## Actions

1. Get project up on GitHub - DONE
2. Add CA CSV in git - DONE
3. Creation of notebooks
   1. ET (ExtractTransform)
   2. L (Load)
4. Git Set Up
   1. Resources Folder - DONE
   2. Rdme - DONE
   3. gitignore - DONE
   4. Report (as a markdown) - DONE
5. Change numbers to INT

Use COVID data: <https://covidtracking.com/data/api>

Use both API and CSV

Linda: California - CSV

John: New York – API

Missing casesViral

DeJuan: Set up, manage GitHub

Emory: SQL, Slack Channel ETL project. Group 5; create ERD, Entity Relationship Document

Siddarth: Create Report outline/framework

## Data Fields to Keep

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Column | Data Field Name | Data Type | Rename | NOTES |
| A | Date (**Primary Key**) | integer | Date | Date on which data was collected by The COVID Tracking Project. |
| B | state | string | State | Two-letter abbreviation for the state or territory. |
| E | deathIncrease | integer | Deaths | Daily increase in death, calculated from the previous day’s value.  Returns Null if no data is available. |
| I | hospitalizedIncrease | integer | daily\_hospitalization | Daily increase in hospitalizedCumulative, calculated from the previous day’s value.  Returns Null if no data is available |
| L | inIcuCurrently | integer | ICU Hospitalizations | Individuals who are currently hospitalized in the Intensive Care Unit with COVID-19. Definitions vary by state / territory, and it is not always clear whether pediatric patients are included in this metric. Where possible, we report patients in the ICU with confirmed or suspected COVID-19 cases.  Returns Null if no data is available |
| U | PositiveCasesViral | integer | Positive Viral Cases | Confirmed Cases  Total number of unique people with a positive PCR or other approved nucleic acid amplification test (NAAT), as reported by the state or territory. This is equivalent to a confirmed case as per the CSTE case definitions of August 5th, 2020 and April 5th, 2020. If we discover a jurisdiction is labeling cases as confirmed using other evidence (e.g. positive antigen tests), we will annotate that state or territory’s data accordingly.  Returns Null if no data is available |
| V | positiveIncrease | integer | Positive Increase | New cases  The daily increase in API field positive, which measures Cases (confirmed plus probable) calculated based on the previous day’s value.  Returns Null if no data is available |
| AF | totalTestResults | integer | Total Test Results | Total test results  At the national level, this metric is a summary statistic which, because of the variation in test reporting methods, is at best an estimate of US viral (PCR) testing. Some states/territories report tests in units of test encounters, some report tests in units of specimens, and some report tests in units of unique people.  Returns Null if no data is available |
| AG | totalTestResultsIncrease | integer | Test Increase | New tests  Daily increase in totalTestResults, calculated from the previous day’s value. This calculation includes all the caveats associated with Total tests/totalTestResults, and we recommend against using it at the state/territory level.  Returns Null if no data is available |

## Type of Transformation

1. Copy of columnar data
2. Rename keys
3. Joining data sets
4. Cleaning: Nulls; treat as zeros “0”, using Python