Code with me: Data Exploration with Pandas using Volcano.csv file

```
1 # Target Age Group: Teens with Python coding background
```

Note: This dataset has no date of last eurption. it is a list of known volcano with location data on on global volcanic hazard, historical events, population exposure, vulnerability, and impact has been provided to GAR15 by Global Volcano Model (GVM) and The International Association of Volcanology and Chemistry of the Earth's Interior (IAVCEI).

Download data from here

Panda Cheatsheet

```
1 # Pandas Basic
2 # http://datacamp-community-prod.s3.amazonaws.com/dbed353d-2757-4617-8206-8767ab379ab3
3
4 # Pandas Data Wrangling
5 # https://pandas.pydata.org/Pandas_Cheat_Sheet.pdf
```

How to use Pandas to read CSV file?

```
1 import pandas as pd
2 df = pd.read_csv("volcano.csv")
3 df
```

Map with Geographical Information

How to embed a map in Notebook?

```
import folium

# The default tiles are set to OpenStreetMap, but Stamen Terrain, Stamen Toner, Mapbox Bright, and Mapbox Control Room, a

m = folium.Map(location=[5.5236, 95.6750], tiles='Stamen Terrain', zoom_start=5)
m
```

How to add a map marker?

```
tooltip = 'Click me!'

# test
folium.Marker([3.170, 98.392], popup='<i>Mt Sinabung in Sumatra</i>', tooltip=tooltip).add_to(m)
folium.Marker([0.070, 127.420], popup='<b>Mt Tigalalu in Halmahera</b>', tooltip=tooltip).add_to(m)
m
```

DataFrame: Subset

How to get the count of records from a DataFrame?

```
indon.VolcanoID.count()
```

	VolcanoID	V_Name	Country	Region	Subregion	Latitude	Longitude	PEI	H_active	VEI_Holoce	hazard
0	210010	West Eifel Volcanic Field	Germany	Mediterranean and W Asia	Western Europe	50.170	6.85	6	0	Unknown VEI	NaN
1	210020	Cha?ne des Puys	France	Mediterranean and W Asia	Western Europe	45.775	2.97	7	0	Unknown VEI	NaN
2	210030	Olot Volcanic Field	Spain	Mediterranean and W Asia	Western Europe	42.170	2.53	5	0	No confirmed eruptions	NaN
3	210040	Calatrava Volcanic Field	Spain	Mediterranean and W Asia	Western Europe	38.870	-4.02	6	0	Unknown VEI	NaN
4	211001	Larderello	Italy	Mediterranean and W Asia	Italy	43.250	10.87	4	0	3	NaN
...	...	...	...	...	...	...	...	...	...	...	...
1541	390100	Candlemas Island	United Kingdom	Antarctica	Antarctica and South Sandwich Islands	-57.080	-26.67	1	1	2	NaN
1542	390110	Hodson	United Kingdom	Antarctica	Antarctica and South Sandwich Islands	-56.700	-27.15	1	0	No confirmed eruptions	NaN

What is a Pandas DataFrame?

```
1 df.info()
```

```
# lets use a for loop
tooltip = 'Click me!'

total = indon.VolcanoID.count()

for index in range(total):
    record = indon.iloc[index]

    lat       = record.Latitude
    lon       = record.Longitude
    vname     = record.V_Name
    subregion = record.Subregion

    formatted = "<i>Mt {} in {}</i>".format(vname, subregion)

    folium.Marker([lat, lon], popup=formatted, tooltip=tooltip).add_to(m)
    if index == 5: break # test 5 markers

m
```

Complete Codes

```
# complete codes
import folium

# The default tiles are set to OpenStreetMap, but Stamen Terrain, Stamen Toner, Mapbox Bright, and Mapbox Control Room, a

m = folium.Map(location=[5.5236, 95.6750], tiles='Stamen Terrain', zoom_start=5)
tooltip = 'Click me!'

total = indon.VolcanoID.count()

for index in range(total):
    record = indon.iloc[index]

    lat       = record.Latitude
```

```
    lon       = record.Longitude
    vname     = record.V_Name
    subregion = record.Subregion

    formatted = "<i>Mt {} in {}</i>".format(vname, subregion)

    folium.Marker([lat, lon], popup=formatted, tooltip=tooltip).add_to(m)
    # if index == 5: break # test 5 markers

m
```

Coding Challenge

```
# we have completed the map for Indonesia.

# challenge
# can you do the map for all the volcanoes in the world?
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1546 entries, 0 to 1545
Data columns (total 13 columns):
 #   Column      Non-Null Count  Dtype
---  ------      --------------  -----
 0   VolcanoID   1546 non-null   int64
 1   V_Name      1546 non-null   object
 2   Country     1546 non-null   object
 3   Region      1546 non-null   object
 4   Subregion   1546 non-null   object
 5   Latitude    1546 non-null   float64
 6   Longitude   1546 non-null   float64
 7   PEI         1546 non-null   int64
 8   H_active    1546 non-null   int64
 9   VEI_Holoce  1546 non-null   object
 10  hazard      328 non-null    float64
```

How to get the list of unique values from a Column?

```
1 # how many volcano in each region
2 df["Region"].value_counts()
```

```
South America                197
Indonesia                    145
```

How to select data subset from DataFrame?

```
Alaska                       92
```

```
1 df[df.Country == 'Indonesia']
```

	VolcanoID	V_Name	Country	Region	Subregion	Latitude	Longitude	PEI	H_active	VEI_Holoce	hazard	class
384	261020	Seulawah Agam	Indonesia	Indonesia	Sumatra	5.448	95.658	4	1	2	NaN	U-HHR
385	261030	Peuet Sague	Indonesia	Indonesia	Sumatra	4.914	96.329	3	1	2	2.0	NaN
386	261050	Telong, Bur ni	Indonesia	Indonesia	Sumatra	4.769	96.821	5	1	2	1.0	NaN
387	261070	Sibayak	Indonesia	Indonesia	Sumatra	3.230	98.520	5	1	Unknown VEI	NaN	U-HHR
388	261080	Sinabung	Indonesia	Indonesia	Sumatra	3.170	98.392	5	1	2	NaN	U-HHR
...	...	...	...	...	...	...	...	...	...	...	...	...
521	268063	Moti	Indonesia	Indonesia	Halmahera	0.450	127.400	3	0	No confirmed eruptions	NaN	U-NHHR
522	268070	Makian	Indonesia	Indonesia	Halmahera	0.320	127.400	3	1	4	3.0	NaN

How to save to CSV file using Pandas?

```
1 indon = df[df.Country == 'Indonesia']
2 indon.to_csv("indonesia_v.csv")
```

```
142
```

How to get each row of record by the index?

```
indon.iloc[0]
```

```
VolcanoID            261020
V_Name         Seulawah Agam
Country            Indonesia
Region             Indonesia
Subregion            Sumatra
Latitude               5.448
Longitude             95.658
PEI                        4
H_active                   1
VEI_Holoce                 2
hazard                   NaN
class                  U-HHR
risk                     NaN
Name: 384, dtype: object
```

How to get the attribute from a Python object?

```
indon.iloc[0].Latitude
```

```
5.448
```

Use a FOR loop

Test with 5 Markers