

Отримання даних

Бібліотека - urllib, import urllib

1. Визначити змінну з посиланням та парметризувати її для отримання інформації для відповідного регіону. Краще робити це через окрему функцію, або можна використати цикл

for i in range(1,29):

```
url='https://www.star.nesdis.noaa.gov/smcd/emb/vci/VH/get_TS_admin.php?country=UKR&provinceID={}&year1=1981&year2=2020&type=Mean'.format(i)
```

Можна створити цикл і в циклі викликати функцію для отримання даних

2. Відкриття веб сторінки можна проводити так:

```
wp = urllib.request.urlopen(url)
```

```
text = wp.read()
```

3. Після відкриття у вас у змінній text записаний текст із веб сторінки, який ви можете записати у файл

```
out=open('obl_{}.csv'.format(i), 'wb')
```

```
out.write(text)
```

```
out.close()
```

При цьому в назву вам також необхідно додати дату та час завантаження файла. Це можна зробити за допомогою конкатенації (+) або формату

Так можна отримати поточну дату і час

```
now = datetime.now()
```

Згенерувати строку з поточною датою і часом та необхідним форматуванням можна за допомогою процедури strftime

```
date_and_time_time = now.strftime("%d%m%Y%H%M%S")
```

Ви можете створити новий файл за допомоги функції open

```
out = open('NOAA_ID'+str(province_n)+'_'+date_and_time_time+'.csv','wb')
```

Відкриття та робота з файлами у вигляді датафреймів

Бібліотека – pandas , import pandas as pd

1. Визначення хедерів для ваших колонок

```
headers = ['Year', 'Week', 'SMN', 'SMT', 'VCI', 'TCI', 'VHI', 'empty']
```

2. Відкриття файлу

```
df = pd.read_csv(filename, header = 1, names = headers)
```

3. Відкинути всі пап. При цьому враховуйте, що пап у файлі позначаються як -1

```
df = df.drop(df.loc[df['VHI'] == -1].index)
```

4. Додайте стовбчик з індексами ваших регіонів
`df['area'] = i`
5. Для зміни індексів можна використати `replace`
`df["area"].replace({old:new}, inplace = True)`
6. Для отримання значень VHI для обраного району та року використовуйте кондішени
`df[(df["area"] == index) & (df["Year"] == year)]['VHI']`
7. Так само кондішени використовуйте для отримання інформації щодо посух
`df_drought = df[(df.VHI <= 15) & (df.VHI != -1)]`
8. Характеристики такі як екстремуми дуже просто визначаються за допомогою функціоналу `pandas`
`min_v = df[(df.Year.astype(str)==str(i)) & (df.VHI != -1)]['VHI'].min()`