Uninsured in New Jersey

Team Pushing P Breakers:
Sharif Rakhimov, Luis Rivera, Marjea Mckoy, Ahad Hussain, Parth Patel

Objective

- Predict the number of uninsured people in New Jersey

Overview

- Questions and Hypothesis
- ETL
- Data Platform
- Spatial Area Interpolation
- Dash Dashboard

Questions

1. Which city has the highest uninsurance rate?
2. How many counties have a population of uninsured people of 8% or more?
3. How does income change the number of people being uninsured?
4. Is there a race that has a higher uninsured population?
5. Which sex has a higher population of uninsured people?

Questions (cont.)

6. What age cohort has the most uninsured people?

7. Does employment affect the amount of the uninsured population?

8. Hypothesis test: the larger the population, the higher the uninsured rate.

9. What area and demographic group can Prudential Financial target for insurance sales in NJ?

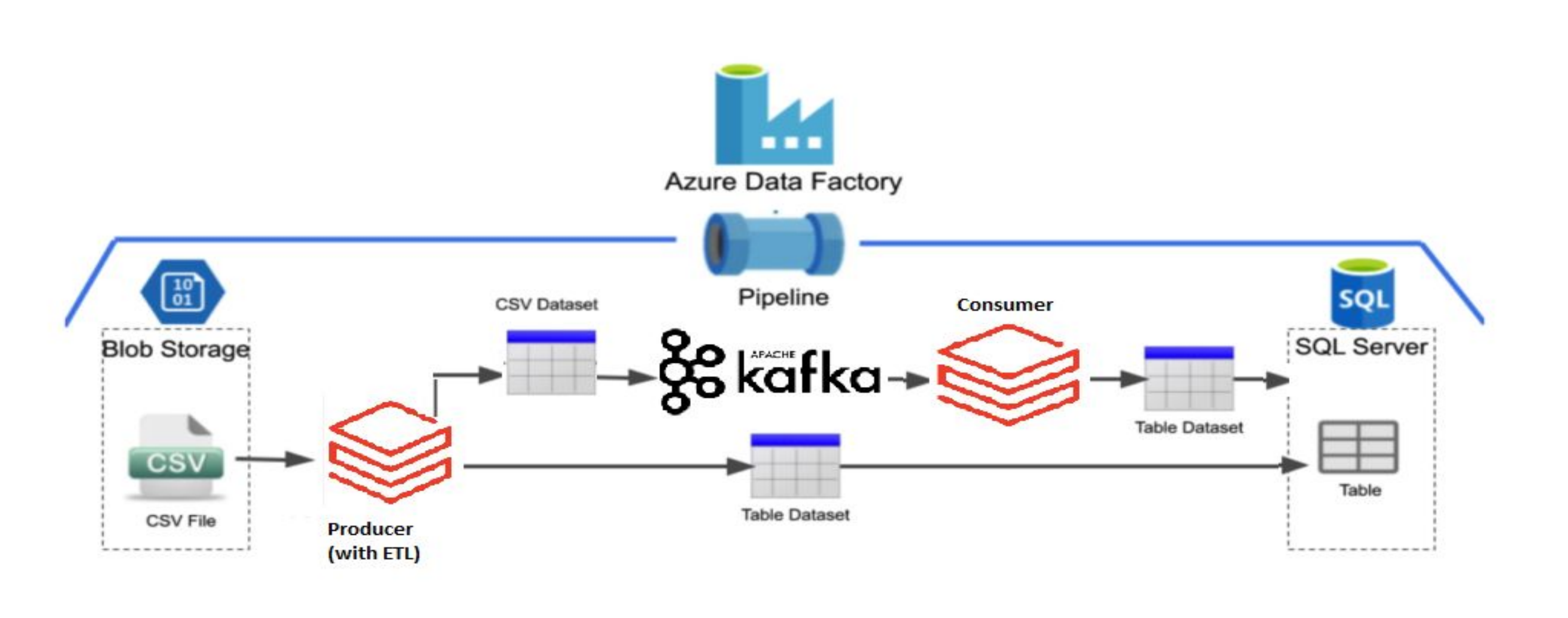
Datasets

1. Small Area Health Insurance Estimates 2019 (SAHIE)
2. NJ Uninsured
3. NJ Unemployed
4. NJ Income
5. Cartographic Boundary Files
6. Coverage for the Household Population by States
7. Unemployment by Counties
8. Census Tract in NJ

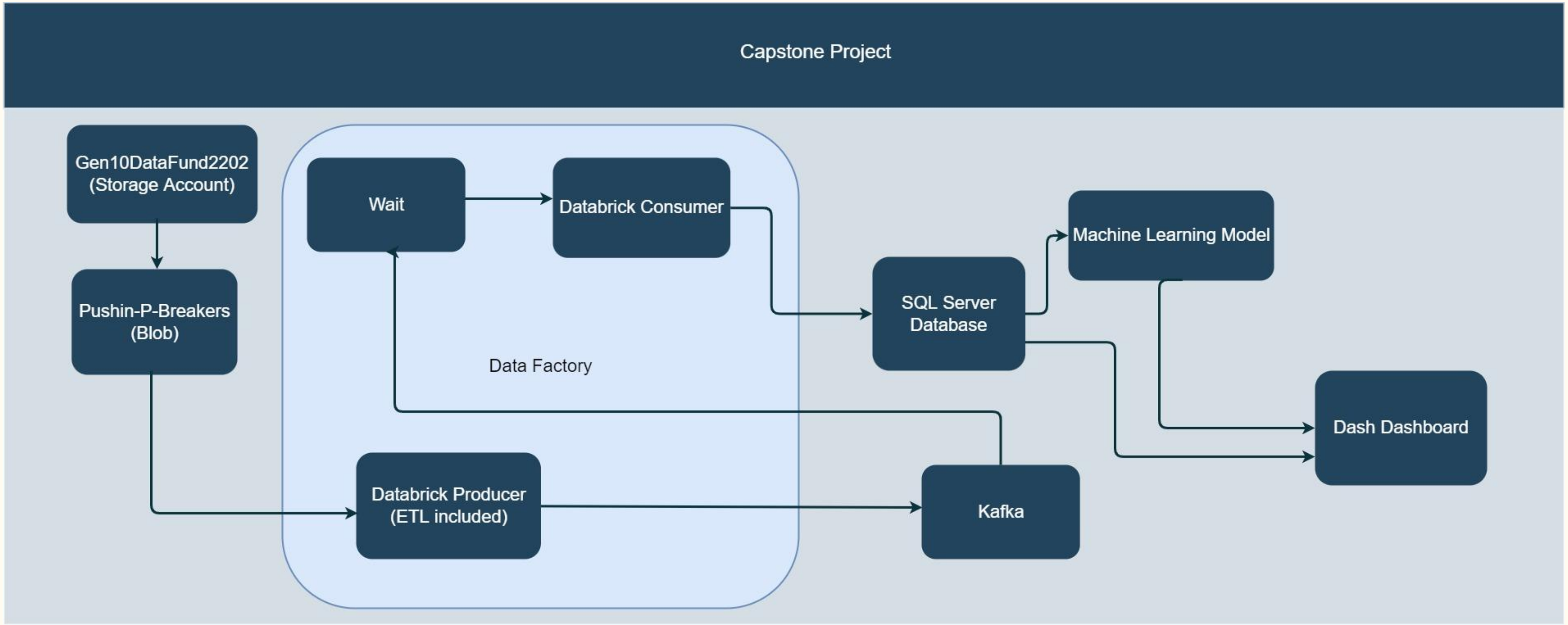
ETL

- Cleaned the datasets and select the columns needed
- Joined all the tables together based on State, County, and City
- Created three separate data frames
  - NJ Cities
  - NJ County
  - Census Tracts
- Using Kafka, each data frame would write into an SQL database

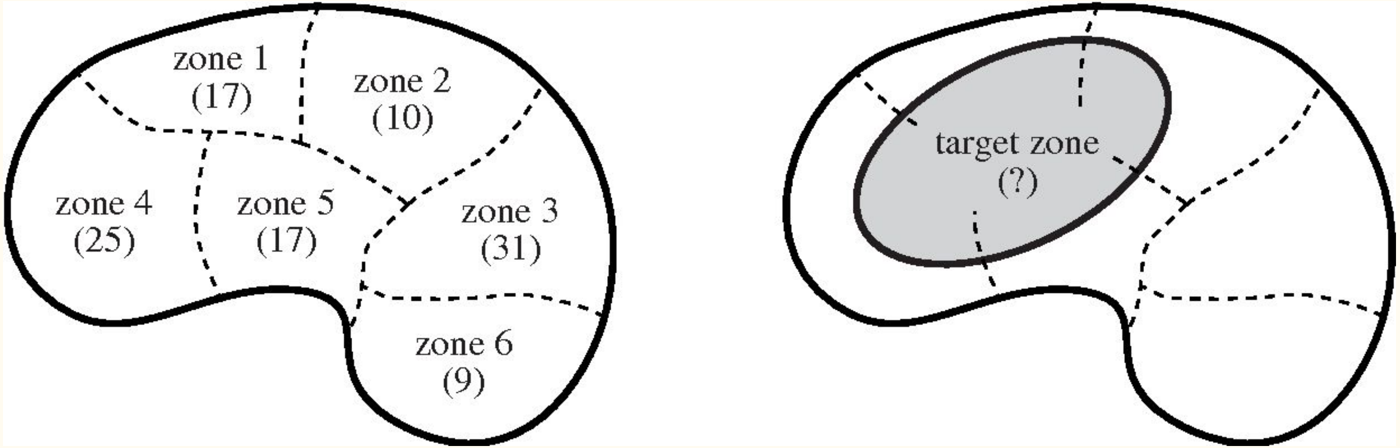
Data Factory



Data Platform



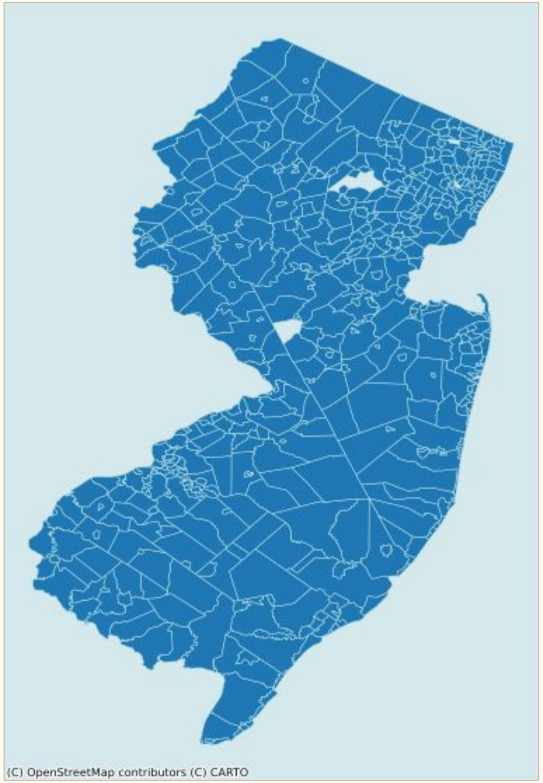
Machine Learning: Spatial Area Interpolation



Predict Target Zone using its overlap with Zones with Known Values

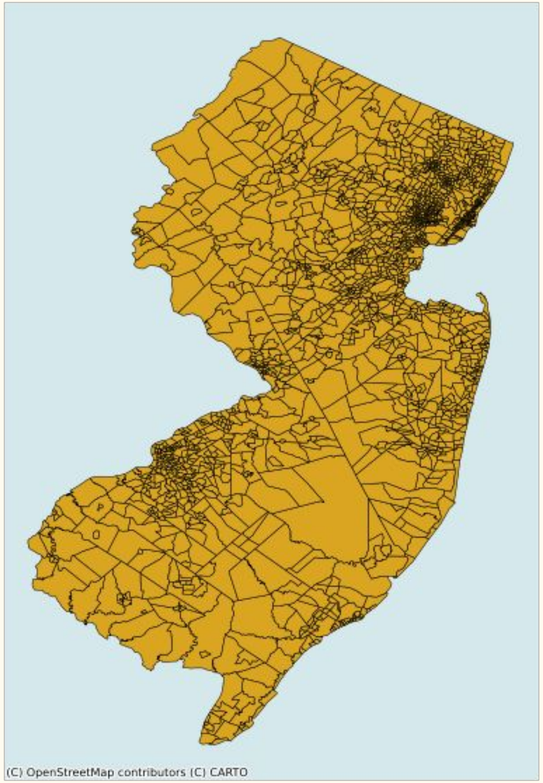
Model Goal

Cities in New Jersey



(C) OpenStreetMap contributors (C) CARTO

Census Tracts in New Jersey



(C) OpenStreetMap contributors (C) CARTO

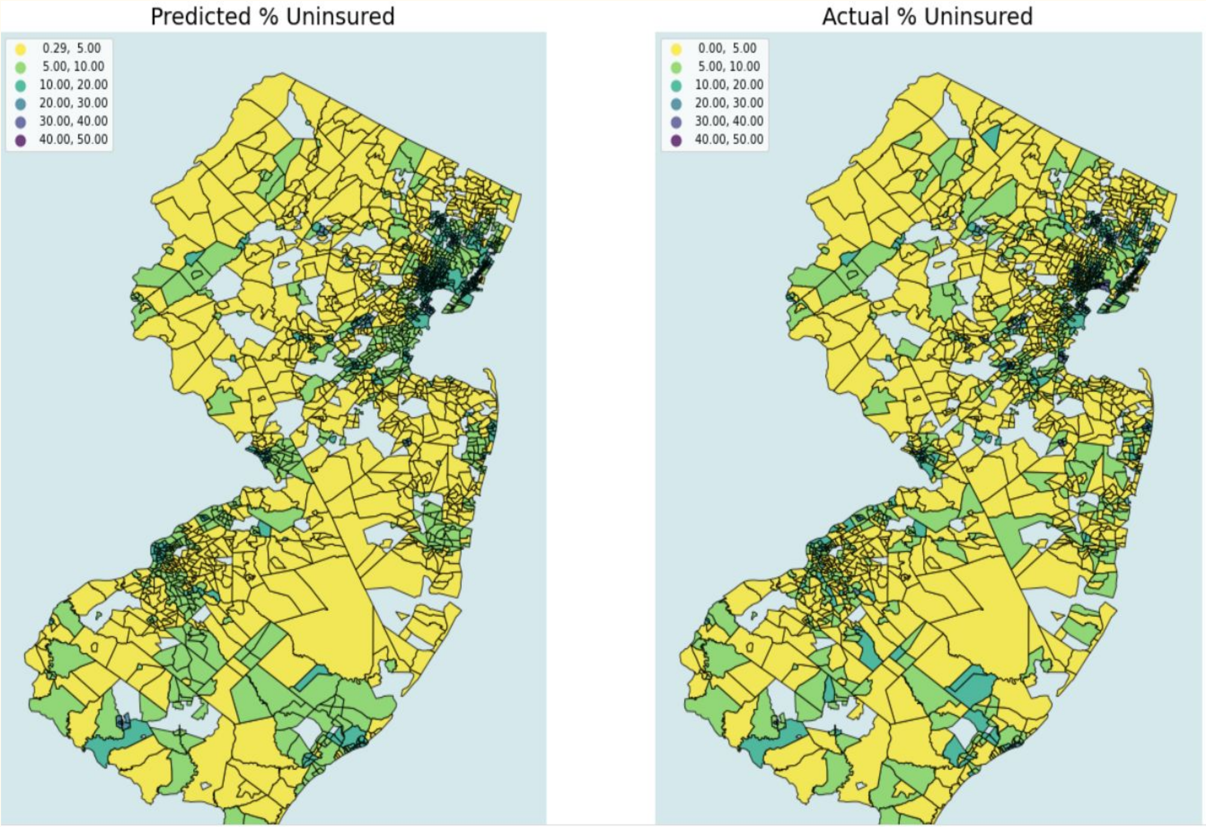
Model Parameters

- Source DataFrame - geometries with known values
- Target DataFrame - contains target geometries
- Intensive Variables List - independent of the size of the system (Population Density, Concentration, Melting Point, etc.)
- Extensive Variables List - dependent on the size of the system (Population Count, Mass, Volume, etc.)

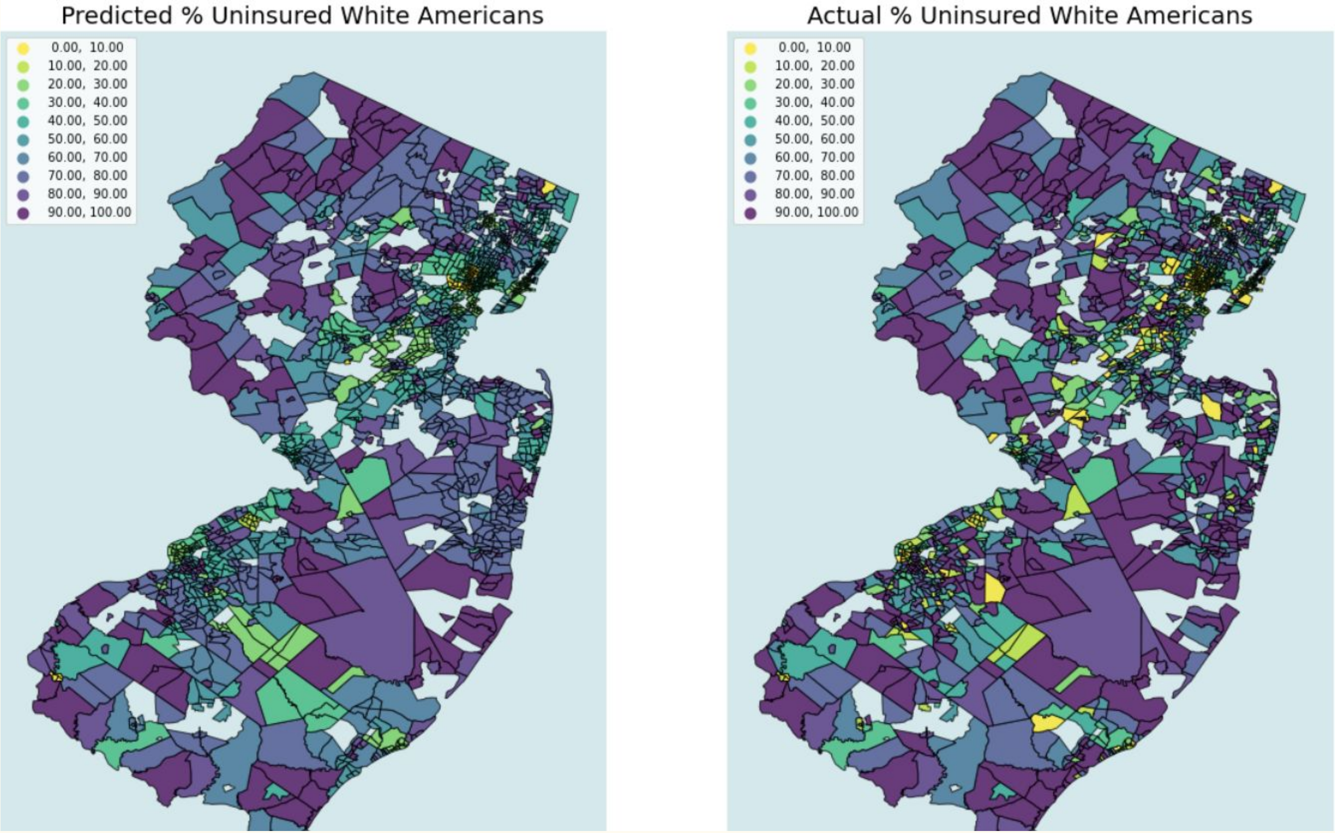
Model Tuning

- Model has no hyperparameters, no score, and no accuracy metrics of any kind
- Of the 2010 Census Tracts, only about 1850 of them contained healthcare data
- In order to improve the accuracy of the model, the empty census tracts were removed from our analysis

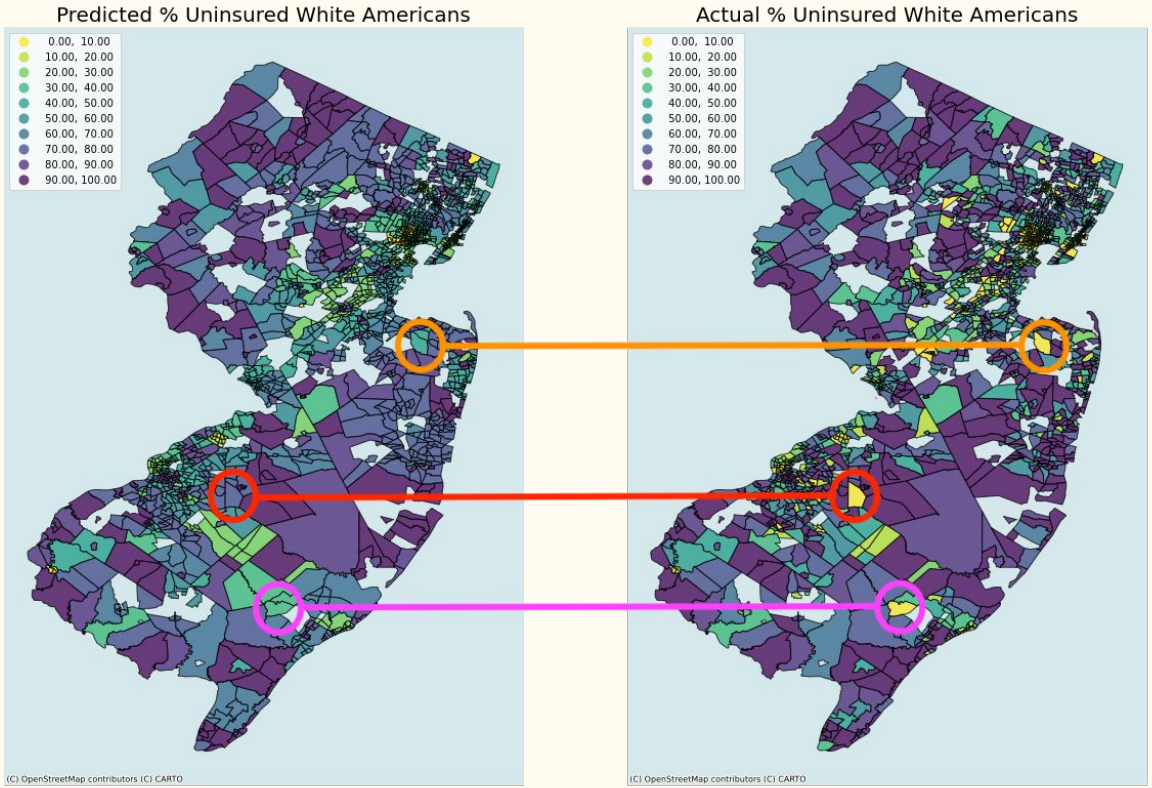
Predicted % Uninsured vs. Actual in NJ Census Tracts



Predicted % Uninsured, White vs. Actual in NJ Census Tracts



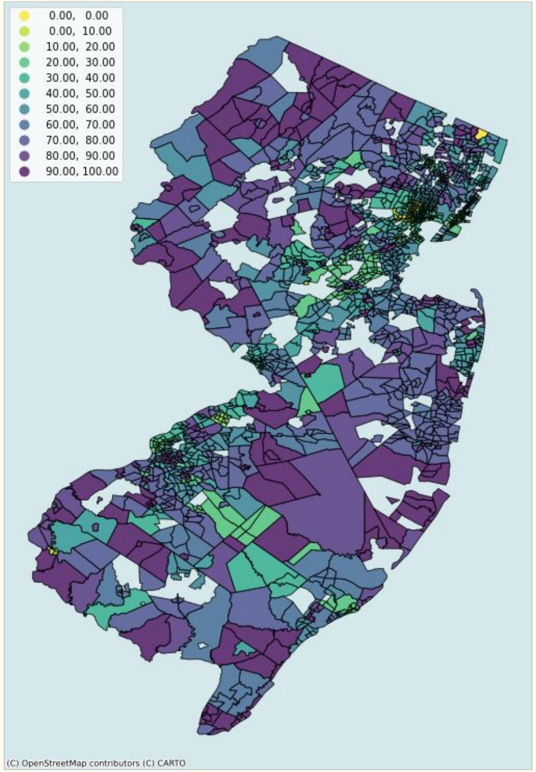
Disadvantage: Model struggles with Local Outliers


Predicted % Uninsured White Americans
Actual % Uninsured White Americans
0.00, 10.00
10.00, 20.00
20.00, 30.00
30.00, 40.00
40.00, 50.00
50.00, 60.00
60.00, 70.00
70.00, 80.00
80.00, 90.00
90.00, 100.00
(C) OpenStreetMap contributors (C) CARTO
(C) OpenStreetMap contributors (C) CARTO

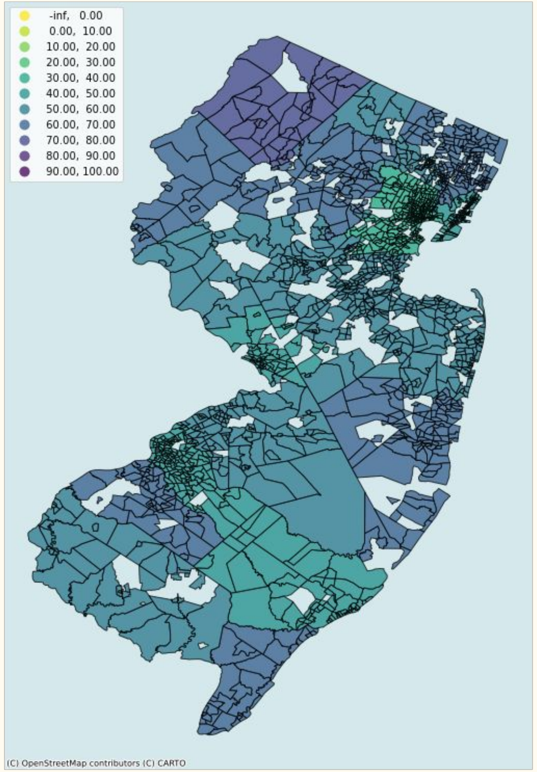
Disadvantage: Modifiable Areal Unit Problem (MAUP)

- MAUP: data tabulated for different spatial scale levels or according to different zonal systems for the same region will not provide consistent analysis results.
- In other words, in the absence of known values, what geography level (state, county, city) do we choose to make our predictions at the census tract level?

MAUP Illustrated



Predicted % Uninsured (City Data)



Predicted % Uninsured (County Data)

Recommendations

- There are two approaches to solve the issues of Local Outliers and the MAUP:
- Approach 1: Dasymetric Mapping
  - Use Geographic Distribution data to remove areas where Uninsured people do not live and run the spatial interpolation model
- Approach 2: Model-based Interpolation
  - Use a spatial model (such as a regression) to model the distribution of Uninsured people within each City

Dash Dashboard

Theme: black background, bright colors

How it works:

Broken down by **County**:

Page 1:

- Total number of uninsured
- Rate of uninsured by age
- Rate of uninsured by race
- Top 10 uninsured cities
- Rate of insured/uninsured
- Rate of employed/unemployed

Broken down by **Census Geographical Tract**:

Page 2:

- % Uninsured dropdown (filter by race, age,sex)
- % Uninsured Map divided by county results
- % Uninsured Map showing county city results
- Predicted vs. Actual results of ML

Background, Team Name:

Page 3:

- About Us