

## Project Proposal – Team ETL-06

1. **Project Title:** A Comparative Demographic Study of Race/Ethnicity Distribution: Population, Education, COVID and Economy
2. **Team Members**
  - Neil Vora
  - Aaron Paul Lewis (lewisaaronpaul)
3. **Project Description/Outline**
  - a. An analysis of demographic distribution of race/ethnicity.
  - b. County level focus: Population, COVID, Business, Education.
  - c. A look at the demographic impact.
4. **Questions to Answer**
  - a. What is the distribution of the population by race/ethnicity?
  - b. What is the distribution of the COVID cases/death by race/ethnicity?
  - c. What is the county distribution/breakdown of various businesses/industries?
  - d. Which counties receive more or less governmental funds/aid and how is this money allocated across businesses?
5. **Datasets to be used (tentative)**
  - a. United States Department of Agriculture: Economic Research Service
    - <https://www.ers.usda.gov/data-products/atlas-of-rural-and-small-town-america/download-the-data/>
    - Excel and csv files
  - b. US Census
    - Excel files
  - c. USA facts.org:
    - Coronavirus Locations: COVID-19 Map by County and State
    - <https://usafacts.org/visualizations/coronavirus-covid-19-spread-map/>
    - Web scraping
    - County level data
    - Totals: cases, deaths
  - d. OpenWeatherMap API
    - JSON file
  - e. Wikipedia: List of U.S. state abbreviations
    - [https://en.wikipedia.org/wiki/List\\_of\\_U.S.\\_state\\_abbreviations](https://en.wikipedia.org/wiki/List_of_U.S._state_abbreviations)WHO Dashboard
  - f. Other demographics by county?
  - g. Business/Economic
    - i. [www.bea.gov](http://www.bea.gov)
    - ii. <https://www.census.gov/econ/geo-county.html>
6. **Rough breakdown of tasks.**
  - a. Data Exploration and Extraction
    - i. *Download files: Excel, csv, JSON*
    - ii. *Use at least one API*
    - iii. *Web scraping*
  - b. Data Cleanup:
    - i. *Use Jupyter notebook and Python*
  - c. Database
    - i. *PostgreSQL*
7. **Additional Notes:** Explore cloud storage???