***Business Understanding***

The United States constitution had made mandatory the decennial census. The census provided a headcount which did not provide enough details to make data driven decisions about the use of government funds. A long form with necessary details was introduced in the 20th century. Only one in six households received the additional questions on the long form. In the 21st century, the American Community Survey (ACS) was established to conduct ongoing surveys with the details previously included on the decennial long form census. The ACS survey results provide data to determine the appropriate distribution of more than $400 billion in federal and state funds. The funds are for social programs as well as physical and service infrastructure.

<https://www.census.gov/programs-surveys/acs/>

Annually 3.5 million homes are surveyed. A random sample was performed on both the housing data as well as the personal data to allow for manageable data sets while including a macro vision of the data. We will consider the data the training sets.

The validation of the data can be done in several ways. A statistically sound method would be to validate the data using another random sampling which would be the test set. Another potential option would be the validate the federal spending. Due to the difficulty in finding the specifics spending details, the spending validation model would be prohibited by the time constraints as well as data availability constraints.

***Data Description***

The housing data includes 231 attributes. The personal data included 283 attributes. The full description of the data can be found in the data dictionary