practical-exam-15

May 23, 2023

[ ]:
```
# Imports
```

[ ]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from importlib import reload
plt=reload(plt)
```

[ ]:
```
# Taking input file
```

[ ]:
```
df=pd.read_csv("/home/anshu/Downloads/weatherHistory.csv")
df.head()
```

[ ]:
```
                  Formatted Date        Summary Precip Type  Temperature (C)  \
0  2006-04-01 00:00:00.000 +0200  Partly Cloudy        rain         9.472222
1  2006-04-01 01:00:00.000 +0200  Partly Cloudy        rain         9.355556
2  2006-04-01 02:00:00.000 +0200  Mostly Cloudy        rain         9.377778
3  2006-04-01 03:00:00.000 +0200  Partly Cloudy        rain         8.288889
4  2006-04-01 04:00:00.000 +0200  Mostly Cloudy        rain         8.755556

   Apparent Temperature (C)  Humidity  Wind Speed (km/h)  \
0                  7.388889      0.89            14.1197
1                  7.227778      0.86            14.2646
2                  9.377778      0.89             3.9284
3                  5.944444      0.83            14.1036
4                  6.977778      0.83            11.0446

   Wind Bearing (degrees)  Visibility (km)  Loud Cover  Pressure (millibars)  \
0                   251.0          15.8263         0.0               1015.13
1                   259.0          15.8263         0.0               1015.63
2                   204.0          14.9569         0.0               1015.94
3                   269.0          15.8263         0.0               1016.41
4                   259.0          15.8263         0.0               1016.51

                        Daily Summary
0  Partly cloudy throughout the day.
```

```
1  Partly cloudy throughout the day.
2  Partly cloudy throughout the day.
3  Partly cloudy throughout the day.
4  Partly cloudy throughout the day.
```

[ ]:
```
# Data Curation
```

[ ]:
```
df[["Date-Time","TZ"]]=df["Formatted Date"].str.split("+",expand=True)
df1=df.drop(columns="Formatted Date")
df1.head()
```

[ ]:
```
         Summary Precip Type  Temperature (C)  Apparent Temperature (C)  \
0  Partly Cloudy        rain         9.472222                  7.388889
1  Partly Cloudy        rain         9.355556                  7.227778
2  Mostly Cloudy        rain         9.377778                  9.377778
3  Partly Cloudy        rain         8.288889                  5.944444
4  Mostly Cloudy        rain         8.755556                  6.977778

   Humidity  Wind Speed (km/h)  Wind Bearing (degrees)  Visibility (km)  \
0      0.89            14.1197                   251.0          15.8263
1      0.86            14.2646                   259.0          15.8263
2      0.89             3.9284                   204.0          14.9569
3      0.83            14.1036                   269.0          15.8263
4      0.83            11.0446                   259.0          15.8263

   Loud Cover  Pressure (millibars)                     Daily Summary  \
0         0.0               1015.13  Partly cloudy throughout the day.
1         0.0               1015.63  Partly cloudy throughout the day.
2         0.0               1015.94  Partly cloudy throughout the day.
3         0.0               1016.41  Partly cloudy throughout the day.
4         0.0               1016.51  Partly cloudy throughout the day.

                  Date-Time    TZ
0  2006-04-01 00:00:00.000   0200
1  2006-04-01 01:00:00.000   0200
2  2006-04-01 02:00:00.000   0200
3  2006-04-01 03:00:00.000   0200
4  2006-04-01 04:00:00.000   0200
```

[ ]:
```
columns_order=["Date-Time","TZ","Summary","Precip Type","Temperature␣
↪(C)","Apparent Temperature (C)",
               "Humidity","Wind Speed (km/h)","Wind Bearing␣
↪(degrees)","Visibility (km)","Loud Cover",
               "Pressure (millibars)", "Daily Summary"]
df2=df1.reindex(columns=columns_order)
df3=df2.drop(columns="TZ")
df3.head()
```

[ ]:

```
                  Date-Time        Summary Precip Type  Temperature (C)  \
0  2006-04-01 00:00:00.000   Partly Cloudy        rain         9.472222
1  2006-04-01 01:00:00.000   Partly Cloudy        rain         9.355556
2  2006-04-01 02:00:00.000   Mostly Cloudy        rain         9.377778
3  2006-04-01 03:00:00.000   Partly Cloudy        rain         8.288889
4  2006-04-01 04:00:00.000   Mostly Cloudy        rain         8.755556

   Apparent Temperature (C)  Humidity  Wind Speed (km/h)  \
0                  7.388889      0.89            14.1197
1                  7.227778      0.86            14.2646
2                  9.377778      0.89             3.9284
3                  5.944444      0.83            14.1036
4                  6.977778      0.83            11.0446

   Wind Bearing (degrees)  Visibility (km)  Loud Cover  Pressure (millibars)  \
0                   251.0          15.8263         0.0                1015.13
1                   259.0          15.8263         0.0                1015.63
2                   204.0          14.9569         0.0                1015.94
3                   269.0          15.8263         0.0                1016.41
4                   259.0          15.8263         0.0                1016.51

                       Daily Summary
0  Partly cloudy throughout the day.
1  Partly cloudy throughout the day.
2  Partly cloudy throughout the day.
3  Partly cloudy throughout the day.
4  Partly cloudy throughout the day.
```

[ ]:

```
df3["Date-Time"]=pd.to_datetime(df3["Date-Time"])
df3.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 96453 entries, 0 to 96452
Data columns (total 12 columns):
 #   Column                    Non-Null Count  Dtype
---  ------                    --------------  -----
 0   Date-Time                 96453 non-null  datetime64[ns]
 1   Summary                   96453 non-null  object
 2   Precip Type               95936 non-null  object
 3   Temperature (C)           96453 non-null  float64
 4   Apparent Temperature (C)  96453 non-null  float64
 5   Humidity                  96453 non-null  float64
 6   Wind Speed (km/h)         96453 non-null  float64
 7   Wind Bearing (degrees)    96453 non-null  float64
 8   Visibility (km)           96453 non-null  float64
 9   Loud Cover                96453 non-null  float64
 10  Pressure (millibars)      96453 non-null  float64
```

```
 11  Daily Summary             96453 non-null  object
dtypes: datetime64[ns](1), float64(8), object(3)
memory usage: 8.8+ MB
```

[ ]:
```
df3["Year"]=pd.DatetimeIndex(df3["Date-Time"]).year
df3["Month"]=df3["Date-Time"].dt.month_name()
df3["day"]=df3["Date-Time"].dt.day
df3.head()
```

[ ]:

	Date-Time	Summary	Precip Type	Temperature (C)	Apparent Temperature (C)	Humidity	Wind Speed (km/h)	Wind Bearing (degrees)	Visibility (km)	Loud Cover	Pressure (millibars)	Daily Summary	Year	Month	day
0	2006-04-01 00:00:00	Partly Cloudy	rain	9.472222	7.388889	0.89	14.1197	251.0	15.8263	0.0	1015.13	Partly cloudy throughout the day.	2006	April	1
1	2006-04-01 01:00:00	Partly Cloudy	rain	9.355556	7.227778	0.86	14.2646	259.0	15.8263	0.0	1015.63	Partly cloudy throughout the day.	2006	April	1
2	2006-04-01 02:00:00	Mostly Cloudy	rain	9.377778	9.377778	0.89	3.9284	204.0	14.9569	0.0	1015.94	Partly cloudy throughout the day.	2006	April	1
3	2006-04-01 03:00:00	Partly Cloudy	rain	8.288889	5.944444	0.83	14.1036	269.0	15.8263	0.0	1016.41	Partly cloudy throughout the day.	2006	April	1
4	2006-04-01 04:00:00	Mostly Cloudy	rain	8.755556	6.977778	0.83	11.0446	259.0	15.8263	0.0	1016.51	Partly cloudy throughout the day.	2006	April	1

[ ]:
```
# Wind Speed
```

[ ]:
```
df3["Wind Speed (km/h)"].describe()
```

[ ]:
```
count    96453.000000
mean        10.810640
std          6.913571
min          0.000000
25%          5.828200
50%          9.965900
```

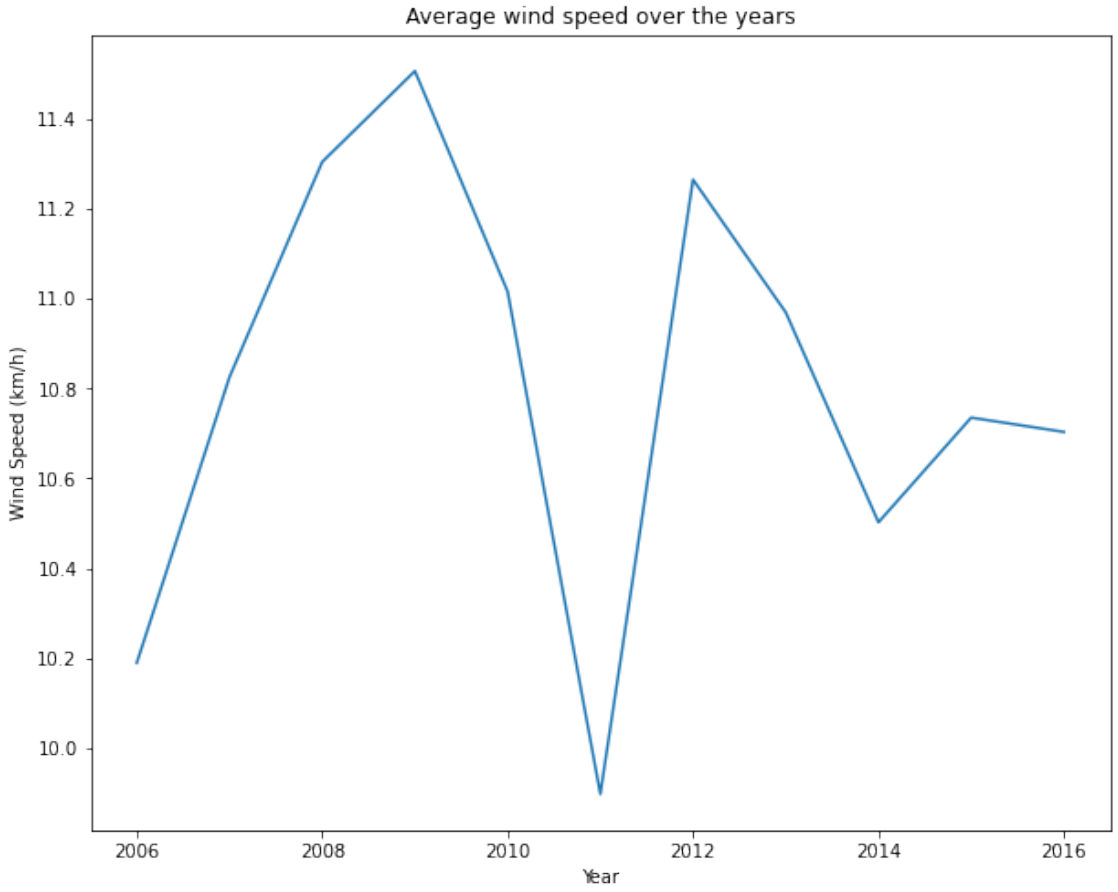
```
75%         14.135800
max         63.852600
Name: Wind Speed (km/h), dtype: float64
```

[ ]:
```
avg_wind_Speed=pd.DataFrame(df3.groupby("Year")["Wind Speed (km/h)"].mean())
avg_wind_Speed
```

[ ]:
```
      Wind Speed (km/h)
Year
2006          10.189852
2007          10.825392
2008          11.303897
2009          11.505948
2010          11.015628
2011           9.898262
2012          11.264545
2013          10.969389
2014          10.502473
2015          10.735247
2016          10.703441
```

[ ]:
```
fig,ax=plt.subplots(figsize=(10,8))
sns.lineplot(x=avg_wind_Speed.index,y=avg_wind_Speed["Wind Speed (km/h)"])
plt.title("Average wind speed over the years")
```

[ ]:
```
Text(0.5, 1.0, 'Average wind speed over the years')
```



[ ]:
```
month_avg_wind_Speed=pd.DataFrame(df3.groupby("Month")["Wind Speed (km/h)"].
  ↪mean())
order=["January","February","March","April","May","June","July","August","September",
            "October","November","December"]
monthly_wind_speed=month_avg_wind_Speed.reindex(index=order)
monthly_wind_speed
```

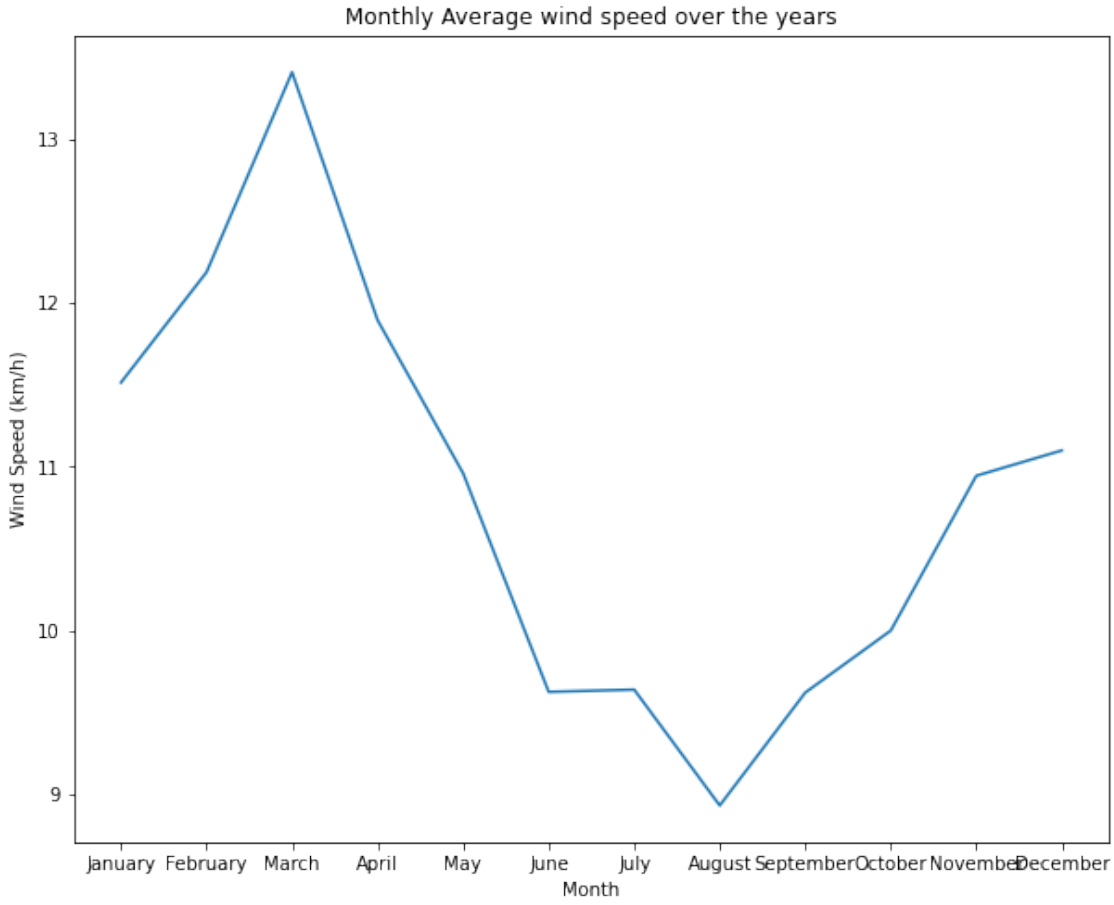
[ ]:

Month	Wind Speed (km/h)
January	11.512816
February	12.185543
March	13.405461
April	11.893094
May	10.959337
June	9.626471
July	9.639907
August	8.933431
September	9.621813
October	10.000153

```
November            10.944266
December            11.098682
```

[ ]:
```
fig,ax=plt.subplots(figsize=(10,8))
sns.lineplot(x=monthly_wind_speed.index,y=monthly_wind_speed["Wind Speed (km/
↪h)"])
plt.title("Monthly Average wind speed over the years")
```

[ ]:
```
Text(0.5, 1.0, 'Monthly Average wind speed over the years')
```



[ ]:
```
# Dew and Precipitation Count
```

[ ]:
```
df3["Summary"].value_counts()
```

[ ]:
```
Partly Cloudy                   31733
Mostly Cloudy                   28094
Overcast                        16597
Clear                           10890
```

```
Foggy                                  7148
Breezy and Overcast                     528
Breezy and Mostly Cloudy                516
Breezy and Partly Cloudy                386
Dry and Partly Cloudy                    86
Windy and Partly Cloudy                  67
Light Rain                               63
Breezy                                   54
Windy and Overcast                       45
Humid and Mostly Cloudy                  40
Drizzle                                  39
Breezy and Foggy                         35
Windy and Mostly Cloudy                  35
Dry                                      34
Humid and Partly Cloudy                  17
Dry and Mostly Cloudy                    14
Rain                                     10
Windy                                     8
Humid and Overcast                        7
Windy and Foggy                           4
Windy and Dry                             1
Dangerously Windy and Partly Cloudy       1
Breezy and Dry                            1
Name: Summary, dtype: int64
```

[ ]:
```
weather_cond=pd.DataFrame(df3.groupby("Year")["Summary"].describe(include="O").
↪top)

weather_cond.rename(columns={"top":"most frequent weather"})
```

[ ]:
```
     most frequent weather
Year
2006         Partly Cloudy
2007         Partly Cloudy
2008         Partly Cloudy
2009         Partly Cloudy
2010         Partly Cloudy
2011         Partly Cloudy
2012         Partly Cloudy
2013         Partly Cloudy
2014         Mostly Cloudy
2015         Partly Cloudy
2016         Mostly Cloudy
```

[ ]:
```
month_avg_visibility=pd.DataFrame(df3.groupby("Month")["Visibility (km)"].
↪mean())
order=["January","February","March","April","May","June","July","August","September",
```

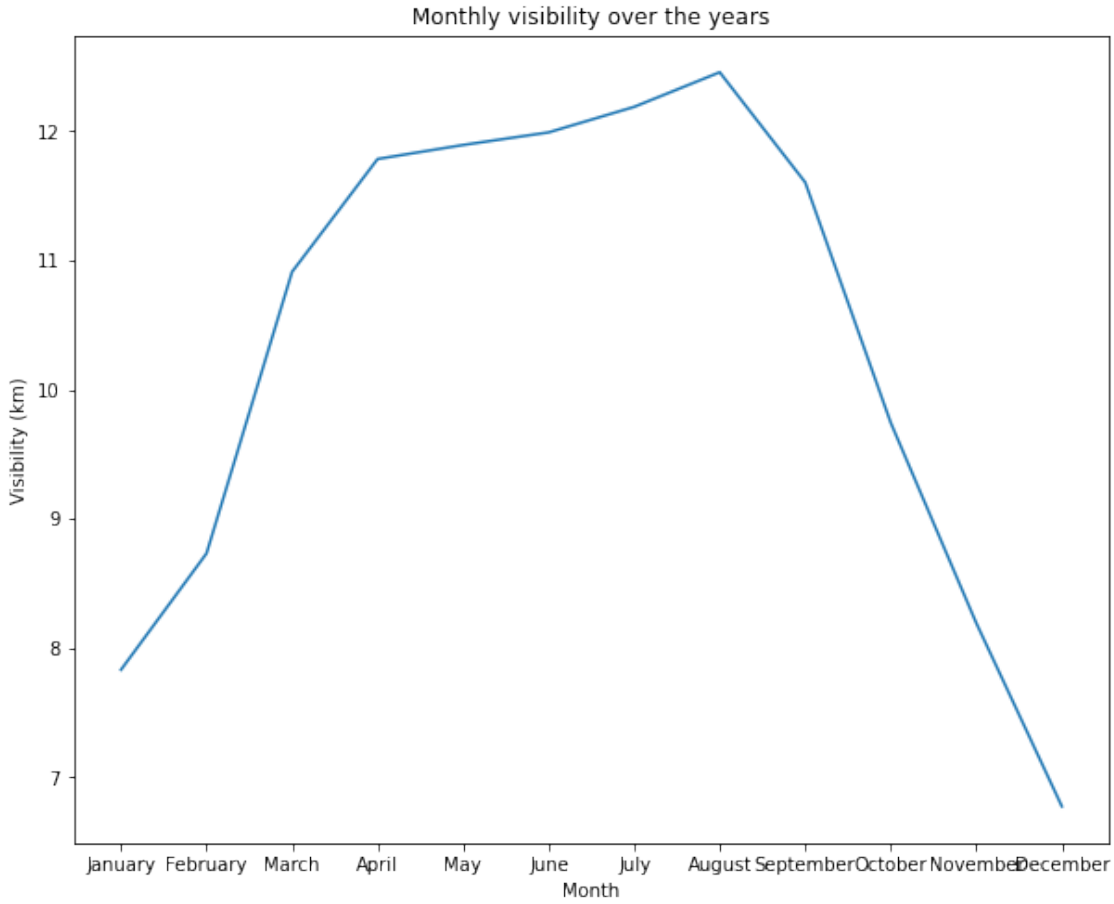
```
            "October","November","December"]
monthly_visibility=month_avg_visibility.reindex(index=order)
monthly_visibility
```

[ ]:

Month	Visibility (km)
January	7.830584
February	8.731368
March	10.910450
April	11.784224
May	11.892754
June	11.990266
July	12.187820
August	12.455549
September	11.602874
October	9.741691
November	8.191229
December	6.773288

[ ]:
```
fig,ax=plt.subplots(figsize=(10,8))
sns.lineplot(x=monthly_visibility.index,y=monthly_visibility["Visibility (km)"])
plt.title("Monthly visibility over the years")
```

[ ]:
```
Text(0.5, 1.0, 'Monthly visibility over the years')
```



[ ]:
```
percip=pd.DataFrame(df3.groupby("Month")["Precip Type"].describe(include="O").
  ↪top)
order=["January","February","March","April","May","June","July","August","September",
            "October","November","December"]
m_p=percip.rename(columns={"top":"Precip Type"})
monthly_percip=m_p.reindex(index=order)
monthly_percip
```

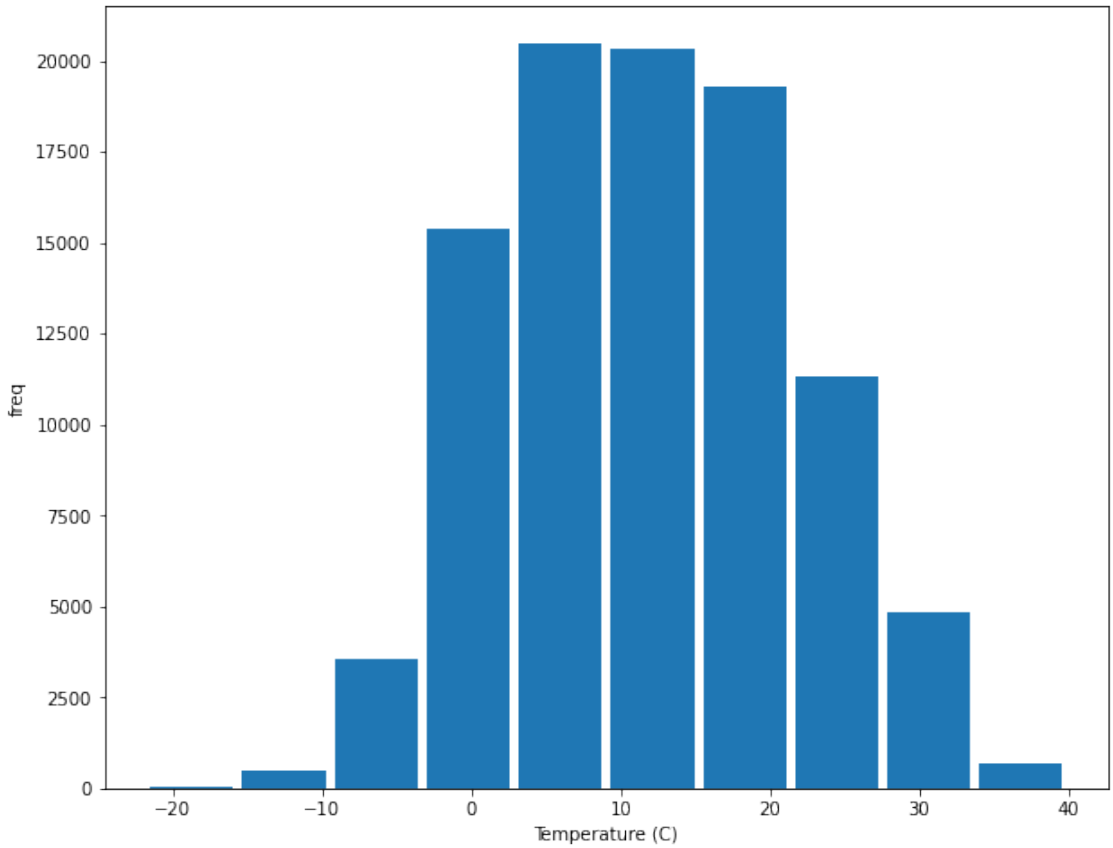
[ ]:
```
           Precip Type
Month
January           rain
February          rain
March             rain
April             rain
May               rain
June              rain
July              rain
August            rain
```

```
September       rain
October         rain
November        rain
December        rain
```

[ ]:
```
# Temperature
```

[ ]:
```
fig,ax=plt.subplots(figsize=(10,8))
plt.hist(df3["Temperature (C)"],bins=10,rwidth=0.9)
plt.xlabel("Temperature (C)")
plt.ylabel("freq")
```

[ ]:
```
Text(0, 0.5, 'freq')
```



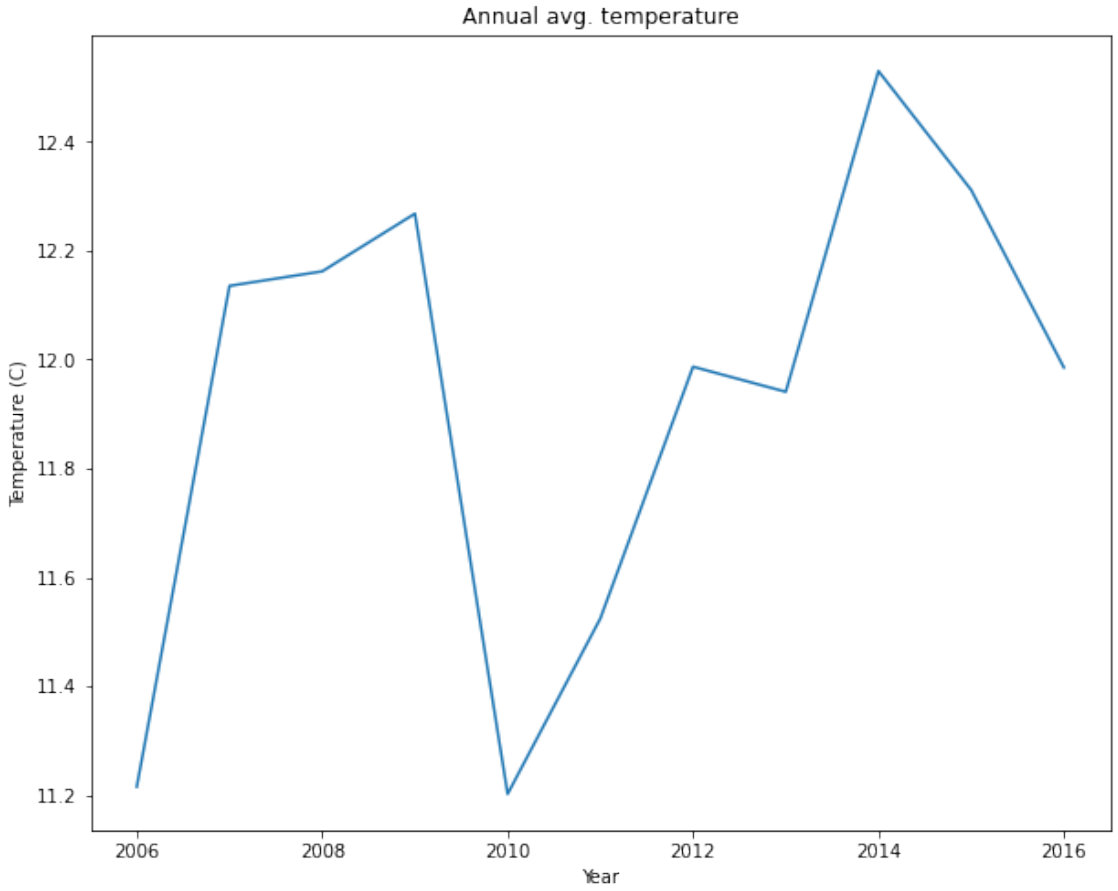
[ ]:
```
year_avg_temp=pd.DataFrame(df3.groupby("Year")["Temperature (C)"].mean())
year_avg_temp
```

[ ]:
```
      Temperature (C)
Year
2006        11.215365
```

```
2007        12.135239
2008        12.161876
2009        12.267910
2010        11.202061
2011        11.524453
2012        11.986726
2013        11.940719
2014        12.529737
2015        12.311370
2016        11.985292
```

[ ]:
```
fig,ax=plt.subplots(figsize=(10,8))
sns.lineplot(x=year_avg_temp.index,y=year_avg_temp["Temperature (C)"])
plt.title("Annual avg. temperature")
```

[ ]:
```
Text(0.5, 1.0, 'Annual avg. temperature')
```



[ ]:
```
month_temp=pd.DataFrame(df3.groupby("Month")["Temperature (C)"].mean())
order=["January","February","March","April","May","June","July","August","September",
```

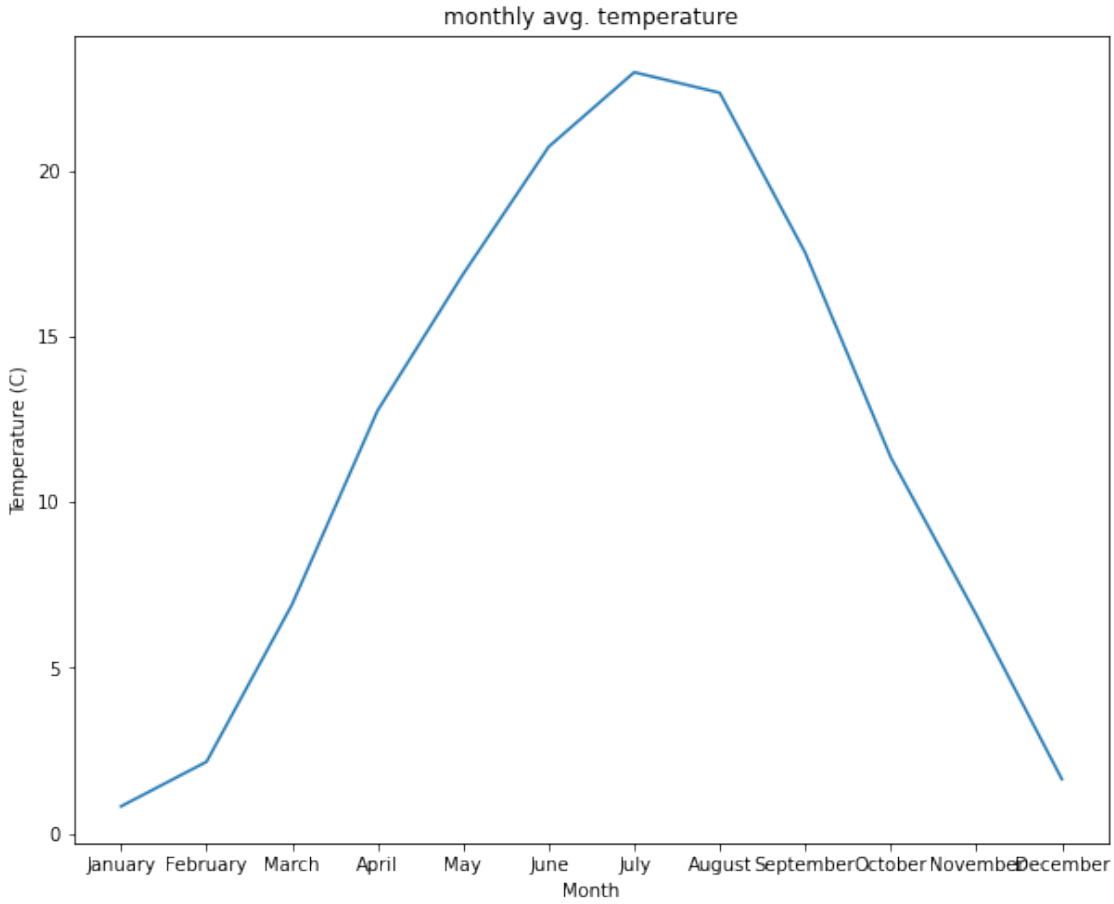
```
            "October","November","December"]

monthly_avg_temp=month_temp.reindex(index=order)
monthly_avg_temp
```

[ ]:
```
           Temperature (C)
Month
January           0.813890
February          2.159699
March             6.906599
April            12.756417
May              16.873692
June             20.715617
July             22.963943
August           22.345031
September        17.516790
October          11.342247
November          6.589907
December          1.633742
```

[ ]:
```
fig,ax=plt.subplots(figsize=(10,8))
sns.lineplot(x=monthly_avg_temp.index,y=monthly_avg_temp["Temperature (C)"])
plt.title("monthly avg. temperature")
```

[ ]:
```
Text(0.5, 1.0, 'monthly avg. temperature')
```


monthly avg. temperature
Temperature (C)
0
5
10
15
20
January
February
March
April
May
June
July
August
September
October
November
December
Month