**INCOME QUALIFICATION CLASSIFICATION**

The problem to be addressed is to classify the poverty level of families based on many features. The classes are divided into 4 types and we need to predict the right class of a family based on the features.

I have done some data analysis to just understand about the data and then pre-process it.

The following queries have been answered in this project

1. Identify the output variable.
2. Understand the type of data.
3. Check if there are any biases in your dataset.
4. Check whether all members of the house have the same poverty level.
5. Check if there is a house without a family head.
6. Set poverty level of the members and the head of the house within a family.
7. Count how many null values are existing in columns.
8. Remove null value rows of the target variable.
9. Predict the accuracy using random forest classifier.
10. Check the accuracy using random forest with cross validation.

We have used random forest classifier on train test data and got an accuracy of 91% and also used cross validation method and had an accuracy of 89%.

Below are the images captured as a part of the project















