CS 506

Vaccine Equity Team 2
Deliverable 0
Jingyi Li, Zhou Ye, Kush Vekeria, Liyu Qu

1. What is the project focus/overall goal?
   Our group wants to compare the vaccination rates of two cities, one which doesn't have a strong grassroots organization and one that does: Revere and Chelsea. Large hospitals have clinics in both cities that provide non-reimbursed services in order to maintain their non-for-profit tax status (it is their obligation under tax law). The goal is to compare the vaccination rates in these cities - if the hospital clinics were providing the critical services then the vaccine rate in Revere should be similar to that in Chelsea.

2. What type of data will you collect or be analyzing?

   Percentage by Age Group with a First Dose
   Percentage of Sex and Race/Ethnicity groups that have recived a First Dose
   Percentage by Age Group who are Partially Vaccinated
   Percentage of Sex and Race/Ethnicity groups that have been Partally Vaccinated
   Percentage by Age Group who are Fully Vaccinated
   Percentage of Sex and Race/Ethnicity groups that have been Fully Vaccinated

   COVID-19 Case Counts and Rates for 20 Prioritized Communities

   Mostly are discrete data, and some of them are also nominal data.

3. What are potential limitations of the project?
   Limitations: (1) The dataset we need to process contains data from the time when the covid-19 began till now, so it's a big dataset and may take us a long time to deal with and optimize the data. (2)The required output asked us to give a graph to show the effect of some grassroots organizations. But which kind of graph is better and what parameters should be given into consideration is still not decided yet. (3) Our data is concerning the Boston area and while the dataset provides rich features to scale for the entire country, given the time constraints, we cannot process data representing the entire country.

4. Why is this project important or Why did you choose this project?
   Zhou: Because I have some experience with IoT devices that are linked with healthcare, and I found I'm interested in such health care topics. Besides, covid-19 and equity

problems are the most-concerned problems in the past few years, which also interests me a lot.

Jingyi: I have done some data projects during my undergraduate years, but all of them are not related to medicine. I wish I can do analysis for medical related projects. Medicine is a very meaningful field in the world, and covid-19 is a hot topic that everyone is concerned about now.

Kush: I'm just starting out in data science and machine learning and I really enjoy it so far. My passions lie in medicine and healthcare as I hope to make it easier for low-income and minority communities to have access to healthcare. I felt that this project was important to me not only because it bridges the fields of medicine and computer science, but it sheds light on the substantial gap that exists in access to healthcare and it provides a way for me to help my community.

Liyu: As a beginner in the Data Science field, I think the real first project for something related to me is very important, and medicine is that field. Covid 19 affects many people all around the world, and equity problems are the same. Moreover, I think data analysis for vaccines is important and meaningful. By doing the project about these two themes, we can find if there is something that needs to be concerned, and we can also help the community as well.

5. What are your team's next steps? (include action items/tasks)
    - Grabbing datasets from the state of Massachusetts and contacts affiliated with local hospitals
    - Cleaning the data (removing any points that do not have sufficient features to ensure unbiased results).
    - Determining method to process the data (K-means, regression, etc.)
    - Producing charts/graphs so that audience can understand data
    - Provide analysis on the results