Predicting Earthquakes at Volcanoes

EMILY SNELL- STATS 112

Introduction- Terms and Goal

Terms

Eruption: an event where lava is extruded from a volcanic structure

VEI: (Volcanic Explosivity Index) analogous to the Richter scale, the magnitude of an eruption. Log scale from 0-8

Magnitude: (earthquake), in the Richter scale

Type: Refers to the classification of the volcano. This term groups volcanoes based on physical and chemical attributes (ex. Stratovolcano, shield volcano, caldera…)

Extinct: A volcano which has not erupted in 10,000 years and is suspected to no longer have a magmatic supply

Goal:

- Predict seismicity associated with eruptions prior to the invention of seismometers

- Input earthquake magnitudes, volcanic type, location, volcanic name, and VEI

- Output earthquake magnitude

Collecting Data

USGS
science for a changing world

Smithsonian Institution
National Museum of Natural History
Global Volcanism Program

Oregon State University

Volcano World

Data Collection

- Earthquake Data collected from USGS earthquake catalog in CSV format (1990-2023)

- Volcanic locations and eruption data collected from Smithsonian Global Volcanism Program CSV format (1990-2023)

```
#lets add some more data from various regions to make the data more robust
other_earthquakes = ['Hawaii_2015.csv', 'aleutians_2000.csv', 'aleutians_2015.csv', 'NewZealand_2000.csv',
                     'NewZealand_2010.csv', 'Japan_2001.csv', 'indonesia_2014.csv', 'indonesia_2018.csv', 'hawaii_2023.csv',
                     'central_america_1989.csv', 'central_america_1995.csv', 'central_america_2000.csv','columbia_1990.csv',
                     'new_zealand_1990.csv', 'Russia.csv', 'Indonesia.csv', 'Chile_1990.csv']
for earthquake in other_earthquakes:
  new_df = pd.read_csv(earthquake)
  earthquake_df = pd.concat([earthquake_df, new_df], axis=0)
```

	time	latitude	longitude	depth	mag	magType	ns
0	2022-01-15T04:14:45.000Z	-20.546000	-175.390000	0.00	5.8	ms_20	Nal
1	2018-08-02T21:55:12.060Z	19.411167	-155.283167	0.55	5.3	mw	34.
2	2018-07-31T17:59:46.000Z	19.410333	-155.285667	0.57	5.3	mw	36.
3	2018-07-29T22:10:25.570Z	19.406333	-155.282333	1.29	5.3	mw	39.
4	2018-07-28T12:37:25.390Z	19.396833	-155.271833	0.21	5.3	mw	37.
...	...	...	...	...	...	...	...
2323	1990-03-15T09:07:01.160Z	-21.143000	-68.748000	120.00	4.6	mb	Nal
2324	1990-03-10T02:04:13.390Z	-22.671000	-68.267000	123.20	4.8	mb	Nal
2325	1990-03-08T03:59:35.230Z	-24.625000	-70.263000	59.20	4.5	mb	Nal
2326	1990-03-04T16:30:52.060Z	-20.069000	-69.311000	125.00	4.2	mb	Nal
2327	1990-03-04T13:24:06.860Z	-21.247000	-67.758000	199.60	4.6	mb	Nal

116671 rows × 22 columns

Data Collection pt. 2

- Eruptions predicted earthquakes for were scraped from Oregon State University website

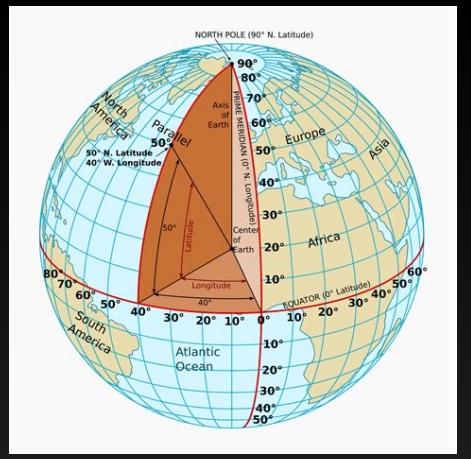
- Eruptions span roughly 1400 AD to early 1960's

- Contains volcano name, VEI, date, Country, and Type

Cleaning Data

pandas.to_datetime()

Convert Strings To Dates



Data Cleaning- Long Process!

Data was in csv format but the time and name columns were not standardized. Further I had to generate multiple columns to create a unique key which I could merge my earthquake and eruptions data frames on

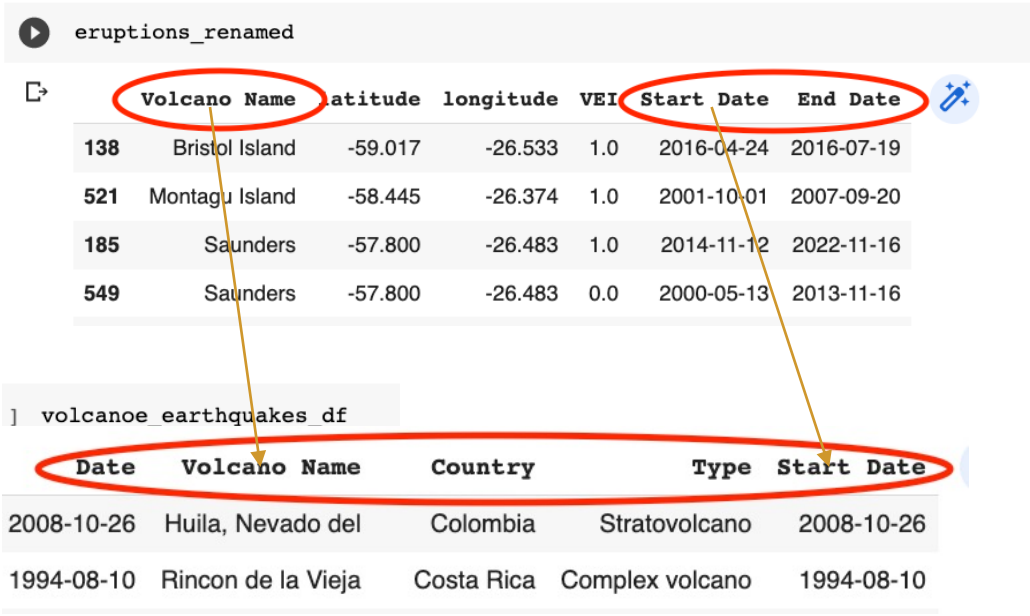
Major steps:

- Normalizing time data to pandas datetime (both df's)

- Normalizing volcanic names and mapping earthquakes to volcanic centers using latitude and longitude

- Mapping earthquakes to eruptions with latitude and longitude (next slide)

- Merging (outer merge)



```
eruptions_renamed
```

	Volcano Name	Latitude	longitude	VEI	Start Date	End Date
138	Bristol Island	-59.017	-26.533	1.0	2016-04-24	2016-07-19
521	Montagu Island	-58.445	-26.374	1.0	2001-10-01	2007-09-20
185	Saunders	-57.800	-26.483	1.0	2014-11-12	2022-11-16
549	Saunders	-57.800	-26.483	0.0	2000-05-13	2013-11-16

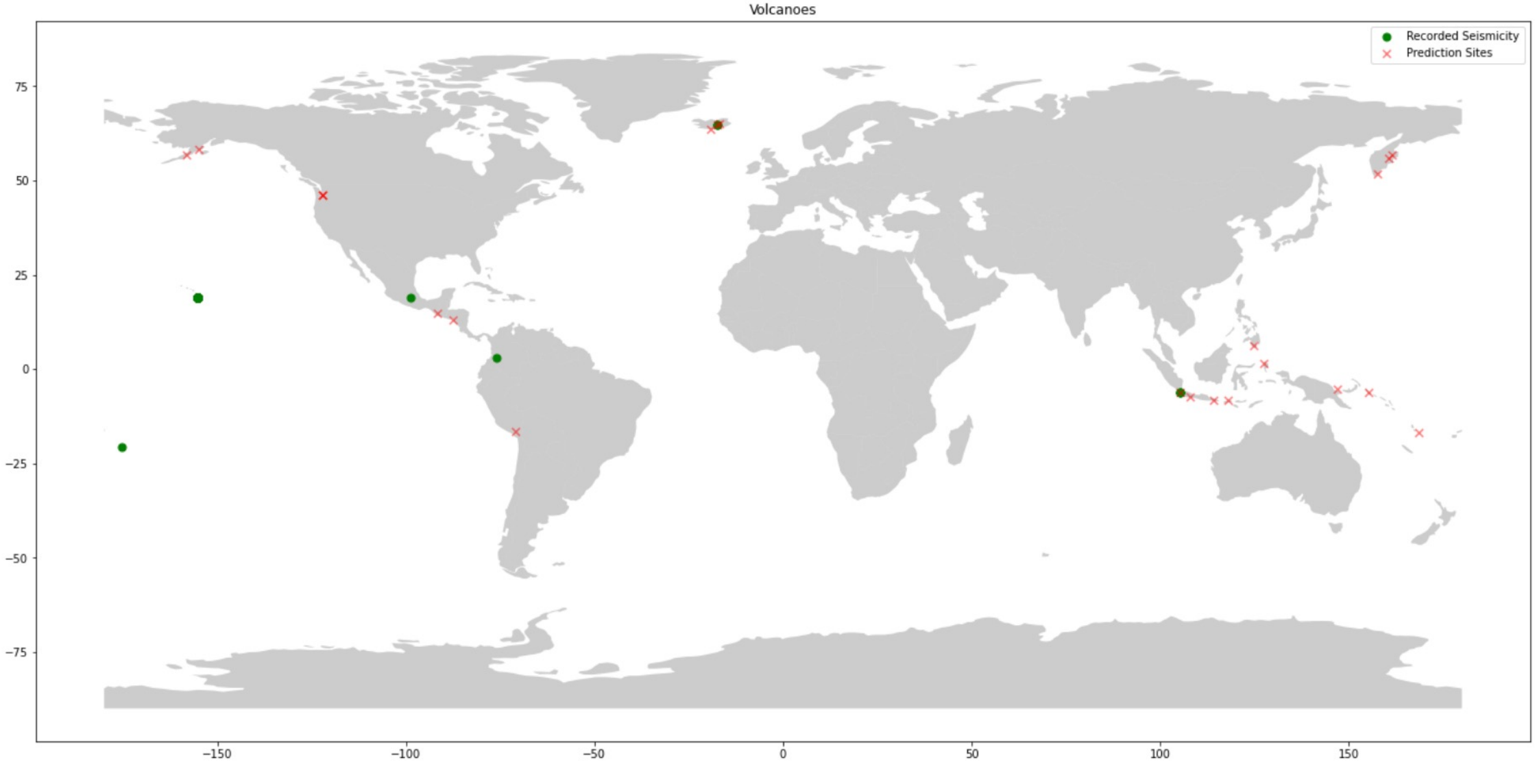
```
volcanoe_earthquakes_df
```

Date	Volcano Name	Country	Type	Start Date
2008-10-26	Huila, Nevado del	Colombia	Stratovolcano	2008-10-26
1994-08-10	Rincon de la Vieja	Costa Rica	Complex volcano	1994-08-10

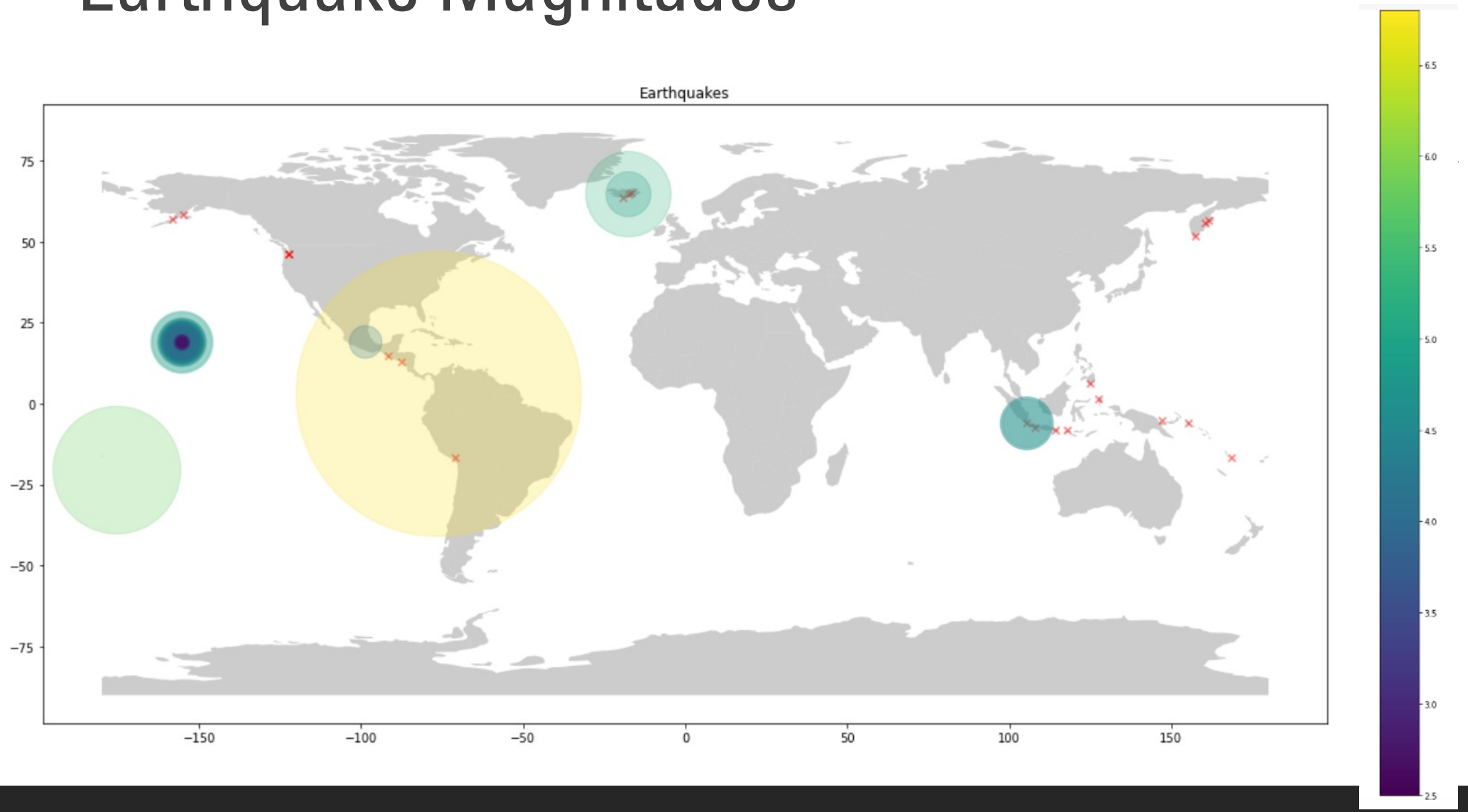
```
## this function takes in eruption and earthquake data
## using latitude, longitude, and time it rules if a volcanic earthquake happened during an eruption
## if the earthquake happened during an eruption, the time data is re-assigned to the start date of the eruption
## this is because I will merge my data sets on time and volcanic name
## if a volcanic earthquake does not correspond to an eruption it is still kept in the data set
## just the time data is not altered so when I merge it is not matched to an eruption
def eruption_matching(earthquake_df, eruptions_df):
  ind = {}
  #empty data frame
  final_df = pd.DataFrame()
  #create a list of unique volcanoes that have eruptions
  volcanoes = eruptions_df['Volcano Name'].unique().tolist()
  #loop through
  for volcano in volcanoes:
    #creating mini df's with just earthquakes and eruptions for that specific volcano
    mini_df = earthquake_df[earthquake_df['Volcano Name']==volcano]
    mini_df=mini_df.reset_index()
    eruption_df = eruptions_df[eruptions_df['Volcano Name']==volcano]
    eruption_df=eruption_df.reset_index()
    #creating an artifical eruptions series with the earthquake dates which I will overwrite soon
    eruption_series = pd.Series(mini_df['Date'])
    #then find the index of earthquakes that match an eruption time
    #but make sure the volcano exists in both data sets
    if len(mini_df) > 0 and len(eruption_df) > 0:
      #print('yes')
      for i in range(len(mini_df)):
        for p in range(len(eruption_df)):
          if eruption_df['End Date'][p]>=mini_df['Date'][i] and eruption_df['Start Date'][p]<=mini_df['Date'][i]:
          #assigning start date here, may change this to something else later depending on how it looks merged
            eruption_series[i] = eruption_df['Start Date'][p]
            #print(eruption_df['Start Date'][p])
      #make column in our masked data frame
      mini_df['Start Date'] = pd.to_datetime(eruption_series, yearfirst='True', utc=False, unit='ns')
      final_df = pd.concat([final_df, mini_df], axis=0)
  return final_df

dummy = eruption_matching(volcanoe_earthquakes_df, eruptions_renamed)
```

Exploring Data

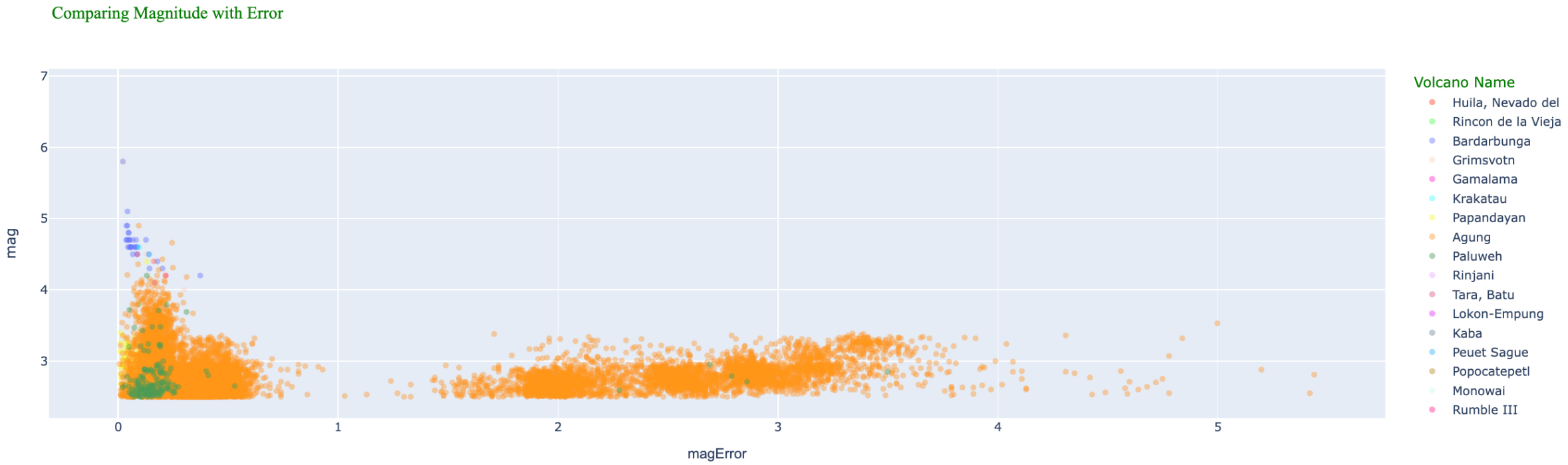
Volcanic Centers



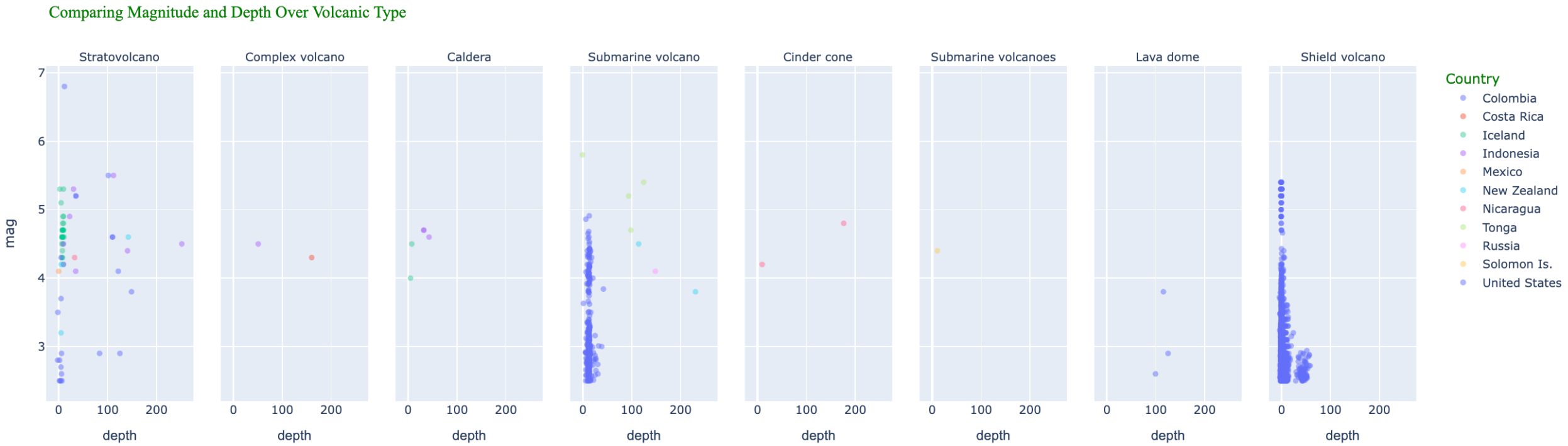
Earthquake Magnitudes



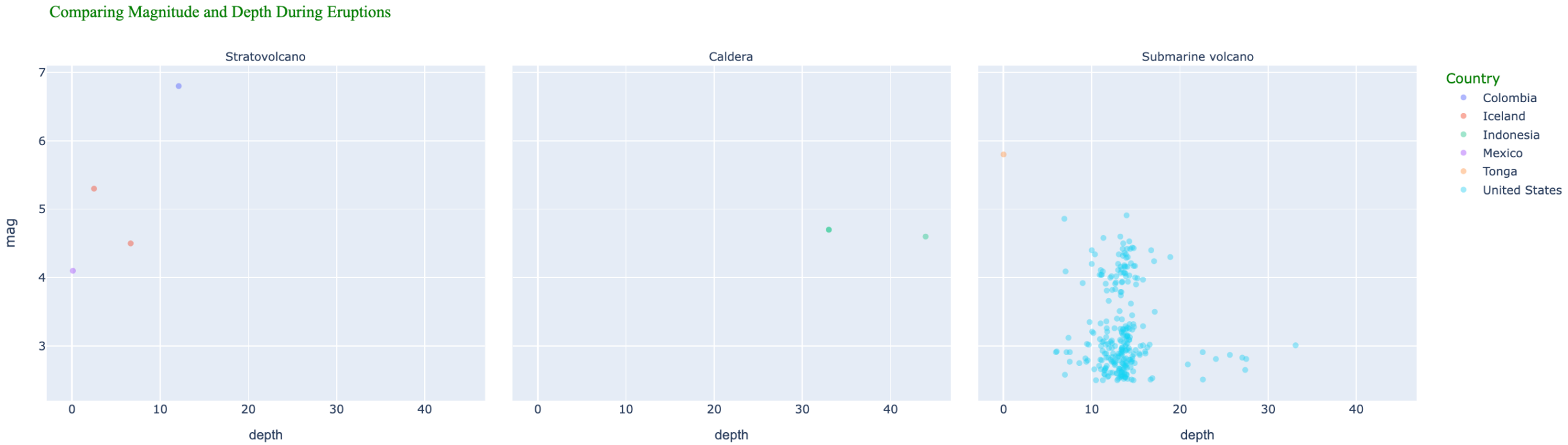
Magnitude vs Error



Earthquake Qualities vs Type

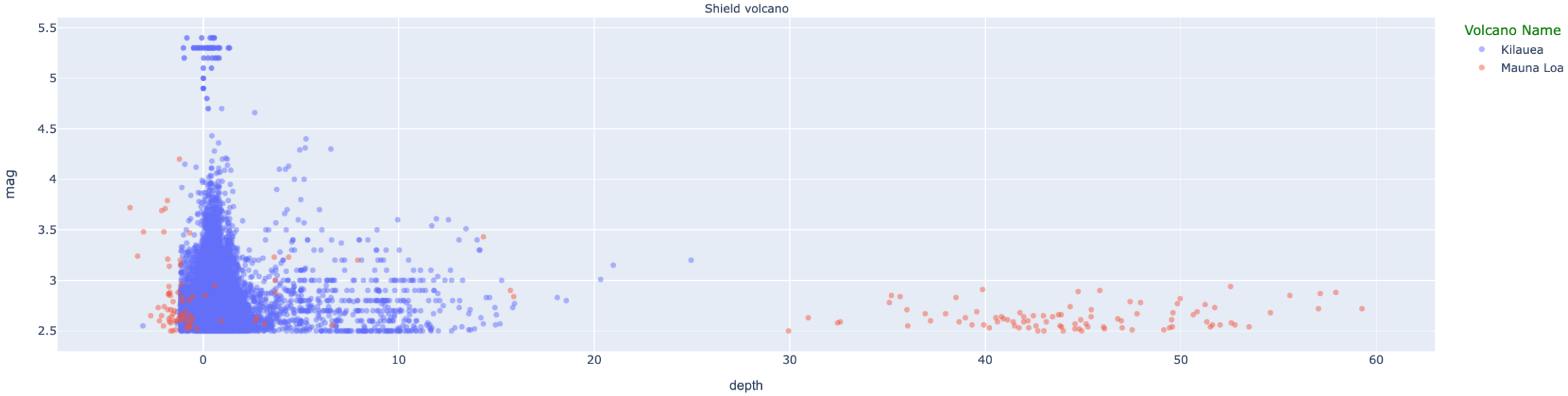


Earthquakes During Eruption?

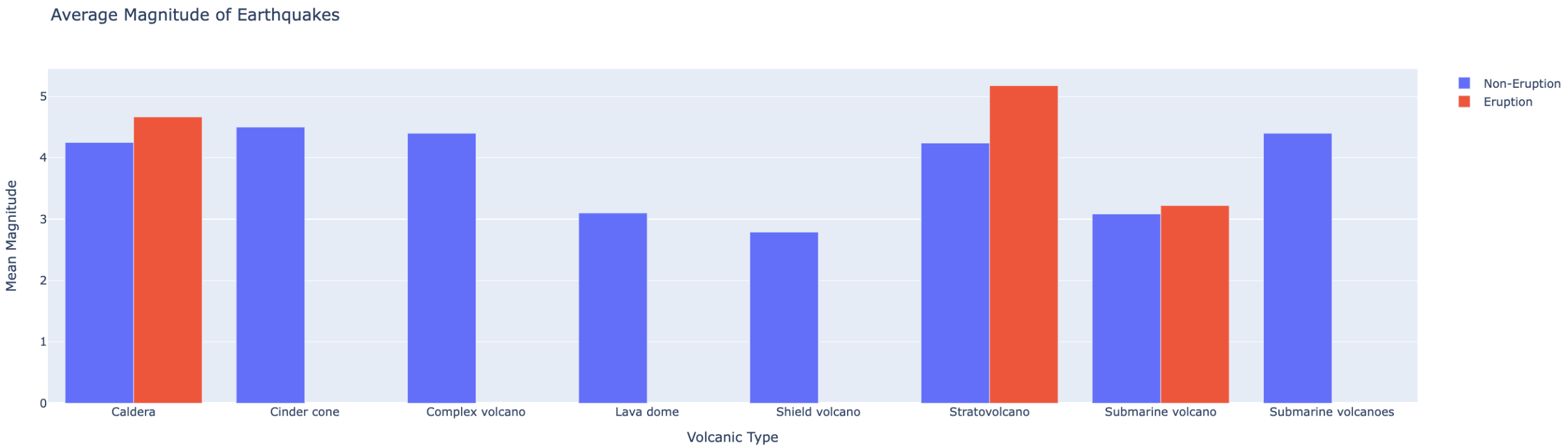


Shield Volcano Pattern?

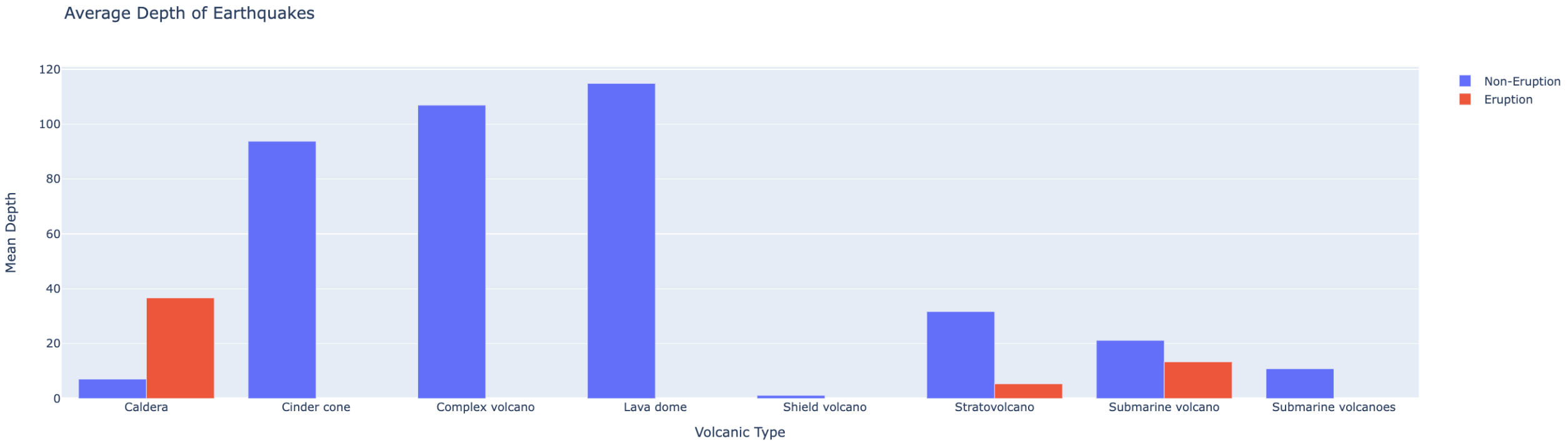
Comparing Depth of Shield Volcanoes



Understanding Eq Behavior



Cont.


Average Depth of Earthquakes
Mean Depth
120
100
80
60
40
20
0
Caldera
Cinder cone
Complex volcano
Lava dome
Shield volcano
Stratovolcano
Submarine volcano
Submarine volcanoes
Volcanic Type
Non-Eruption
Eruption

Most Dangerous Volcano?

Volcano Name	mag	VEI	Eruption length
Huila, Nevado del	6.800000	5.0	1175 days 00:00:00
Hunga Tonga-Hunga Ha'apai	5.800000	5.0	26 days 00:00:00
Popocatepetl	4.100000	3.0	2818 days 00:00:00
Krakatau	4.666667	2.0	487 days 08:00:00
Kama'ehuakanaloa	3.147677	2.0	166 days 00:00:00

Machine Learning

Models and RSME

Models Used:

- Kneighbors (k = 15, metric = cosine) **RSME : 0.144**

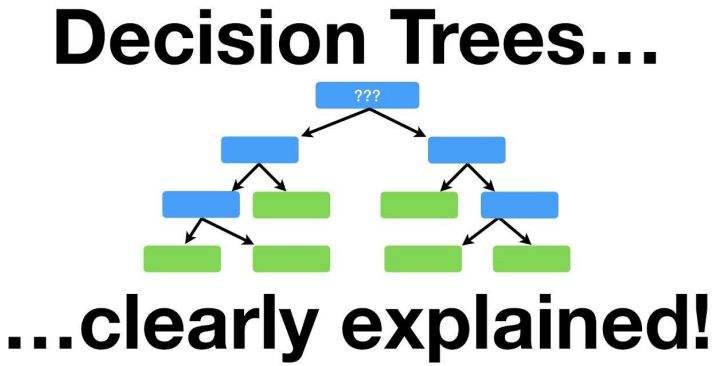
- Linear Regression **RSME: 8.1819 x 10^24**

- Random Forest Regressor (n_estimators = 31) **RSME: 0.367**

- Decision Tree Regressor (depth = 2, leaf = 2) **RSME: 0.136**

- Stacking (all models but LR) **RSME: 0.365**

- Voting (all models but LR) **RSME: 0.367**



Final Predictions!

	Date	VEI	Volcano Name	Country	Type	Earthquake Mag
0	1450C	5	Aniakchak	United States	Caldera	3.170264
1	1452	6	Kuwae	Vanuatu	Caldera	6.300000
2	1477	5	Bardarbunga	Iceland	Stratovolcano	6.300000
3	1480D	5	St. Helens	United States	Stratovolcano	3.170264
4	1540	5	St. Helens	United States	Stratovolcano	3.170264
5	18-May-80	5	St. Helens	United States	Stratovolcano	3.170264
6	1580C	6	Billy Mitchell	Papua New Guinea	Ash shield	6.300000
7	1593	5	Raung	Indonesia	Stratovolcano	6.300000
8	1600	6	Huaynaputina	Peru	Explosion crater	6.300000
9	Jan. 4, 1641	6	Parker	Philippines	Stratovolcano	6.300000
10	1660	6	Long Island	Papua New Guinea	Complex volcano	6.300000
11	1673	5	Gamkonora	Indonesia	Stratovolcano	6.300000
12	Oct. 17, 1755	5	Katla	Iceland	Subglacial	6.300000
13	April 5, 1815	7	Tambora	Indonesia	Stratovolcano	6.300000
14	Oct. 8, 1822	5	Galunggung	Indonesia	Stratovolcano	6.300000
15	June 20, 1835	5	Cosiguina	Nicaragua	Stratovolcano	6.300000
16	Feb. 17, 1854	5	Sheveluch	Russia	Stratovolcano	6.300000
17	March 29, 1875	5	Askja	Iceland	Stratovolcano	6.300000
18	Aug. 26, 1883	6	Krakatau	Indonesia	Caldera	6.300000
19	Oct., 24, 1902	6	Santa Maria	Guatemala	Stratovolcano	6.300000
20	March 28, 1907	5	Ksudach	Russia	Shield	6.300000
21	June 6, 1911	6	Katmai	United States	Stratovolcano	3.170264
22	March 30, 1956	5	Bezymianny	Russia	Stratovolcano	6.300000

Final Notes & Questions?

- Limited data supply- emerging field and more work needed

- Predicted 2 different magnitudes (6.3 and 3.17)