[Jupyter Notebook](https://serversm43d9ju-dev-machine-server-8000.ind.hackerrank.com/tree)

EDA\_question(unsaved changes)

Python 3

Not Trusted

* [File](https://serversm43d9ju-dev-machine-server-8000.ind.hackerrank.com/notebooks/EDA_question.ipynb)
* [Edit](https://serversm43d9ju-dev-machine-server-8000.ind.hackerrank.com/notebooks/EDA_question.ipynb)
* [View](https://serversm43d9ju-dev-machine-server-8000.ind.hackerrank.com/notebooks/EDA_question.ipynb)
* [Insert](https://serversm43d9ju-dev-machine-server-8000.ind.hackerrank.com/notebooks/EDA_question.ipynb)
* [Cell](https://serversm43d9ju-dev-machine-server-8000.ind.hackerrank.com/notebooks/EDA_question.ipynb)
* [Kernel](https://serversm43d9ju-dev-machine-server-8000.ind.hackerrank.com/notebooks/EDA_question.ipynb)
* [Help](https://serversm43d9ju-dev-machine-server-8000.ind.hackerrank.com/notebooks/EDA_question.ipynb)

Run



**Directions**

- The data required for this task has been provided in the file 'data.csv'

- Read the questions provided for each cell and assign your answers to respective variables provided in the following cell.

- If answers are floating point numbers round of updo two floating point after the decimal

- for example 10.546 should be read as 10.55, 10.544 as 10.54 and 10.1 as 10.10

- pandas and numpy packages are preinstalled for this task which should be sufficient to complete this task.

- If you need any other additional package run !pip3 install <package\_name> --user in a new cell.

- You can either try out the solution in the same notebook or free to create additional notebook, but make sure you come back to this notebook to answer the questions.

- Please dont change variable name meant to assign your answers.

- Dont leave any of the answers blank for you test cases run smoothly

In [72]:



**import** pandas **as** pd

**import** numpy **as** np

In [73]:



*### Read the data (this will not be graded)*

​

data **=** pd.read\_csv('data.csv')

data.head(22)

​

​

​

​

Out[73]:

|  | **Day** | **Average temperature (°F)** | **Average humidity (%)** | **Average dewpoint (°F)** | **Average barometer (in)** | **Average windspeed (mph)** | **Average gustspeed (mph)** | **Average direction (°deg)** | **Rainfall for month (in)** | **Rainfall for year (in)** | **Maximum rain per minute** | **Maximum temperature (°F)** | **Minimum temperature (°F)** | **Maximum humidity (%)** | **Minimum humidity (%)** | **Maximum pressure** | **Minimum pressure** | **Maximum windspeed (mph)** | **Maximum gust speed (mph)** | **Maximum heat index (°F)** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **0** | 1/01/2009 | 37.8 | 35 | 12.7 | 29.7 | 26.4 | 36.8 | 274 | 0.00 | 0.00 | 0.00 | 40.1 | 34.5 | 44 | 27 | 29.762 | 29.596 | 41.4 | 59.0 | 40.1 |
| **1** | 2/01/2009 | 43.2 | 32 | 14.7 | 29.5 | 12.8 | 18.0 | 240 | 0.00 | 0.00 | 0.00 | 52.8 | 37.5 | 43 | 16 | 29.669 | 29.268 | 35.7 | 51.0 | 52.8 |
| **2** | 3/01/2009 | 25.7 | 60 | 12.7 | 29.7 | 8.3 | 12.2 | 290 | 0.00 | 0.00 | 0.00 | 41.2 | 6.7 | 89 | 35 | 30.232 | 29.260 | 25.3 | 38.0 | 41.2 |
| **3** | 4/01/2009 | 9.3 | 67 | 0.1 | 30.4 | 2.9 | 4.5 | 47 | 0.00 | 0.00 | 0.00 | 19.4 | -0.0 | 79 | 35 | 30.566 | 30.227 | 12.7 | 20.0 | 32.0 |
| **4** | 5/01/2009 | 23.5 | 30 | -5.3 | 29.9 | 16.7 | 23.1 | 265 | 0.00 | 0.00 | 0.00 | 30.3 | 15.1 | 56 | 13 | 30.233 | 29.568 | 38.0 | 53.0 | 32.0 |
| **5** | 6/01/2009 | 24.8 | 42 | 4.6 | 29.8 | 16.0 | 23.9 | 276 | 0.00 | 0.00 | 0.00 | 29.5 | 19.7 | 57 | 27 | 29.879 | 29.637 | 29.9 | 48.0 | 32.0 |
| **6** | 7/01/2009 | 34.2 | 60 | 21.6 | 29.7 | 20.4 | 30.0 | 276 | 0.00 | 0.00 | 0.00 | 39.1 | 27.3 | 86 | 46 | 29.860 | 29.602 | 38.0 | 54.0 | 39.1 |
| **7** | 8/01/2009 | 42.1 | 41 | 20.0 | 29.8 | 17.5 | 25.2 | 265 | 0.00 | 0.00 | 0.00 | 51.3 | 36.9 | 53 | 28 | 29.883 | 29.627 | 35.7 | 49.0 | 51.3 |
| **8** | 9/01/2009 | 30.3 | 46 | 11.4 | 30.0 | 6.9 | 10.6 | 292 | 0.00 | 0.00 | 0.00 | 41.5 | 19.2 | 82 | 27 | 30.446 | 29.575 | 24.2 | 36.0 | 41.5 |
| **9** | 10/01/2009 | 26.2 | 38 | 3.6 | 30.4 | 18.2 | 24.6 | 258 | 0.00 | 0.00 | 0.00 | 31.9 | 22.5 | 48 | 29 | 30.489 | 30.291 | 31.1 | 46.0 | 32.0 |
| **10** | 11/01/2009 | 30.1 | 41 | 9.0 | 30.3 | 17.0 | 24.6 | 270 | 0.00 | 0.00 | 0.00 | 36.3 | 26.0 | 52 | 29 | 30.421 | 30.112 | 31.1 | 44.0 | 36.3 |
| **11** | 12/01/2009 | 23.4 | 60 | 11.0 | 30.3 | 7.4 | 11.0 | 351 | 0.03 | 0.03 | 0.01 | 31.6 | 16.1 | 87 | 41 | 30.524 | 30.012 | 25.3 | 33.0 | 32.0 |
| **12** | 13/01/2009 | 32.4 | 47 | 14.6 | 30.2 | 16.3 | 22.6 | 254 | 0.03 | 0.03 | 0.00 | 38.0 | 28.5 | 54 | 34 | 30.340 | 30.046 | 32.2 | 47.0 | 38.0 |
| **13** | 14/01/2009 | 33.1 | 45 | 14.2 | 30.2 | 9.9 | 14.0 | 259 | 0.03 | 0.03 | 0.00 | 39.3 | 28.7 | 53 | 34 | 30.271 | 30.115 | 29.9 | 32.0 | 39.3 |
| **14** | 15/01/2009 | 35.7 | 42 | 14.5 | 30.3 | 7.9 | 11.8 | 270 | 0.03 | 0.03 | 0.00 | 43.6 | 27.9 | 59 | 28 | 30.405 | 30.201 | 19.6 | 26.0 | 43.6 |
| **15** | 16/01/2009 | 41.6 | 28 | 9.7 | 30.3 | 11.9 | 16.7 | 269 | 0.03 | 0.03 | 0.00 | 53.0 | 32.1 | 46 | 14 | 30.432 | 30.127 | 26.5 | 37.0 | 53.0 |
| **16** | 17/01/2009 | 40.9 | 21 | 3.1 | 30.3 | 6.7 | 9.7 | 276 | 0.03 | 0.03 | 0.00 | 54.1 | 28.1 | 28 | 11 | 30.435 | 30.093 | 21.9 | 26.0 | 54.1 |
| **17** | 18/01/2009 | 43.4 | 15 | -1.7 | 30.3 | 9.6 | 14.7 | 288 | 0.03 | 0.03 | 0.00 | 52.9 | 27.3 | 27 | 8 | 30.425 | 30.190 | 24.2 | 36.0 | 52.9 |
| **18** | 19/01/2009 | 46.6 | 19 | 3.6 | 30.3 | 7.0 | 10.5 | 293 | 0.03 | 0.03 | 0.00 | 57.5 | 36.9 | 39 | 7 | 30.333 | 30.154 | 20.7 | 28.0 | 57.5 |
| **19** | 20/01/2009 | 40.0 | 27 | -1.0 | 30.4 | 4.8 | 7.4 | 285 | 0.03 | 0.03 | 0.00 | 61.0 | 16.1 | 69 | 4 | 30.607 | 30.016 | 18.4 | 26.0 | 61.0 |
| **20** | 21/01/2009 | 48.2 | 11 | -4.6 | 30.0 | 12.4 | 16.8 | 257 | 0.03 | 0.03 | 0.00 | 55.7 | 41.6 | 19 | 6 | 30.143 | 29.871 | 25.3 | 35.0 | 55.7 |
| **21** | 22/01/2009 | 44.8 | 40 | 20.8 | 29.9 | 11.8 | 16.3 | 264 | 0.03 | 0.03 | 0.00 | 52.3 | 38.9 | 63 | 17 | 29.911 | 29.784 | 31.1 | 44.0 | 52.3 |

**What is the standard deviation of maximum windspeed across all the days**

In [74]:



ws\_std **=** round(np.std(data['Maximum windspeed (mph)']),2)

print(ws\_std)

13.06

**What is the difference between 50th percentile and 75th percentile of average temperature**

In [75]:



p\_range **=** format((np.percentile(data['Average temperature (°F)'],75)**-**np.percentile(data['Average temperature (°F)'],50)),'.2f')

print(p\_range)

12.20

In [12]:



**!**pip3 install scipy **--**user

Collecting scipy

Downloading <https://files.pythonhosted.org/packages/7a/0e/3781e028d62a8422244582abd8f084e6314297026760587c85607f687bf3/scipy-1.3.1-cp35-cp35m-manylinux1_x86_64.whl> (25.1MB)

|████████████████████████████████| 25.1MB 639kB/s eta 0:00:01

Requirement already satisfied: numpy>=1.13.3 in /home/user/.local/lib/python3.5/site-packages (from scipy) (1.14.1)

Installing collected packages: scipy

Successfully installed scipy-1.3.1

In [77]:



**from** scipy.stats **import** pearsonr

**What is the pearson correlation between average dew point and average temperature**

In [21]:



corr, \_ **=** pearsonr(data['Average dewpoint (°F)'],data['Average temperature (°F)'])

corr **=** format(corr,'.2f')

0.76

**Out of all the available records which month has the lowest average humidity.**

- Assign your answer as month index, for example if its July index is 7

In [78]:



dew\_month **=** 1

**Which month has the highest median for maximum\_gust\_speed out of all the available records. Also find the repective value**

- hint: group by month

In [101]:



*#days=[d.split('/')[0] for d in data.Day]*

*#data['Days'] = days*

Month**=**[d.split('/')[1] **for** d **in** data.Day]

data['Month'] **=** Month

grp\_mnth **=** data.groupby('Month') [['Maximum gust speed (mph)']].median()

​

max\_gust\_value **=** grp\_mnth[['Maximum gust speed (mph)']].idxmax()

max\_gust\_month **=** grp\_mnth[['Maximum gust speed (mph)']].max()

Maximum gust speed (mph) 02

dtype: object

**Determine the average temperature between the months of March 2010 to May 2012 (including both the months)**

In [106]:



Year**=**[d.split('/')[2] **for** d **in** data.Day]

data['Year'] **=** Year

data.head()

*#avg\_temp =*

Out[106]:

|  | **Day** | **Average temperature (°F)** | **Average humidity (%)** | **Average dewpoint (°F)** | **Average barometer (in)** | **Average windspeed (mph)** | **Average gustspeed (mph)** | **Average direction (°deg)** | **Rainfall for month (in)** | **Rainfall for year (in)** | **...** | **Minimum temperature (°F)** | **Maximum humidity (%)** | **Minimum humidity (%)** | **Maximum pressure** | **Minimum pressure** | **Maximum windspeed (mph)** | **Maximum gust speed (mph)** | **Maximum heat index (°F)** | **Month** | **Year** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **0** | 1/01/2009 | 37.8 | 35 | 12.7 | 29.7 | 26.4 | 36.8 | 274 | 0.0 | 0.0 | ... | 34.5 | 44 | 27 | 29.762 | 29.596 | 41.4 | 59.0 | 40.1 | 01 | 2009 |
| **1** | 2/01/2009 | 43.2 | 32 | 14.7 | 29.5 | 12.8 | 18.0 | 240 | 0.0 | 0.0 | ... | 37.5 | 43 | 16 | 29.669 | 29.268 | 35.7 | 51.0 | 52.8 | 01 | 2009 |
| **2** | 3/01/2009 | 25.7 | 60 | 12.7 | 29.7 | 8.3 | 12.2 | 290 | 0.0 | 0.0 | ... | 6.7 | 89 | 35 | 30.232 | 29.260 | 25.3 | 38.0 | 41.2 | 01 | 2009 |
| **3** | 4/01/2009 | 9.3 | 67 | 0.1 | 30.4 | 2.9 | 4.5 | 47 | 0.0 | 0.0 | ... | -0.0 | 79 | 35 | 30.566 | 30.227 | 12.7 | 20.0 | 32.0 | 01 | 2009 |
| **4** | 5/01/2009 | 23.5 | 30 | -5.3 | 29.9 | 16.7 | 23.1 | 265 | 0.0 | 0.0 | ... | 15.1 | 56 | 13 | 30.233 | 29.568 | 38.0 | 53.0 | 32.0 | 01 | 2009 |

5 rows × 22 columns

**Find the range of averange temperature on Dec 2010**

In [8]:



temp\_range **=**

**Out of all available records which day has the highest difference between maximum\_pressure and minimum\_pressure**

- assign the date in string format as 'yyyy-mm-dd'. Make sure you enclose it with single quote

In [9]:



max\_p\_range\_day **=**

**How many days falls under median of barrometer reading.**

In [10]:



median\_b\_days **=**

**Out of all the available records how many days are within one standard deviation of average temperaturem**

In [11]:



num\_days\_std **=**

**Once you are done with your solution make sure you have saved the notebook (ctrl + s)**