Bryan D'Amico
IST652-Scripting for Data Analysis
Homework 2 Report: Semistructured Data

Analysis of the USGS Earthquakes Data

Data Set

The data set used for this analysis was the USGS real-time earthquakes data feed featuring earthquakes with a magnitude of 4.5 or greater over the past 30 days. The real-time feed was pulled on February 28th so the data set reflects the span from 01/29/2023 to 02/28/2023. There were 585 magnitude 4.5 or greater earthquakes over this one month period. This feed, among others, can be found on the GeoJSON Summary Format page on the earthquake.usgs.gov website. The GeoJSON data was imported into a Jupyter notebook using the urllib and json packages. From there the data was converted to a structured pandas DataFrame. The following table summarizes the data that was converted from the GeoJSON format to the data frame.

Feature Name	DataFrame Name	Description
mag	Magnitude	The magnitude of the earthquake with typical values between -1 and 10.
place	Place	The closest name geographical region (like a city) to the earthquake.
time	Date	The time in milliseconds since Jan 1, 1970 when the earthquake occurs. This will be converted to a date in the DataFrame.
felt	Felt	The number of reports submitted to the "Did You Feel It?" system by people who felt the earthquake.
cdi	CDI	The maximum reported intensity. Intensity is a measure of shaking caused by the earthquake.
mmi	MMI	The maximum estimated instrumental intensity.
alert	Alert	The alert level from the PAGER earthquake impact scale which measures fatalities and economic loss.
status	Status	This indicates whether an earthquake's data has been reviewed by a human.
tsunami	Tsunami	A value of 1 means the earthquake occured in an oceanic region, 0 if not. This does not mean there was a tsunami caused.
sig	Significance	A measure of how significant the earthquake was.
nst	NST	The number of seismic stations that were used to determine the earthquake's location.
type	Type	The type of the seismic event.
latitude	Latitude	The latitude measurement of the earthquake.
longitude	Longitude	The longitude measurement of the earthquake.
depth	Depth	The depth of the earthquake in kilometers.

The descriptions in the table were taken from the ANSS Comprehensive Earthquake Catalog Event Terms Documentation on the USGS website.
(URL to real time feeds: https://earthquake.usgs.gov/earthquakes/feed/v1.0/geojson.php)
(URL to Terms Documentation: https://earthquake.usgs.gov/data/comcat/data-eventterms.php)

Description of Data Cleaning Steps to Create the pandas DataFrame

Creating the pandas DataFrame

To create the pandas DataFrame, the magnitude, place, time, felt, cdi, mmi, alert, status, tsunami, sig, nst, type, latitude, longitude, and depth features were captured in lists with each entry in the lists corresponding to a different earthquake. The time values were recorded in the GeoJSON data as the number of milliseconds since the epoch (January 1, 1970). These values were converted to human readable dates. First, the time values were converted to seconds, then those strings were converted to datetime objects. Then the year, month, and date were isolated and datetime objects containing just those values were used as the date column in the pandas DataFrame. After converting the time values to dates, a dictionary of all the lists were created and passed to pd.DataFrame to create the pandas DataFrame of the earthquake data.

Data Cleaning Steps

Investigating the number of missing values in the DataFrame revealed that there were significant numbers of missing values for the Felt, CDI, MMI, and Alert columns. There were also three missing values for the NST column. The three rows with missing values for NST (the number of seismic stations used to determine the earthquake's location) were dropped because there were so few, and because it was not clear how to replace those missing values. The missing values for the Felt, CDI, and MMI columns were replaced by 0. It was assumed that if these values were left blank then there were either no reports made to the Did You Feel It? system, no maximum intensity was calculated, or no maximum estimated instrumental intensity was calculated, respectively. Lastly, if the value for the Alert column was missing it was assumed that no alert was issued, and these were replaced by the string "No Alert".

After replacing all the missing values, the data types of the columns were checked. All the numeric columns were of a numeric type so those did not need to be addressed. The columns that contained categorical values such as Alert, having values of No Alert, green, yellow, orange, and red, were of the object data type. These were all converted to be categorical.

The finalized pandas DataFrame was written to an external csv file so anyone with access to the data could run analyses of their choosing.

Questions

Four questions were developed based on this data set.

Question 1: How strong are the earthquakes as measured by their magnitudes and significance?

Unit of Analysis: Magnitude and Significance

Summarization: The distributions of the two variables are visualized using boxplots and the median and mean are explicitly calculated.

Question 2: Where have the earthquakes of different strengths occurred in the world over the last month?

Unit of Analysis: Bins of magnitudes (4.5 to 5, 5 to 6, and 6 or greater)

Filtering/Grouping: The data is separated into bins based on their magnitude and three data frames are created. The data frames are saved as external files so anyone can explore the locations of these earthquakes.

Question 3: How has seismic activity changed over the last month?

Unit of Analysis: Date

Comparison: For each day in the data set, the mean magnitude, significance, depth, CDI, and Felt are calculated. The change in mean magnitude, significance, and depth are visualized using line plots.

Question 4: How are the different properties of an earthquake, such as magnitude, depth, or significance, related to one another?

Unit of Analysis: Pearson correlation coefficients

Comparison: For each pair of variables in the group Magnitude, Significance, Felt, Depth, CDI, MMI, and NST, the Pearson correlation coefficient is calculated and visualized using a heatmap.

Description of the Program

The program was created in the form of a report style Jupyter notebook. It guides the reader through the data importing, DataFrame creation, and data cleaning steps previously described. The program then answers each of the four questions by generating data frames, visualizations, and external files that contain the information needed to answer those questions.

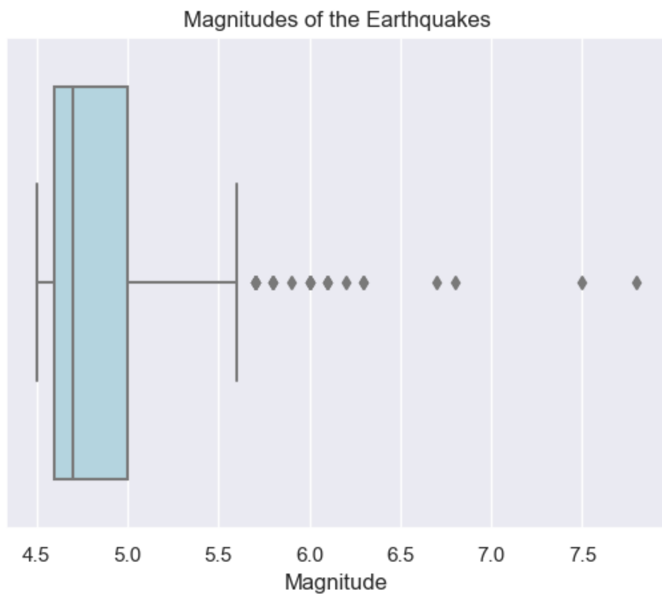
Boxplots are created to visualize the distribution of the magnitudes and significance values of the earthquakes over the one month span. These are created to better understand the size of the earthquakes and begin to understand how common larger earthquakes are versus smaller earthquakes based on this sample. This leads into examining the most significant earthquakes to see where they happened. Then, the earthquakes are broken into three groups based on their magnitudes and a pandas DataFrame and external csv file are created from each. In the Jupyter notebook, samples of the entries in each data frame are shown to discuss some of the locations that are present. To understand the changes in seismic activity over the month, a new data frame is created where the rows are grouped by date and then the mean values of several of the numeric attributes of the earthquakes are calculated. This data frame is written as an external csv file and used to create line plots which show the changes in the mean magnitude, significance, and depth of the earthquakes occurring on each day of the month. Lastly, a correlation matrix is created based on the values of magnitude, significance, the number of felt reports, depth, maximum reported intensity, maximum estimated intensity, and the number of seismic stations used to determine the earthquake's location. The lower triangle correlation matrix is represented using a heatmap with a diverging color scheme.

Output and Analysis

Question 1: How strong are the earthquakes as measured by their magnitudes and significance?

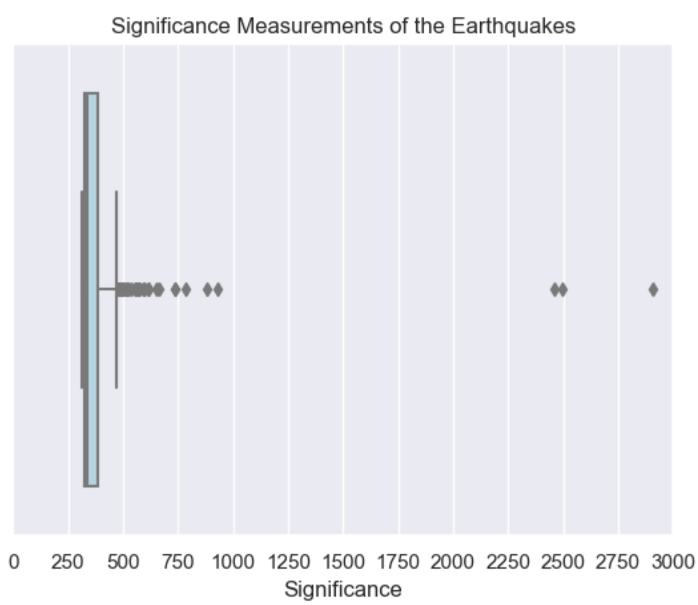
This question will be answered by investigating the distributions of the Magnitude and Significance variables while also calculating the mean and median values of these columns. Both boxplots are saved as external pdf files in the Jupyter notebook program.

The following boxplot displays the distribution of the magnitudes of the earthquakes. It is important to remember that this data set only contains the information on earthquakes that had a magnitude of 4.5 or higher. Of these earthquakes having a magnitude of 4.5 or higher, the mean magnitude is 4.84 and the median magnitude is 4.7.



Approximately 75% of these earthquakes have a magnitude of 5 or less. Magnitudes greater than 5 are rare when compared to the total number of earthquakes occurring around the world. Even when only looking at earthquakes whose magnitudes are 4.5 or higher, the great majority of those earthquakes have a magnitude close to that minimum value. While the highest magnitude earthquakes are the ones that make the news, there are many, many more earthquakes occurring every day than we might realize.

The following boxplot shows the distribution of the significance ratings of the earthquakes. Significance is calculated using a combination of many factors including magnitude, the number of felt reports, its estimated impact, and the maximum estimated instrumental intensity. (Reference: https://earthquake.usgs.gov/data/comcat/index.php#sig)




```
The average significance of the earthquakes is 376.64
The median significance of the earthquakes is 340.0
```

Almost all these earthquakes have a significance calculated to be between 250 and 500. There are three earthquakes that are extreme outliers, all with significances greater than 2250. These may have been especially damaging earthquakes. This leads into the next question which discusses where the earthquakes are taking place.

Question 2: Where have the earthquakes of different strengths occurred in the world over the last month?

This question will be answered by creating three DataFrames. One will contain the largest earthquakes with magnitudes greater than or equal to 6, one will contain the smallest earthquakes with magnitudes less than 5 but greater than or equal to 4.5, and one will contain the medium earthquakes with magnitudes between 5 and 6. Each of these DataFrames will be saved as an external csv file so those interested in where each magnitude grouping of earthquakes took place can review the data to see those locations.

First, the largest earthquakes from the previous section will be highlighted. Those earthquakes were particularly severe with significance ratings of greater than 2250. Did those earthquakes happen across the globe or where they concentrated in one area? Also, did they happen in succession or were they spread out over the course of the past month?

	Magnitude	Place	Date	Felt	CDI	MMI	Alert	Status	Tsunami	Significance	NST	Type	Latitude	Longitude	Depth
127	6.3	3 km SSW of Uzunbağ, Turkey	2023-02-20	543.0	9.1	8.841	red	reviewed	0	2494	140.0	earthquake	36.1094	36.0165	16.0
441	7.5	4 km SSE of Ekinözü, Turkey	2023-02-06	503.0	9.1	9.030	red	reviewed	0	2458	135.0	earthquake	38.0235	37.2030	10.0
479	7.8	26 km ENE of Nurdağı, Turkey	2023-02-05	1946.0	9.1	9.419	red	reviewed	0	2910	230.0	earthquake	37.2251	37.0209	10.0

Interestingly, all those most significant earthquakes occurred in Turkey. They each have high magnitudes and occurred in close proximity to one another. Two of them occurred on back-to-back days on February 5, 2023, and February 6, 2023. All three of these earthquakes also had the highest alert level and were reported as being felt by many people.

Also, looking at the magnitude values reveals that these three earthquakes were among those with the greatest magnitudes. By isolating the earthquakes with magnitudes greater than or equal to 6, it is possible to see where else in the world the large earthquakes occurred. The entries for the earthquakes will be sorted by descending magnitude and then in chronological order for each magnitude rating. This way, the data can be filtered for any particular magnitude and see the earthquakes in the order which they occurred. The following is a screenshot of the external csv file created from this data frame.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Date	Place	Magnitude	Number of F	Max Reporte	Max Est. Inst	Alert Level	Review Statu	Tsunami	Significance	Number of S	Type	Latitude	Longitude	Depth (km)	
2	2/5/23	26 km ENE of Nurdafüf±, Turkey	7.8	1946	9.1	9.419	red	reviewed	0	2910	230	earthquake	37.2251	37.0209	10	
3	2/6/23	4 km SSE of Ekinv∂zvº, Turkey	7.5	503	9.1	9.03	red	reviewed	0	2458	135	earthquake	38.0235	37.203	10	
4	2/22/23	67 km W of Murghob, Tajikistan	6.8	32	8.7	7.004	green	reviewed	0	739	128	earthquake	38.0726	73.2077	20.522	
5	2/5/23	18 km E of Nurdafüf±, Turkey	6.7	124	7.8	7.817	yellow	reviewed	0	787	192	earthquake	37.1784	36.9468	10.708	
6	2/20/23	3 km SSW of Uzunbafü, Turkey	6.3	543	9.1	8.841	red	reviewed	0	2494	140	earthquake	36.1094	36.0165	16	
7	2/23/23	177 km N of Tobelo, Indonesia	6.3	6	3.3	4.594	green	reviewed	0	613	116	earthquake	3.3237	128.1497	97.145	
8	2/25/23	29 km ENE of Kandrian, Papua New Guinea	6.2	7	4.2	5.536	green	reviewed	1	594	92	earthquake	-6.111	149.7927	38.158	
9	2/13/23	Kermadec Islands, New Zealand	6.1	0	0	3.269	green	reviewed	0	572	47	earthquake	-29.5218	-177.9727	374.033	
10	2/15/23	11 km NNE of Miaga, Philippines	6.1	53	7.7	7.382	green	reviewed	0	613	193	earthquake	12.3255	123.868	8	
11	2/17/23	130 km SW of Tual, Indonesia	6.1	8	3.4	4.977	green	reviewed	0	575	119	earthquake	-6.5986	132.0763	38.615	
12	2/1/23	Mindanao, Philippines	6	51	8.2	6.998	green	reviewed	0	596	220	earthquake	7.74	126.056	19	
13	2/6/23	Central Turkey	6	22	7.5	7.567	yellow	reviewed	0	666	138	earthquake	38.0546	36.5099	8.116	
14	2/6/23	10 km SE of Dofüan≈üehir, Turkey	6	3	8.2	6.788	green	reviewed	0	556	59	earthquake	38.0302	37.9636	20.094	
15	2/25/23	61 km ESE of Kushiro, Japan	6	17	7.7	4.629	green	reviewed	0	567	157	earthquake	42.7801	145.0736	50.181	
16																

Looking at the 14 largest earthquakes over the last month reveals that they occurred in a variety of locations and were fairly spread out over the course of the month. The dates range from February 1st until February 25th.

Surprisingly, 6 of the 14 largest earthquakes occurred in Turkey. This may imply that seismic activity tends to occur in groups or that this region is particularly sensitive to having earthquakes. Other locations affected by earthquakes include New Zealand, Japan, and Indonesia.

There were many more smaller earthquakes, with 419 quakes with a magnitude less than 5 occurring over the course of the last month. This data frame will be ordered by magnitude and then by date. The following screenshot is of the first 10 rows of the external csv file created from this data frame.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Date	Place	Magnitude	Number of F	Max Reporte	Max Est. Inst	Alert Level	Review Statu	Tsunami	Significance	Number of S	Type	Latitude	Longitude	Depth (km)
2	1/30/23	107 km SSW of ≈ªurrieq, Malta	4.5	3	4.1	0	No Alert	reviewed	0	313	132	earthquake	34.8794	14.2319	10
3	1/30/23	95 km SE of Ishinomaki, Japan	4.5	1	2	0	No Alert	reviewed	0	312	66	earthquake	37.8497	142.1119	26.488
4	1/30/23	258 km ESE of Tadine, New Caledonia	4.5	0	0	0	No Alert	reviewed	0	312	37	earthquake	-22.4116	170.2024	10
5	1/30/23	18 km ENE of Jac√≥, Costa Rica	4.5	14	3.9	0	No Alert	reviewed	0	317	59	earthquake	9.6771	-84.4773	46.047
6	1/30/23	53 km W of Palauig, Philippines	4.5	0	0	0	No Alert	reviewed	0	312	50	earthquake	15.5066	119.4087	26.504
7	1/31/23	103 km SE of Lakatoro, Vanuatu	4.5	0	0	0	No Alert	reviewed	0	312	12	earthquake	-16.6312	168.2143	10
8	1/31/23	245 km NNE of Tobelo, Indonesia	4.5	0	0	0	No Alert	reviewed	0	312	43	earthquake	3.9003	128.454	35
9	1/31/23	49 km NNW of Smirnykh, Russia	4.5	0	0	0	No Alert	reviewed	0	312	77	earthquake	50.1156	142.4599	10
10	2/1/23	86 km NNE of Antofagasta, Chile	4.5	1	1	0	No Alert	reviewed	0	312	24	earthquake	-22.9766	-69.9712	46.485
11	2/1/23	Kermadec Islands, New Zealand	4.5	0	0	0	No Alert	reviewed	0	312	15	earthquake	-29.7465	-178.9779	303.595

Looking at the first 10 smallest earthquakes in this data set reveals that these smaller seismic events were happening all over the world at the end of January, from Japan and the Philippines to Costa Rica and Chile.

The medium sized earthquakes with magnitudes between 5 and 6 will be isolated next. Again, this data frame will be ordered by magnitude and then date.

As a sample of the data, the smaller magnitude earthquakes at the beginning of the one month period and the larger magnitude earthquakes will be displayed. The following screenshot is from the external csv file created from this data set.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Date	Place	Magnitude	Number of F	Max Reporte	Max Est. Inst	Alert Level	Review Statu	Tsunami	Significance	Number of S	Type	Latitude	Longitude	Depth (km)
2	1/30/23	105 km ENE of Port Blair, India	5	2	2.6	0	No Alert	reviewed	0	385	93	earthquake	12.1363	93.5873	107.401
3	1/30/23	152 km S of Hotan, China	5	0	0	0	No Alert	reviewed	0	385	122	earthquake	35.7381	79.8912	10
4	1/31/23	10 km NNE of Malango, Solomon Islands	5	1	5.3	0	No Alert	reviewed	0	385	104	earthquake	-9.6025	159.7462	29.152
5	1/31/23	77 km E of Hihifo, Tonga	5	0	0	0	No Alert	reviewed	0	385	87	earthquake	-15.8651	-173.0788	43.64
6	2/1/23	32 km S of Matanzas, Dominican Republic	5	629	5.6	0	No Alert	reviewed	0	737	235	earthquake	17.948	-70.3792	46.555
7	2/2/23	62 km NW of Port-Olry, Vanuatu	5	0	0	0	No Alert	reviewed	0	385	91	earthquake	-14.5964	166.7198	37.319
8	2/2/23	86 km NNE of Tobelo, Indonesia	5	0	0	0	No Alert	reviewed	0	385	108	earthquake	2.4828	128.2209	81.032
9	2/2/23	46 km NNE of Beausv©jour, Guadeloupe	5	2	3.1	0	No Alert	reviewed	0	385	72	earthquake	16.6999	-60.9339	19.945
10	2/3/23	North Island of New Zealand	5	12	4.1	3.406	green	reviewed	0	390	61	earthquake	-38.8911	175.8758	100.216
11	2/3/23	northern Qinghai, China	5	0	0	0	No Alert	reviewed	0	385	78	earthquake	37.0695	96.9059	10

Looking at the first 10 earthquakes with a magnitude of 5 reveals that these earthquakes occurred in some different locations compared to the small and large earthquakes. Two of these earthquakes took place in China, with other earthquakes occurring in New Zealand, India, and the Dominican Republic.

The following screenshot shows the 10 largest medium-sized earthquakes.

	Magnitude	Place	Date	Felt	CDI	MMI	Alert	Status	Tsunami	Significance	NST	Type	Latitude	Longitude	Depth
25	5.6	Fiji region	2023-02-27	0.0	0.0	1.691	green	reviewed	0	482	129.0	earthquake	-20.5315	-178.3629	548.130
578	5.7	110 km ESE of Aral, China	2023-01-29	8.0	4.1	5.605	green	reviewed	0	503	132.0	earthquake	40.0166	82.3702	28.000
478	5.7	central Turkey	2023-02-05	0.0	0.0	7.294	yellow	reviewed	0	650	97.0	earthquake	37.2241	36.9749	10.000
438	5.7	eastern Turkey	2023-02-06	4.0	6.6	6.582	green	reviewed	0	502	88.0	earthquake	38.1236	38.0529	12.336
231	5.7	Cook Strait, New Zealand	2023-02-15	91.0	5.8	4.754	green	reviewed	0	553	81.0	earthquake	-40.5465	174.5709	74.320
162	5.7	South Sandwich Islands region	2023-02-18	0.0	0.0	3.504	green	reviewed	0	500	104.0	earthquake	-55.4328	-27.8877	20.395
494	5.8	Izu Islands, Japan region	2023-02-04	1.0	1.0	2.808	green	reviewed	0	518	213.0	earthquake	32.6612	141.6609	11.000
501	5.8	eastern New Guinea region, Papua New Guinea	2023-02-04	19.0	4.7	4.559	green	reviewed	0	526	124.0	earthquake	-6.4048	146.2298	121.000
439	5.8	13 km SW of Doğanşehir, Turkey	2023-02-06	5.0	6.4	6.793	green	reviewed	0	521	67.0	earthquake	38.0084	37.7505	10.000
283	5.9	Pulau-Pulau Talaud, Indonesia	2023-02-11	2.0	3.4	5.303	green	reviewed	0	536	201.0	earthquake	3.6086	126.7339	22.941

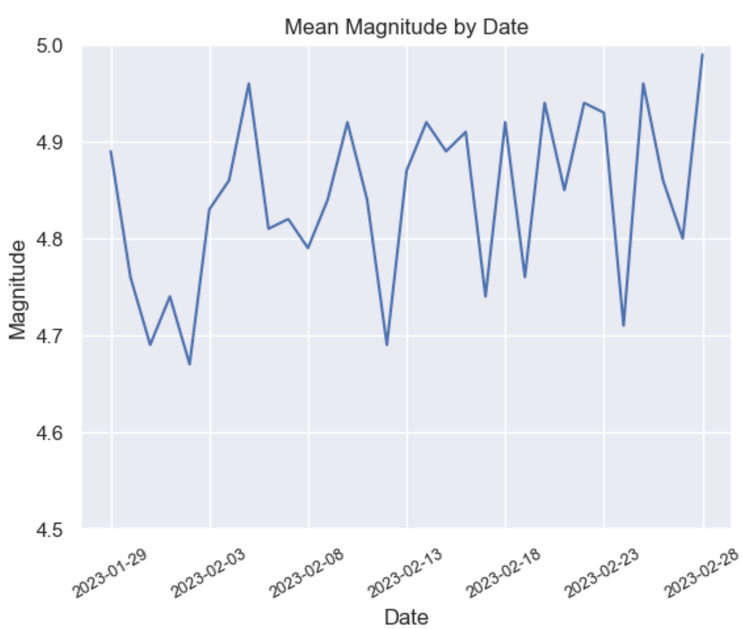
Looking at the 10 medium earthquakes with the greatest magnitudes for this category reveals more seismic activity in Turkey, Japan, and Indonesia. Three of these ten earthquakes occurring in Turkey reveals that a lot of fairly large earthquakes occurred there over this month period.

Question 3: How has seismic activity changed over the last month?

The following report contains the mean values of the magnitude, significance, depth, CDI (max reported intensity), and number of felt reports submitted to the Did You Feel It? system for each day over the month period. This data frame can be used to visualize how the mean values are changing over time over the course of the last month. This data frame will be saved to an external csv file and a few sample visualizations will be provided. The screenshot below shows a portion of the rows from that csv file.

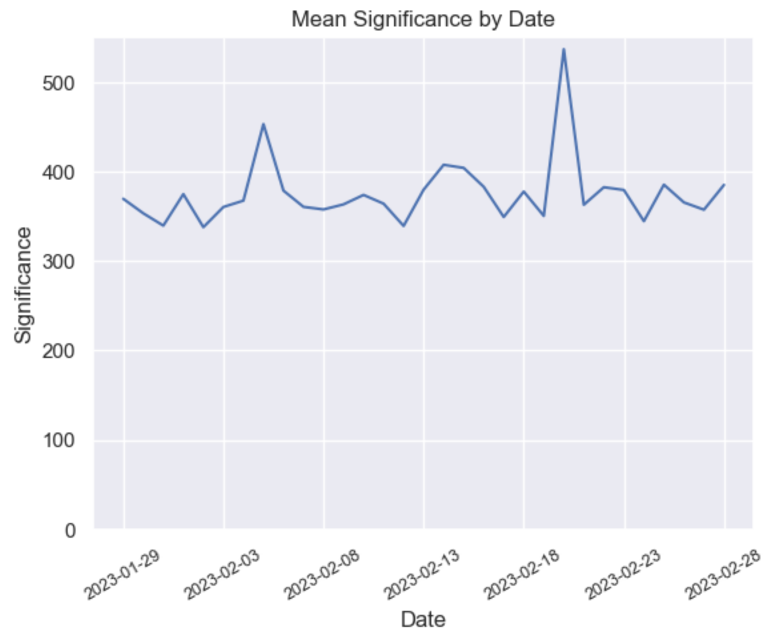
	A	B	C	D	E	F
1	Date	Mean Magnitude	Mean Significance	Mean Depth (km)	Mean Max Reported Intensity	Mean Number of Felt Reports
2	1/29/23	4.89	369.43	25.5	0.59	1
3	1/30/23	4.76	353.23	37.15	1.72	9
4	1/31/23	4.69	339.46	75.19	1.01	0
5	2/1/23	4.74	374.67	69.47	1.63	46
6	2/2/23	4.67	337.71	130.55	1.05	3
7	2/3/23	4.83	360.29	40.33	1.51	2
8	2/4/23	4.86	367.41	31.27	1.15	3
9	2/5/23	4.96	452.94	56.39	2.2	64
10	2/6/23	4.81	378.59	21.33	1.59	7
11	2/7/23	4.82	360.44	46.11	1.39	2
12	2/8/23	4.79	357.53	48.56	2.08	5
13	2/9/23	4.84	363.22	67.93	0.96	1
14	2/10/23	4.92	373.78	17.21	0.82	0
15	2/11/23	4.84	364	80.87	0.65	0
16	2/12/23	4.69	339.12	64.01	1.06	1
17	2/13/23	4.87	379.67	59.72	1.22	17
18	2/14/23	4.92	407.46	122.84	0.78	43
19	2/15/23	4.89	404	130.32	2.41	86
20	2/16/23	4.91	382.86	39.32	2.03	18
21	2/17/23	4.74	349.11	71.84	1.38	2
22	2/18/23	4.92	377.56	74.96	1.39	5
23	2/19/23	4.76	350.33	71.25	0.52	0
24	2/20/23	4.94	536.67	118.74	1.01	46

The following screenshot shows the line plot of the mean magnitude by date.



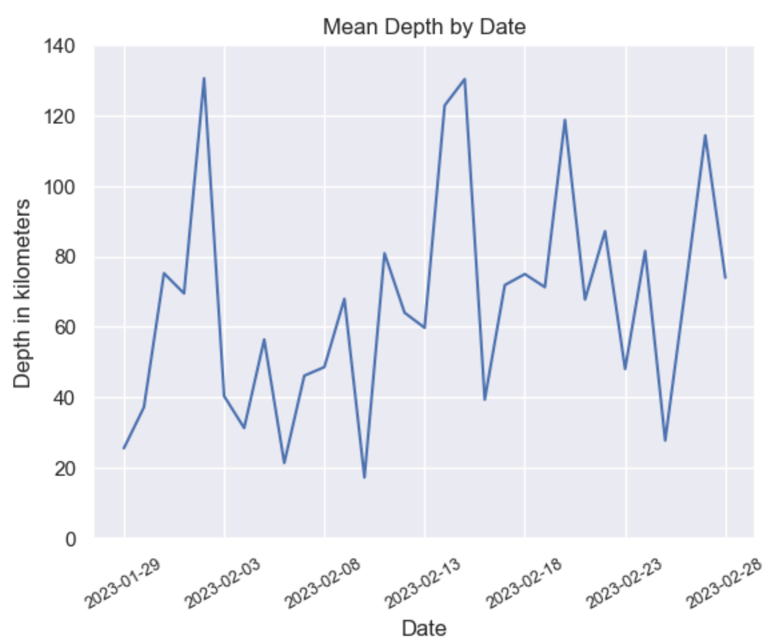
There is a large amount of variation in the mean magnitude of the earthquakes occurring each day in the data, although a slight overall upward trend can be seen. Looking at just the peaks of the line plot, there is a fairly consistent increase in those peaks from after the 13th of February until the end of the month.

The following screenshot shows the line plot of the mean significance by date.



The mean significance of the earthquakes occurring each day also shows some variation from day to day, although there is less of an overall general trend when compared to the mean magnitude values. There are two large spikes in the graph. Recall that the most significant earthquakes occurred in Turkey on February 5th, 6th, and 20th. These especially impactful earthquakes are the cause of these spikes.

Lastly, the following screenshot shows the line plot of the mean depth by date.



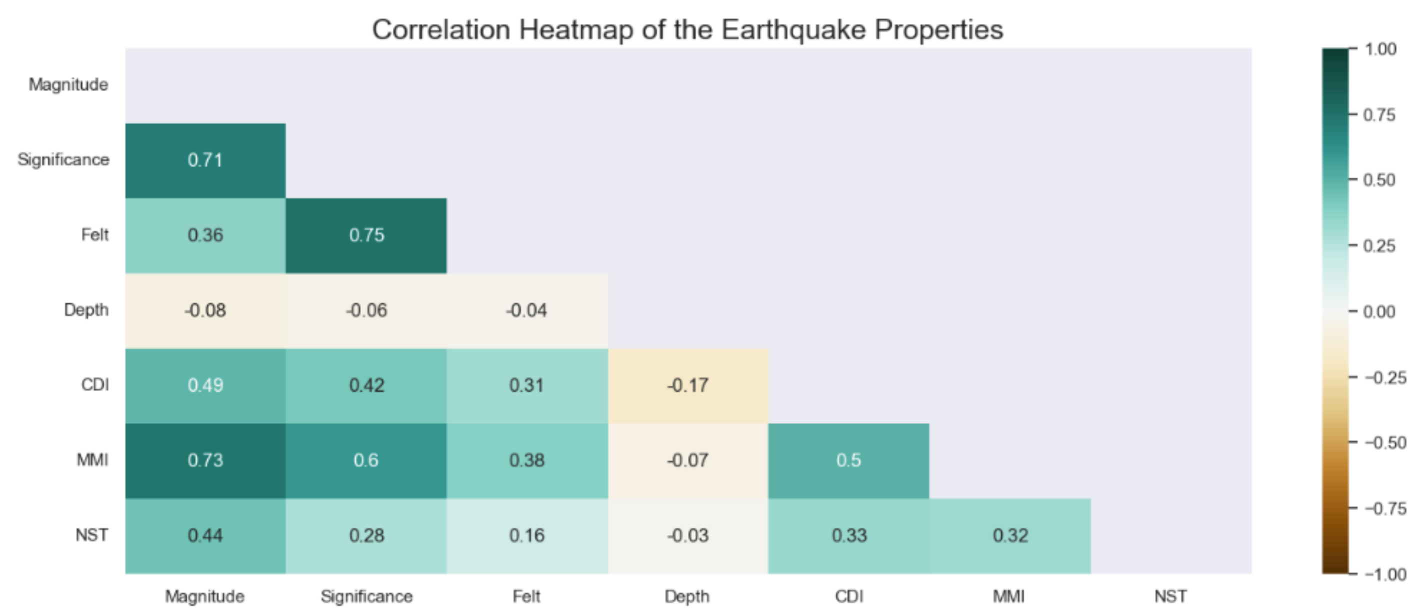
Plotting the mean depth of the earthquakes measured in kilometers for each date in the data set does not reveal any distinct patterns. The mean depth is quite low around the 5th and 6th of February when the large earthquakes occurred in Turkey, but then is much higher around the 20th of February when the third large earthquake occurred in Turkey. This could suggest that deeper earthquakes might not be those earthquakes that are the most significant or those with the highest magnitudes.

This leads to the next question, where we hope to increase our understanding of the relationship between the earthquake properties such as the depth, magnitude, and significance.

Question 4: How are the different properties of an earthquake, such as magnitude, depth, or significance, related to one another?

If the magnitude of an earthquake increases as the depth of an earthquake increases, then we would expect to see a strong positive correlation between these variables. To explore the associations among the different properties, a matrix of Pearson correlation coefficients will be created using the earthquake DataFrame. The numeric properties, Magnitude, Significance, Felt, Depth, CDI, MMI, and NST will be focused on for this question. The following screenshots show the external csv file of the correlation matrix and the heatmap created within the Jupyter notebook.

	A	B	C	D	E	F	G	H
1		Magnitude	Significance	Felt	Depth	CDI	MMI	NST
2	Magnitude	1	0.71	0.36	-0.08	0.49	0.73	0.44
3	Significance	0.71	1	0.75	-0.06	0.42	0.6	0.28
4	Felt	0.36	0.75	1	-0.04	0.31	0.38	0.16
5	Depth	-0.08	-0.06	-0.04	1	-0.17	-0.07	-0.03
6	CDI	0.49	0.42	0.31	-0.17	1	0.5	0.33
7	MMI	0.73	0.6	0.38	-0.07	0.5	1	0.32
8	NST	0.44	0.28	0.16	-0.03	0.33	0.32	1
9								



Viewing the correlation heatmap reveals that magnitude is most strongly correlated to significance and MMI (the maximum estimated instrumental intensity). This means that earthquakes with greater magnitudes also tend to feature greater significances and greater intensities. Based on the correlation values associated with depth, it appears that the depth of an earthquake is not correlated with any of the other variables since all of these values are very close to 0. The number of people who report feeling an earthquake is most strongly positively correlated with the significance of the earthquake, more so than the magnitude.

Final Conclusions

The most important takeaway is how much everyone can learn from sources of data that are not already contained in neatly arranged columns and rows. By having the means to represent semistructured data in a more readable format and one that allows for easy computations and aggregations, it is possible to spread important knowledge to a much wider audience. The analysis for each of the questions provided more insight into the data set and into seismic activity occurring on earth in general. Most people are not well-educated about earthquakes as there is a strong tendency to only hear about the most destructive ones that make the nightly news. There are far more earthquakes occurring than one might suspect. Just in the last 30 days there were over 500 earthquakes with a magnitude of 4.5 or greater, and if we were to expand our analysis to include all of the earthquakes regardless of magnitude we would find there were well over 10,000!

Examining the data in this report can provide many insights into earthquakes that are not heard about in our daily lives. Many people living in the United States are familiar with earthquakes happening in California, but this real time data feed has revealed the prevalence of earthquakes in Asia, in countries such as Turkey, Japan, China, and the Philippines. The earthquake data also helps us to understand how the different metrics are related to one another. Over the last month there was almost no correlation between depth and magnitude and significance, but very strong positive correlations between magnitude and significance and magnitude and the maximum estimated instrumental intensity.

This analysis is just the beginning of what can be learned from the USGS real time data feeds. As potential next steps, one could create map visualizations showing different aggregations of magnitude and significance for each region the world, one could continue to collect monthly data over the course of a whole year to gain a greater understanding of how the levels of seismic activity are trending, or one could analyze what aspects of an earthquake make it more likely to receive a certain category of alert rating. Understanding this data better could inform future policy decisions for those countries most affected by earthquakes. If governments around the world have a very deep understanding of these earthquake patterns, they could act in ways that would make their populations safer when these seismic events do occur.