

Subject: Regarding the Data File that You Provided

Hello Linda,

I hope this message finds you well. I wrote the Python script that reads the file that you provided to me and created a Pandas DataFrame to store/display the data in a tabular form. I have attached the code to this file, so please take a look at it! Oh, and furthermore, the code also generates various visualizations and insights from the dataset.

Below is a further explanation of the code and the resulting plots:

### **Histogram of SS and S1**

Overview: These histograms offer visualizations and insights into the distribution of the value (intensity) of the short period 0.2-second ground motion at the location (SS) and the intensity of the 1.0-second ground motion at the location (S1).

Analysis: At first glance, we can see that the histogram of SS is more spread out than that of S1, which suggests that the intensity of 0.2-second ground motions varies a lot compared to the intensity of 1-second ground motions. Furthermore, the frequency of the values higher than 100 in the SS histogram is pretty high while that of the S1 histogram is pretty low. This suggests that there are many cases where strong (intensity>100) 0.2-second ground motions occur quite often while strong 1-second ground motions do not occur that often.

### **Boxplots of SS and S1**

Overview: This plot visualizes the spread and central tendency of SS and S1. These plots show the median, quartiles, and potential outliers for these variables.

Analysis: Same as the histogram above, these plots also suggest that the intensity of 0.2-second ground motions tends to be stronger than that of 1-second ground motions (median of SS = around 20 while median of S1 = around 10). Also, the intensity of 0.2-second ground motions is more spread compared to the number of 1.0-second ground motions (compare two boxes' lengths). However, we should keep in mind that both SS and S1 have many outliers.

### **Correlation matrix heatmap**

Overview: This shows the correlations between Longitude, Latitude, SS, and S1. If the colour is red, it means that there exists a high correlation, and if the colour is blue, it means that there is no correlation.

Analysis: The heatmap suggests that there is a high correlation between SS and S1. This means that if there is a high SS, there is a high chance that there will be a high S1, and vice versa. However, the correlation does not really exist between other pairs of variables.

## **Scatter plots colored by SS and S1**

Overview: These plots visualize geographic data with points on a 2D plane. Each point represents a location with longitude and latitude, and the colour of the points corresponds to SS and S1 values.

Analysis: The scatter plots show how the 'SS' and 'S1' values are distributed across different geographic locations (latitude and longitude). Patterns in the scatter plot, such as a gradient in colours can suggest a potential correlation or spatial pattern between the geographic coordinates and the 'SS' or 'S1' values. In this case, we cannot really find a notable correlation between latitude/longitude and SS/S1. However, based on the fact that two scatter plots (SS and S1 plots) have a similar gradient, we can assume that S1 and SS have a strong correlation.

If you have any further questions or would like additional details about the analysis, please feel free to reach out. I'm here to assist with any questions you may have!

Best Regards,  
Yunseo Heo.