# **Gathering Data Status Report**

## Name: Alyssa Basile

## Date: November 28, 2021

## Status Report for Gathering Data

| **Status** | Progress |
| --- | --- |
| **Variables of Interest** | * HLTHPLN1 – The targeted audience is patients with no health coverage so this data will show whether they have health coverage or not. * ASTHMA3 – One of the groups is patients that have been told that they have asthma so this data will show if they have been told that they have asthma or not. * SMOKDAY2 – One of the groups is patients that smoke regularly so data will show whether they smoke regularly or not. |
| **Data Set Preparation** | * HLTHPLN1 – INT * ASTHMA3 – INT * SMOKDAY2 - INT * The revised data set is attached. |
| **Gathering Data\*** | * The new table in SQL is shown below. |
| **Process for Gathering Data** | * Before gathering data, obtain the data needed for analysis and select the data that is needed to be used (prepping data). Then save it in CSV format. * Before importing data into the table, a new table needs to be created. Column names needs to be selected and the data type with the columns. * Import the CSV file into the table. After the data has been imported, click “Apply” to MySQL database. I came across an issue where the error message popped up stating “incorrect integer value: ‘1’ for column ‘ID’ at row 1.” I finally got it fixed by manually fixing the first row ID number (1) and changing it to 0. Then I clicked apply and it applied to the database. Then I changed it back to 1 and updated the database and I was successful. * After applying the data to the database, it is ready for analysis. |

## \*Gathering Data

* ID – the column with primary key
* IMONTH and IYEAR – the month and year the interview had occurred, I kept these two columns because if one wants to update the data after Jan 2017 – Jan 2018, they can see that this data would be outdated.
* HLTHPLN1 – the column that will state whether they have health coverage or not (based on codebook)
* ASTHMA3 – the data that contains whether they have been told that they have asthma or not (based on codebook)
* SMOKDAY2 – the data that contains whether they smoke regularly, occasionally, or not at all (based on codebook)

A picture containing text, screenshot, monitor

Description automatically generated