2025

[1]:
```
#practical 1
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

[2]:
```
weather = pd.read_csv("weatherAUS.csv")
```

[4]:
```
weather.head()
```

[4]:

	Date	Location	MinTemp	MaxTemp	Rainfall	Evaporation	Sunshine	\
0	01-12-2008	Albury	13.4	22.9	0.6	NaN	NaN	
1	02-12-2008	Albury	7.4	25.1	0.0	NaN	NaN	
2	03-12-2008	Albury	12.9	25.7	0.0	NaN	NaN	
3	04-12-2008	Albury	9.2	28.0	0.0	NaN	NaN	
4	05-12-2008	Albury	17.5	32.3	1.0	NaN	NaN	

	WindGustDir	WindGustSpeed	WindDir9am	…	Humidity3pm	Pressure9am	\
0	W	44.0	W	…	22.0	1007.7	
1	WNW	44.0	NNW	…	25.0	1010.6	
2	WSW	46.0	W	…	30.0	1007.6	
3	NE	24.0	SE	…	16.0	1017.6	
4	W	41.0	ENE	…	33.0	1010.8	

	Pressure3pm	Cloud9am	Cloud3pm	Temp9am	Temp3pm	RainToday	RISK_MM	\
0	1007.1	8.0	NaN	16.9	21.8	No	0.0	
1	1007.8	NaN	NaN	17.2	24.3	No	0.0	
2	1008.7	NaN	2.0	21.0	23.2	No	0.0	
3	1012.8	NaN	NaN	18.1	26.5	No	1.0	
4	1006.0	7.0	8.0	17.8	29.7	No	0.2	

	RainTomorrow
0	No
1	No
2	No
3	No
4	No

```
[5 rows x 24 columns]
```

[5]:
```
weather.tail()
```

[5]:
```
             Date Location  MinTemp  MaxTemp  Rainfall  Evaporation  \
142188  20-06-2017    Uluru      3.5     21.8       0.0          NaN
142189  21-06-2017    Uluru      2.8     23.4       0.0          NaN
142190  22-06-2017    Uluru      3.6     25.3       0.0          NaN
142191  23-06-2017    Uluru      5.4     26.9       0.0          NaN
142192  24-06-2017    Uluru      7.8     27.0       0.0          NaN

        Sunshine WindGustDir  WindGustSpeed WindDir9am  … Humidity3pm  \
142188       NaN           E           31.0        ESE  …        27.0
142189       NaN           E           31.0         SE  …        24.0
142190       NaN         NNW           22.0         SE  …        21.0
142191       NaN           N           37.0         SE  …        24.0
142192       NaN          SE           28.0        SSE  …        24.0

        Pressure9am  Pressure3pm  Cloud9am  Cloud3pm  Temp9am  Temp3pm  \
142188       1024.7       1021.2       NaN       NaN      9.4     20.9
142189       1024.6       1020.3       NaN       NaN     10.1     22.4
142190       1023.5       1019.1       NaN       NaN     10.9     24.5
142191       1021.0       1016.8       NaN       NaN     12.5     26.1
142192       1019.4       1016.5       3.0       2.0     15.1     26.0

        RainToday  RISK_MM  RainTomorrow
142188         No      0.0            No
142189         No      0.0            No
142190         No      0.0            No
142191         No      0.0            No
142192         No      0.0            No

[5 rows x 24 columns]
```

[6]:
```
weather.isnull()
```

[6]:
```
          Date  Location  MinTemp  MaxTemp  Rainfall  Evaporation  Sunshine  \
0        False     False    False    False     False         True      True
1        False     False    False    False     False         True      True
2        False     False    False    False     False         True      True
3        False     False    False    False     False         True      True
4        False     False    False    False     False         True      True
…        …         …        …        …         …             …         …
142188   False     False    False    False     False         True      True
142189   False     False    False    False     False         True      True
142190   False     False    False    False     False         True      True
```

```
142191  False     False    False    False     False         True      True
142192  False     False    False    False     False         True      True

        WindGustDir  WindGustSpeed  WindDir9am  …  Humidity3pm  Pressure9am  \
0             False          False       False  …        False        False
1             False          False       False  …        False        False
2             False          False       False  …        False        False
3             False          False       False  …        False        False
4             False          False       False  …        False        False
…               …              …           …    …          …            …
142188        False          False       False  …        False        False
142189        False          False       False  …        False        False
142190        False          False       False  …        False        False
142191        False          False       False  …        False        False
142192        False          False       False  …        False        False

        Pressure3pm  Cloud9am  Cloud3pm  Temp9am  Temp3pm  RainToday  RISK_MM  \
0             False     False      True    False    False      False    False
1             False      True      True    False    False      False    False
2             False      True     False    False    False      False    False
3             False      True      True    False    False      False    False
4             False     False     False    False    False      False    False
…               …         …         …        …        …          …        …
142188        False      True      True    False    False      False    False
142189        False      True      True    False    False      False    False
142190        False      True      True    False    False      False    False
142191        False      True      True    False    False      False    False
142192        False     False     False    False    False      False    False

        RainTomorrow
0              False
1              False
2              False
3              False
4              False
…                …
142188         False
142189         False
142190         False
142191         False
142192         False

[142193 rows x 24 columns]
```

[7]:
```
weather.isnull().sum()
```

[7]:
```
Date                0
Location            0
MinTemp           637
MaxTemp           322
Rainfall         1406
Evaporation     60843
Sunshine        67816
WindGustDir      9330
WindGustSpeed    9270
WindDir9am      10013
WindDir3pm       3778
WindSpeed9am     1348
WindSpeed3pm     2630
Humidity9am      1774
Humidity3pm      3610
Pressure9am     14014
Pressure3pm     13981
Cloud9am        53657
Cloud3pm        57094
Temp9am           904
Temp3pm          2726
RainToday        1406
RISK_MM             0
RainTomorrow        0
dtype: int64
```

[8]:
```
weather.describe()
```

[8]:

	MinTemp	MaxTemp	Rainfall	Evaporation	\
count	141556.000000	141871.000000	140787.000000	81350.000000	
mean	12.186400	23.226784	2.349974	5.469824	
std	6.403283	7.117618	8.465173	4.188537	
min	-8.500000	-4.800000	0.000000	0.000000	
25%	7.600000	17.900000	0.000000	2.600000	
50%	12.000000	22.600000	0.000000	4.800000	
75%	16.800000	28.200000	0.800000	7.400000	
max	33.900000	48.100000	371.000000	145.000000	

	Sunshine	WindGustSpeed	WindSpeed9am	WindSpeed3pm	\
count	74377.000000	132923.000000	140845.000000	139563.000000	
mean	7.624853	39.984292	14.001988	18.637576	
std	3.781525	13.588801	8.893337	8.803345	
min	0.000000	6.000000	0.000000	0.000000	
25%	4.900000	31.000000	7.000000	13.000000	
50%	8.500000	39.000000	13.000000	19.000000	
75%	10.600000	48.000000	19.000000	24.000000	
max	14.500000	135.000000	130.000000	87.000000	

```
         Humidity9am    Humidity3pm    Pressure9am    Pressure3pm  \
count  140419.000000  138583.000000  128179.000000  128212.000000
mean       68.843810      51.482606    1017.653758    1015.258204
std        19.051293      20.797772       7.105476       7.036677
min         0.000000       0.000000     980.500000     977.100000
25%        57.000000      37.000000    1012.900000    1010.400000
50%        70.000000      52.000000    1017.600000    1015.200000
75%        83.000000      66.000000    1022.400000    1020.000000
max       100.000000     100.000000    1041.000000    1039.600000

          Cloud9am      Cloud3pm        Temp9am        Temp3pm        RISK_MM
count  88536.000000  85099.000000  141289.000000  139467.000000  142193.000000
mean       4.437189      4.503167      16.987509      21.687235       2.360682
std        2.887016      2.720633       6.492838       6.937594       8.477969
min        0.000000      0.000000      -7.200000      -5.400000       0.000000
25%        1.000000      2.000000      12.300000      16.600000       0.000000
50%        5.000000      5.000000      16.700000      21.100000       0.000000
75%        7.000000      7.000000      21.600000      26.400000       0.800000
max        9.000000      9.000000      40.200000      46.700000     371.000000
```

[9]:
```
weather.shape
```

[9]:
```
(142193, 24)
```

[10]:
```
weather.dtypes
```

[10]:
```
Date              object
Location          object
MinTemp          float64
MaxTemp          float64
Rainfall         float64
Evaporation      float64
Sunshine         float64
WindGustDir       object
WindGustSpeed    float64
WindDir9am        object
WindDir3pm        object
WindSpeed9am     float64
WindSpeed3pm     float64
Humidity9am      float64
Humidity3pm      float64
Pressure9am      float64
Pressure3pm      float64
Cloud9am         float64
Cloud3pm         float64
Temp9am          float64
```

```
Temp3pm          float64
RainToday         object
RISK_MM          float64
RainTomorrow      object
dtype: object
```

[11]:
```
weather.dtypes.value_counts()
```

[11]:
```
float64    17
object      7
dtype: int64
```

[12]:
```
weather['Date'] = pd.to_datetime(weather['Date'])
```

```
C:\Users\GAURI\AppData\Local\Temp\ipykernel_25204\1942266597.py:1: UserWarning:
Parsing dates in DD/MM/YYYY format when dayfirst=False (the default) was
specified. This may lead to inconsistently parsed dates! Specify a format to
ensure consistent parsing.
  weather['Date'] = pd.to_datetime(weather['Date'])
```

[13]:
```
weather.dtypes
```

[13]:
```
Date             datetime64[ns]
Location                 object
MinTemp                 float64
MaxTemp                 float64
Rainfall                float64
Evaporation             float64
Sunshine                float64
WindGustDir              object
WindGustSpeed           float64
WindDir9am               object
WindDir3pm               object
WindSpeed9am            float64
WindSpeed3pm            float64
Humidity9am             float64
Humidity3pm             float64
Pressure9am             float64
Pressure3pm             float64
Cloud9am                float64
Cloud3pm                float64
Temp9am                 float64
Temp3pm                 float64
RainToday                object
RISK_MM                 float64
RainTomorrow             object
dtype: object
```

[14]:
```
weather['RainToday'] = weather['RainToday'].map({'No' : 0, 'Yes' : 1})
```

[15]:
```
weather['RainToday'].head()
```

[15]:
```
0    0.0
1    0.0
2    0.0
3    0.0
4    0.0
Name: RainToday, dtype: float64
```

[16]:
```
weather.dtypes
```

[16]:
```
Date             datetime64[ns]
Location                 object
MinTemp                 float64
MaxTemp                 float64
Rainfall                float64
Evaporation             float64
Sunshine                float64
WindGustDir              object
WindGustSpeed           float64
WindDir9am               object
WindDir3pm               object
WindSpeed9am            float64
WindSpeed3pm            float64
Humidity9am             float64
Humidity3pm             float64
Pressure9am             float64
Pressure3pm             float64
Cloud9am                float64
Cloud3pm                float64
Temp9am                 float64
Temp3pm                 float64
RainToday               float64
RISK_MM                 float64
RainTomorrow             object
dtype: object
```

[18]:
```
categorical_cols = weather.select_dtypes(include = 'object').columns
print("Categorical Columns :\n",categorical_cols)
```

```
Categorical Columns :
 Index(['Location', 'WindGustDir', 'WindDir9am', 'WindDir3pm', 'RainTomorrow'],
dtype='object')
```

[21]:
```
from sklearn.preprocessing import LabelEncoder

le = LabelEncoder()

for col in categorical_cols:
    weather[col] = weather[col].astype(str)
    weather[col] = le.fit_transform(weather[col])

weather.head()
```

[21]:

	Date	Location	MinTemp	MaxTemp	Rainfall	Evaporation	Sunshine	\
0	2008-01-12	12	13.4	22.9	0.6	NaN	NaN	
1	2008-02-12	12	7.4	25.1	0.0	NaN	NaN	
2	2008-03-12	12	12.9	25.7	0.0	NaN	NaN	
3	2008-04-12	12	9.2	28.0	0.0	NaN	NaN	
4	2008-05-12	12	17.5	32.3	1.0	NaN	NaN	

	WindGustDir	WindGustSpeed	WindDir9am	…	Humidity3pm	Pressure9am	\
0	5	44.0	5	…	22.0	1007.7	
1	6	44.0	13	…	25.0	1010.6	
2	7	46.0	5	…	30.0	1007.6	
3	11	24.0	16	…	16.0	1017.6	
4	5	41.0	1	…	33.0	1010.8	

	Pressure3pm	Cloud9am	Cloud3pm	Temp9am	Temp3pm	RainToday	RISK_MM	\
0	1007.1	8.0	NaN	16.9	21.8	0.0	0.0	
1	1007.8	NaN	NaN	17.2	24.3	0.0	0.0	
2	1008.7	NaN	2.0	21.0	23.2	0.0	0.0	
3	1012.8	NaN	NaN	18.1	26.5	0.0	1.0	
4	1006.0	7.0	8.0	17.8	29.7	0.0	0.2	

	RainTomorrow
0	0
1	0
2	0
3	0
4	0

[5 rows x 24 columns]

[23]:
```
weather.dtypes
```

[23]:
```
Date             datetime64[ns]
Location                  int32
MinTemp                 float64
MaxTemp                 float64
Rainfall                float64
```

```
Evaporation             float64
Sunshine                float64
WindGustDir               int32
WindGustSpeed           float64
WindDir9am                int32
WindDir3pm                int32
WindSpeed9am            float64
WindSpeed3pm            float64
Humidity9am             float64
Humidity3pm             float64
Pressure9am             float64
Pressure3pm             float64
Cloud9am                float64
Cloud3pm                float64
Temp9am                 float64
Temp3pm                 float64
RainToday               float64
RISK_MM                 float64
RainTomorrow              int32
dtype: object
```

[ ]: