**Project Introduction**

Our data science term project aims to model the impact of COVID-19 in Chicago areas, using sociodemographic data (age, income, education etc.) and COVID-19 statistics (hospitalizations, deaths etc.). The sociodemographic and COVID-19 data used here is for each ZIP code within Chicago. An existing COVID-19 model, the CCVI score, ranks neighborhoods in Chicago by vulnerability based on several sociodemographic factors, and develops an overall vulnerability score for each neighborhood based on these rankings. An issue with this ranking system is that the difference in factors is not accurately reflected by a difference in ranks; for example, two neighborhoods could have a relatively big difference in income but only be a rank apart, and vice versa (relatively small difference in income and a rank apart). The ranking-based score would normalize all differences in sociodemographic factors when calculating the vulnerability score, which could be a potential issue when deciding on how to accurately and precisely distribute resources to areas based on their COVID-19 vulnerability. Our model seeks to improve upon the CCVI score by assessing areas in a more absolute way, such that it can predict COVID-19 risk of an area in terms of cases, hospitalizations or deaths, given the sociodemographic factors of that area.

We chose to develop the model for Chicago both because of our proximity to the city, and because Chicago has been a big COVID-19 hotspot in America during the last two years. Being a big, densely populated city, Chicago has had 567,470 COVID-19 cases, 41,694 hospitalizations and 7,635 deaths as of today. Despite these high numbers, the impact of COVID-19 has been unevenly distributed over the city of Chicago, and is primarily dictated by the sociodemographic factors that we are inspecting in our project. Since our model aims to predict COVID-19 impact from sociodemographic data, it can potentially help organizations distribute resources, healthcare and vaccination efforts in the city more accurately, as well as help policy-makers understand how the pandemic impact is affected by disparities in the city, and plan accordingly in the future to remedy this.