
Predicting Deaths from **COVID-19** via Global Obesity Percentage Data

Perri Solhjou: Wave 13 DS5K



Introduction: Kaggle Dataset

Motivation

Research Question

Data Sources (5)

Exploratory Data Analysis

Linear Regression & Results

Conclusions & Limitations

Next Steps & Recommendations

Introduction

Written by the owner of the Kaggle dataset:

"In this dataset, I have combined data of different types of food, world population obesity and undernourished rate, and global COVID-19 cases counted from around the world [170 countries] in order to learn more about how a healthy eating style could help combat COVID-19 [from 2020-2021].

From the dataset, we can gather information regarding diet patterns from countries with lower COVID mortality rate, and adjust our own diet [and lifestyle] accordingly."



WARNING

**536,000 deaths in the
United States alone***

T March 17, 2021

* NY

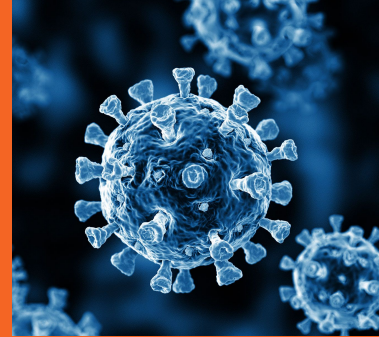
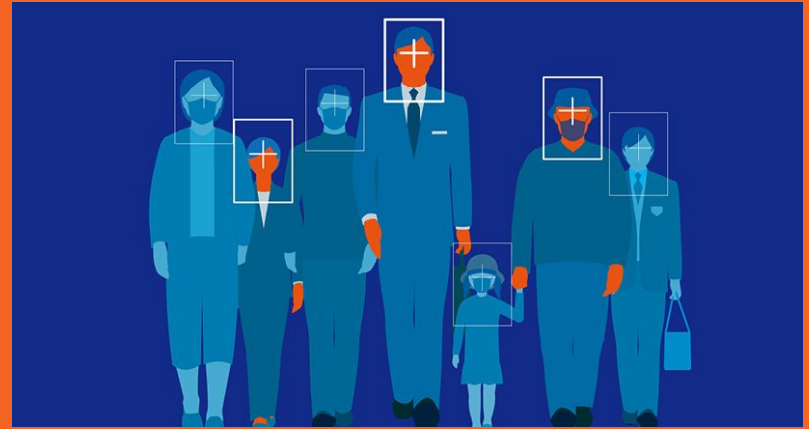
Motivation

Is Obesity a Risk Factor for COVID-19?

According to the CDC, “**overweight** and **obesity** are defined as abnormal or excessive fat accumulation that presents a risk to health. A body mass index (BMI) over 25 is considered **overweight**, and over 30 is **obese**.”**

*Even worse...“having obesity may triple the risk of hospitalization due to a COVID-19 infection.” ****

Risk factors for COVID-19 can cause a need for special care, and can require more hospitalizations for acute and long-term symptoms.



Need to Know

**Obesity nation-wide
is increasing
rapidly...**


***<https://www.cdc.gov/coronavirus/2019-ncov/covid-data/investigations-discovery/assessing-risk-factors.html>
https://www.google.com/search?q=obesity+definition&qq=obesity+definition&ags=chrome_69i57j0i5j0i20j263j0i3.3821j0j7&sourceid=chrome&ie=UTF-8 **

Research Question



—

“Can I *predict* mortality from
COVID-19 based on global
obesity percentage data via
nutritional categories of food?”



5 Datasources- 170 Countries each

- **Fat_Supply_Quantity_Data.csv:** percentage of fat intake from different food groups for 170 different countries.
- **Food_Supply_Quantity_kg_Data.csv:** percentage of food intake(in *kg*) from different food groups for 170 different countries.
- **Food_Supply_kcal_Data.csv:** percentage of energy intake (in *kcal*) from different food groups for 170 different countries.
- **Protein_Supply_Quantity_Data.csv:** percentage of protein intake from different food groups for 170 different countries.
- **Supply_Food_Data_Descriptions.csv:** This dataset is obtained from FAO.org, and is used to show the specific types of food that belongs to each category for the above datasets.



Datasets Chosen

All of these files have obesity, undernourishment and COVID-19 cases (recovered vs. mortality) columns as percentages of the total population.

Predicting: Obesity (global) against undernourished (global) via nutritional food categories.

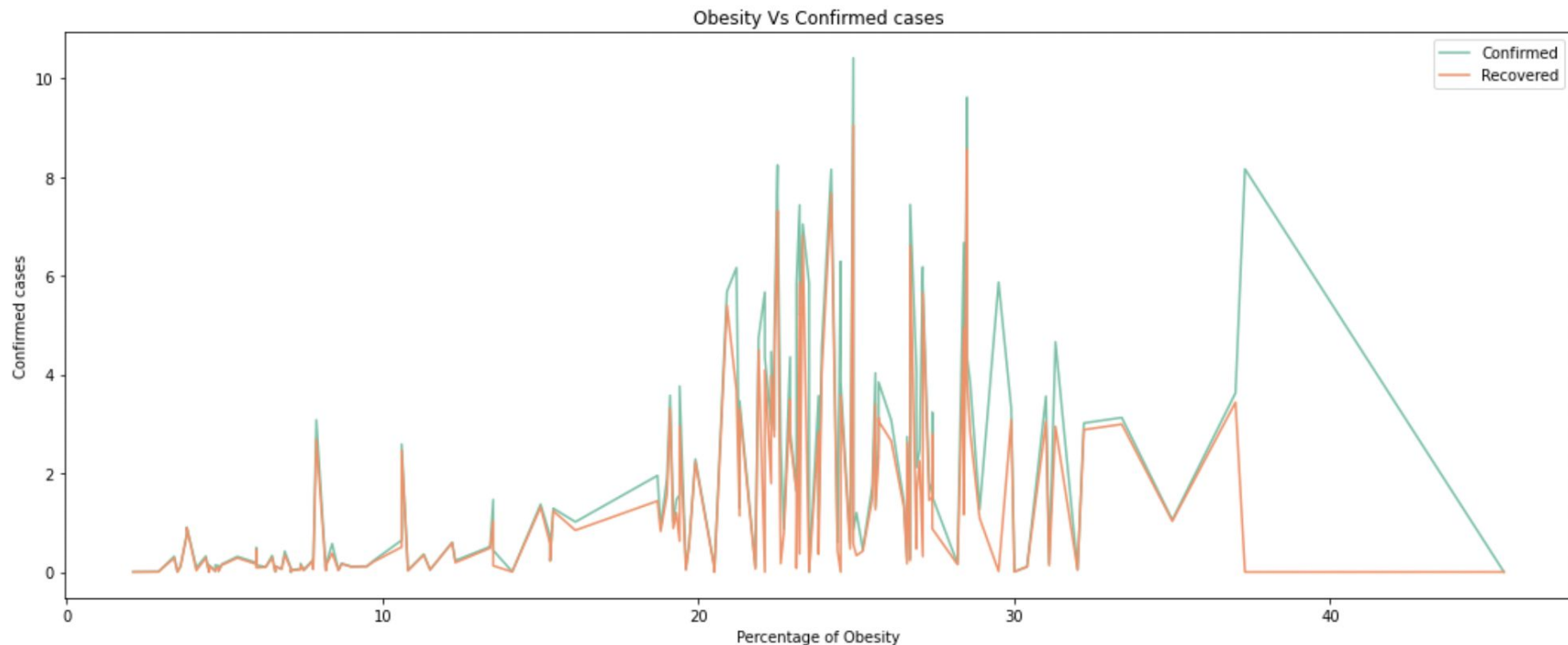
Metric: Recovered vs. Mortality Rates of COVID-19

Emphasis: fat_supply_quantity & supply_food_data_descriptions

The background of the slide is a vibrant green with a pattern of concentric circles and intersecting lines, creating a network-like effect. Scattered across this background are various stylized human figures of different ages, genders, and ethnicities. Some figures are wearing face masks, and one person is in a wheelchair. The overall theme suggests a diverse and interconnected community.

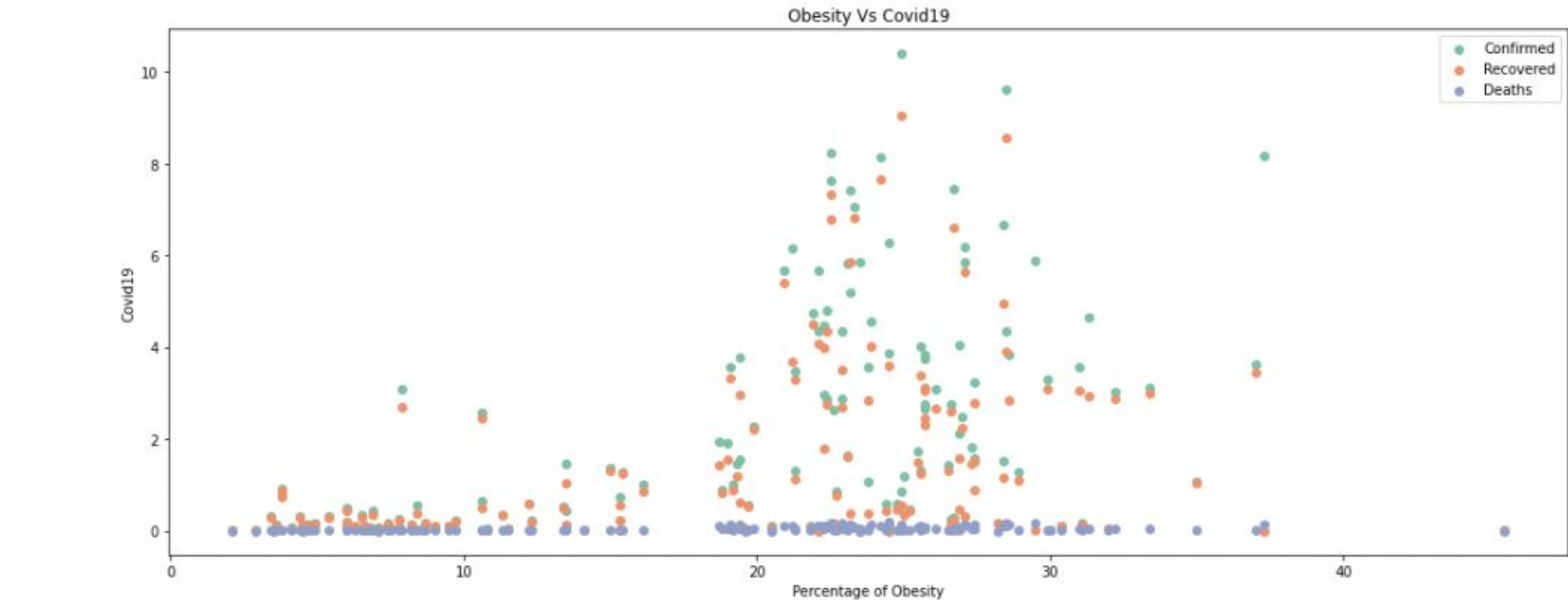
Exploratory Data Analysis

Looking into Relationship between Obesity and Mortality from COVID-19: EDA Example 1: Confirmed Case Deaths vs Recovered Cases in 170 countries



A Closer Look: Looking into Relationship between Obesity and Mortality from COVID-19:

EDA Example 2: Confirmed Cases, Confirmed Deaths vs Recovered Cases per Top 10 out of 170 Highest Obesity Countries



Looking into Food Categories per 170 Countries

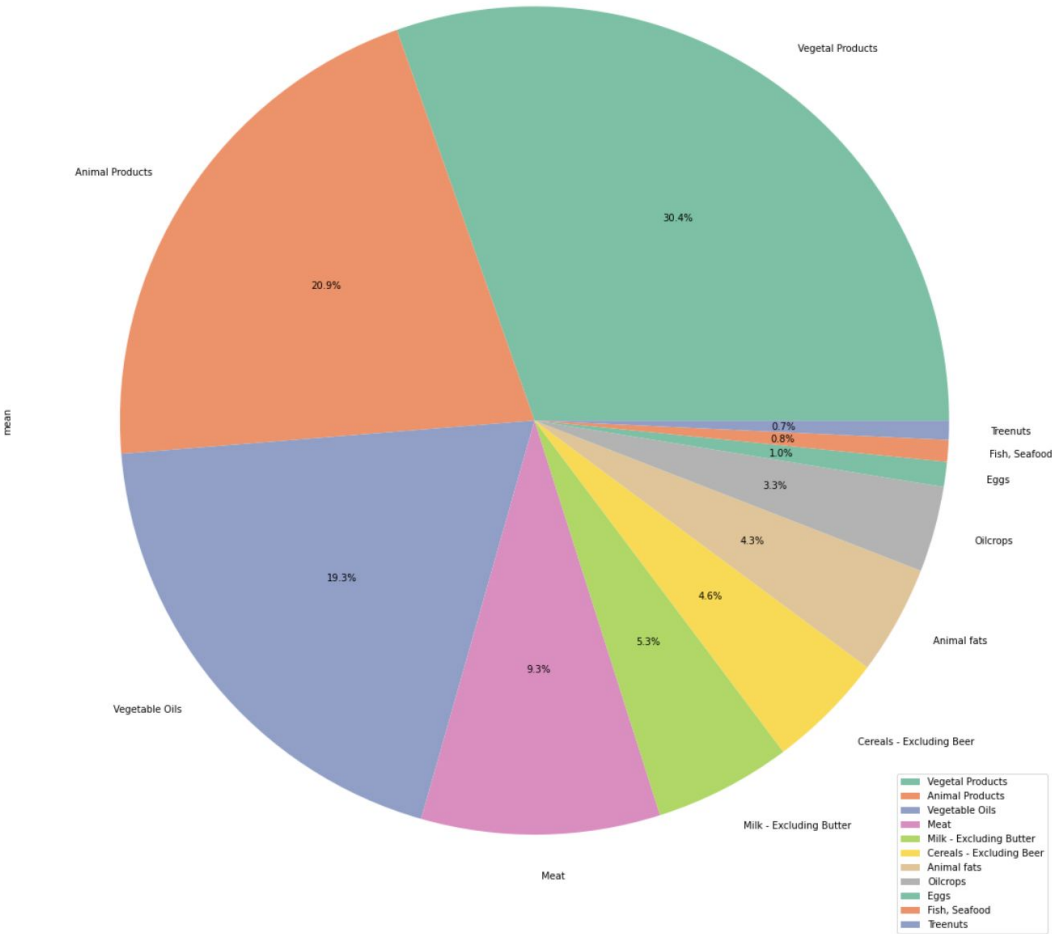
EDA Example 3: Pie Chart

Data Source:

- Supply_Food_Data_Descriptions.csv: This dataset is obtained from FAO.org, and is used to show the specific types of food that belongs to each category for the above datasets.

Observations:

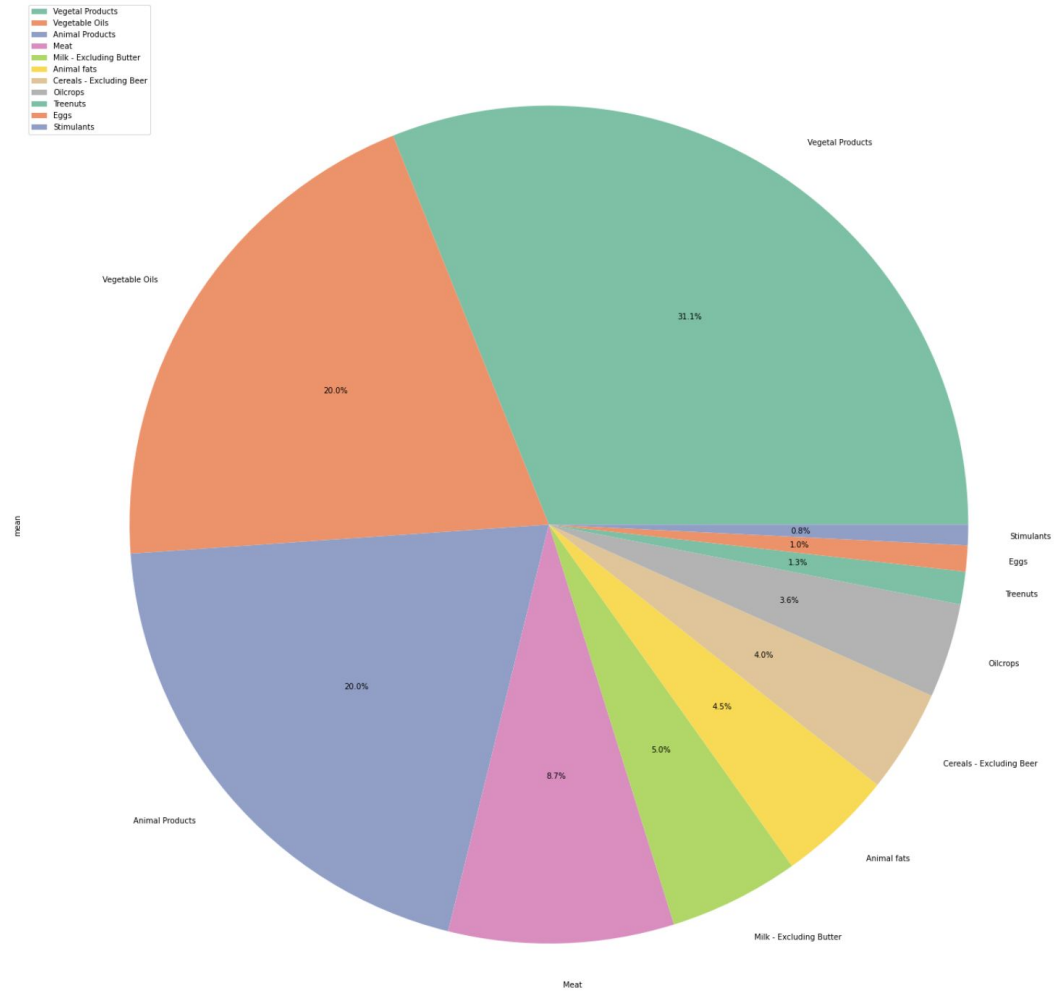
- **Large in Vegetal Products, Vegetal Oils, & Animal Products** (potentially starches and unhealthy foods)



Looking into Food Categories per 10 / 170 Countries with Highest Obesity

EDA Example 4: Pie Chart

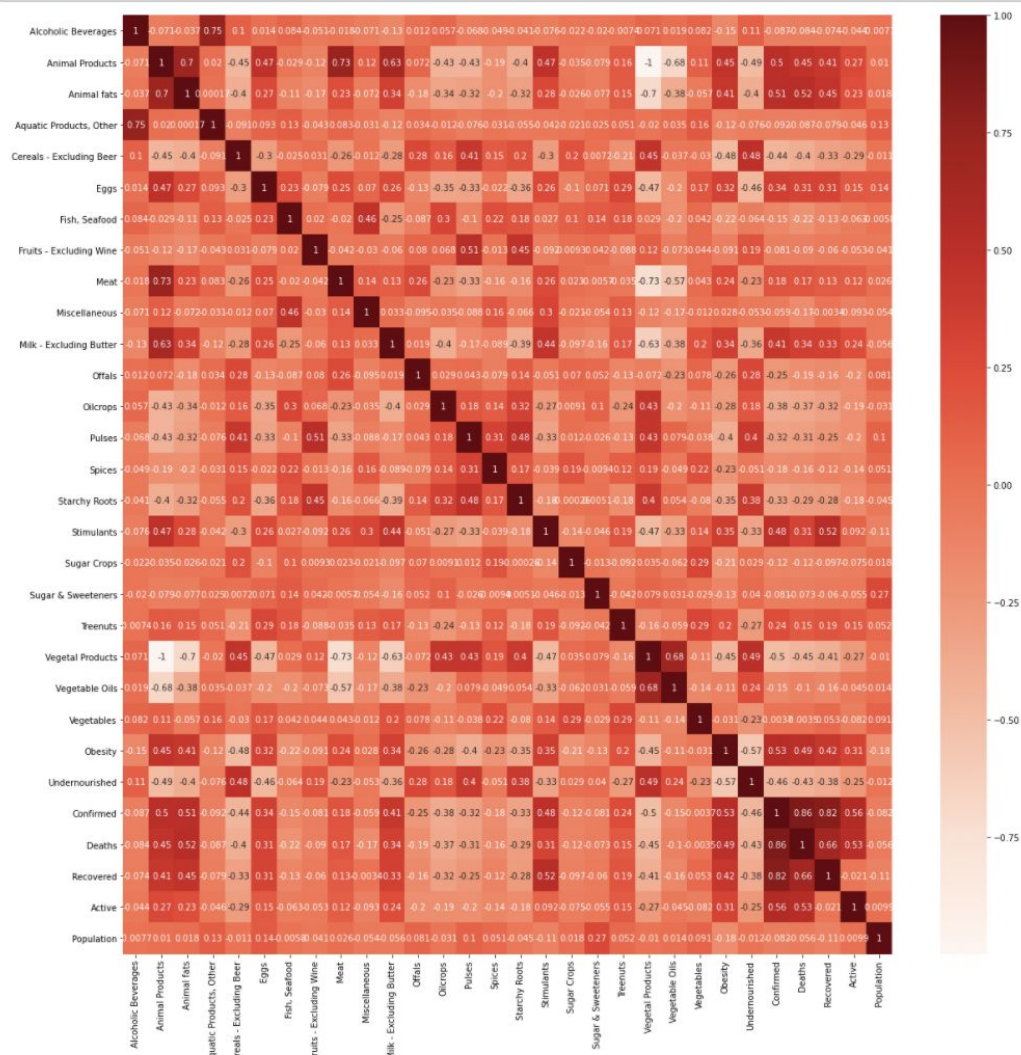
- Countries with higher obesity percentage rates seem to consumer more starchy vegetables than less obesity-high nations on average.
- **Higher in Vegetal Products and Vegetal Oils** than global food consumption data



EDA Example 5: Visual Heatmap

(Assumption: obesity and unhealthy foods = higher mortality rates of COVID-19)

- **Animal Fats + Starchy Roots = Higher Correlation Value with COVID-19 confirmed cases and mortality cases**
- Value of 1 = Higher Correlation Value
- **In this supervised model, obesity, undernourishment and COVID-19 (confirmed, deaths, and recovered) cases are all being evaluated in the visual columns.**
- I used data from food intake by 170 countries along with mortality data.



Model Chosen - Linear Regression - Results

Simple to explain + Highly interpretable + Model training and prediction are fast + No tuning is required (excluding regularization) + No Feature Scaling

feature selection: Four Feature_names = ['animal_products', 'cereals_excluding_beer', 'obesity', 'vegetal_products']

Results:

- **MSE:** 0.0014986303806756888

- **RMSE:** 0.03871214771458294

Conclusions & Limitations



Variable Outliers that could have influenced the **RMSE/MSE** of obesity & COVID-19 mortality **BESIDES** data in food_fat_quantity and supply_food_quantity:

- Neighborhood design
- Access to healthy, affordable foods and beverages
- Access to safe and convenient places for physical activity

conclusions & Limitations

Racial and Ethnic Disparities in Obesity based on the social determinants of health:

- We need to address the social determinants of health such as poverty, education, and housing to remove barriers to healthcare access and funding.
- At the Policy and Governmental Level nationally, the United States needs to ensure that obesity prevention and management starts early. The US government needs to ensure that and each person has access to good nutrition, and has access to safe places to work out/stay active.

Obesity is a complex disease with numerous contributing elements.



Next Steps

Try different models (Logistic Regression, KNN, and Random Forests).

Research the other 3 datasets in more depth in terms of the relationships between them.

Add more features to my feature selection.

Recommendations

***Aim* to eat healthy, maintain a healthy lifestyle, and eat less starchy/fatty foods, eat less vegetable oils, and have more lower-cholesterol sources of protein.**

Take precaution and social distance
(minimum of 6 feet apart), wash hands, and wear a mask!

Better understand how an underlying medical condition could put you at risk to COVID-19

Anticipate medical treatment that you might need if you get sick & clear it with your health insurance provider

Thank you!

Perri Solhjou: Wave 13 DS5K

Q/A

Perri Solhjou: Wave 13 DS5K
