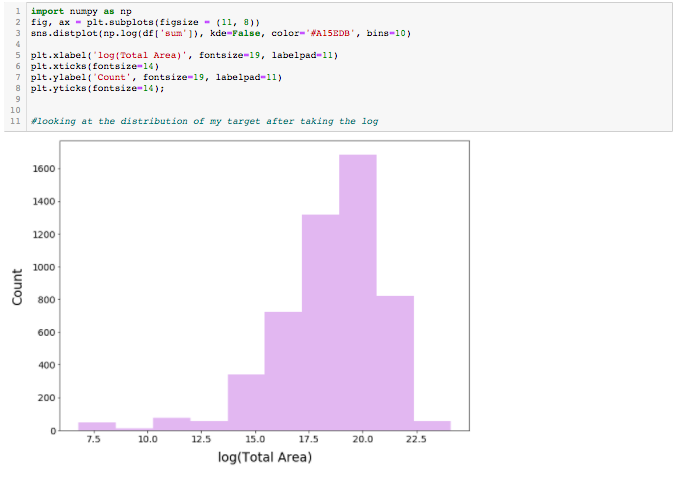
**CELL ANALYSIS**

Exploratory Data Analysis:

1. ‘Sum’ distribution:

Looking at the distribution of my target variable ‘Sum’ after taking the log



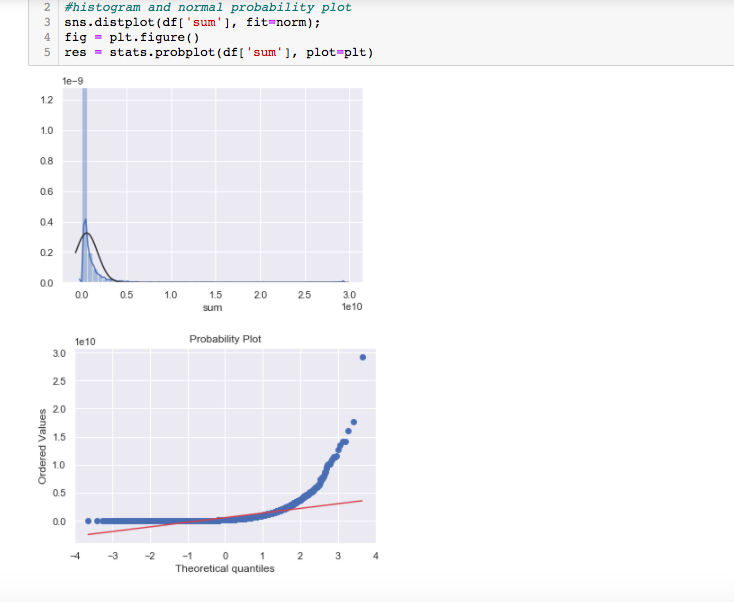
1. Facts about 'Sum' and check for normality:

-Deviates from the normal distribution

-Have appreciable positive skewness

-Show peakedness

'Sum' is not normal. It shows 'peakedness', positive skewness and does not follow the diagonal line.



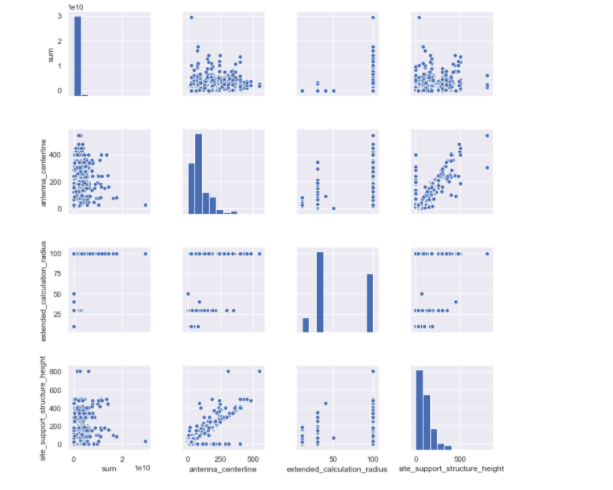
1. Correlation matrix:

We can see 'antenna\_centerline','extended\_calculation\_radius', 'site\_support\_structure\_height' and 'max\_power' (to some extent) are important features and are correlated with the 'sum'.



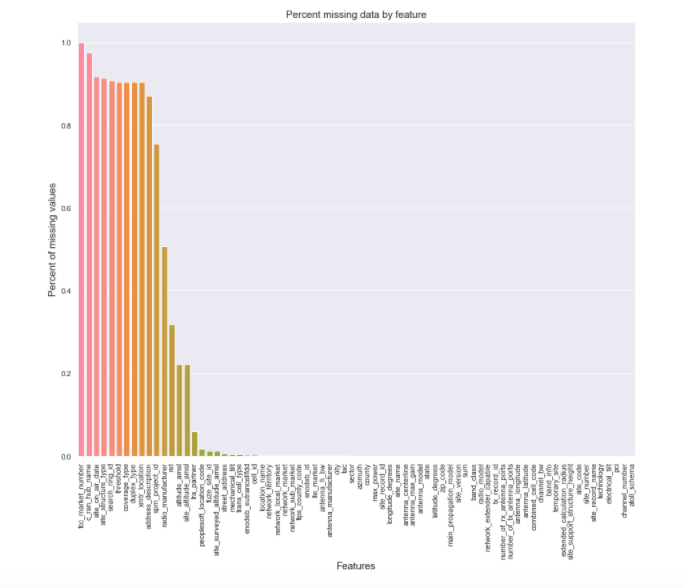
1. Scatterplot analysis

The plot concerning 'Sum' and other features can also make us think. We see what almost appears to be a shy exponential function.



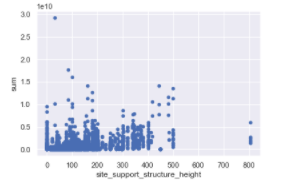
1. Missing Data:

We'll consider that when more than 15% of the data is missing, we should delete the corresponding variable and pretend it never existed. This means that we will not try any trick to fill the missing data in these cases.We can delete fcc\_market\_number,c\_ran\_hub\_name, site\_structure\_type, threshold,coverage\_type, search\_ring\_id,duplex\_type xmtr\_location,site\_on\_air\_date,address\_description,spm\_project\_id and radio\_manufacturer,ret,site\_altitude\_amsl and altitude\_amsl



1. Remove outliers

The max value with highest 'sum' seem strange and they are not following the crowd. We can speculate why this is happening.Therefore, we'll define them as outliers and delete them.



\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*