HW01 – Introduction to Python and Pandas

This assignment is an introduction to Python and Pandas to get familiar with how to get started with Data Science.

File

weather.csv – two months of weather data from 3/1/17 to 4/30/17 in Sturgis, KY containing:

- Date
- MaxTemp
- MaxTime
- MinTemp
- MinTime
- cloudCover
- precipProbability
- precipAmount
- sunriseTime
- sunsetTime
- windSpeed

Format of this Homework

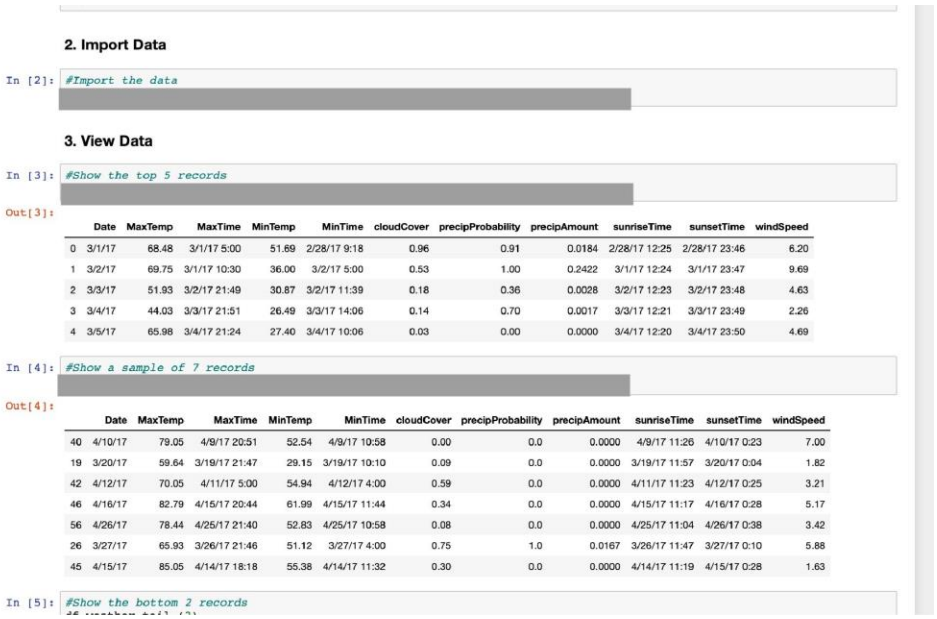
IMPORTANT – Be sure your Jupyter Notebook is formatted correctly with markdown, comments, and code that works. Significant points will be deducted for missing section titles and summaries (markdown)

You are to do the following for each section (See Figure 1 below for title examples):

- Include a title as markdown Heading 2, for example: "Section 3. View Data"
- Include a description of the section detailing its purpose (*markdown*)
- Include your code and make sure it is executable and correct. (*code*)
- At the end of the section, include a brief summary describing results. Your summaries should get more detailed as the course progresses. (*markdown*)

How to turn it in:

- Your Jupyter notebook file must be named HW01_LastnameFirstInitial.ipynb. For example, HW01_ApigianC.ipynb.
- You are to turn in your Jupyter notebook file only. No data files and no folders.
- It is assumed that you created your Jupyter notebook in a folder named HW01_student and in that folder is a data folder. It is expected the path for importing data is in "data" folder, for example '**data/weather.csv'**.



2. Import Data

```
In [2]: #Import the data
```

3. View Data

```
In [3]: #Show the top 5 records
```

Out[3]:

	Date	MaxTemp	MaxTime	MinTemp	MinTime	cloudCover	precipProbability	precipAmount	sunriseTime	sunsetTime	windSpeed
0	3/1/17	68.48	3/1/17 5:00	51.69	2/28/17 9:18	0.96	0.91	0.0184	2/28/17 12:25	2/28/17 23:46	6.20
1	3/2/17	69.75	3/1/17 10:30	36.00	3/2/17 5:00	0.53	1.00	0.2422	3/1/17 12:24	3/1/17 23:47	9.69
2	3/3/17	51.93	3/2/17 21:49	30.87	3/2/17 11:39	0.18	0.36	0.0028	3/2/17 12:23	3/2/17 23:48	4.63
3	3/4/17	44.03	3/3/17 21:51	26.49	3/3/17 14:06	0.14	0.70	0.0017	3/3/17 12:21	3/3/17 23:49	2.26
4	3/5/17	65.98	3/4/17 21:24	27.40	3/4/17 10:06	0.03	0.00	0.0000	3/4/17 12:20	3/4/17 23:50	4.69

```
In [4]: #Show a sample of 7 records
```

Out[4]:

	Date	MaxTemp	MaxTime	MinTemp	MinTime	cloudCover	precipProbability	precipAmount	sunriseTime	sunsetTime	windSpeed
40	4/10/17	79.05	4/9/17 20:51	52.54	4/9/17 10:58	0.00	0.0	0.0000	4/9/17 11:26	4/10/17 0:23	7.00
19	3/20/17	59.64	3/19/17 21:47	29.15	3/19/17 10:10	0.09	0.0	0.0000	3/19/17 11:57	3/20/17 0:04	1.82
42	4/12/17	70.05	4/11/17 5:00	54.94	4/12/17 4:00	0.59	0.0	0.0000	4/11/17 11:23	4/12/17 0:25	3.21
46	4/16/17	82.79	4/15/17 20:44	61.99	4/15/17 11:44	0.34	0.0	0.0000	4/15/17 11:17	4/16/17 0:28	5.17
56	4/26/17	78.44	4/25/17 21:40	52.83	4/25/17 10:58	0.08	0.0	0.0000	4/25/17 11:04	4/26/17 0:38	3.42
26	3/27/17	65.93	3/26/17 21:46	51.12	3/27/17 4:00	0.75	1.0	0.0167	3/26/17 11:47	3/27/17 0:10	5.88
45	4/15/17	85.05	4/14/17 18:18	55.38	4/14/17 11:32	0.30	0.0	0.0000	4/14/17 11:19	4/15/17 0:28	1.63

```
In [5]: #Show the bottom 2 records
```

Figure 1: Screenshot of Section 2 and 3, with titles

Create a new Jupyter Notebook

1. **Import Libraries** *(Notebook Intro to Pandas)*
   - numpy as np, pandas as pd, and datetime as dt
2. **Import Data** *(Notebook Intro to Pandas – Code Block 01)*
   - weather.csv as df_weather with index_col = None and header=0
3. **View Data** *(Notebook Intro to Pandas & Video Module 02_03)*
   - Create a code block to view the top 5 rows of df_weather (see Figure 2 below)
   - Create a code block to view a sample of 7 rows of df_weather
   - Create a code block to view the bottom 2 rows of df_weather

Hint: Code block refers to creating a block of code to execute something. In your code, there should be some syntax used to create the output and the output should occur when executed.

Out[4]:

	Date	MaxTemp	MaxTime	MinTemp	MinTime	cloudCover	icon	precipProbability	precipType	summary	sunriseTime	sunsetTime	windSpeed
0	3/1/17	68.48	3/1/17 5:00	51.69	2/28/17 9:18	0.96	rain	0.91	rain	Light rain in the morning.	2/28/17 12:25	2/28/17 23:46	6.20
1	3/2/17	69.75	3/1/17 10:30	36.00	3/2/17 5:00	0.53	rain	1.00	rain	Rain in the morning.	3/1/17 12:24	3/1/17 23:47	9.69
2	3/3/17	51.93	3/2/17 21:49	30.87	3/2/17 11:39	0.18	clear-day	0.36	rain	Clear throughout the day.	3/2/17 12:23	3/2/17 23:48	4.63
3	3/4/17	44.03	3/3/17 21:51	26.49	3/3/17 14:06	0.14	rain	0.70	rain	Drizzle in the morning and afternoon.	3/3/17 12:21	3/3/17 23:49	2.26
4	3/5/17	65.98	3/4/17 21:24	27.40	3/4/17 10:06	0.03	clear-day	0.00	NaN	Clear throughout the day.	3/4/17 12:20	3/4/17 23:50	4.69

Figure 2: Screenshot of viewing the top 5 rows of df_weather

4. Create a variable and a way to convert from Celsius to Fahrenheit *(Video Module 01_02)*

- Create a variable for Fahrenheit (f) that is equal to 32
- Create a variable for Celius (c) that is equal to 0
- Print f and c in the same code block.
- Create a formula that takes Fahrenheit (f) and converts to Celsius (c)
    - Change Fahrenheit to 86
    - (Fahrenheit − 32) × 5/9 = Celsius
    - Print Celsius
- Create a statement that will state:
    - "Fahrenheit 86 converts to 30 Celsius"
    - You will need to convert f and c to a string.
    - Your Celsius or Fahrenheit may or may not have decimal places.

5. Convert Objects to Dates using datetime *(Video Module 02_06)*

- For the following variables, convert them from objects to datetime
    - df_weather['MaxTime']
    - df_weather['MinTime']
    - df_weather['sunriseTime']
    - df_weather[' sunsetTime']

6. Create additional Date Features *(Video Module 02_06)*

- For df_weather:
    - Create a ['MaxDay'] feature with number for dayofweek
    - Create a ['MaxMonth'] feature with number for month
    - Create a ['MaxDayName'] feature for day_name( )
- Show the top 5 records for df_weather.

7. Create a new column *(Video Module 01 prep & Module 02_12)*

- Create a new column named ['TempRange] by subtracting MaxTemp and MinTemp.
- Show the top 5 records for df_weather.

8. Slice and Filter Data *(Video Module 02_03)*

- Show the MaxDay column only by specifying the label name. Show only top 5 records.

Out[35]:

	MaxDay
0	2.0
1	2.0
2	3.0
3	4.0
4	5.0

Figure 3: Screenshot of viewing the top 5 rows of MaxDay

- Show the MaxMonth column only by using iloc. Show only top 5 records.

Out[37]:

	MaxMonth
0	3.0
1	3.0
2	3.0
3	3.0
4	3.0

Figure 4: Screenshot of viewing the top 5 rows of MaxMonth using iloc

- Create a new DataFrame named df_temp that includes ['MaxTemp'], ['MinTemp'], ['TempRange'], and ['precipAmount']. Use any technique to create df_temp. *(Notebook – Intro to Pandas – Code Block 11)*

```
In [45]: df_temp.head()
```

Out[45]:

	MaxTemp	MinTemp	TempRange	precipAmount
0	68.48	51.69	16.79	0.0184
1	69.75	36.00	33.75	0.2422
2	51.93	30.87	21.06	0.0028
3	44.03	26.49	17.54	0.0017
4	65.98	27.40	38.58	0.0000

Figure 5: Screenshot of viewing the top 5 rows of df_temp

- Show the descriptive statistics by using .describe(). *(Notebook Intro to Pandas – Code Block 14)*

Out[47]:

	MaxTemp	MinTemp	TempRange	precipAmount
count	61.000000	61.000000	61.000000	62.000000
mean	65.899672	44.517869	21.381803	0.029766
std	13.686520	12.397472	7.482015	0.056639
min	29.530000	16.390000	8.220000	0.000000
25%	56.630000	36.000000	15.750000	0.000000
50%	68.700000	45.700000	20.890000	0.000000
75%	76.190000	52.690000	25.610000	0.020400
max	86.360000	66.100000	38.580000	0.242200

Figure 6: Screenshot of viewing describe() for df_temp

Submission

Save your file as HW01_LastNameFirstInitial.ipynb and turn it in to D2L per the Dropbox instructions. (file only, no data)