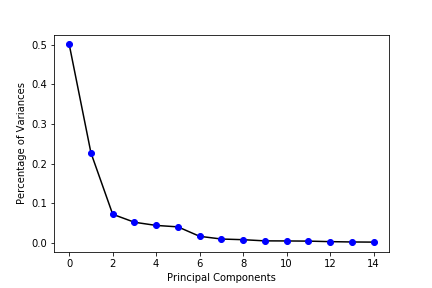
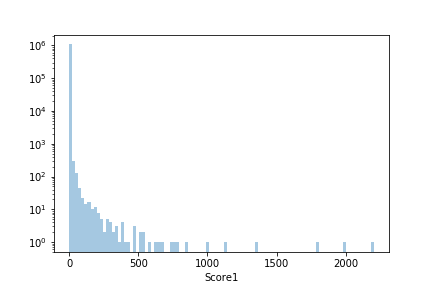
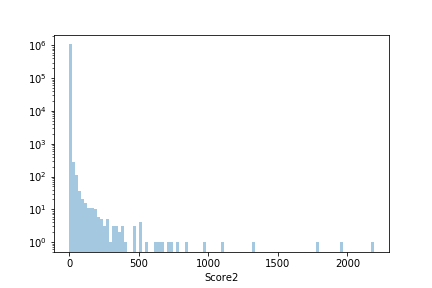
**Results and Observations**

Starting with over 45 variables, we reduced the number of variables to 6 through PCA (dimensionality reduction). These variables explained more than 93.7% of the variance in the data. A plot of number of principal components against cumulative variance explained (courtesy of PCA) has been shown below.

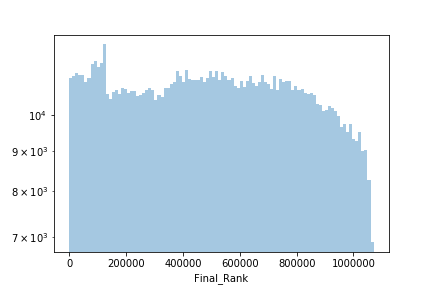
Fraud score calculated based on outlier detection via z-scores algorithm exhibited a right-skewed distribution (refer to the illustrations below). As expected, most of the records had low fraud scores and there were few outliers relative to the number of total records.



The distribution of fraud scores calculated using autoencoder algorithm also had a right skew (refer to the illustrations below). Once again, there were few records, relatively speaking, which had high fraud scores.



The final score



We manually examined the records with high fraud scores. We highlight the government properties and park& recreations and original missing value because although these properties are anomalies on several fronts, we know that government properties, parks and universities have very low risk of tax fraud. As for the missing value, we highlight the instances since the rules we applied is not sophisticated enough and may need further confirmation.

The top candidates for potential tax fraud have been listed below.



Clearly, for some of these records, the full value is exceptionally high. Further scrutiny reveals that such records have no information about lot front, lot depth, etc. For some other records, the full value of the property per build area is either excessively high or low. Since, a lot of these properties are owned by real estate firms, we can infer that either the properties owned by them are significantly different from an average property or they might be exploiting loopholes (and/or committing potential tax fraud) in the property tax law.

Top 10 Candidates for Potential Fraud



For this property, the BLDFRONT and the BLDDEPTH are just 1, which is so low and definitely needs further investigation.



For this property, the FULLVAL, the AVLAND and the AVTOT are so high, compared to the average. Since the owner is US GOVERNMENT, it may make sense.



For this property, the FULLVAL, the AVLAND and the AVTOT are so high, compared to the average.



For this property, the LTFRONT and the LTDEPTH are just 1, which is so low and definitely needs further investigation.



For this property, the LTFRONT and the LTDEPTH are just 1, which is so low and definitely needs further investigation.



For this property, the LTFRONT and the LTDEPTH are just 1, which is so low and definitely needs further investigation.



For this property, the BLDFRONT and the BLDDEPTH are just 1, which is so low and definitely needs further investigation.



For this property, the BLDFRONT and the BLDDEPTH are just 1, which is so low and definitely needs further investigation.



For this property, the FULLVAL, the AVLAND and the AVTOT are so high, compared to the average. Since the owner is PARKS AND RECREATION, it may make sense.



For this property, the FULLVAL, the AVLAND and the AVTOT are so high, compared to the average. Since the owner is PARKS AND RECREATION, it may make sense.