

Wrangle Report

The gather phase of the project was done in the following steps: I read the file 'acs2017_census_tract_data.csv' into a dataframe called 'df_original' I then created a copy called 'df' for wrangling purposes.

The assess and clean phases of the project were done in the following steps: I found 158 rows where only one column was null. 133 of these were for the column 'ChildPoverty'. I isolated these rows and used a random forest regressor to predict their value. I don't expect this to have any meaningful impact on the results of any analysis done on this dataset. I used this as an opportunity to sharpen and demonstrate some of my knowledge of machine learning. The second mistake I saw was in the 'County' column. Each county name was followed by the word 'County', which was obviously redundant. I removed this word, leaving only the name. I then created a new CSV file from this updated dataframe called 'acs2017_census_tract_data_master.csv'