

2024

# Research & Summarize

BY

IVONNE ASPILCUETA

## **US Census Population Data by Geography**

**Data Source:** The data is external since it is taken from the U.S. Census Bureau, a principal agency of the U.S. Federal Statistical System (A government agency). Data is trustworthy since they are dedicated and are the leading provider of quality data about America's people and economy.

**Data Collection:** The data is administrative. They collect the data by providing surveys; depending on the survey, they have many options for participating online, by mail, in-person, or over the phone or by fax.

**Data Contents:** This dataset contains variables like: County, Year, Total population, Gender, and Age. It is a detailed breakdown to make it easy to analyze the US population.

### **Data Limitations:**

The dataset is from the years 2009 to 2017, It is not updated which will not reflect a recent demographic population or trends. This outdated data can affect the accuracy of current population estimations and projections. I see there is more current data, since the last census was in 2020. Possibly pull the data sets to that year of 2020.

Margin of Errors since the data is collected manually through surveys; the risk of manual entry can lead to misclassification, especially in large datasets.

**Data Relevance:** The data is relevant since it provides the data I need to fulfill my hypothesis on regards to the vulnerable population and covers all 50 states, data I need for the scope of the project overview.

## **CDC Influenza Laboratory Tests and Patient Visits data sets**

**Data Source:** The Influenza Division at CDC (Center for Disease Control and Prevention) collects, compiles, and analyzes information on influenza activity year-round in the United States. The U.S. influenza surveillance system is a collaborative effort between CDC and its many partners in state, local, and territorial health departments, public health and clinical laboratories, vital statistics offices, health care providers. I can conclude that Data will be trustworthy since is coming from a government agency and a health public institution rule by the government.

**Data Collection:** The data is collected by Flue View, a weekly influenza surveillance report, and Flue View Interactive, an online application that allows for more in-depth exploration of influenza surveillance data, are updated each week.

### **Data Contents:**

**Influenza Visits:** This dataset contains variables like region listing all 50 states, year from 2010 until 2019, weeks 53 per year, percentage of Unweighted ILI, ages group, the total of ILI, numbers of providers and the total patients.

**Lab Tests:** this dataset contains variables like region listing all 50 states, year from 2010 until 2015, weeks 53 per year, total specimens, percentage positive, and positive tests by influenza type (A, B) and subtypes like (2009 H1N1), (H1), (H3), and H3N2v.

### **Data Limitations:**

Since the data comes from a list of partners and the reports are voluntary, will tend to not cover the entire scope across the United States. The weekly collection of data may change as more data is received and may affect the timeliness of real-time decision-making. Data set columns marked with X which indicates missing data or not reported, that could limit the analysis of the data.

**Data Relevancy:** I can use this information for Influenza Visits and Lab Tests to address the project goal and the objectives. We can learn when the peak of

the flu season starts and for how long, to determine how many personnel they will need to accommodate during the flu season. Also, can help me answer one of my hypotheses as well.

## **Children Flu Shots data set**

**Data Source:** The National Immunization Surveys (NIS) are conducted by NORC at the University of Chicago under the direction of CDC. This data is trustworthy since the collection is supervised by CDC, a government agency.

**Data Collection:** The dataset collection is mainly phone surveys, collected through telephone interviews. The surveys collect data through telephone interviews with parents or guardians in all 50 states. Also, after the parents or guardians' permission, a questionnaire is mailed to the children's vaccination providers to collect some additional information.

**Data Contents:** This dataset contains variables like year for only 2017, age group from 19 months to 35 months, race by two categories: Hispanic or Non-Hispanic, race type, state, financial status, education among other variables.

**Data Limitations:** Since the data is collected by telephone survey and a questionnaire mailed to providers, can lead to inaccurate personal information or not cover all the desired population since cell phone numbers are randomly selected and can be subject to Bias.

**Data Relevancy:** The data is highly relevant for one of hypotheses, If the vulnerable population under 5 years old are low income, then they are a mortality rate group. This dataset is relevant for the project objectives as well.