

# WHERE TO RETIRE IN WEST VIRGINIA

ECM

Applied Data Science

Capstone Presentation

2020

# A MAJOR CONSIDERATION FOR RETIREES IS ACCESSIBILITY

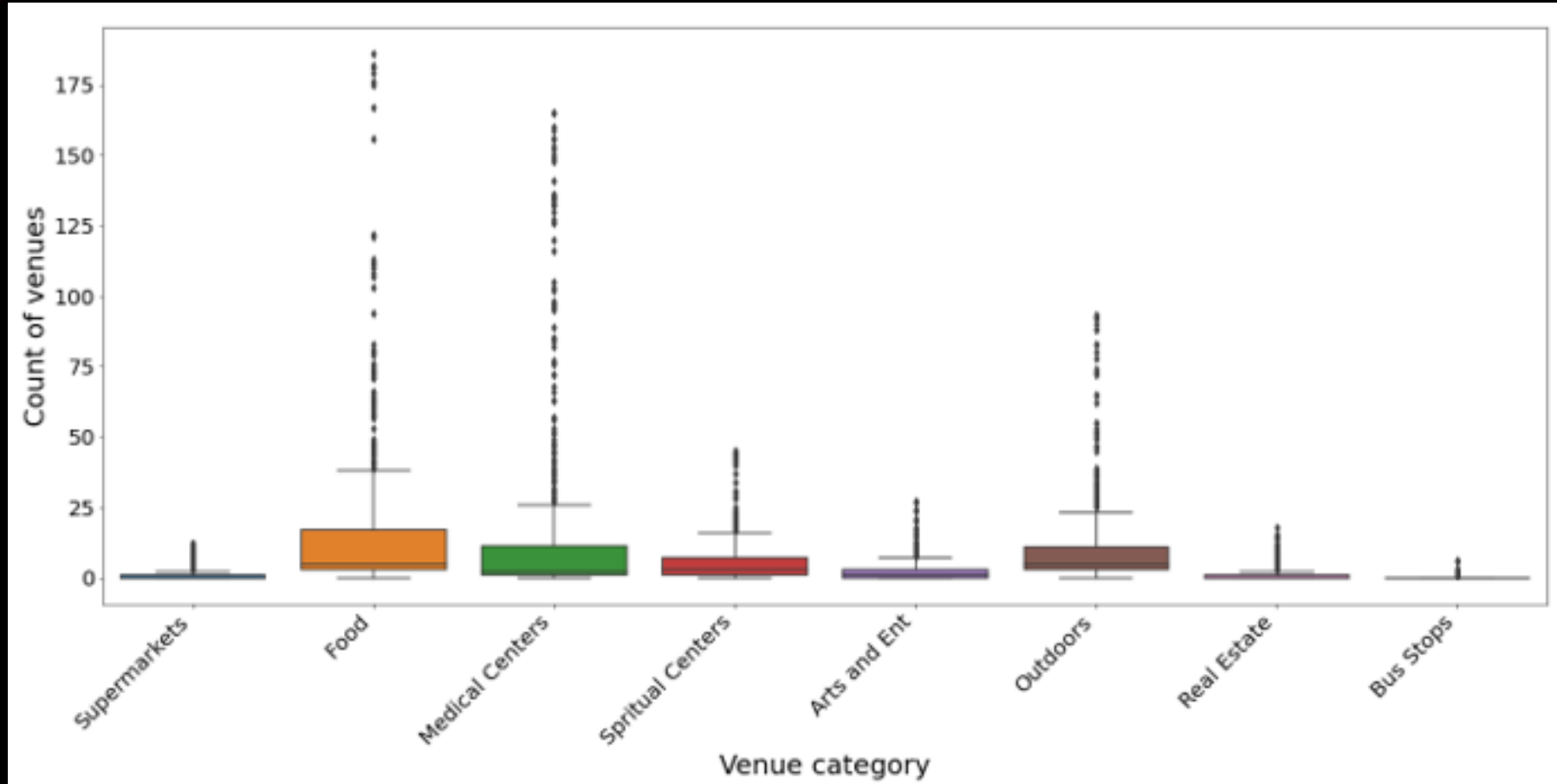
As people get older, they make sure that certain needs and wants can be easily met. In particular, retirees want ready access to:

- Supermarkets
- Restaurants, or food
- Medical centers
- Spiritual centers
- Arts and entertainment venues
- Outdoors and recreations
- Bus stops

# DATA ACQUISITION AND CLEANING

- The complete list of ZIPcodes can be found at <https://www.unitedstateszipcodes.org/wv/#zips-list>.
- Roughly a quarter of ZIPcodes are for Post Office Boxes only, and so do not represent potential retirement areas, leaving 580 possibilities out of a total of 708 ZIPcodes.
- The information for food, medical care, entertainment, and real estate, etc. can be gotten through the Foursquare API.
- A radius of approximately 6 miles (10,000 meters) will be used when searching for the number of venues, as this should be comparable to a 15-20 minute drive on local roads, a reasonable short drive.
- For the hospitals and universities, the maximum radius of 100,000 meters (approximately 60 miles) will be used. The algorithm will find the minimum distance to these types of venues.

# VENUE COUNT DISTRIBUTIONS



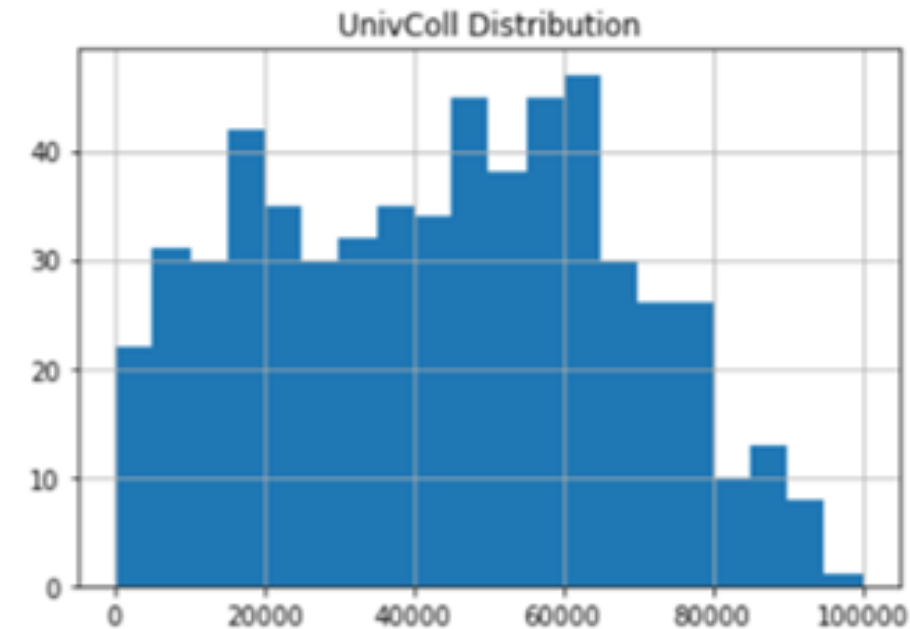
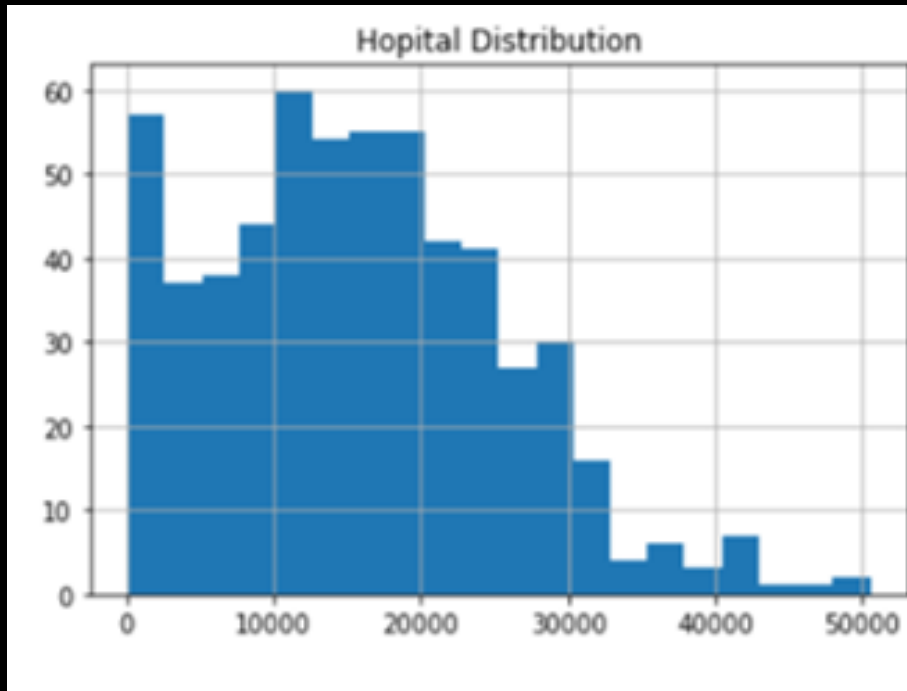
Most of the values reported are low, with the upper fence being 25 or less in every category except food.

# VENUE COUNT STATISTICS

	Super-markets	Food	Medical Centers	Spiritual Centers	Arts and Ent	Outdoors	Real Estate	Bus Stops
mean	1.12	16.54	16.99	5.82	2.64	10.95	1.31	0.13
min	0	0	0	0	0	0	0	0
25%	0	2.75	1	1	0	3	0	0
50%	0	5	2	3	1	5	0	0
75%	1	17	11.25	7	3	11	1	0
max	12	186	165	45	27	93	18	6

- Over half of the ZIPcodes don't have any supermarkets, arts and entertainment venues, or real estate agents nearby.
- Three-quarters of the ZIPcodes have no bus stops, indicating that mass transit is not widely available.
- The 75<sup>th</sup> percentile indicates that about a quarter of the zip codes are more metropolitan, having plenty of food places, medical centers and spiritual centers to choose from.

# VENUE DISTANCE DISTRIBUTIONS



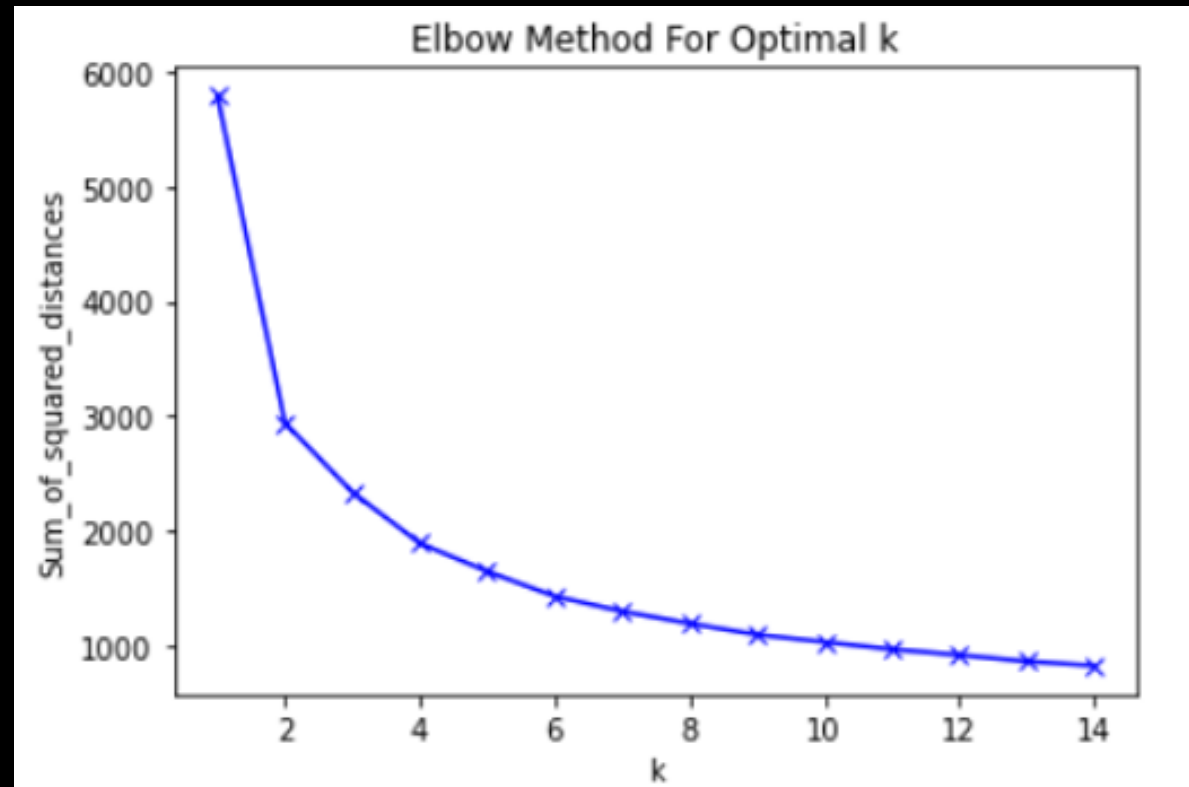
- The distances to the nearest hospital show a distribution similar to a truncated normal distribution. (Distances cannot go below zero, so that part of the distribution is compressed at the low end. )
- The distances to the nearest university or college show closer to a uniform distribution over the distances between 0 and 80,000 meters with a few ZIPcodes as far away as 100,000 meters from any venue of higher education.



# METHODOLOGY

- Will use K-means method to sort into clusters.
- K-means clustering is one of the simplest and popular unsupervised machine learning algorithms.
- A cluster refers to a collection of data points aggregated together because of certain similarities.
- The goal is to find ZIPcodes with similar types of venues that should be equally appealing to retirees.

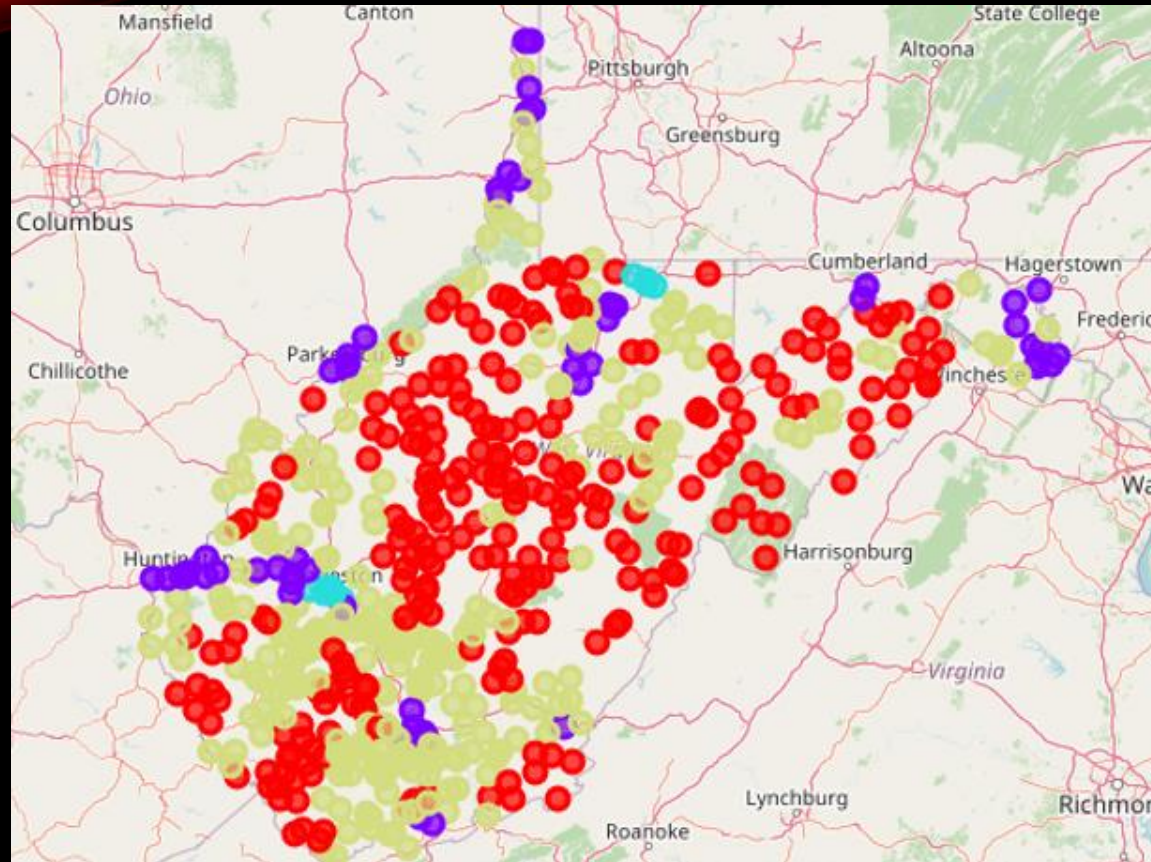
# BEST VALUE OF K



- There are bends, or “elbows” at  $k = 2$ ,  $4$ , and  $6$ .
- Using  $k = 6$  created groups that only consisted of 1 zip code, so it added no information.
- Analysis will proceed using  $k = 4$ .

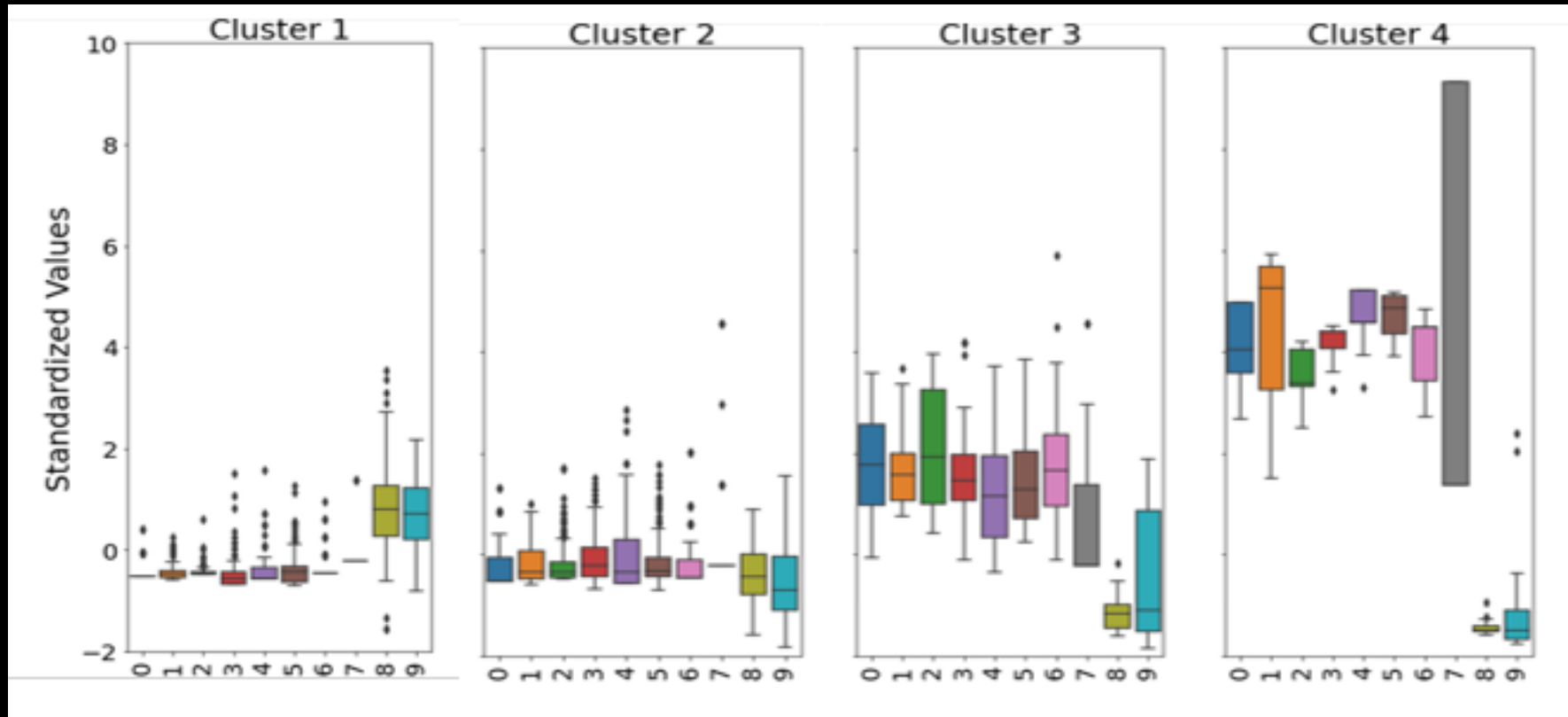


# RESULTS



- RED: Cluster 1 – Isolated / Rural ZIPcodes
- YELLOW: Cluster 2 – Outer Suburban ZIPcodes
- PURPLE: Cluster 3 – Inner Suburban / Population Center ZIPcodes
- LIGHT BLUE: Cluster 4 – Urban ZIPcodes

# BOXPLOTS FOR VENUES BY CLUSTER



It is easy to see the increases in the number of venues available as we move from one cluster to another. What is not easy to see here is the difference in cluster sizes. Many more ZIPcodes (88%) fall in clusters 1 and 2 than in 3 or 4.

# DISCUSSION OF CLUSTERS

## Cluster 1

- Has the fewest venues of any type nearby.
- Suitable only for retirees that value isolation more than access to basic needs.

## Cluster 2

- Little access to many amenities; has more variety in the venue offerings.
- Good locations for retirees that like less dense areas; not too far from necessities and a variety of entertainment.

## Cluster 3

- Very close to the larger cities in West Virginia.
- Appeals to those retirees who don't mind driven; quick drive to any venue.

## Cluster 4

- Many venues within walking distance.
- Best for retirees that do not wish to drive; great diversity in venues available.

# FUTURE IMPROVEMENTS

- Include socio-demographic information as well as venues to allow for further differentiation between ZIPcodes.
  - Housing prices
  - Population density
  - Average population age
  - Real estate taxes
- Choose more targeted venue categories.
  - State / National Parks instead of all Outdoors
  - Doctor's Office instead of Medical Center (which includes hospitals)
  - Museums or Performing Arts Venue instead of Arts and Entertainment