Introduction/Business Problem

In the city of Los Angeles, the COVID-19 pandemic has been very concerning as it was listed as the epicenter of COVID of the world (1). As misinformation continued to spread, the city was inundated with COVID-19 cases. Although there has been some suggestion of literacy levels playing a factor, no specific studies have been performed in Los Angeles (2). We hypothesize that education level also plays a factor in Los Angeles's pandemic struggles. However, it is unclear how healthy literacy plays a role, and this study will help determine whether to further investigate this topic (health literacy vs misinformation?) and whether we need to work on public education health efforts to combat the Pandemic in Los Angeles.

We will first create a map of the current COVID cases in Los Angeles County. We will apply a heatmap of the case load and also plot the deaths to demonstrate the density of the cases. We will also use FourSquare API to pull data on nearby super markets in the area identified as the most cases/deaths to see if there was a heavy amount of essential workers to contribute to that area. We will determine this based on looking at the heatmap with high death count (markers will be created to indicate high death count). We will then examine if income and education level play a role in the spread of COVID-19. We will plan to use Pearson's test to determine if there is any correlation with these factors.

This data can help assist the Mayor of Los Angeles County, the Department of Los Angeles Public Health and community leaders to develop a readjustment to their response plan as the pandemic has worsen during this time.

Describe the Data

We will obtain the LA County Health Department data on COVID-19 cases and deaths. This data is updated every weekend, which makes it fairly updated. The data also includes case rates and death rates. We will also pull Education Status and Income from the LA County Health Department. The education dataset will be the Highest Level of Education Attained Among Adults (Ages 25 Years and Older). The data categorizes different LA county areas into four educational attainment status: Less Than High School, High School Graduate, Some College, and Bachelor Degree or higher. These categories are presented in percentiles The Income dataset will be Median Household Income. Median household income is defined as the amount of that divides the household income distribution of a population into two equal groups. Each city will have an estimated dollar amount associated with each. I will ask a PhD scientist to review the data and make adjustments for normalization recommendations.

The data will allow me to perform a Pearson's correlation as well as graph trends to answer these questions. I will also perform a heatmap to help visualize the Cases/Deaths to help allow the reader to be able to understand any geographical reasons for the different case loads.

(1)L.A. County on verge of becoming COVID-19 epicenter - Los Angeles Times (latimes.com)
(2) Paakkari L, Okan O. COVID-19: health literacy is an underestimated problem. *Lancet Public Health*. 2020;5(5):e249-e250. doi:10.1016/S2468-2667(20)30086-4

Tools used:
Web Scraping html table from Wiki | by Priya Raja | Analytics Vidhya | Medium