lab^6

May 25, 2023

1 LAB №6-Pandas

1.1 The lab was fulfilled by Khilko Victoria (group K-13) , Yefremov M.S.

1.1.1 Variant 129

[1]:
```
import pandas as pd
import matplotlib.pyplot as plt
import re
```

Specify the versions of the pandas, matplotlib, and python libraries used, specify the period, get the file and convert it.

Versions

[2]:
```
pd.__version__
```

[2]:
```
'1.5.3'
```

[3]:
```
import matplotlib
matplotlib.__version__
```

[3]:
```
'3.6.3'
```

[4]:
```
import sys
sys.version
```

[4]:
```
'3.11.1 (tags/v3.11.1:a7a450f, Dec  6 2022, 19:58:39) [MSC v.1934 64 bit
(AMD64)]'
```

Period

[5]:
```
P= 129%12+1
P #23.10
```

[5]:
```
10
```

File-reading

[6]:
```
data = pd.read_csv('weather.csv',encoding='utf-8',names␣
 ↪=('day','cloudiness','rainfall','day air t',
'night air t','wind strength','period'),
```

```
header = 0)
data
```

[6]:

	day	cloudiness	rainfall	day air t	night air t	wind strength	period
0	1	72%	-	13.5°C	6.0°C	4 /	2023-04
1	2	100%	1 . .	8.0°C	8.5°C	3 /	2023-04
2	3	100%	2 . .	6.0°C	6.0°C	4 /	2023-04
3	4	100%	-	5.0°C	2.0°C	3 /	2023-04
4	5	87%	4 . .	7.5°C	3.0°C	5 /	2023-04
..	…	…	…	…	…	…	…
360	27	64%	1 . .	19.5°C	13.0°C	6 /	2022-05
361	28	39%	-	17.0°C	10.0°C	4 /	2022-05
362	29	38%	-	18.5°C	10.5°C	3 /	2022-05
363	30	86%	4 . .	15.5°C	12.5°C	4 /	2022-05
364	31	39%	-	17.0°C	13.5°C	3 /	2022-05

[365 rows x 7 columns]

Converting

[7]:
```
data['cloudiness'] = data['cloudiness'].str.findall(r'([\d.]+)%').str[0].
 ↪astype(float)
data['rainfall'] = data['rainfall'].str.findall(r'([\d.]+)  . ').str[0].
 ↪fillna('0').astype(float)
data['day air t'] = data['day air t'].str.findall(r'([\d.]+)').str[0].
 ↪astype(float)
data['night air t'] = data['night air t'].str.findall(r'([\d.]+)').str[0].
 ↪astype(float)
data['wind strength'] = data['wind strength'].str.findall(r'([\d.]+)  / ').
 ↪str[0].astype(float)
data
```

[7]:

	day	cloudiness	rainfall	day air t	night air t	wind strength	period
0	1	72.0	0.0	13.5	6.0	4.0	2023-04
1	2	100.0	1.0	8.0	8.5	3.0	2023-04
2	3	100.0	2.0	6.0	6.0	4.0	2023-04
3	4	100.0	0.0	5.0	2.0	3.0	2023-04
4	5	87.0	4.0	7.5	3.0	5.0	2023-04
..	…	…	…	…	…	…	…
360	27	64.0	1.0	19.5	13.0	6.0	2022-05
361	28	39.0	0.0	17.0	10.0	4.0	2022-05
362	29	38.0	0.0	18.5	10.5	3.0	2022-05
363	30	86.0	4.0	15.5	12.5	4.0	2022-05
364	31	39.0	0.0	17.0	13.5	3.0	2022-05

[365 rows x 7 columns]

1.2 Task 1

For the period, linear graphs of changes in daytime temperature, nighttime temperature, cloudiness and wind strength by day were constructed. The graphs were combined into one picture.

[8]:
```
data_10period = data[data['period'] == '2022-10']

data_10period
```

[8]:

	day	cloudiness	rainfall	day air t	night air t	wind strength	period
181	1	48.0	0.0	19.0	14.5	3.0	2022-10
182	2	97.0	4.0	12.0	14.0	6.0	2022-10
183	3	98.0	2.0	12.0	8.0	7.0	2022-10
184	4	100.0	2.0	9.5	7.0	5.0	2022-10
185	5	71.0	0.0	11.0	7.5	4.0	2022-10
186	6	10.0	0.0	19.5	9.5	4.0	2022-10
187	7	7.0	0.0	18.0	11.0	3.0	2022-10
188	8	1.0	0.0	18.5	8.5	2.0	2022-10
189	9	62.0	0.0	14.5	9.0	3.0	2022-10
190	10	9.0	0.0	16.0	7.5	5.0	2022-10
191	11	9.0	0.0	14.5	5.5	2.0	2022-10
192	12	42.0	0.0	14.0	9.0	3.0	2022-10
193	13	63.0	0.0	12.5	9.0	1.0	2022-10
194	14	38.0	0.0	12.0	9.0	2.0	2022-10
195	15	3.0	0.0	13.0	5.5	3.0	2022-10
196	16	3.0	0.0	13.0	4.5	3.0	2022-10
197	17	20.0	0.0	15.0	4.5	3.0	2022-10
198	18	7.0	0.0	19.5	9.5	4.0	2022-10
199	19	59.0	0.0	14.5	11.0	5.0	2022-10
200	20	99.0	0.0	4.5	4.0	6.0	2022-10
201	21	22.0	0.0	9.5	2.0	2.0	2022-10
202	22	63.0	0.0	11.5	5.0	4.0	2022-10
203	23	99.0	2.0	11.0	8.5	3.0	2022-10
204	24	87.0	0.0	14.0	10.5	3.0	2022-10
205	25	74.0	1.0	14.5	11.0	2.0	2022-10
206	26	76.0	1.0	13.0	11.0	4.0	2022-10
207	27	73.0	0.0	11.5	9.5	4.0	2022-10
208	28	26.0	0.0	13.5	8.5	3.0	2022-10
209	29	26.0	0.0	16.5	9.5	4.0	2022-10
210	30	46.0	0.0	12.5	12.5	6.0	2022-10
211	31	18.0	0.0	11.0	8.5	5.0	2022-10

[9]:
```
fig, axs = plt.subplots(2, 2, figsize=(11, 11))

data_10period.plot(x='day', y='day air t', ax=axs[0,0], color="salmon")
axs[0,0].set_title('Daytime Air Temperature, °C')
axs[0,0].set_ylabel("°C")
```

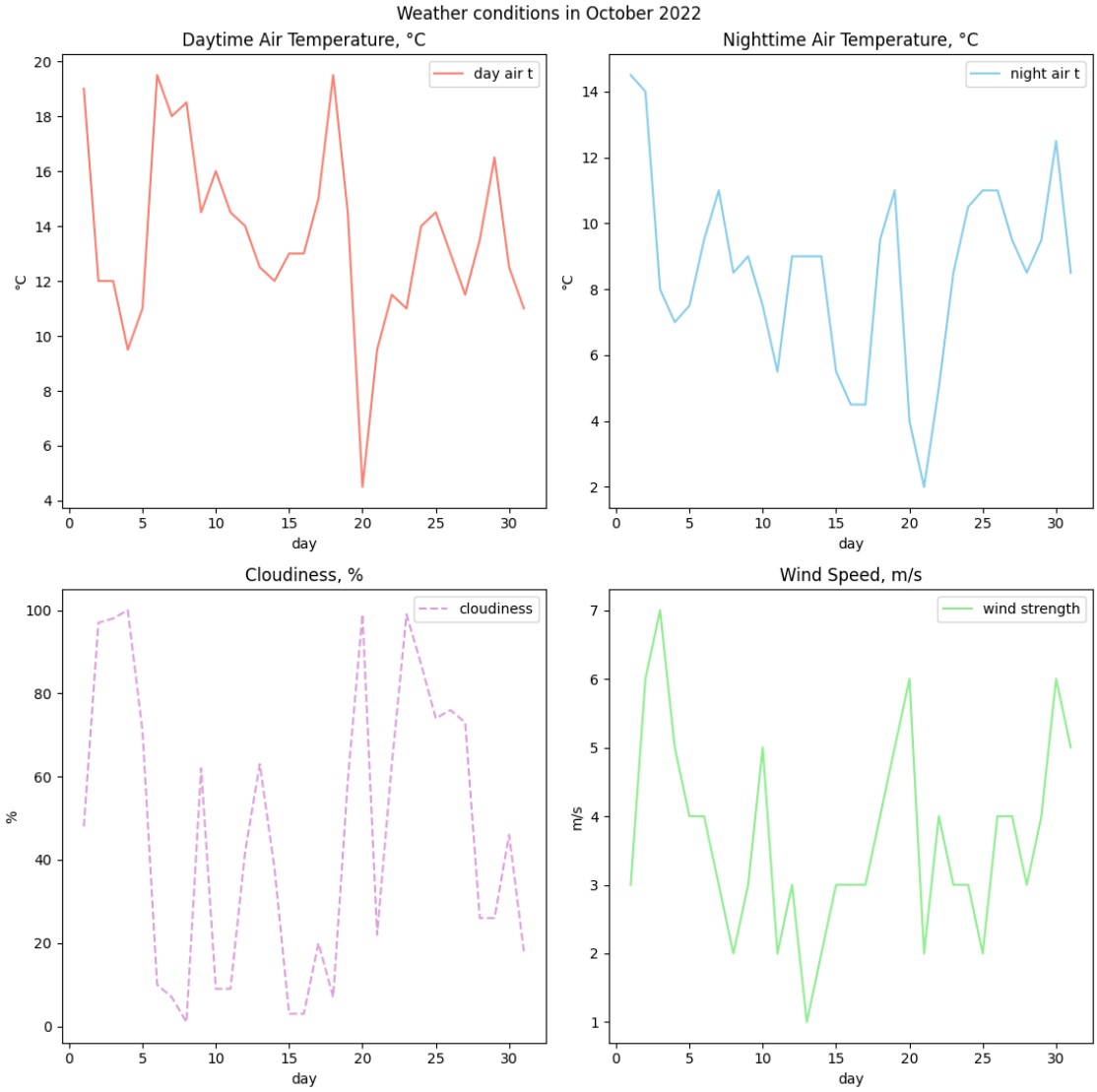
```
data_10period.plot(x='day', y='night air t', ax=axs[0,1], color="skyblue")
axs[0,1].set_title('Nighttime Air Temperature, °C')
axs[0,1].set_ylabel("°C")

data_10period.plot(x='day', y='cloudiness', ax=axs[1,0], color="plum",␣
↪linestyle="--")
axs[1,0].set_title('Cloudiness, %')
axs[1,0].set_ylabel("%")

data_10period.plot(x='day', y='wind strength', ax=axs[1,1], color="lightgreen")
axs[1,1].set_title('Wind Speed, m/s')
axs[1,1].set_ylabel("m/s")

fig.suptitle('Weather conditions in October 2022')
plt.tight_layout()


plt.show()
```

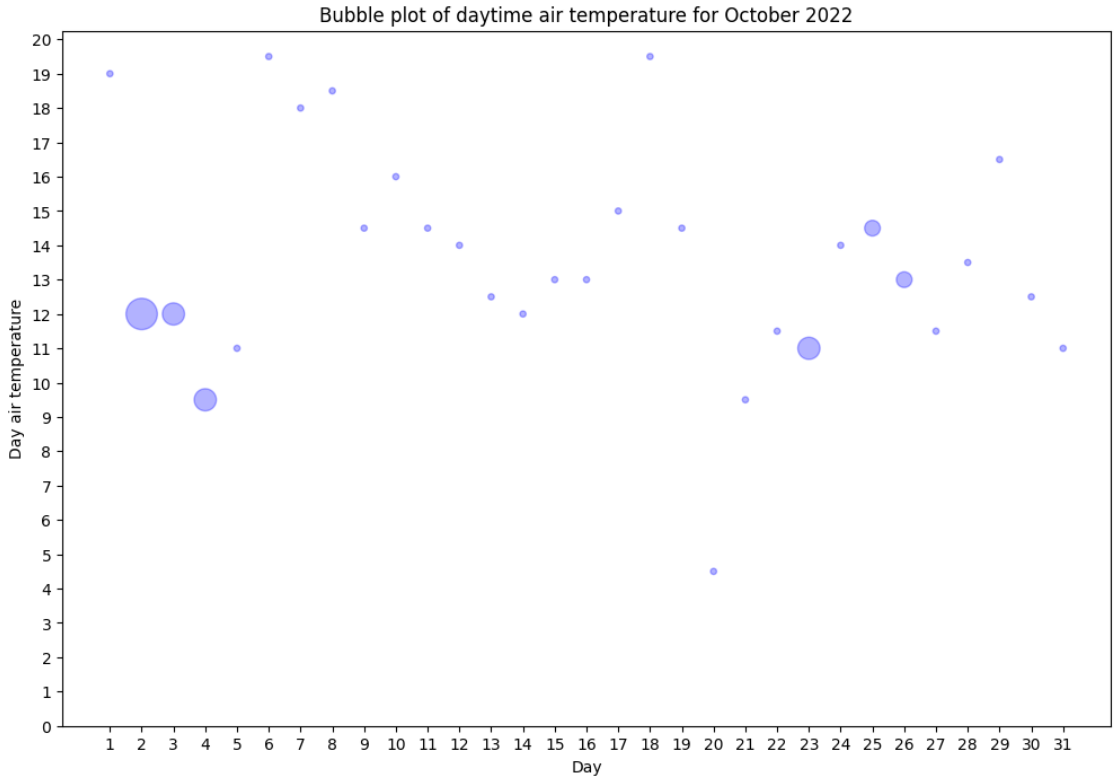


1.3 Task 2

For period P, a bubble (scatter) graph of daytime temperature was constructed, the diameter of the "bubble" depends on the amount of precipitation. If there was no precipitation, there is just a noticeable point.

[10]:
```
plt.figure(figsize=(12, 8))
plt.scatter(data_10period['day'], data_10period['day air t'],
            s=data_10period['rainfall'].apply(lambda x: 15 if x==0 else x *␣
 ↪100), alpha=0.3, color='blue')
y = range(0,21)
x = range(1,32)
plt.title("Bubble plot of daytime air temperature for October 2022")
```

```
plt.xlabel("Day")
plt.ylabel("Day air temperature")
plt.xticks(x)
plt.yticks(y)
plt.show()
```



1.4 Task 3

Average monthly deviation of night temperature from day temperature

[11]:
```
data['difference t'] =abs(data['day air t'] - data['night air t'])
data
difference_temp= data.groupby('period')['difference t'].mean()
pd.DataFrame(difference_temp)
```

[11]:

period	difference t
2022-05	7.451613
2022-06	8.316667
2022-07	8.693548
2022-08	9.467742
2022-09	5.416667

```
2022-10      5.177419
2022-11      3.133333
2022-12      2.596774
2023-01      2.048387
2023-02      2.625000
2023-03      4.338710
2023-04      4.416667
```

1.5 Task 4

All days when the largest difference between day and night temperatures was recorded per month, and all available information about them

[12]:
```
filt = data.groupby('period')['difference t'].transform(max)
largest_deviation = data[data['difference t'] == filt]
largest_deviation
```

[12]:

	day	cloudiness	rainfall	day air t	night air t	wind strength	\
21	22	21.0	0.0	16.0	7.0	3.0	
48	19	2.0	0.0	11.0	1.5	2.0	
68	8	4.0	0.0	2.0	9.0	2.0	
70	10	16.0	0.0	0.0	7.0	4.0	
89	1	47.0	0.0	11.5	6.5	6.0	
130	11	100.0	4.0	9.0	2.0	5.0	
160	10	48.0	0.0	13.5	6.0	4.0	
197	17	20.0	0.0	15.0	4.5	3.0	
218	7	6.0	0.0	18.5	6.5	2.0	
267	26	9.0	0.0	36.5	20.5	4.0	
274	2	0.0	0.0	29.5	17.5	3.0	
293	21	28.0	0.0	26.0	14.0	6.0	
295	23	11.0	0.0	29.5	17.5	3.0	
298	26	25.0	0.0	27.5	15.5	3.0	
323	20	0.0	0.0	32.0	18.5	4.0	
345	12	36.0	0.0	27.0	12.0	6.0	

	period	difference t
21	2023-04	9.0
48	2023-03	9.5
68	2023-02	7.0
70	2023-02	7.0
89	2023-01	5.0
130	2022-12	7.0
160	2022-11	7.5
197	2022-10	10.5
218	2022-09	12.0
267	2022-08	16.0
274	2022-07	12.0

```
293  2022-07           12.0
295  2022-07           12.0
298  2022-07           12.0
323  2022-06           13.5
345  2022-05           15.0
```

The day with the largest difference between day and night temperatures

[13]:
```
difference_max= (data['difference t'] == data['difference t'].max())
data.loc[difference_max]
```

[13]:
```
     day  cloudiness  rainfall  day air t  night air t  wind strength  \
267   26         9.0       0.0       36.5         20.5            4.0

      period  difference t
267  2022-08          16.0
```

1.6 Task 5

The 4 windiest days are all available information for them.

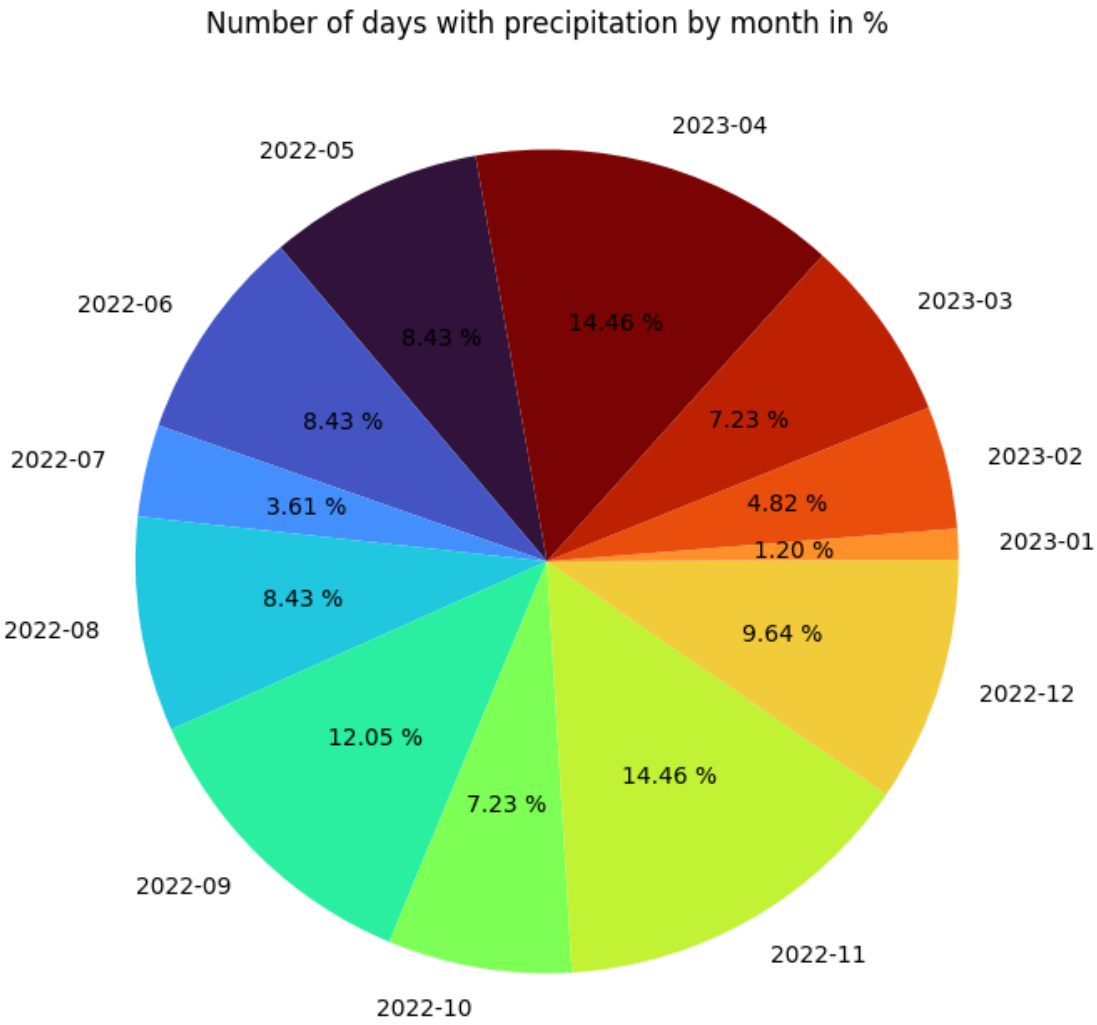
[14]:
```
wind_max_days = data_10period.nlargest(4, 'wind strength')
wind_max_days
```

[14]:
```
     day  cloudiness  rainfall  day air t  night air t  wind strength   period
183    3        98.0       2.0       12.0          8.0            7.0  2022-10
182    2        97.0       4.0       12.0         14.0            6.0  2022-10
200   20        99.0       0.0        4.5          4.0            6.0  2022-10
210   30        46.0       0.0       12.5         12.5            6.0  2022-10
```

1.7 Task 6

Pie chart of the number of days with precipitation per month

[15]:
```
days_rainfall = data[data['rainfall'] > 0].groupby('period')['day'].count()
days_rainfall.plot(kind='pie', figsize=(8, 8),autopct='%.2f␣
 ↪%%',cmap='turbo',startangle=100)
plt.title('Number of days with precipitation by month in %')
plt.ylabel("")
plt.show()
```

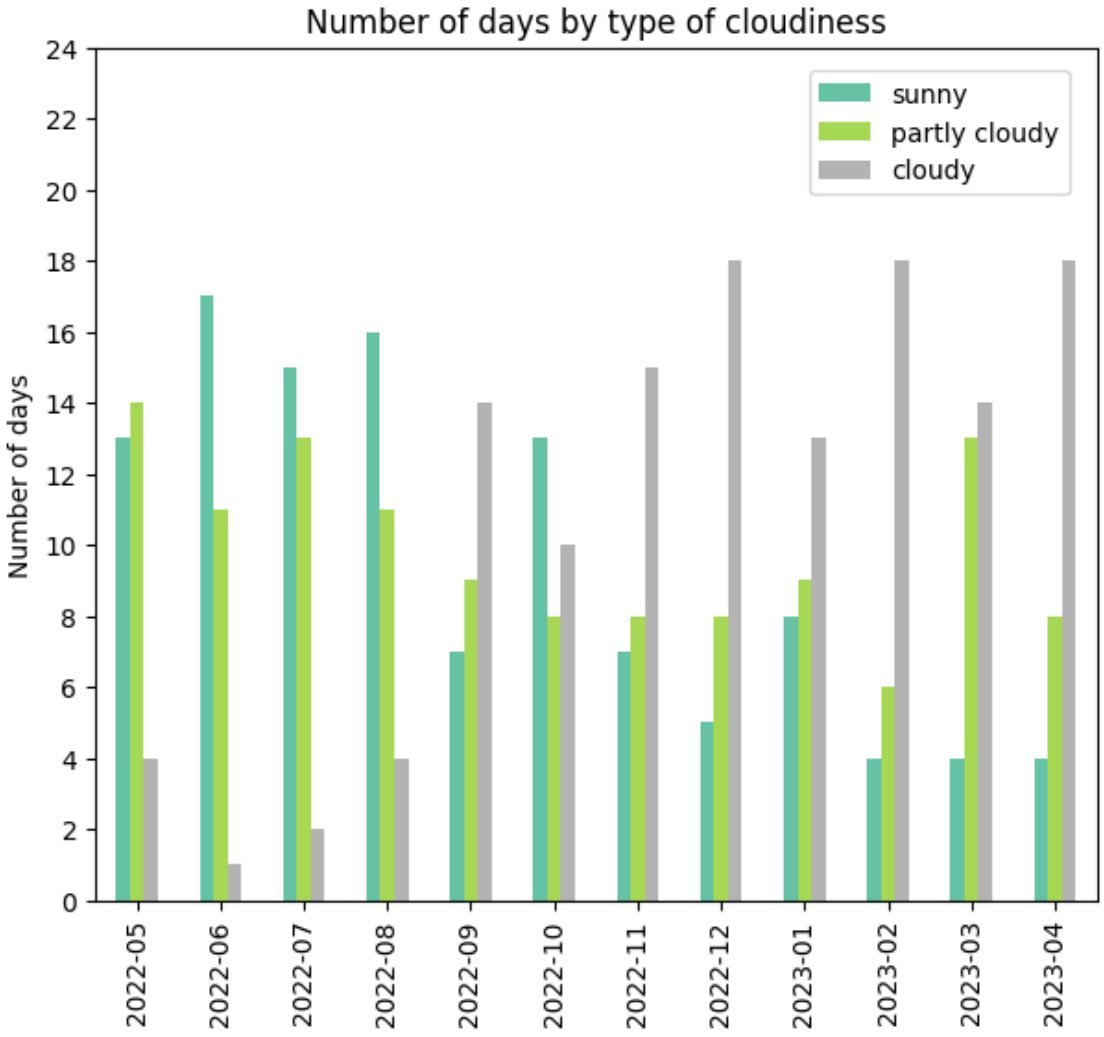


1.8 Task 7

If the cloudiness is more than 70%, then we consider the day cloudy. If the cloud cover is less than 35%, the day is considered sunny. In the rest of the cases, we believe that there was changeable cloudiness that day. We build a bar chart with accumulation by the number of days of each type per month

[16]:
```
categories = ['sunny', 'partly cloudy', 'cloudy']
thresholds = [0, 35, 70, 100]
data['cloud type'] = pd.cut(data['cloudiness'], bins=thresholds,␣
 ↪labels=categories)
group = data.groupby(['period', 'cloud type']).size().unstack()
y = range(0, 26,2)
group.plot(kind='bar',figsize=(7, 6),cmap="Set2")
```

```
plt.ylabel('Number of days')
plt.xlabel('')
plt.yticks(y)
plt.title('Number of days by type of cloudiness')
plt.legend(loc='center left', bbox_to_anchor=(0.7, 0.9))
plt.show()
```



1.9 Task 8

All months in which there were more sunny days than days with at least some precipitation

[17]:
```
sunny_days = data[data['cloud type'] == 'sunny'].groupby('period').size()
rainfall_days = data[data['rainfall'] > 0].groupby('period').size()
months = sunny_days[sunny_days > rainfall_days].index.tolist()
print("Months in which there were more sunny days than rainy days")
```

```
pd.Series(months, name='months sunny days')
#pd.DataFrame(months,columns=['months sunny days'])
```

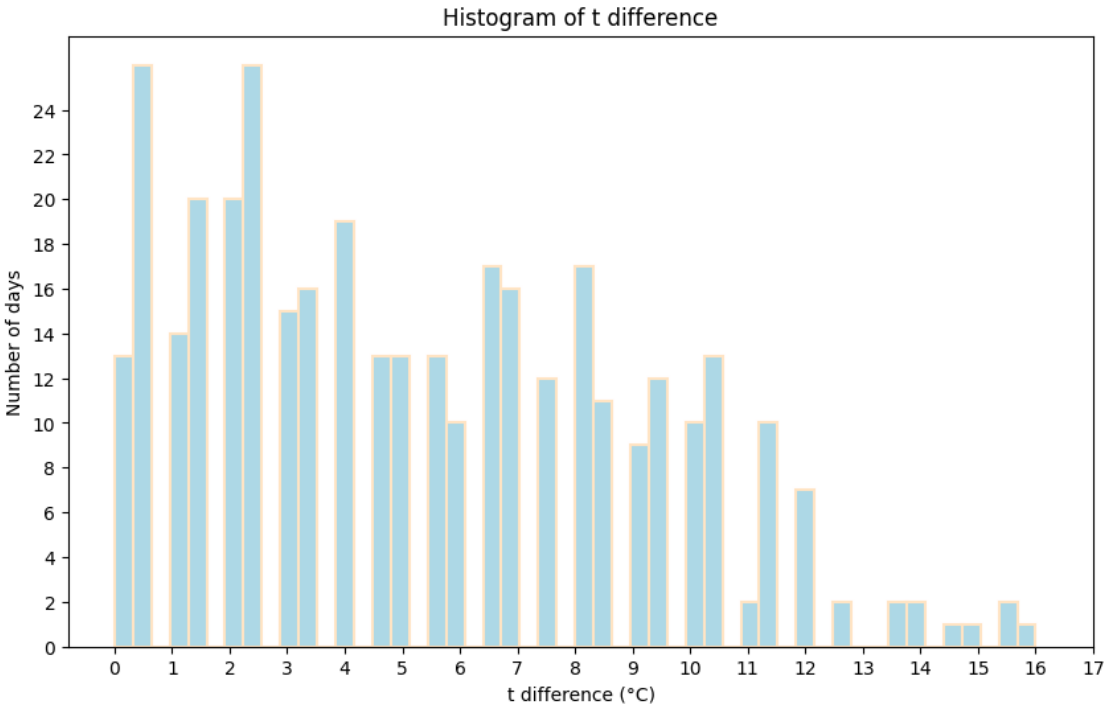
Months in which there were more sunny days than rainy days

[17]:
```
0    2022-05
1    2022-06
2    2022-07
3    2022-08
4    2022-10
5    2023-01
Name: months sunny days, dtype: object
```

1.10 Task 9

Histogram of the deviation of the night temperature from the daytime (for the number of days with a specific deviation)

[18]:
```
plt.figure(figsize=(10, 6))
plt.hist(data['difference t'],bins=50, edgecolor =␣
 ↪'bisque',color="lightblue",linewidth = 1.5)
plt.xlabel('t difference (°C)')
y = range(0, 25, 2)
x = range(0, 18)
plt.xticks(x)
plt.yticks(y)
plt.ylabel('Number of days')
plt.title('Histogram of t difference')
plt.show()
```


Histogram of t difference
Number of days
t difference (°C)
0
2
4
6
8
10
12
14
16
18
20
22
24
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17