Data Understanding and Preparation

Group 2

Vigna Nalliah, Charles Turley, Anthony Grannis

Final Project Summary

Our project objective is to evaluate how the City of Chicago is progressing with COVID-19 cases, tests, and deaths at a ZIP code level. With nearly 2.7 million people living throughout the city, we will be able to see which areas are proving to be hit the hardest with the ongoing pandemic.

The dataset that we are using is directly from the cityofchicago.org website and provides 21 columns and ~2900 rows of data to be analyzed and broken down. Here is the link to the dataset for use:

COVID-19 Cases, Tests, and Deaths by ZIP Code
https://data.cityofchicago.org/Health-Human-Services/COVID-19-Cases-Tests-and-Deaths-by-ZIP-Code/yhhz-zm2v/data

Below is an example of the tables that we have broken out using this data. This will be the framework for our database creation. Additionally, there is an example EER diagram showing the table relations within.

Week Table	Locations Table	Cases Table	Deaths Table	Tests Table
Week Number (PK)	Location ID (PK)	Week Number (PK)	Week Number (PK)	Week Number (PK)
Week Start	ZIP Code	Cases - Weekly	Deaths - Weekly	Tests - Weekly
Week End	Population	Cases - Cumulative	Deaths - Cumulative	Tests - Cumulative
	Row ID	Case Rate - Weekly	Death Rate - Weekly	Test Rate - Weekly
	ZIP Code Location	Case Rate - Cumulative	Death Rate - Cumulative	Test Rate - Cumulative
		Location ID (FK)	Location ID (FK)	Percent Tested Positive - Weekly
				Percent Tested Positive - Cumulative
				Location ID (FK)