**Census Income**

**Project Description**

This data was extracted from the 1994 Census bureau database by Ronny Kohavi and Barry

Becker (Data Mining and Visualization, Silicon Graphics). A set of reasonably clean records was

extracted using the following conditions: ((AAGE&gt;16) &amp;&amp; (AGI&gt;100) &amp;&amp; (AFNLWGT&gt;1) &amp;&amp;

(HRSWK&gt;0)). The prediction task is to determine whether a person makes over $50K a

year.

Description of fnlwgt (final weight)

The weights on the Current Population Survey (CPS) files are controlled to independent

estimates of the civilian non-institutional population of the US. These are prepared monthly for

us by Population Division here at the Census Bureau. We use 3 sets of controls. These are:

1. A single cell estimate of the population 16+ for each state.

2. Controls for Hispanic Origin by age and sex.

3. Controls by Race, age and sex.

We use all three sets of controls in our weighting program and &quot;rake&quot; through them 6 times so

that by the end we come back to all the controls we used. The term estimate refers to

population totals derived from CPS by creating &quot;weighted tallies&quot; of any specified socio-

economic characteristics of the population. People with similar demographic characteristics

should have similar weights. There is one important caveat to remember about this statement.

That is that since the CPS sample is actually a collection of 51 state samples, each with its own

probability of selection, the statement only applies within state.

Dataset Link-

https://raw.githubusercontent.com/dsrscientist/dataset1/master/census\_income.csv