Memo

|  |  |
| --- | --- |
| To: | Dr. Landowski |
| From: | Maria Ng |
| Date: | 021821 |
| Re: | Project Proposal |

**Topic: Analyze Covid19 World Vaccination Progress**

**Data Description:**The main data set, Covid-19 World Vaccination Progress, is from the Kaggle website. The data is consisted of 15 columns (Table 1) and 3082 rows now. The data is constantly updated to reflect the current count of the day, while new rows will be added to the data set.

Table 1: Column Description

|  |  |
| --- | --- |
| Variables | Data Type |
| Country | object |
| iso\_code | object |
| Date | object |
| total\_vaccination | float64 |
| people\_vaccinated | float64 |
| people\_fully\_vaccinated | float64 |
| daily\_vaccinations\_raw | float64 |
| daily\_vaccinations | float64 |
| total\_vaccinations\_per\_hundred | float64 |
| people\_vaccinated\_per\_hundred | float64 |
| people\_fully\_vaccinated\_per\_hundred | float64 |
| daily\_vaccinations\_per\_million | float64 |
| Vaccines | object |
| source\_name | object |
| source\_website | object |

Additional data will be collected to track on number of vaccination sites open per day(Twitter, New Channels, Newspaper websites). The data will be depended on how readily available for analysis from US, California, or county. Also, California CDC data will be used.

**Research Questions:**

* Which vaccine did each country use?
* What is the percentage for total vaccination for US up to date in comparison to certain countries (TBD) vs. time?
* With California vaccination site open does the California trending reflects the US report from world vaccination progress data?
* How the Northern California counties (Contra Costa, Alameda, Marin, San Francisco..) performed among themselves?

**Data Preparation Plan**

1. Data download into Python3 spyder
2. Data cleaning
3. Data will be grouped by country, vaccine, and county.
4. Normalize data with population for ease of comparison.
5. Perform exploratory data analysis.
6. Then proceed to answering research questions.