Group #4 - ETL Project Proposal   4/21/20

Chris Bock, Val Lobas, Goutham Polavaram, Frankie Wong

Project Topic

Coronavirus has spread across the globe in recent months. This virus has upended how people work, learn, and interact. Our group plans to obtain data related to COVID-19 for the US by state and date. We will obtain and clean our datasets and place them in a database so analysis can be performed to identify new trends.

Analysts will be able to evaluate current hospital resource usage by state and plan resource utilization. They will also be able to evaluate the effectiveness of policy dates related to restricting contact between people at work, school, and in public places.

Potential Data Sources:

Source	Format	Description
Covid19.healthdata.org	Web scrape	State, mass gathering restriction date, initial business closure date, educational facility closure date, nonessential business closure date, stay-at-home order date, travel severely limited date, hospital resource usage
Kaggle.com	csv	Total confirmed cases and total deaths by state and county
Ourworldindata.org	Web scrape	Total tests per day by country
Worldometers.info	Web scrape	Total cases, new cases(per day), total deaths, new deaths, active cases, total cases/1M pop, deaths/1M pop, total tests, tests/1M pop

Do we have a high-level idea of how to accomplish our tasks?

Yes

How the team plans to split up the project work to start:

Frankie – web scraping of covid19, set up Github repo

Chris – web scraping of ourworldindata, start Github pages

Goutham – scrape worldometers and look for additional APIs

Val – ERD diagram and pull in the Kaggle csv