Initial Data Exploration

## Purpose

This document does a very cursory exploration of the data included for the health care (HC) data files. There are three and each will have its own section.

# HC Chronic

Below is a truncated table to help visualize the raw data that is provided.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ID | Dataset | Label | Characteristic | Chronic Conditions | 2002 | 2003 | 2004 | 2005 | 2006 |
| 1 | Chronic | Age | 18-64 years | 0-1 chronic conditions | 85.1 | 84.9 | 84.5 | 84.0 | 83.7 |
| 2 | Chronic | Age | 18-44 years | 0-1 chronic conditions | 93.3 | 93.2 | 93.0 | 93.0 | 92.9 |
| 3 | Chronic | Age | 18-24 years | 0-1 chronic conditions | 96.4 | 96.7 | 96.8 | 96.6 | 96.2 |
| 4 | Chronic | Age | 25-44 years | 0-1 chronic conditions | 92.3 | 92.0 | 91.7 | 91.7 | 91.8 |
| 5 | Chronic | Age | 45-64 years | 0-1 chronic conditions | 71.4 | 71.6 | 70.9 | 70.2 | 69.9 |
| 6 | Chronic | Age | 45-54 years | 0-1 chronic conditions | 78.4 | 79.1 | 77.6 | 77.8 | 77.1 |

The columns included are: - ID - Dataset - Label - Characteristic - Chronic Conditions - 2002 - 2003 - 2004 - 2005 - 2006 - 2007 - 2008 - 2009 - 2010 - 2011 - 2012

The columns with year indicators likely have the calculation ‘*characteristic multiplier factors*’ which is the value we will eventually be forecasting. The ID column is a unique identifier for each row and won’t really be of use during the analysis. Data set is an indicator for which data set it comes from Chronic, Insurance, or Vaccinations. This will only become useful should the analysis require all three data sets to be appended. This leaves Label, Characteristic, and Chronic Condidtions as the dimensions that will drive the analytic.

To that end, each of these dimension contains the following values:

|  |
| --- |
| Label |
| Age |
| Sex |
| Pct Poverty |
| InsStatus |
| Geo |

|  |
| --- |
| Characteristic |
| 18-64 years |
| 18-44 years |
| 18-24 years |
| 25-44 years |
| 45-64 years |
| 45-54 years |
| 55-64 years |
| 65 years and over |
| 65-74 years |
| 75 years and over |
| Male |
| Female |
| Below 100% |
| 100%-199% |
| 200%-399% |
| 400% or more |
| Insured |
| Private |
| Medicaid |
| Uninsured |
| Northeast |
| Midwest |
| South |
| West |

|  |
| --- |
| Chronic Conditions |
| 0-1 chronic conditions |
| 2-3 chronic conditions |
| 4 or more chronic conditions |