**Project 1 Group 3**

# Project Proposal

1. Project Title
2. Team Members
   1. Phil
   2. Rob
   3. Diane
   4. Gurupdesh
3. Project Description/Outline
   1. Alcohol consumption has been an important part of American culture since inception, data is available that allows us to analyze consumer preference-based on a variety of variables, race, income, region etc. During this time of COVID liquor stores were identified early on as essential services and remained open. We are interested in understanding the impact shutdowns due to covid spread have had on alcohol sales, consumption and/or distribution.
   2. Scope Statement: Compare alcohol consumption stats from NIH.gov (see below) for 16 states, years 2017-2020, months March, April, May and/or June 2020.
4. Research Questions to Answer
   1. Alcohol Consumption Data:
      1. Identify seasonality patterns of alcohol consumption for prior years, by month.
      2. **Run statistical analysis on pandemic-specific months**, (Mar – Jun 2020) to identify pandemic-related changes in alcohol consumption.
   2. Shutdown Data:
      1. **Create a score for level of shutdown by state (see: NYTimes link)**
         1. Compare two states with the most and least level of shutdown?
   3. ~~Per census data, what are the consumption trends in the U.S.?~~
   4. COVID related Data:
      1. Testing Rates by State
      2. Positive Test Rates by States over time
      3. **Total Cases by State – Infection Rates**
   5. Overall Project Questions/Data:
      1. **What, if any, non-Pandemic year alcohol consumption rate changes occur with a given year? (i.e. Seasonality)**
      2. **Did alcohol consumption changes in 2020/Pandemic months from the prior year?**
      3. **How did COVID shutdowns, infections rates and alcohol consumption?**
      4. **Which regions of the U.S. saw increases in alcohol tax revenue? For the 16 states of data available from NIH.**
      5. **Did regions not affected by shutdowns see any significant changes in alcohol consumption/tax revenue?**
      6. **Alcohol consumption rates associated with infection rates? Or vice versa.**
5. Datasets to be Used
   1. Alcohol Consumption Data (source/links)
      1. National Institutes of Health | National Institute on Alcohol Abuse and Alcoholism: <https://pubs.niaaa.nih.gov/publications/surveillance-covid-19/COVSALES.htm>
   2. COVID Infection Rate Data (source/links)
      1. The COVID Tracking Project | Data API: <https://covidtracking.com/data/api>
      2. The COVID Tracking Project | Our Data: <https://covidtracking.com/data>
   3. COVID Shutdowns by State Information/Article (source/links)
      1. New York Times - See How All 50 States Are Reopening (and Closing Again): <https://www.nytimes.com/interactive/2020/us/states-reopen-map-coronavirus.html>
   4. Other possible Beer, Wine, Alcohol APIs (sources/links)
      1. <https://www.thecocktaildb.com/api.php>
      2. <https://www.openbrewerydb.org/>
      3. Others as needed!
6. Rough Breakdown of Tasks
   1. **Task 1: Identify data sources – first thing Saturday [DONE]**
   2. **Task 2: Describe Data [DONE]**
      1. [**NIH data on Pandemic Alcohol Sales:**](https://pubs.niaaa.nih.gov/publications/surveillance.htm)This file contains data on per capita alcohol sales from 16 states (Alaska, Arkansas, Colorado, Connecticut, Florida, Illinois, Kansas, Kentucky, Louisiana, Massachusetts, Missouri, North Dakota, Oregon, Texas, Virginia, and Wisconsin) by type of alcoholic beverage from January 2017 through June 2020. For 2020, currently available for the months of March, April, May and/or June 2020.
      2. [The COVID Tracking Project](https://covidtracking.com/data/api): COVID, Jan 22, 2020 thru Oct 9, 2020 (updated daily) deaths, hospitalization, positivity rates, testing rates
      3. New York Times shutdown by state:
   3. **Task 3: Assign data review by state (4 states each for 16 states total) for NIH/Consumption, COVID Tracking and New York Times Shutdowns**
      1. Phil
         1. Alaska,
         2. Arkansas,
         3. Colorado,
         4. Connecticut,
      2. Diane
         1. Florida,
         2. Illinois,
         3. Kansas,
         4. Kentucky,
      3. GG
         1. Louisiana,
         2. Massachusetts,
         3. Missouri,
         4. North Dakota,
      4. Rob
         1. Oregon,
         2. Texas,
         3. Virginia,
         4. Wisconsin
   4. **Task 4: Pull, clean, merge data sources in Jupyter notebooks**
      1. Jupyter Notebook 1: describing data exploration and cleanup process
      2. Jupyter Notebook 2: illustrating the final data analysis
   5. Task 5: Perform statistical analysis on merged data, look for correlations in Jupyter Notebook
   6. Task 6: Pull insights relevant to research questions from data
   7. Task 7: Create visualizations, including map to illustrate the final data analysis with Jupyter/Matplotlib
   8. Task 8: Prepare/draft presentation, i.e. PowerPoint/PDF to preserve work product outside the Jupyter Notebooks