**Function:**

Get\_height(x): This function uses regular expression to get float(height of monument) out of a sequence of strings.

Convert\_height(l): This function converts the height of monuments from Feet and Inches to meters.

**Link to Original Data:** [311 Service Requests from 2010 to Present | NYC Open Data (cityofnewyork.us)](https://data.cityofnewyork.us/Social-Services/311-Service-Requests-from-2010-to-Present/erm2-nwe9)

**Difficulties Encountered:** The main difficulty for this project is that all the data are strings instead of floats and Integers while we are told to do numerical analysis out of this dataset. Even for these columns with Integer values, they are not really Integers but more like Indicators or qualitative variables instead of quantitative. Hence the biggest difficult is then how to get numerical values from the strings. In addition, another difficulty is that there are a lot of missing values in the dataset making it extremely difficult to generate very meaningful results.

**Resources that were helpful:** I truely think this is a very hard project that involves many techniques that we do not cover in class. However a few useful resources are stackoverflow and also python documentation. They have very detailed information on how to solve specific issues or how to use specific functions and really helped me through the whole process. Below are some reference links:

<https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.groupby.html>

<https://stackoverflow.com/questions/4289331/how-to-extract-numbers-from-a-string-in-python>

**How you might improve or expand your work in future?**

One thing that I can expand is that I only examined the distribution of the height of different monuments while there are different other parameters as well such as width and so on. However, the dimension of the monument is arranged in a random way so it’s hard to get any other values other than the height but this is definitely something I can work on in future.