

1-introduction

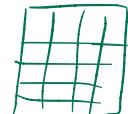
March 8, 2023

1 What is pandas-python? Introduction and Installation

Pandas is python module that makes data science easy and effective

Whenever we see a table we use 'Pandas'.

Weather dataset



Questions?

1. What was the maximum temperature in new york in the month of january?
2. On which days did it rains?
3. What was the average speed of wind during the month?

First we will see in python code

```
[ ]: # now we will see in pandas
      import pandas as pd
      df = pd.read_csv("https://drive.google.com/file/d/
      ↪1KxwFsl6IF7OD_XN28kjx10-amnELhZ8/view?usp=sharing")
      df
```

most imp. data structure
numpy → array
Pandas → dataframe

dataframe ←

localhost:8888/notebooks/Google%20Drive/OnlineVideos/2/python-basics/pandas/1_introduction/1_introduction.ipynb

File Edit View Insert Cell Kernel Help

First we will see in python code

In [4]: # now we will see in pandas

```
import pandas as pd
df = pd.read_csv('nyc_weather.csv')
df
```

Out[4]:

	EST	Temperature	DewPoint	Humidity	Sea Level PressureIn	VisibilityMiles
0	1/1/2016	38	23	52	30.03	10
1	1/2/2016	36	18	46	30.02	10
2	1/3/2016	40	21	47	29.86	10
3	1/4/2016	25	-	-	-	10
4	1/5/2016	20	-3	41	30.57	10

we are reading .csv file.

dataframe ←

710
711 parser_f_name = name

cs/pandas/ x 1_introduction x pandas_intro.py x 2_data_frame_basics x

8888/notebooks/Google%20Drive/OnlineVideos/2/python-basics/pandas/1_introduction/1_introduction.ipynb

dit View Insert Cell Kernel Help Not Trusted Python 3

df

] :

	EST	Temperature	DewPoint	Humidity	Sea Level PressureIn	VisibilityMiles	WindSpeedMPH	Precip
0	1/1/2016	38	23	52	30.03	10		8.0
1	1/2/2016	36	18	46	30.02	10		7.0
2	1/3/2016	40	21	47	29.86	10		8.0
3	1/4/2016	25	9	44	30.05	10		9.0
4	1/5/2016	20	-3	41	30.57	10		5.0
5	1/6/2016	33	4	35	30.50	10		4.0
6	1/7/2016	39	11	33	30.28	10		2.0
7	1/8/2016	39	29	64	30.20	10		4.0
8	1/9/2016	44	38	77	30.16	9		8.0
9	1/10/2016	50	46	71	29.59	4		NaN

ValueError
<ipython-input-6-f83e04cff7df> in <module>()

Traceback (most recent call last)

DELL 305 / 30

```

1 from google.colab import drive
----> 2 drive.mount('https://drive.google.com/file/d/
↳1KxwFsL6IF7OD_XN28kjl0-amnELIhZ8/view?usp=sharing')

/usr/local/lib/python3.6/dist-packages/google/colab/drive.py in
↳mount(mountpoint, force_remount)
    128     # Prompt appearing here means something went wrong with the drive,
↳binary.
    129     d.terminate(force=True)
--> 130     raise ValueError('mount failed')
    131 elif case == 2:
    132     # Not already authorized, so do the authorization dance.

ValueError: mount failed

```

[]: #get the maximum temperature
 df['Temperature'].max()

[]: 50

[]: #to know which day it rains
 df['EST'][df['Events'] == 'Rain']

[]: 8 1/9/2016
 9 1/10/2016
 15 1/16/2016
 26 1/27/2016
 Name: EST, dtype: object

[]: #3. average wind speed
 df['WindSpeedMPH'].mean()

[]: 6.8928571428571432

2 Installation

pip3 install pandas

```

File Edit View Insert Runtime Tools Help Changes will not be saved
+ Code + Text Copy to Drive
{x}
{y}
{z}

def parse_csv():
    import csv
    _file_path = ".../nyc_weather.csv"
    global _parsed_rows
    with open(_file_path, "r") as f:
        reader = csv.reader(f, delimiter=',')
        reader.next()
        for row in reader:
            _parsed_rows.append({
                'date': row[0],
                'temperature': row[1],
                'DewPoint': row[2],
                'Humidity': row[3],
                'Sea_Level_Pressurein': row[4],
                'Visibilitymi': row[5],
                'WindSpeedMPH': row[6],
                'Precipitationin': row[7],
                'CloudCover': row[8],
                'Events': row[9],
                'WindDirDegrees': row[10]
            })
    def get_max_temperature():
        max_temp = 0
        for row in _parsed_rows:
            if int(row['temperature']) > max_temp:
                max_temp = int(row['temperature'])
        return max_temp

    def get_days_of_rain(event):
        days = []
        for row in _parsed_rows:
            if row['Events'] == event:
                days.append(row['date'])
        return days

    def get_avg_wind_speed():
        total = 0
        count = 0
        for row in _parsed_rows:

```

→ if you don't have
 panda then you
 will have to write
 that much of code.
 OR

You can use
 numpy.