Noah Curtis, Nam Nguyen, Lang Xiao
INFO 201 Final Project
Data Nutrition Label

Washington State Revenue to Vaccination Rates

This dataset lists the local revenues by county and lists the relevant vaccination rates for the same areas.

- Intended Domain. Economics
- Intended Use. To measure how vaccination rates have affected business revenue performances

The data for this research is collected from two sources. The Centers for Disease Control and Prevention's COVID-19 Vaccination Rates by County, which contains data collected from vaccination providers partnered with the CDC and the Local Retail Revenues by City (Washington)

Motivation

This dataset was created to analyze the relationship between COVID-19 vaccination rates and local business performance compared to pre and post pandemic times. The dataset was created by the authors of this data nutrition label, no funding was used.

Composition

The dataset is composed of numeric data points of quarterly business performances for counties in the state of washington, including sales and vaccination rates there are 2593 total data points in this set. It is supposed to be a comprehensive overview of the performance of businesses throughout the years since pre and post pandemic. The dataset consists of publicly available data from local and national governments. There is no sensitive data in this dataset. This dataset does not identify any specific subgroups.

Collection

This dataset covers the time range from Q4 2020 to 2022 Q3. The COVID-19 Vaccination Rates by County in the U.S. dataset is collected from vaccination providers that partnered with the

CDC. The Local Retail Revenues by City data is collected through tax returns. The dataset from the CDC was a nation wide dataset, so it had to be narrowed down to the relevant location (WA).

Any unethical data collection processes are unknown to the authors.

Data Values

Data Imputation Protocols:

Create columns for year and quarter in the revenue and vaccine data frames. Then remove rows with missing or unknown data and fill missing values with zero.

Data Manipulation Protocols:

We attach county names to the revenue data frame then convert the taxable units of the revenue data frame and remove unused columns.

Missing Data:

No relevant data was lost during our transformation of the data, if there is any missing data then it would be from the data source itself.

Subpopulation Information

Representation:

Washington businesses and residents

Individual Inferences:

Yes, this dataset does not capture the variations at the individual level. Nor does it include more detailed information into every datapoints situation leaving alot to speculation.

Individual Inferences - Mitigation:

Creating subgroups and avoiding using this data for small generalizations.

Potential Issues

Generalized Inferences:

Limited representativeness, this investigation is very surface level and does not represent all the nuances present in these situations, therefore we should not use this dataset to make important decisions regarding the represented peoples.

Generalized Inferences - Mitigation:

Keeping the dataset relevant to the specified region (Washington State) and removing all other datapoints