CS 620–Introduction to Data Science and Analytics, HW2

For this problem you will use an actual dataset extracted from the US Geological Survey (USGS) earthquake API called "earthquakes.csv" which contains Earthquake data from September 18, 2018 through October 13, 2018. Information on the US Geological Survey (USGS) earthquake API can be found here.

Using the given dataset file and the materials from the python, NumPy, and Pandas, complete the following exercises. You can use any python IDE (e.g., Anaconda/Spyder, PyCharm) or use Google Colab and download the .py file. For Colab, you can directly load the file (url) using read_csv() or upload the file to your Google Drive and load it using drive.mount('/content/drive'). More information here.

What to turn in: You must exactly follow the naming convention: **Lastname-hw2.py**
Your file should contain the following information at the top.

> CS620
> HW2
> @author: <Your Name and UIN>

Submit your.Lastname-hw2.py file to Blackboard.

1. (5pts) Write a code to bring data into a Pandas DataFrame called "df".
2. (5pts) Write a code to inspect your data using DataFrame 'df' attributes 'shape', 'columns', and display the first five rows using method 'head()'
3. (10pts) Write a statement to print unique values (hint: use unique() method) available in the 'alert' column and display them as a list. Write a comment with other alert values available consulting the USGS API documentation for the alert field.
4. (10pts) Write a statement to print the columns "title" and "time" of 'df' in row index 100 through 105.
5. (20pts) Using python list comprehension, and the given 'df.columns', select all of the columns that start with 'mag', along with the 'title' , 'parsed_place' and 'time' columns. Hint: Use python string method "startswith()" to find all the columns that start with the term 'mag'.
6. (15pts) Using the selection in exercise (5) above, find all the earthquakes with the magnitude greater than or equal to 6.0 and display only the 'parsed_place' and the 'mag' columns.
7. (15pts) Using the original DataFrame 'df ', find the earthquakes with a red alert (df.alert == 'red') and a tsunami (df.tsunami ==1) and display only the 'parsed_place' and the 'mag' columns.
8. (10pts) Find the 90th percentile (hint: use 'quantile()' method) of earthquake magnitude in Japan using the magType of 'mb'.
9. (10pts) Get summary statistics (hint: use 'describe()' method) for earthquakes in California.
10. **(Optional: complete this to receive bonus 10pts)** Add a column called 'ring_of_fire' to the 'df' indicating whether or not the earthquake happened in a country or US state that is on the Ring of Fire. Use Bolivia, Chile, Ecuador, Peru, Costa Rica, Guatemala, Mexico (be careful not to select New Mexico), Japan, Philippines, Indonesia, New Zealand, Antarctic, Canada, Fiji, Alaska, Washington, California, Russia, Taiwan, Tonga, and Kermadec Islands.