

What kind of cleaning steps did you perform?

- The data set contained 20 unnamed columns with 0 non null values which I removed.
- The 'Name of Satellite, Alternate Names' column contained 6 satellites that were repeated twice.
 - I initially assumed this might be an error, but further analysis revealed that although the names of the satellites were repeated they were actually to different satellites (for each case) because they had different NORAD and COSPAR numbers (What I believe are satellite id numbers). A couple of those satellites also had slightly different perigee and apogee numbers.
 - I removed the duplicated rows
- I renamed some of the column names to be more easy to use by shortening their length and removing spaces between words
- I removed columns that contained less than 500 entries, and columns with the sources for the satellite information, because they will not be used in the analysis

How did you deal with missing values, if any?

- I dropped entire columns that contained 0 non-null values, using **df.dropna(axis=1, how= 'all')**
- I kept columns that contained at least 1 non null value for further analysis.