

## **Data Mining from US Census**

### **Client**

David Flammer, Data Verity, Inc. [david@dataverity.net](mailto:david@dataverity.net)

### **Background**

For commercial clients, we develop models based on various parameters, such as financial information, many of which are captured in the US Census. The US Census has a large amount of information, which could be used for various kinds of academic or commercial research. We have a current need to improve our ability to pull essentially arbitrary information from the US Census using APIs that are freely available through web services they offer. In particular, the focus is on block level information to help inform demographics of certain locations.

### **Project Goals and Requirements:**

You will develop a (server side) library that will send requests to the US Census servers to request information and then process that information into a usable format. As an example, pull demographics surrounding an address (average income, education level, etc.). This should include the ability to send batch requests (of some number of addresses).

Then you will develop a front end using Javascript/HTML5 that can be used as a research tool to pull the information you request and feed it into a local database for analysis.

If time permits, you can use your tool to ask and then answer some questions you find interesting. For instance, a hot topic right now is wage gaps. You could ask “How much of a wage gap exists between men and women who graduate with STEM undergraduate degrees?”

### **Suggested Team Size:**

3-4 students

**Paid internships are available after Field Session for students that perform well.**

### **Skills/Experience for CSM Students:**

- US Census data, and experience with using web APIs like those used by the US Census.
- Javascript/HTML5
- PHP
- MySQL