Design Transformation Pipeline

1. Outlier visualization in all the columns using Boxplot.
2. The value ‘99999’ is dropped from every column using pandas.
3. The value ‘99999’ is replaced by Nan.
4. Then the Nan values are replaced by the mean of that particular column.
5. Detecting the existing outliers in the dataset using the IQR method.
6. Replacing the outliers with the help of interpolation method. (other method like mean replacement, forward fill were used as trial error method)
7. Finding the top 20 most correlated columns with the target variable.
8. Reducing the dataset to 21 columns (one column of target feature).
9. Applying the algorithm to the transformed dataset.