**Covid Analysis:**

**Cook County Analysis**

**Project Report**



University of Connecticut

Business Analytics & Project Management

OPIM 5272 - Data Management and Business Process Modeling

**Group 6**

Sharanya Kurungati

Donald Bellino

Kishori Kasat

**Table of Contents**

[**Executive Summary**](https://docs.google.com/document/d/1jztI_TQyHcbp7MgWUlr-5YJGgHrLS8nGLyVkMe2FDMI/edit#heading=h.4hsccg9qeye)[**3**](https://docs.google.com/document/d/1jztI_TQyHcbp7MgWUlr-5YJGgHrLS8nGLyVkMe2FDMI/edit#heading=h.4hsccg9qeye)

**Business Objectives 3**

**User Requirements 3**

**Business Rules 4**

**Entity Relationship Diagram 5**

**Data Definition 6**

**Relations in the Third Normal Form 9**

**Tables in SQL 10**

**Sample Queries 10**

**Additional Queries 10**

[**Appendix**](https://docs.google.com/document/d/1jztI_TQyHcbp7MgWUlr-5YJGgHrLS8nGLyVkMe2FDMI/edit#heading=h.az2q3p8jvrdy) **10**

[**References**](https://docs.google.com/document/d/1jztI_TQyHcbp7MgWUlr-5YJGgHrLS8nGLyVkMe2FDMI/edit#heading=h.x2uwgmeafdee)[**11**](https://docs.google.com/document/d/1jztI_TQyHcbp7MgWUlr-5YJGgHrLS8nGLyVkMe2FDMI/edit#heading=h.x2uwgmeafdee)

[**Glossary**](https://docs.google.com/document/d/1jztI_TQyHcbp7MgWUlr-5YJGgHrLS8nGLyVkMe2FDMI/edit#heading=h.dxc0azkg9ncc)[**12**](https://docs.google.com/document/d/1jztI_TQyHcbp7MgWUlr-5YJGgHrLS8nGLyVkMe2FDMI/edit#heading=h.dxc0azkg9ncc)

**Executive Summary:**

The world is facing a global pandemic where real-time developments and the breadth of the

data available have been put to the test current practices.

To address this, we conducting a mixed methods approach where we utilized local data (survey),

public data (behavioral data), and governmental data (IDPH, Cook County, CMS, and Census

data) to see if we can uncover trends related to behavior, cases, deaths, and testing specific to

COVID-19.

**Business Objectives:**

We were curious to see:

• What are the various types of data available during a state-of-emergency like the COVID-19 global pandemic?

• What insights can we derive?

• Understanding the complexity of working with data during such a tumultuous time is important to get a sense of the type of data available.

• Get a better understanding of the types of databases that would be best suited for short-term

Emergencies.

**User Requirements:**

Users: Health data analyst, Health Insurance analyst’s and General public to understand the impact of Covid19.

User Requirements:

• To understand which ethnicity is highly affected.

• To understand what may be the highest medium (place like grocery, public malls etc.) to spread

the virus.

• Insightful information regarding testing results and deaths.

**Business Rules:**

• Business rule: We Only work on covid19 cases in Illinois State, Chicago City and Cook County.

Field constraint: Only Cook county of Chicago IL will be represented in Patient table

• Business rule: Every Patient must supply First name, last name, ethnicity & age.

Relationship constraint: The relationship between the tables must be governed by a participation

constraints wherein a record of information combination in the Patient table must be related to at least one record in the Covid19 impact table.

Data Details:

• Survey news is extracted from the information from individual survey form and general

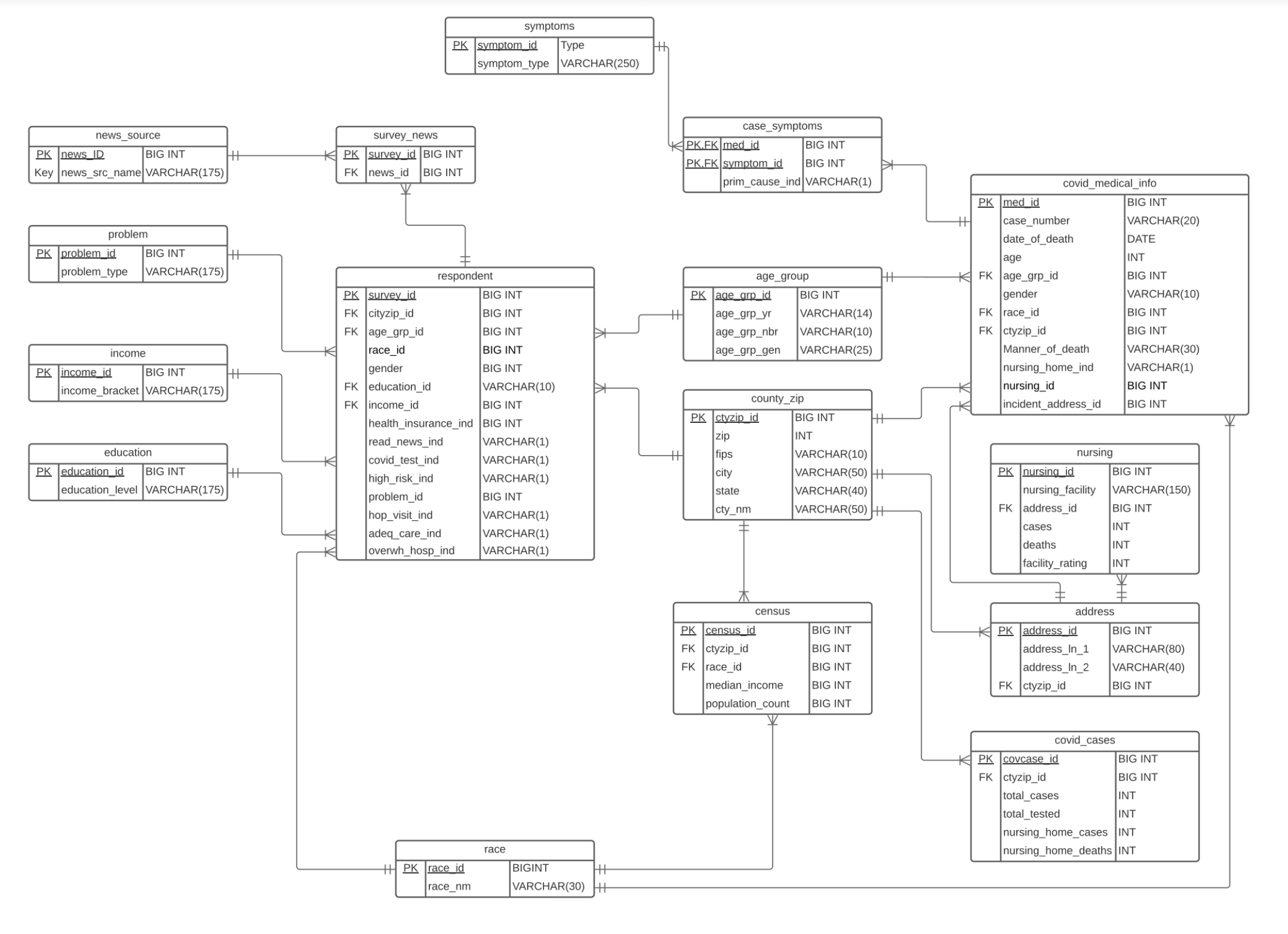
news

• Covid medical information and nursing is limited to data world websites.

•Age group, census, contry\_zip data is collected from public records for year 2020.

• All the social data is gathered from the public records and data world.

**Entity Relationship Diagram:**

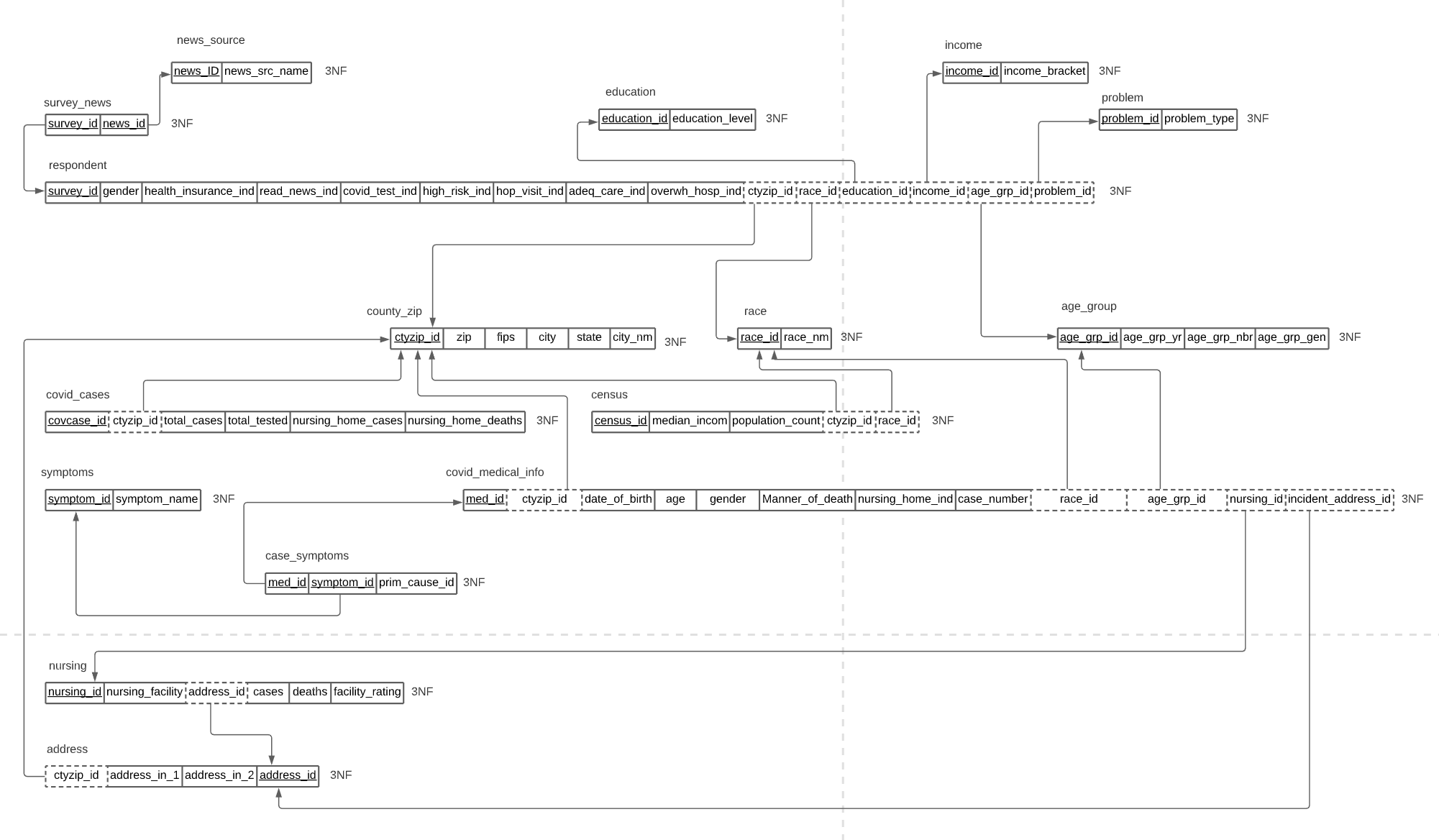
****

**Data Definition:**

|  |  |
| --- | --- |
| **Entity** | **Definition** |
| news\_source | Source of survey data organized by news organization |
| problem | Scores rating of Covid problem according to those surveyed according to stay at home orders |
| income | Grouping of people by income brackets |
| education | Education level of individuals surveyed |
| social\_athome | Distinguishes between people spending time home and people who are social |
| race | Grouping of people by physical characteristics and shared ancestry |
| survey\_news | A survey ID, and a News ID |
| survey | A survey of individuals identified by survey\_id detailing their race, gender, income, health insurance, covid testing, whether they are high risk, problem, hospital visit, adequate care, and whether the hospital was overwhelmed |
| social\_behavior | Illustrates the social behavior of individuals in Cook County in terms of distance travelled, work behaviors, devices used, time home, and dwell time |
| symptoms | Identifies types of symptoms by a unique symptom\_id identifier |
| case\_symptoms | Identification of symptoms and primary causes |
| age\_group | Group of people by similar age and generation |
| county\_zip | The Zip code, City, County name, and FIPS code of individuals |
| census | An official census of population by zip, race, median income and population count |
| covid\_medical\_info | Provides Covid medical information identified by med\_id, provides details with respect to date of death, age, gender, race, manner of death, and nursing home |
| nursing | A nursing facility, the number of covid cases in each facility, deaths, and facility rating |
| address | Street name, number and zip code of nursing home |
| covid\_cases | Illustrates the total number of covid cases, total tested, nursing home cases and deaths by zip code |

|  |  |
| --- | --- |
| **Attribute** | **Definition** |
| news\_id | Primary key that uniquely identifies a news source |
| news\_src\_name | Name of news organization |
| problem\_id | Primary key that uniquely identifies a problem |
| problem\_type | Defines level of Covid problem with regards to stay at home order |
| income\_id | Primary key that uniquely identifies level of income |
| income\_bracket | A bracket of income of either unknown income, less than $20,000, $20,000, $50,000, or $100,000 |
| education\_id | Primary key that uniquely identifies education |
| education\_level | Distinguishes between varying levels of education, those who have less than high school, 2 year associate's degree, some college, some postgraduate or professional schooling, high school, or a 4 year college/university degree |
| athome\_id | Primary key that uniquely identifies those at home |
| socialbehv\_id | Primary key that uniquely identifies social behavior, Foreign key in the social\_athome entity |
| ctyzip\_id | Primary key that uniquely identifies the county\_zip entity, Foreign key in the social\_behavior entity. |
| behv\_date | Date of an occurance of social behavior |
| device\_count | Amount of devices in a particular store or place |
| distance\_traveled | Measure how distance traveled by each individual |
| hour\_val | A value given for hours spent out of home |
| athome\_count | A value given for hours spent at home |
| race\_id | Primary key that uniquely identifies a given race |
| gender | Identifies an individual as either a Male or Female |
| read\_news\_ind | Indicates a Yes(Y) or No(N) if an individual reads the news |
| covid\_test\_ind | Indicates whether an individual has ever had a COVID test |
| hop\_visit\_ind | Indicates whether the individual had a hospital visit |
| adeq\_care\_ind | Indicates whether or not the individual received adequate care |
| overwh\_hosp\_ind | Determines whether not an individual's hospital was overwhelmed by COVID |
| symptom\_id | Primary key that uniquely identifies a symptom, Foreign Key (composite primary key) in case\_symptoms |
| symptom\_type | Details a specific type of symptom |
| med\_id | Primary key that uniquely identifies a covid\_medical\_info, acts as a composite Primary Key (PK,FK) for case\_symptoms |
| prim\_cause\_ind | Determines whether or not COVID was the primary cause of symptoms |
| case\_number | A number given for a medical case |
| date\_of\_death | Date an individual passed away |
| age | Age |
| manner\_of\_death | Way in which an individual passed away, either Natural or Accident |
| nursing\_home\_ind | Nursing Home |
| nursing\_id | Primary key that uniquely identifies nursing |
| incident\_address\_id | Identifies address of an incident |
| nursing\_facility | The name of a nursing facility |
| cases | The amount of COVID cases in a nursing facility |
| deaths | The amount of COVID deaths in a nursing facility |
| facility\_rating | A rating between 0-5 of a nursing facility |
| address\_id | Primary key that uniquely identifies address |
| address\_ln\_1 | First line in an address |
| address\_ln\_2 | Second line in an address |
| covcase\_id | Primary key that uniquely identifies covid\_cases |
| total\_cases | Total amount of cases |
| total\_tested | Total amount of individuals tested for COVID |
| nursing\_home\_cases | Total amount of nursing home COVID cases |
| nursing\_home\_deaths | Total amount of nursing home COVID deaths |
| race\_nm | A specific category of race (American Indian, Asianm Black, Latin, Native Hawaiian, White, Other) |
| census\_id | Primary Key that uniquely identifies census |
| median\_income | Median income of a geographic area by ZIP code |
| population\_count | Given population of a geographic area by ZIP code |
| age\_grp\_id | Primary key that uniquely identifies age\_group, Foreign Key in both covid\_medical\_info and survey |
| age\_grp\_yr | Provides specific year intervals for an age group (1980-1994, 1965-1979,1944-1964,1995-2015, or 1923-1943) |
| age\_grp\_nbr | Provides specific ages for an age group (26-40, 41-55, 56-76, 5-25, or OVER 77) |
| age\_grp\_gen | Provides specific Generations for an age group (Millennial, Gen X, Baby Boomer, Gen Z, The Silent Generation) |
| median\_home\_dwell\_time | Median amount of time spent at home |
| part\_time\_work\_behavior\_devices | A measure of device usage for those working part time |
| full\_time\_work\_behavior\_devices | A measure of device usage for those working full time |
| delivery\_behavior\_devices | A measure of device usage for those performing delivery |
| median\_non\_home\_dwell\_time | The median amount of time for those spent not home |
| candidate\_device\_count | The amount of devices a candidate may have |
| median\_percentage\_time\_home | The median percentage of time for individuals spent at home |

**Relations in the Third Normal Form:**

****