ETL Project Final Report

* The sources of data that you will extract from.
* The type of transformation needed for this data (cleaning, joining, filtering, aggregating, etc).
* The type of final production database to load the data into (relational or non-relational).
* The final tables or collections that will be used in the production database.

Extract: your original data sources and how the data was formatted (CSV, JSON, MySQL, etc).

Transform: what data cleaning or transformation was required.

Load: the final database, tables/collections, and why this was chosen.

Sources of Data:

The sources of data included the following sources:

1. The Census
   1. URL: <https://factfinder.census.gov/faces/nav/jsf/pages/searchresults.xhtml?refresh=t&keepList=t>
   2. Data extracted from census:
      1. Annual State Population Estimate
      2. Estimated Number without Health Insurance
      3. Rate of Uninsured= Number without Health/Population
2. OASIS--stands for Online Analytical Statistical Information System; Is a Georgia State Public Health Database.
   1. URL: <https://oasis.state.ga.us/oasis/webquery/qryER.aspx>
   2. Data extracted from OASIS:
      1. Emergency Room (ER) visitation rates for Georgia for
3. AHRQ—stands for American Healthcare Research and Quality; Is a Healthcare Quality Organization
   1. URL: https://nhqr.net.ahrq.gov/inhqrdr/data/query

Todo List:

All:

* Review requirements to insure nothing missed
* Review code
  + Do one data verification
    - Eg look at data from web site and insure it correlates with final summary df

Cuong

* Do data check on ahrq and oasis data
* First draft of table code

Ted

* Go ahead and submit link
* Create mysql insertion code
* Write 1st draft of 2 sections of paper.
  + 1)sources of data 2)transformation

Jenz

* Data check on census data
* Review teds paper parts and edit as needed
* Do final data part of paper
  + Eg. Using mysql, table code, insertion code
* Send final paper to team for review by 4pm tomorrow