Group 4 Project Proposal

Austin Barrette

Louise Colleen Cachuela

Charles Madison

Chase Lester

COVID-19 Data by County ETL Project

**Data Sources:**

**https://www.kaggle.com/jieyingwu/covid19-us-countylevel-summaries?select=hospital\_visits.csv**

hospital\_visits.csv - Hospital Visits by each day from 3/1/2020-3/21/2020 by County in the United States

counties.csv - County Demographics

**Transformation needed for this data: (cleaning, joining, filtering, aggregating, etc)**

We will be using Postgres for our database

Extract above CSVs into data frames

Transform hospital visits data frame

Create filtered data frame for the hospital\_visits from specific columns

Rename the column headers from int format to text format

Clean the data by dropping columns and setting the index to fip

Transform county data frame

Rename the column headers

Clean the data by dropping column and setting index to fip

Final production database will be **Relational**

The final tables that will be used in the production database are:

Dates from dayone to daytwentyone with fips being our index

In the schema file, there will be a join statement showing how the two tables were related