Data Science as a Field Project Report

Domain: Healthcare

Team Members:

Divya Nallawar

Jeet Choksi

Tanay Shukla

Statement of the question of Interest

Based on different categories of Infants, determine the top 5 states with highest obesity?

Source and short description of the data

The Women, Infant, and Children Participant and Program Characteristics (WIC-PC) dataset contains information on weight status for kids between the ages of 3 months and 4 years. The Data, Trends, and Maps database of the DNPAO uses this information to give data on obesity, nutrition, physical activity, and breastfeeding on a national and state level.

Source of dataset:
https://catalog.data.gov/dataset/nutrition-physical-activity-and-obesity-women-infant-and-child

Possible sources of bias

Sampling Bias:

We can find similar datasets for different age groups or categories by the same publisher but we didn't include all the datasets in our project and focused just on the category of Women, Infant and Child. There may be a possibility that we get different states for highest obesity in different age groups or categories. Following are the different datasets:

- https://catalog.data.gov/dataset/nutrition-physical-activity-and-obesity-behavioral-risk-factor-surveillance-system
- https://catalog.data.gov/dataset/nutrition-physical-activity-and-obesity-american-community-survey

- https://catalog.data.gov/dataset/nutrition-physical-activity-and-obesity-youth-risk-behavior-surveillance-system
- https://catalog.data.gov/dataset/nutrition-physical-activity-and-obesity-national-immunization-survey-breastfeeding
- https://catalog.data.gov/dataset/nutrition-physical-activity-and-obesity-policy-and-environmental-data

Publication Bias:

The data we collected was published by the Centers for Disease Control and Prevention which comes under U.S. Department of Health & Human Services. We may find similar datasets by different publishers as well but it may be having variance in the analysis.

Data Source Bias:

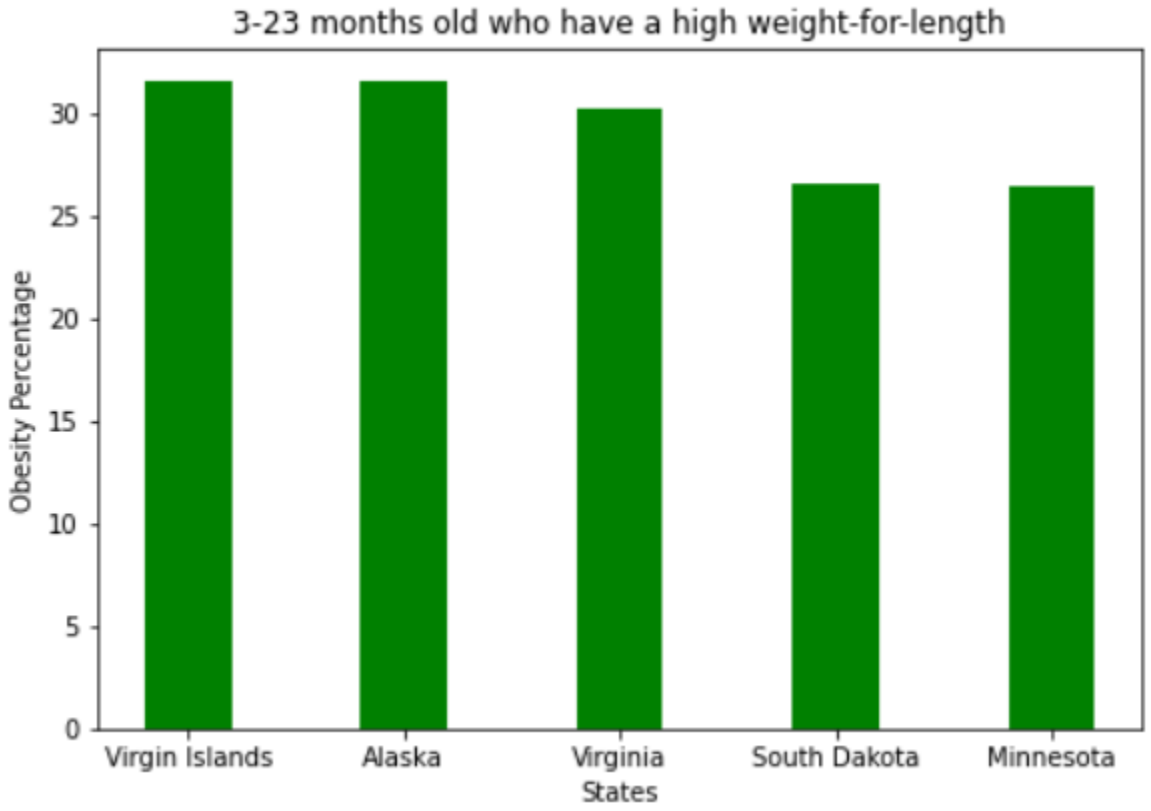
There may be a possibility that the data we collected primarily comes from specific sources, such as hospitals or fitness centers, it may not represent the general population. If the data we used was collected from online sources then it may not capture individuals who do not have internet access.

Visualizations with source code and descriptions

Source Code:
https://github.com/tshukla2001/DTSC-Project/blob/main/DTSC%20Project%20Final.ipynb

Code for the below visualizations for the chosen dataset can be found in the above link.

Visualization 1:

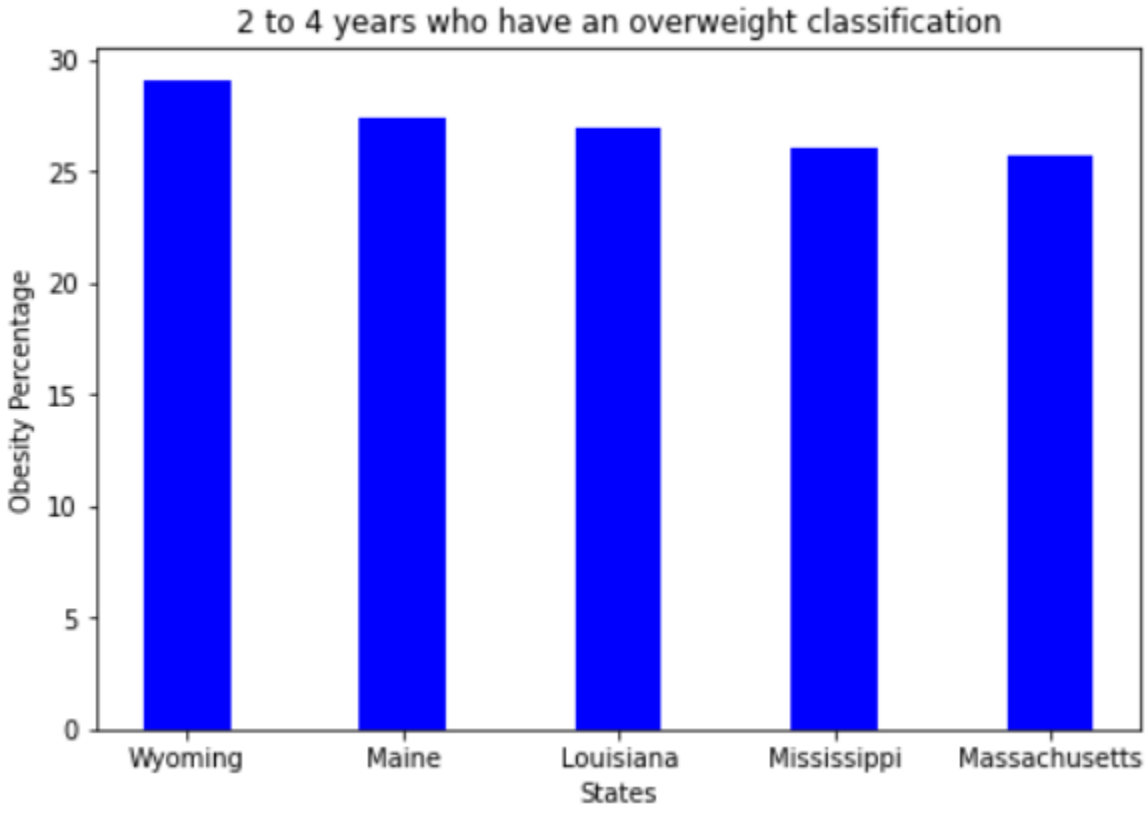


This is a visualization of top 5 states with highest obesity between the age group of 3-23 months old.

Analysis: Between the age group of 3-23 months old, the highest obesity percentage is more than 30% in the Virgin Islands and Alaska states. These states being on the coast, the major source of food is sea. So,

seafood being high in protein could be one of the reasons for obesity. These two states also have a slower pace of life compared to the fast pace life in New York, New Jersey etc. which might be the reason for obesity.

Visualization 2:

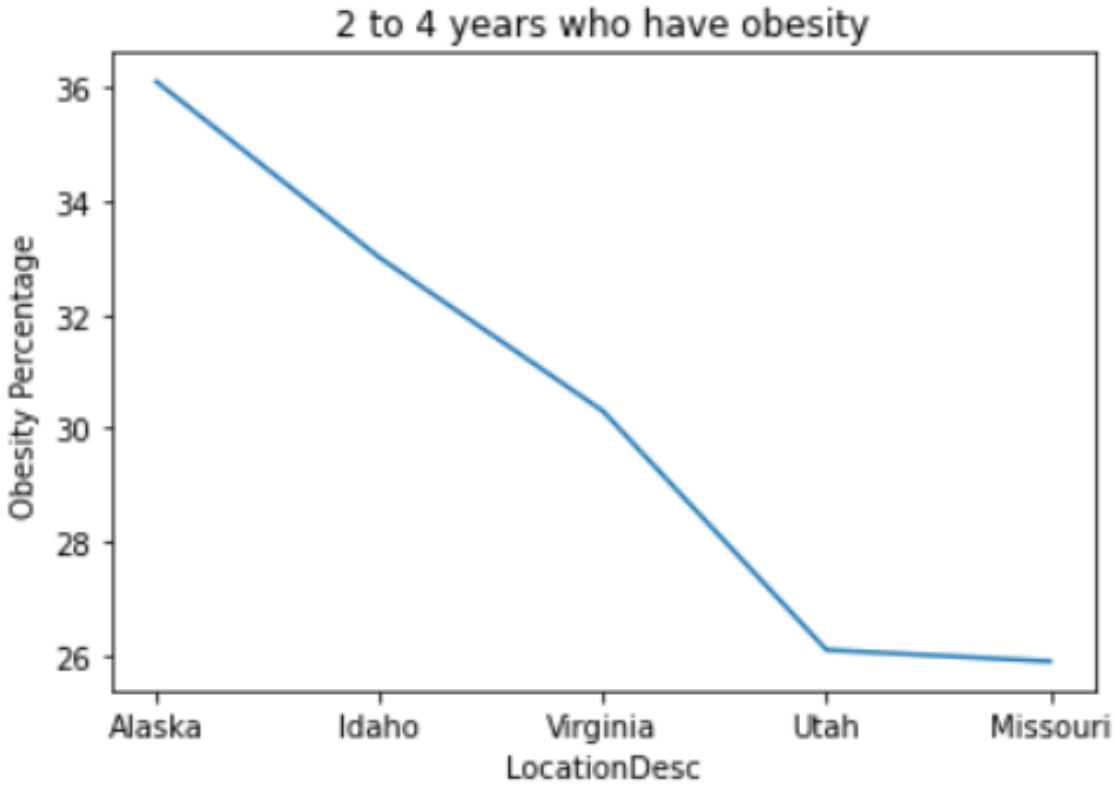


This is a visualization of top 5 states with highest obesity between the age group of 2-4 years old who are overweight.

Analysis: Between the age group of 2-4 years old who have an overweight classification, the highest obesity percentage is approximately 29% in Wyoming. Being a state with less population, the sample size is comparatively low than other states, hence this could be one of the reasons for Wyoming showing the highest overweight group. Wyoming is also a slow paced state where the people are chilled and relaxed, and according to LaramieLive, around 30% of the population does exercise once a month,

this is a case for an adult but if you consider a child, the child might not even exercise. Hence, the highest overweight in this age group.

Visualization 3:



This is a visualization of top 5 states with highest obesity between the age group of 2-4 years old who have obesity.

Analysis: Between the age group of 2-4 years old who have obesity, the lowest obesity percentage is approximately 36% in Alaska. According to the Alaska Department of Health, about 1 out of 3 Alaska children is overweight or obese. Many Alaskans drink too many sugary drinks, eat too few fruits and vegetables, and are not meeting physical activity recommendations. This is the reason for a high obesity percentage in Alaska.