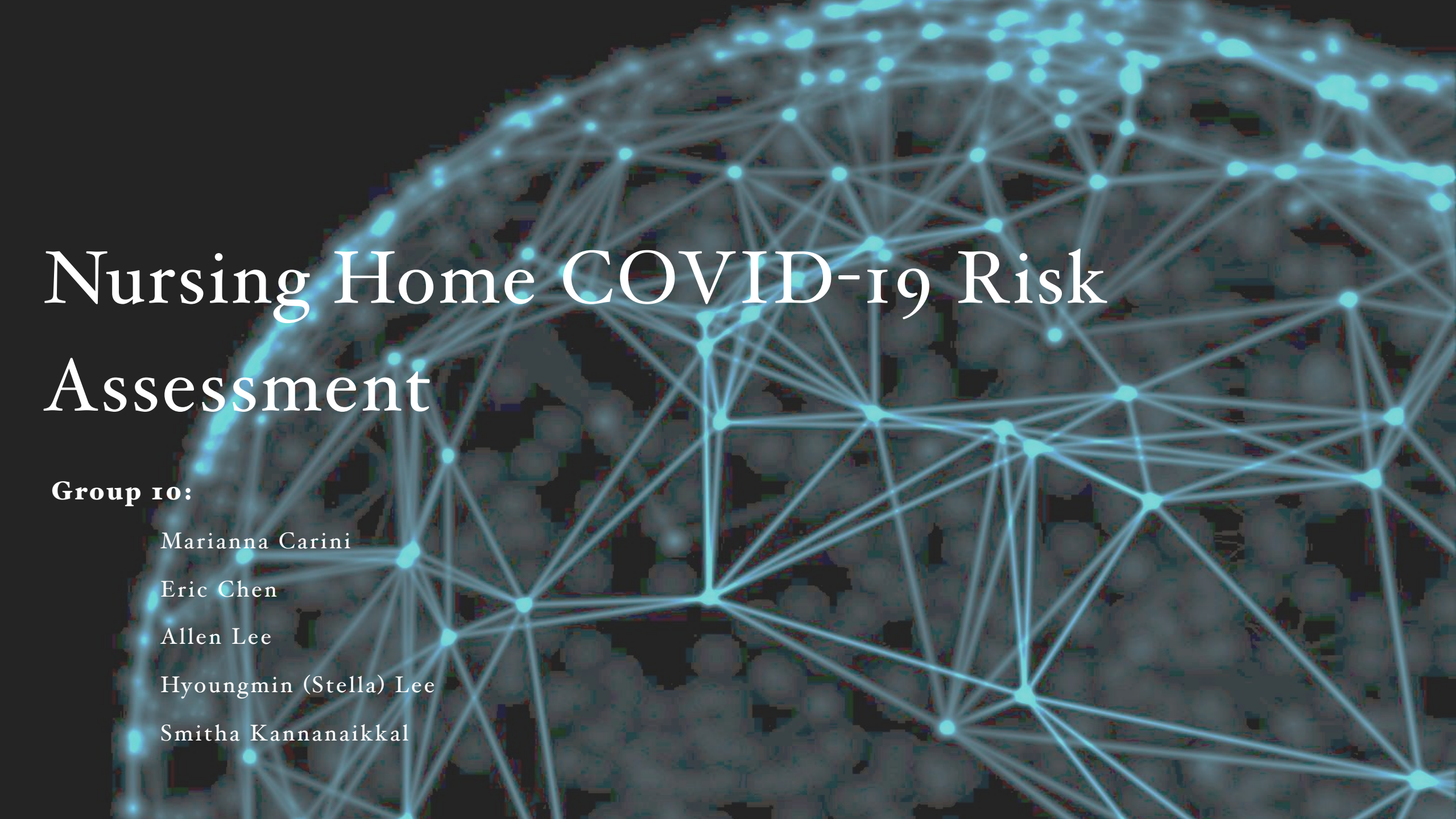


Nursing Home COVID-19 Risk Assessment

Group 10:

Marianna Carini

Eric Chen

Allen Lee

Hyoungmin (Stella) Lee

Smitha Kannanaikkal

Agenda

Problem statement

Data cleaning

K-means cluster model

Data sources

Data overview

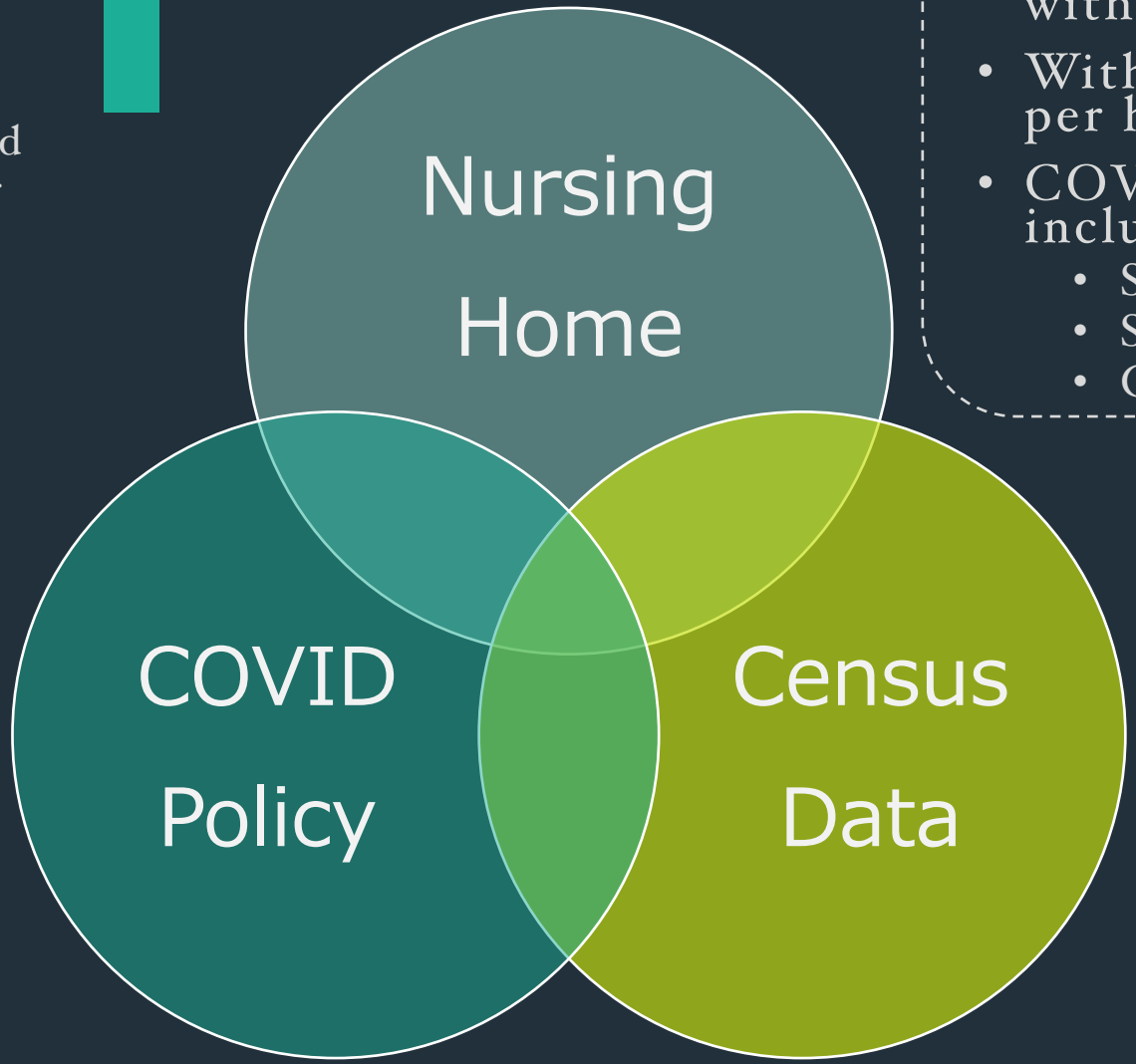
Recommendations

Problem Statement

Study nursing home COVID data to generate insights and recommendations for allocating resources in order to mitigate future risks.

Data Sources

Utilized multiple APIs and data sources to create our dataset



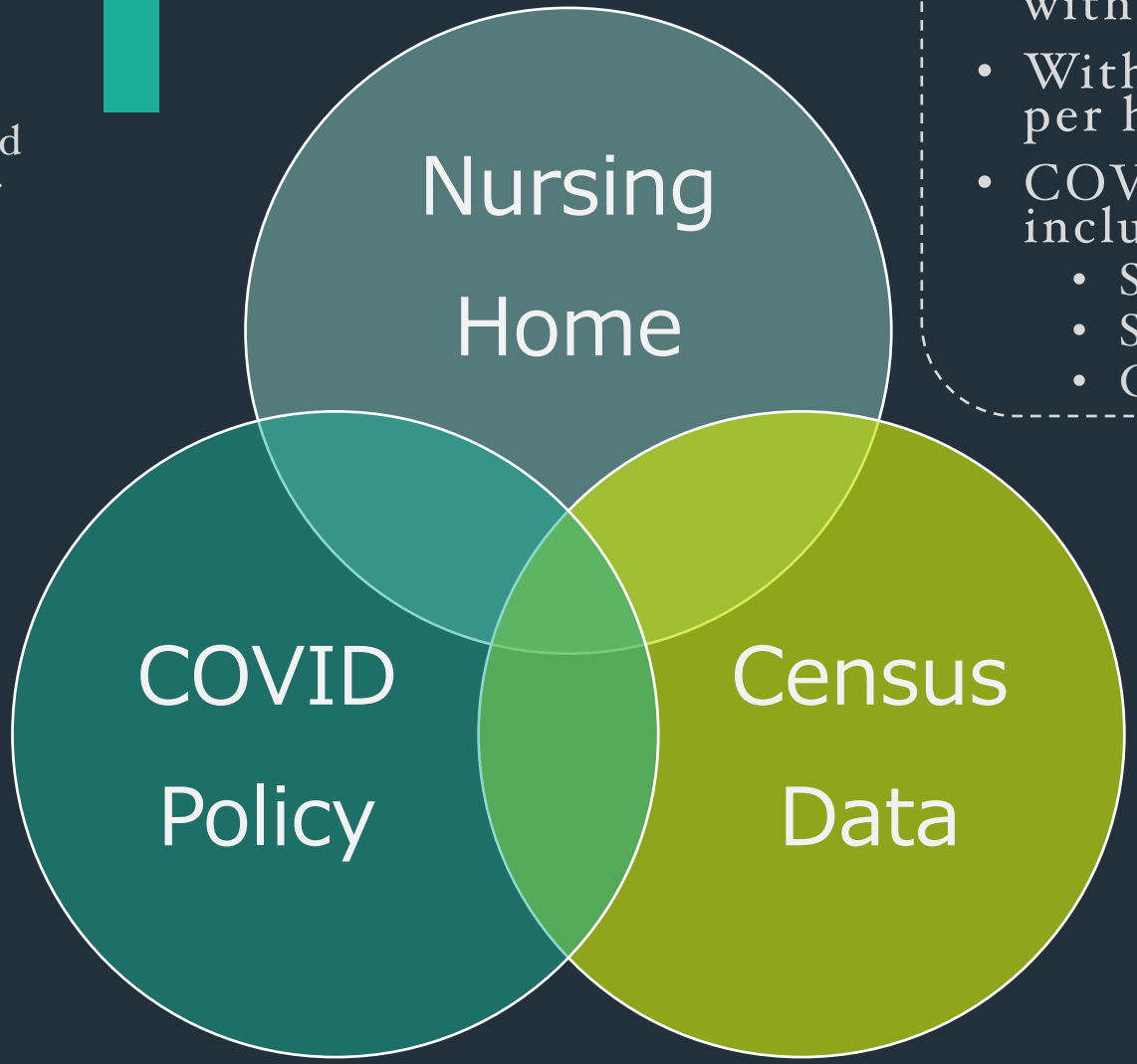
- Connect with Socrata API
- Limit to 100,000 rows without token
- With token, limit 1000 calls per hour
- COVID nursing home report includes
  - Supply levels
  - Staff support
  - Cases and deaths in home

- Connect with Government Health Data API
- Use regex to create dataframe

- Connect with Census API
- Limit to 50 variables per report
- Without key, limit to 50 calls per day
- Built customized query
- Pulled 2019 census report to include county demographics in nursing home dataset

Data Sources

Utilized multiple APIs and data sources to create our dataset



- Connect with Socrata API
- Limit to 100,000 rows without token
- With token, limit 1000 calls per hour
- COVID nursing home report includes
  - Supply levels
  - Staff support
  - Cases and deaths in home

- Co[illegible] wi[illegible] Go[illegible]ealth Data [illegible]
- Us[illegible]ate data[illegible]me

- Limit 100 rows w/o account
- No option to create account
- Manually loaded data

- Connect with Census API
- Limit to 50 variables per report
- Without key, limit to 50 calls per day
- Built customized query
- Pulled 2019 census report to include county demographics in nursing home dataset

Data Cleaning

Everyone's favorite step!

Nursing Home

- Removed attributes which were 95%+ incomplete
- Errors in county name
- Created reference table to connect county state ID to census FIPS code
- Impute missing values for POC machine

Census

- Discovered -9999999.9 represents N/A

COVID Policy

- Extracted county-level policies
- Standardized policy types, ended with 13
- Removed policy types that were not implemented in nursing home locations
- Grouped by FIPS code
- Created binary variable for each policy type to indicate whether that policy type has been implemented

Overview of cleaned data

Mix of nursing home data, COVID policies, county demographics, etc.

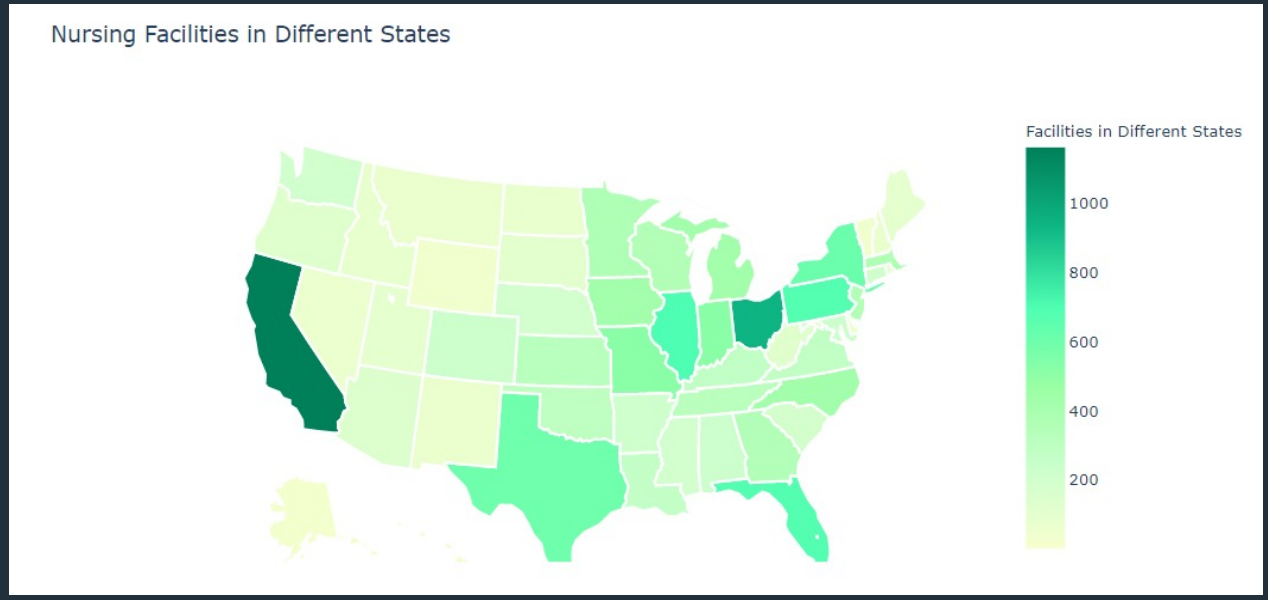
- Shape: 14k+ rows, 113 columns

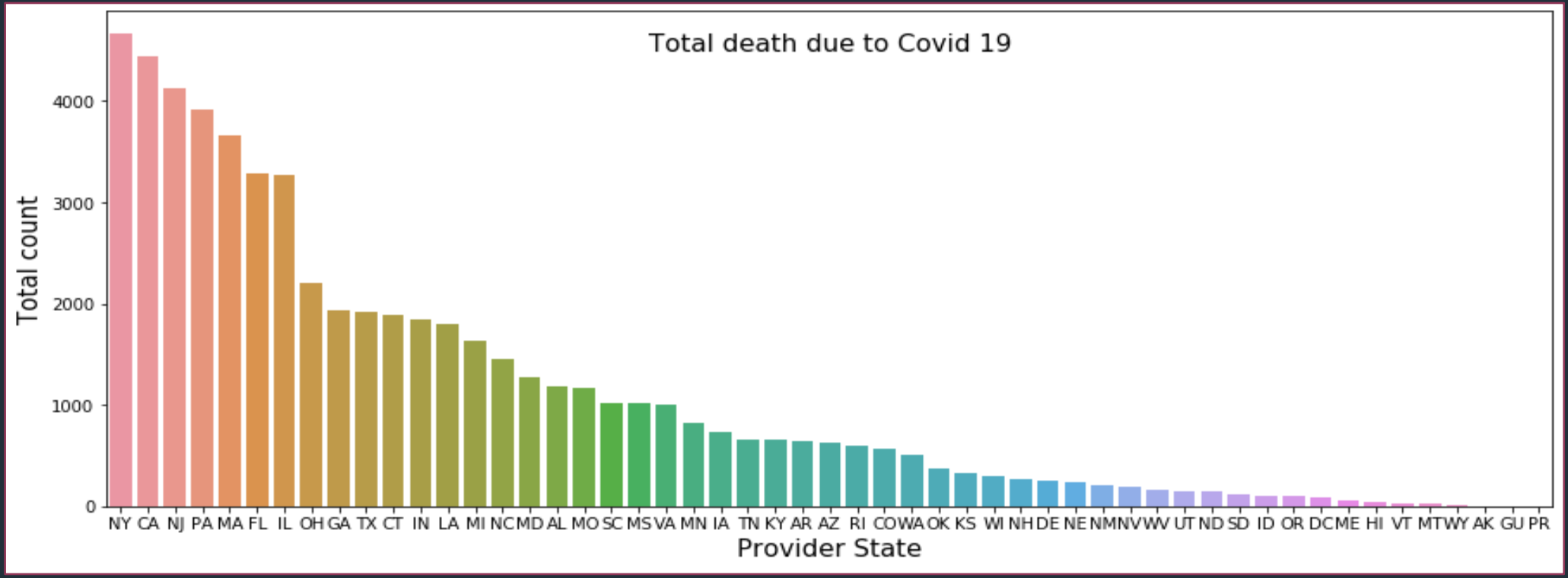
Feature	Example	Non-Null Count	Dtype
Federal Provider Number	11111	14518 non-null	object
Week Ending	5/24/20	14518 non-null	object
Tested Asymptomatic Residents Facility-Wide After a New Case	1	14518 non-null	float64
Shortage of Nursing Staff	0	14518 non-null	float64
One-Week Supply of N95 Masks	1	14518 non-null	float64
Any Current Supply of Hand Sanitizer	1	14518 non-null	float64
Residents Total Confirmed COVID-19	9	14518 non-null	float64
Residents Total COVID-19 Deaths	1	14518 non-null	float64
Occupied Beds	73	14518 non-null	float64
Overall Rating	4.5	14518 non-null	float64
Total Amount of Fines in Dollars	100	14518 non-null	int64
…	…	…	…
avg_household_size	2.5	10122 non-null	float64
median_household_income_dollars	45,000	10122 non-null	float64
civilian_pop_no_health_insurance	12,930	10122 non-null	float64
total_housing_units	1,200,429	10122 non-null	float64
male_pop	43,059	10122 non-null	float64
female_pop	46,023	10122 non-null	float64
pop_65_and_over	9,573	10122 non-null	float64
…	…	…	…
Policy Type: Food and drink	1	10122 non-null	uint8
Policy Type: Mask requirement	0	10122 non-null	uint8

	Federal Provider Number	Week Ending	lab_state_hlth	lab_prvt	lab_othr	tstd_r_sngOsym	tstd_ar_u	tstd_ar_f	t_ar_exp	tstd_sbgrp_r	...	percent_pop_white_y
14513	676213	07/12/2020	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	...	82.1
14514	676266	06/28/2020	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	...	75.7
14515	676306	06/28/2020	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	...	63.6
14516	676453	05/24/2020	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	...	62.1

EDA Visualizations

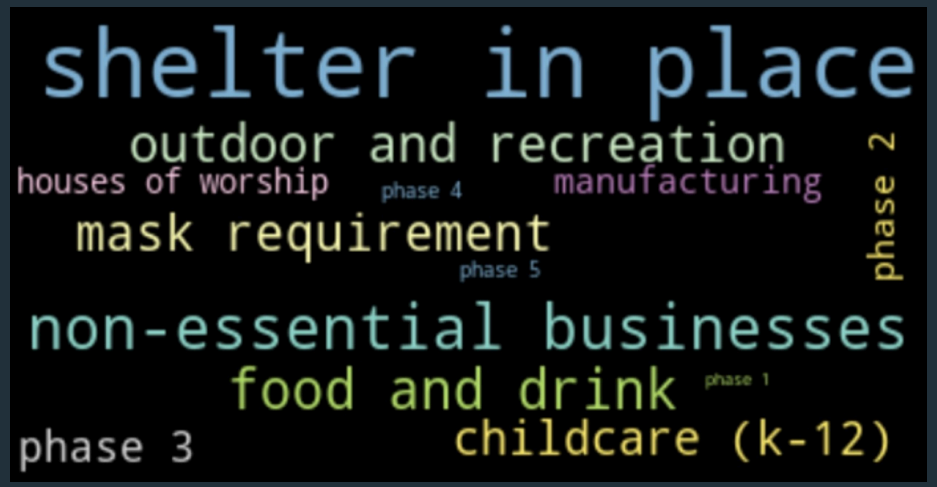
The chart to the right depicts the number of nursing home facilities by state. The chart below shows the mortalities in nursing homes by state.

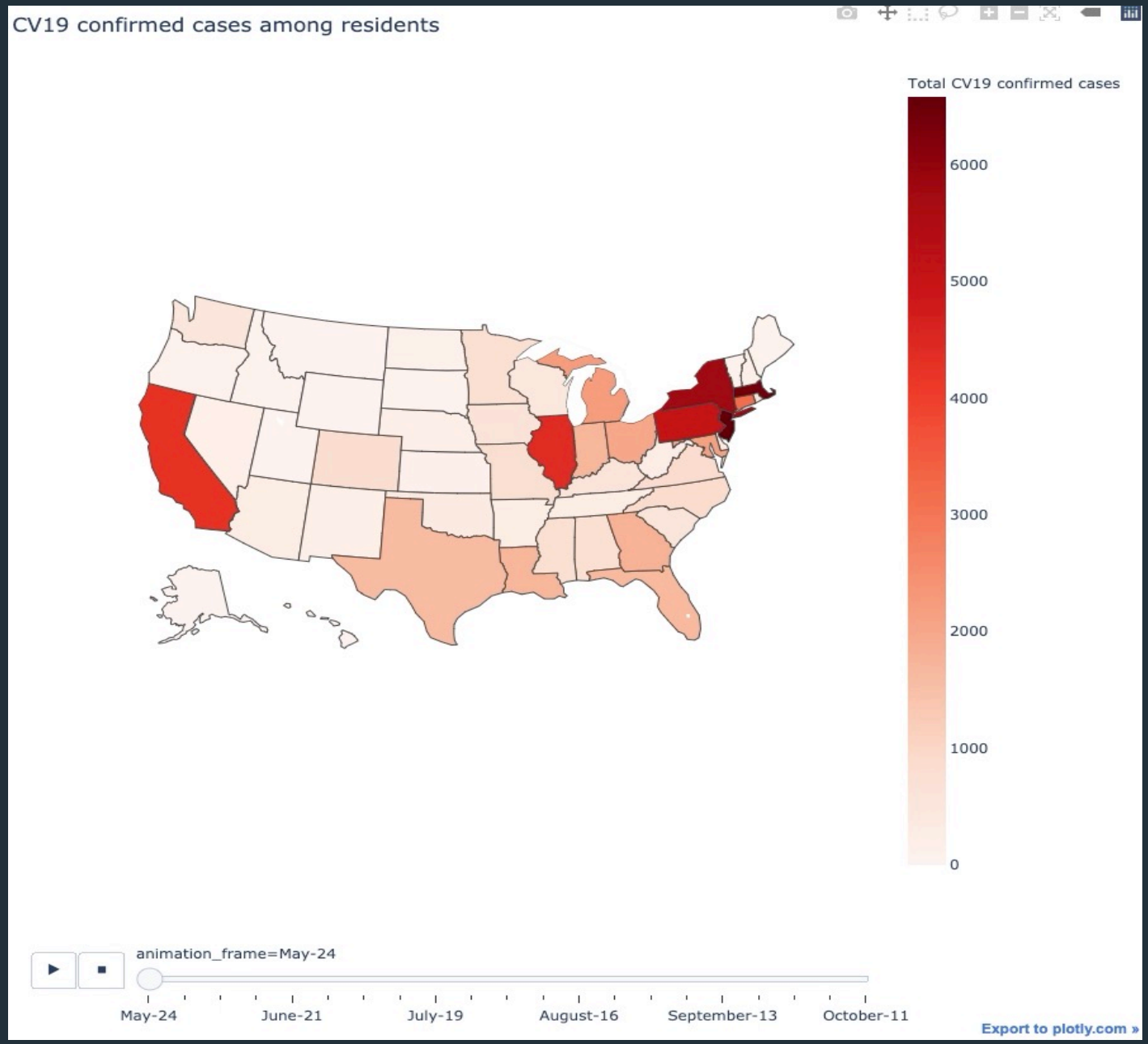




EDA Visualization

Below is a word cloud of the COVID policies implemented in each county. The chart to the right shows the progression of COVID cases by state over time.

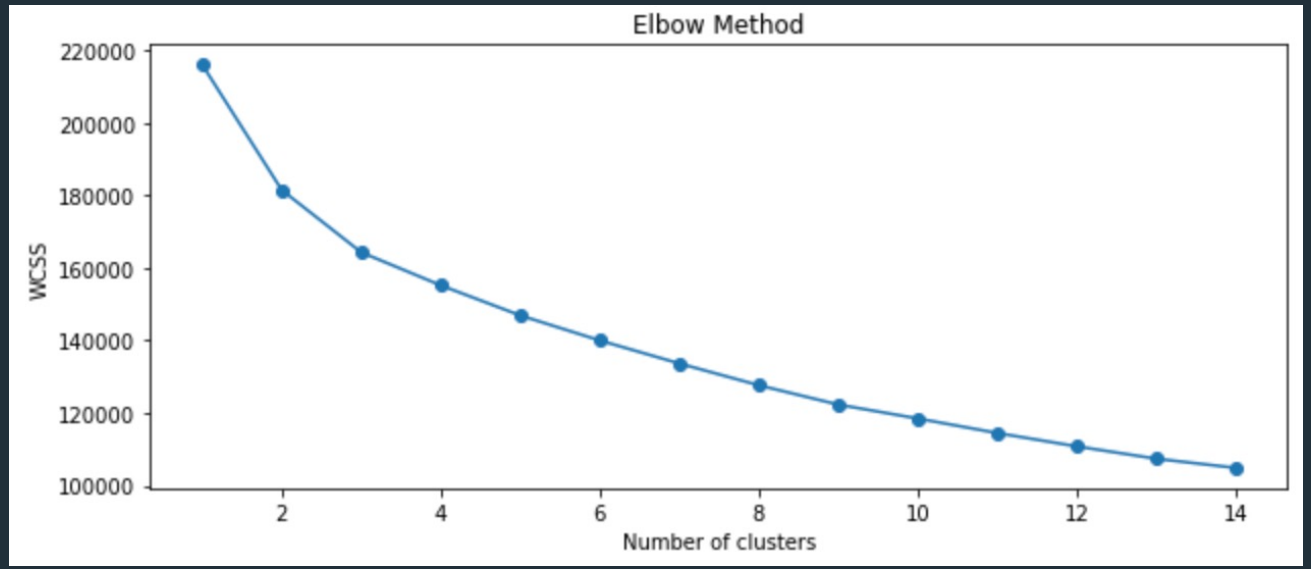


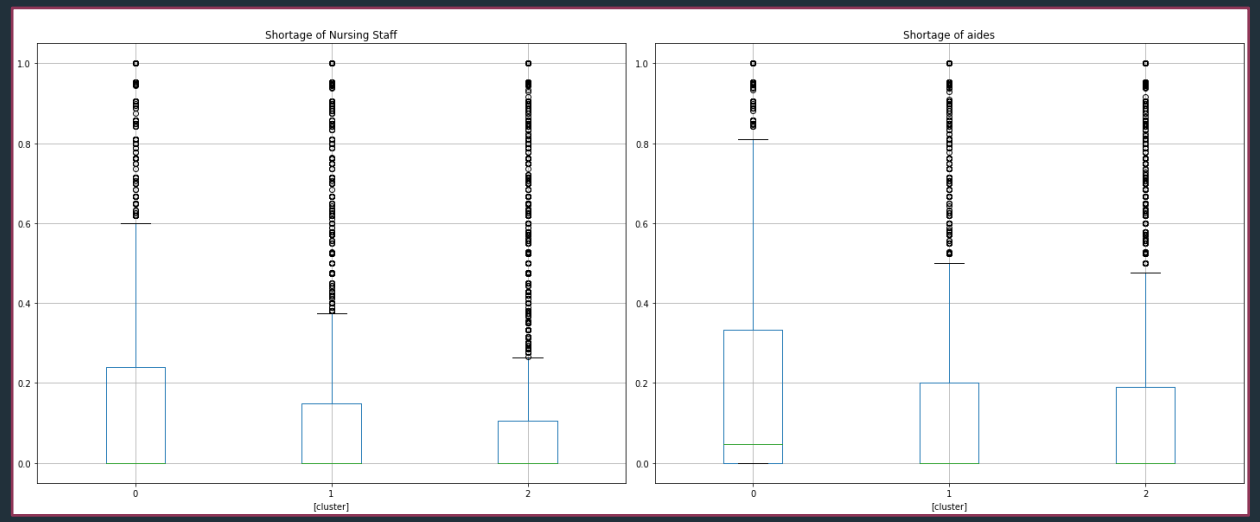


K-means Model

Used K-means to group like nursing homes. Clusters are described by the impact they faced from COVID as well as their risk of issues in the future.

- Used K-means to find similarities
  - Aggregate data by provider for latest week
  - Engineered mortality rate feature
  - Standardized data
- Observed three clusters to be fitting from the elbow plot
  - WSS gradually decreased after three clusters

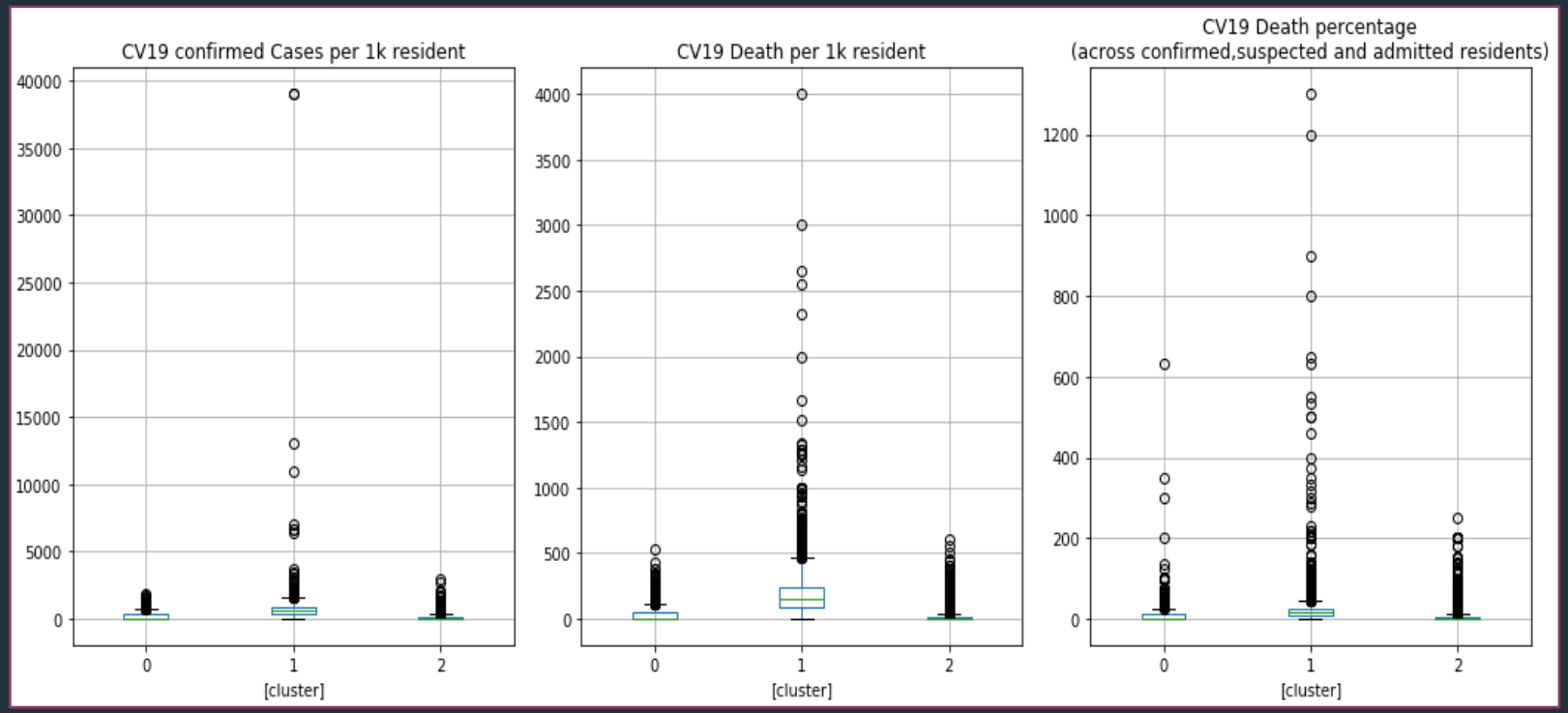


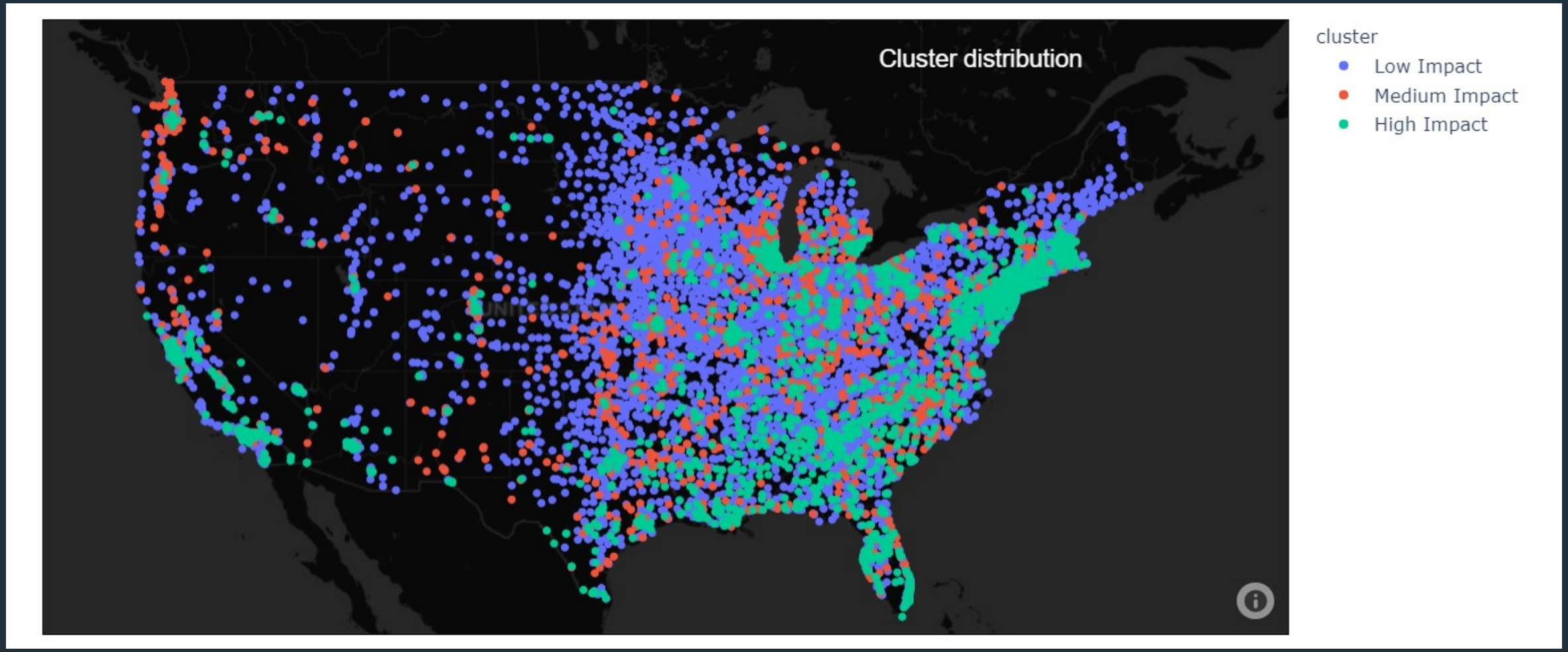


- Three clusters:
  - Cluster 0: Medium impact
  - Cluster 1: High impact
  - Cluster 2: Low impact

K-means Model

K-means model created three distinct clusters.



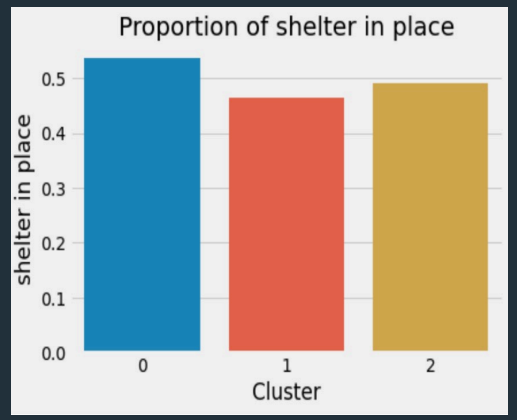


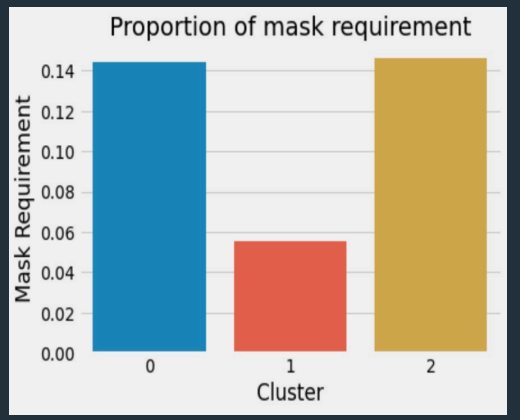
Clusters by location

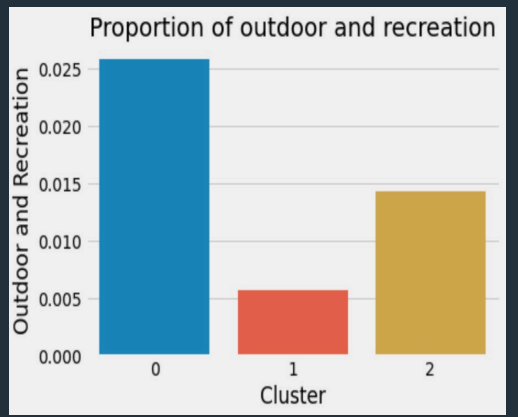
K-means Model

Policies implemented varied across clusters.

- Surprised to see lower mortality rate more shortages in supplies and staff in cluster 0
- Found COVID policies implemented in clusters varied
  - 46% of cluster 1 was under shelter in place restrictions versus 54% in cluster 0
  - 6% of cluster 1 was in a county that had mask requirements versus 14% in cluster 0
  - 0.6% of cluster 1 was under outdoor recreation limitations versus 2.6% in cluster 0

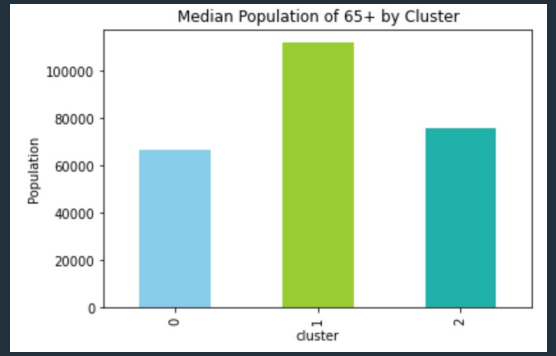


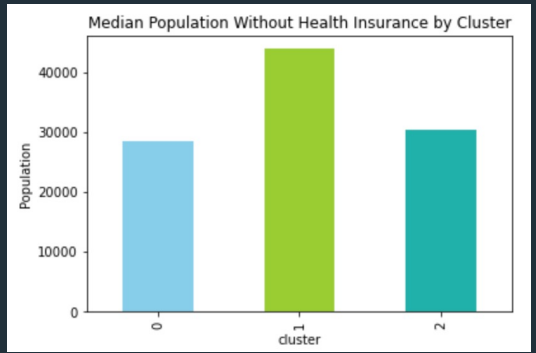


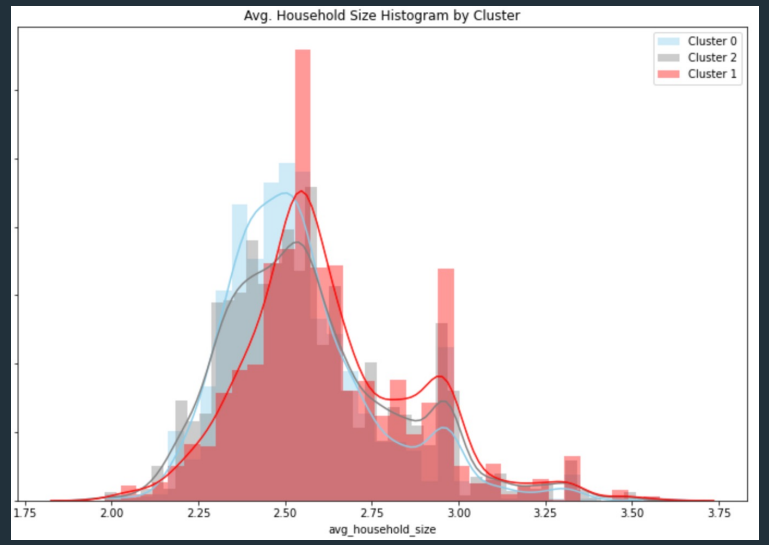


K-means Model

Cluster 1's demographics likely played a role in it's high mortality rate.







- Surprised to see high mortality rate w/ more supplies and staff in cluster 1
- Demographics varied in a few key areas as well for cluster 1
  - Highest density cluster
  - Highest population of 65+ age group
  - Highest population of civilians without health insurance

Recommendation

Send vaccines to nursing homes with highest risk first.

Wave I

- Cluster 1
- # of residents: 295K
- Location: Coasts
- Reasoning: High mortality rate, weaker policies, high at-risk population

Wave II

- Cluster 0
- # of residents: 152K
- Location: Various
- Reasoning: More shortages of supplies

Wave III

- Cluster 2
- # of residents: 652K
- Location: Mid-West, North East
- Reasoning: Lowest mortality rates, adequate supply levels, policies in place

Recommendation

Communicate waves of vaccine shipments with cargo carriers. Dedicate team to oversee logistics of dispersing vaccines.

2009

- HiNi's biggest challenge: dispersing vaccines
  - Cold-chain packing
    - Limited space
    - Handling procedures
  - Volume higher than typical flu season

2020

- COVID vaccines face similar issues
  - High demand
  - Health care workers, nursing homes first
  - Cold-chain packaging

2021

- Study data in order to:
  - Disperse vaccines efficiently and effectively
  - Allow for logistics providers to plan accordingly

Recommendation

Advise counties in cluster one to consider shelter-in-place and mask requirements in order to control mortality rate.

Found higher implementation of these policies likely contributed to the low mortality rates in cluster 0 and cluster 2.

Thank you

Stay healthy!