**Report EDA on Earthquakes in Japan**

**Introduction**

Japan is one of the most earthquake prone regions in the world. I wanted to understand how earthquakes depend on other factors like depth of the rupture, location, etc. I did an Exploratory Data analysis on Earthquake Data in Japan (2001-2018).

**Methods**

I first converted the input data into a Pandas dataframe. I cleaned the data and removed unwanted columns. I also encoded categorical values in the dataframe to 1 or 0 values. I removed the noisy data in the form of earthquakes due to nuclear explosions and only kept natural earthquakes.

I plotted the Earthquake counts on Japanese Map using Folium Library.

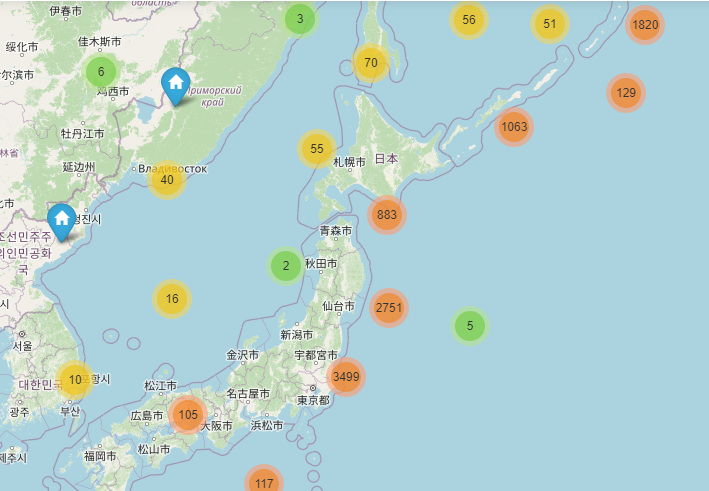
I did further visualisation and analysis using seaborn,matplotlib and pandas to understand relationship between various columns and the magnitude of the Earthquakes. I plotted scatter plots and regression plots for the same in 2D and 3D (for three variables).

I also plotted the date vs magnitude of earthquakes or did “time series analysis” to determine how earthquakes’ magnitude varied in japan with time.

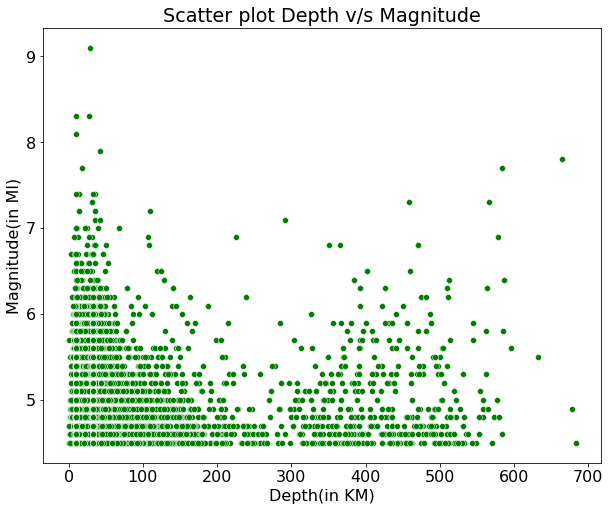
Then I plotted a heatmap over Japanese region using folium.

Finally, using pandas I got the top 20 named regions in and around Japan which have experienced most Earthquakes. Then I plotted them using Horizontal Bar-graph.

**Results and discussions:**

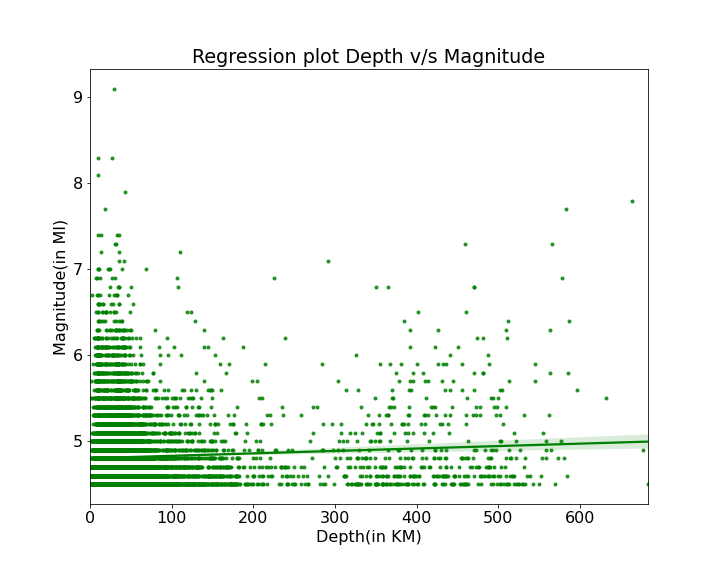


**Figure 1. Earthquakes in Japan 2001-2018**



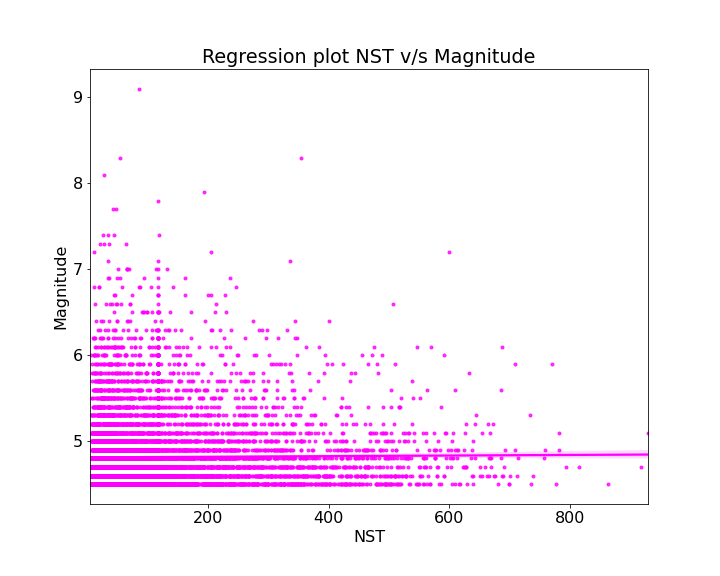
**Figure 2. Scatter plot between Depth of the rupture where earthquake began and Magnitude of the earthquakes.**

Here Depth is the depth where the earthquake begins to rupture and Magnitude is magnitude of the earthquake.The Scatter plot between depth and magnitude does not yeild any kind of relationship among the two prefectly. But we can analyse a linear data between the two.

****

**Figure 3. Regression plot between Depth of the rupture where earthquake began and Magnitude of the earthquakes.**

The regression plot between the depth and the magnitude values shows a linear relationship between the two. As the data is clustered at the start the deviation in the line(denoted by shading) is less while it increases with the value of depth as the data becomes uncertain.



**Figure 4**. **Regression Plot between Number of station required to measure the event versus magnitude**

As we can see, there seems like a relationship between these two variables, however that is a human error as the magnitude of the earthquake should not depend upon the no. of stations used to detect seismic activities.

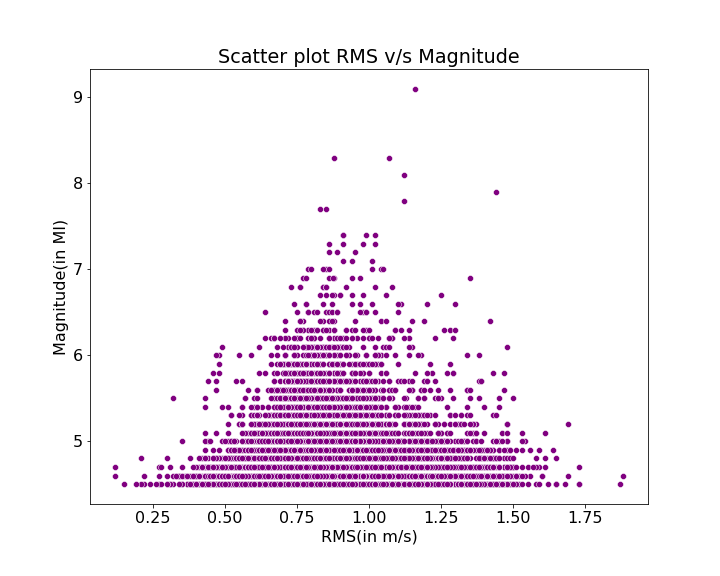
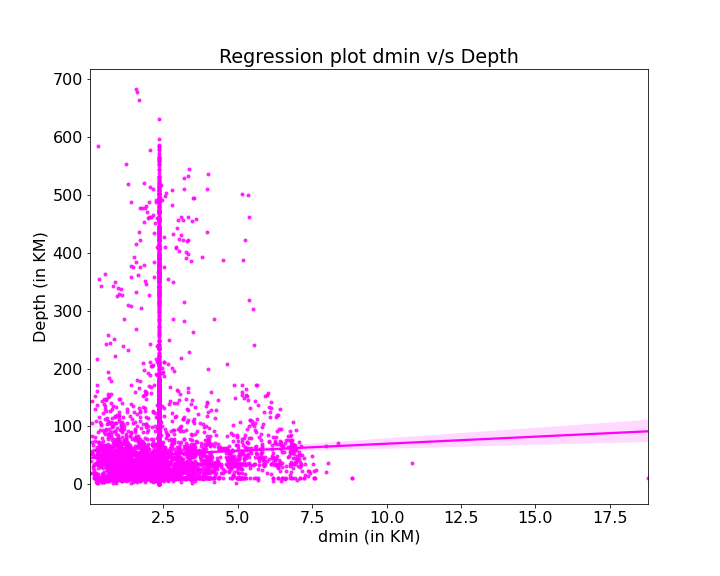


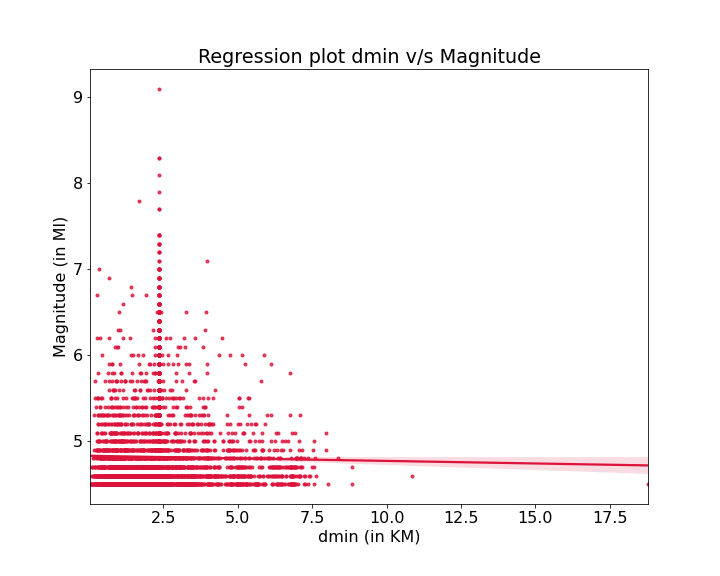
Figure 5. **Scatter plot between the Root Mean square speed of the waves and the magnitude.**

Root mean squared speed of the earthquake waves aren't directly related to the magnitude of the earthquakes in japan. The plot doesn't show any sort of linear regression among the two. There is huge uncertainity in the magnitudes of earthquakes between rms value ranging from 0.5 to 1.25 m/s



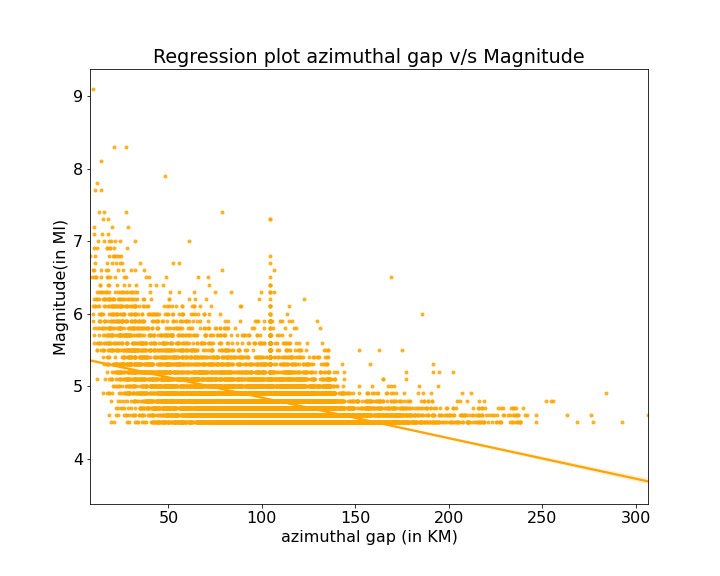
**Figure 6. Regression plot between the horizontal distance of the recorded place(d min) with the depth of the rupture.**

The regression plot shows a large amount of deviation meaning that the relationship among the two is not linear



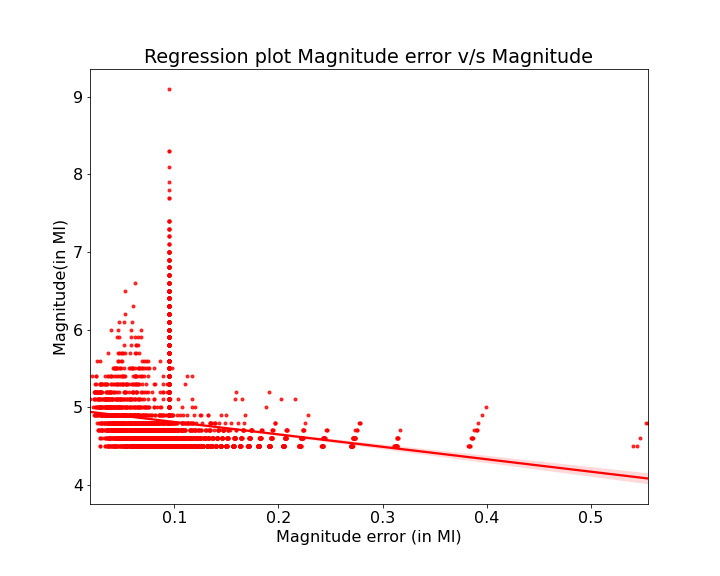
**Figure 7.** **Regression plot between the horizontal distance and the magnitude**

The regression plot between the horizontal distance from the epicenter and the magnitude shows a linear regression plot but with higer deviations, basically implying non-linearity between the two variables.



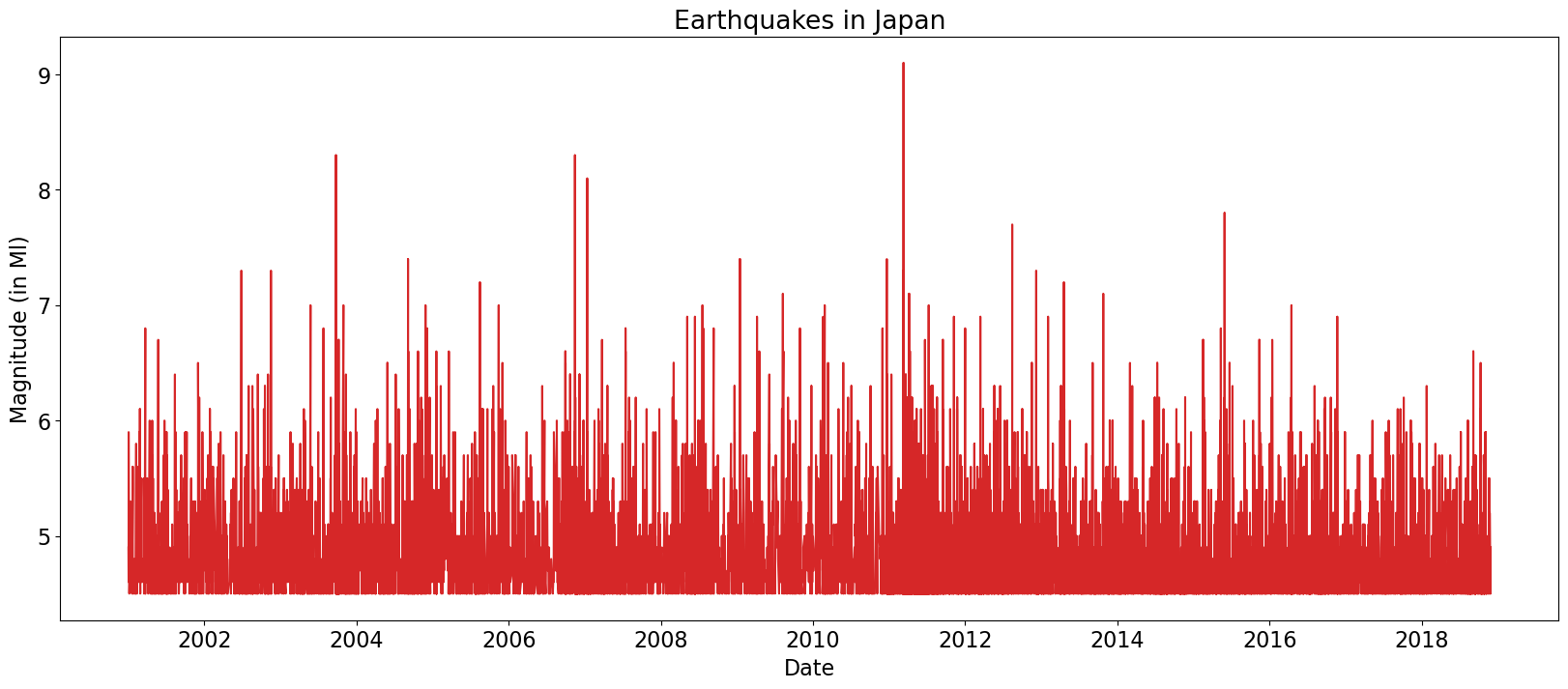
**Figure 8. Regression plot between azimuthal gap versus magnitude**

The regression plot shows a negative linear regression with a minimal deviation simply stating the relationship between the two a negative linear regression. Despite the deviation in the plot, the linearity is noticeable in the plot. On the other hand the negative regression signifies a linear regression relationship between magnitude and the azimuthal gap.



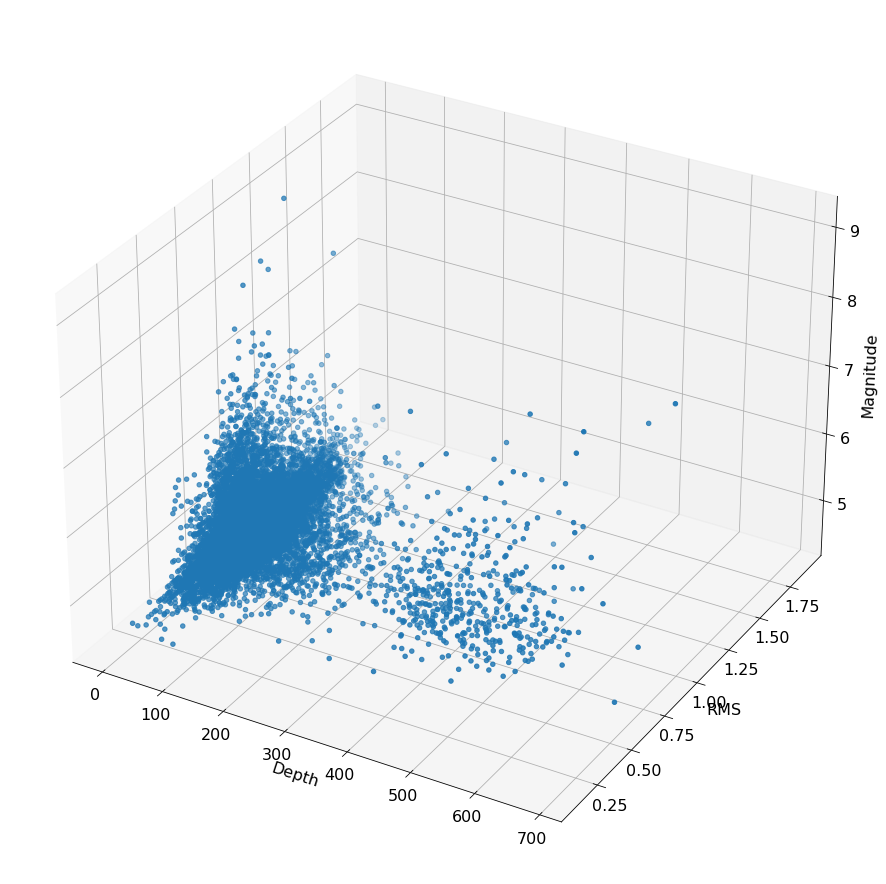
**Figure 9. Regression Plot between error in measuring magnitude**

This plot shows a negative linear regression with a minimal deviation, referring that, if the error in measuring the magnitude increases, the magnitude decreases. This is true and just basically verifying the data in the column.



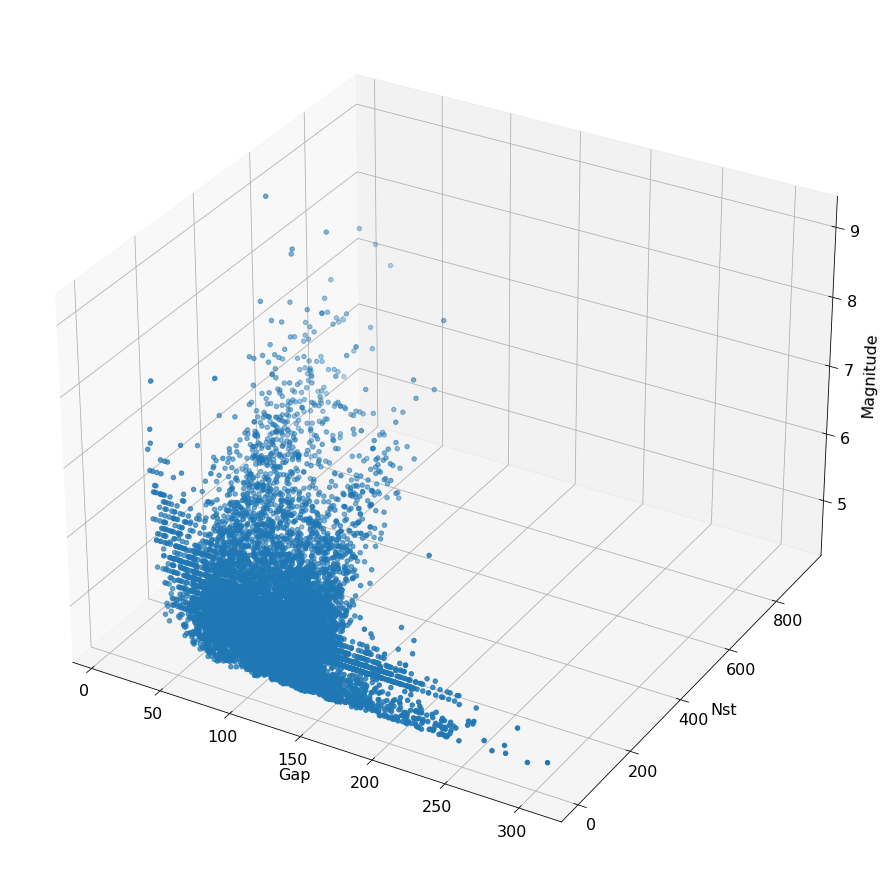
**Figure 10. Date vs Magnitude time series plot.**

from the above plot we can see that japan has been heavily affected by earthquakes in the entire period but peak magnitude was reached in 2011, ofcourse due to massive tokohu earthquake.



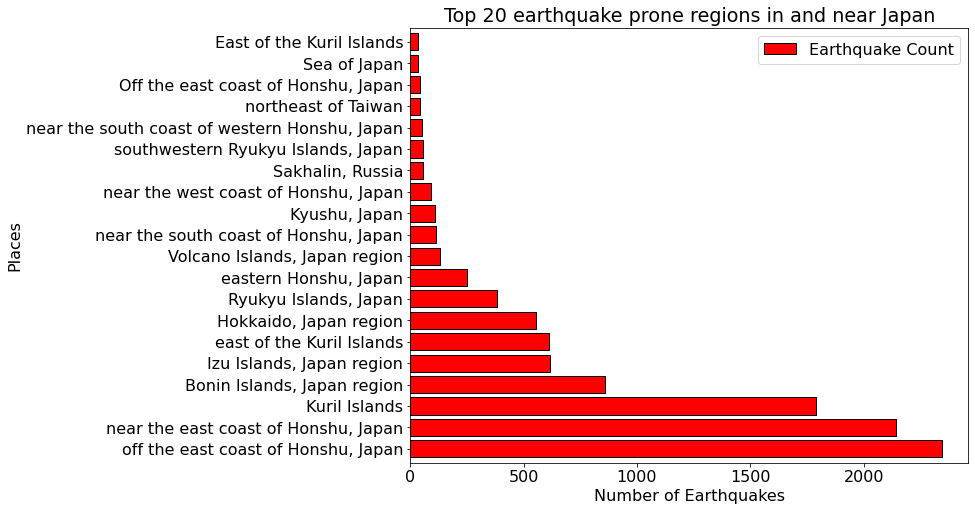
**Figure 11. 3 Dimensional Visualisation of Depth vs. Root Mean Square Speed vs. Magnitude**

Here we plot Depth in the X-Axis, Root Mean Square Speed in the Y-Axis and Magnitude of the recorded event in the Z-Axis to obtain a 3 dimensional plot for better understanding the realtionship between the varaiables.



**Figure 12. 3 Dimensional Visualisation of Gap vs. Number of Station Required vs. Magnitude.**

Here we plot Gap in the X-Axis, number of stations required in the Y-Axis and Magnitude of the recorded event in the Z-Axis to obtain a 3 dimensional plot for better understanding the realtionship between the varaiables.



**Figure 13. a horizontal bar graph to identify top 20 most Earthquake prone regions.**

**Conclusion:**

The project was intended for doing EDA on the Japanese earthquake dataset and also to determine the potentially risky places to earthquakes.