Food Deserts

Data 144 Final Project

Noah McDonald, Raymond Peng, Emma Toon, Kidong Kim

Introduction

- **In the United States, approximately 20 million people live in a “food desert”**
- USDA defines a food desert as “living more than one mile from a supermarket in urban areas, and more than 10 miles from a supermarket in rural areas”
- Food deserts commonly associated with:
  - Large proportions of households with low incomes
  - Inadequate access to transportation
  - Limited number of food retailers providing fresh produce and healthy groceries for an affordable price

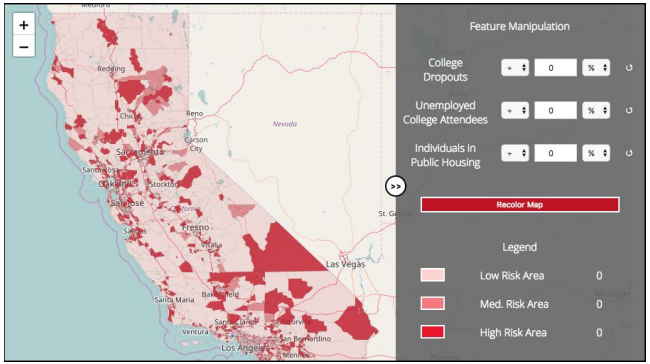
Research Question

What is the relationship between food deserts and demographic/geographic factors such as race, income, rural/urban environment, etc?

How can we utilize different clustering techniques to visualize and communicate trends and relationships between food deserts and demographic/geographic factors?

Discussion Topic

Are there any interesting food desert/food access trends in the Bay Area specifically? Real world applications of food access risk.

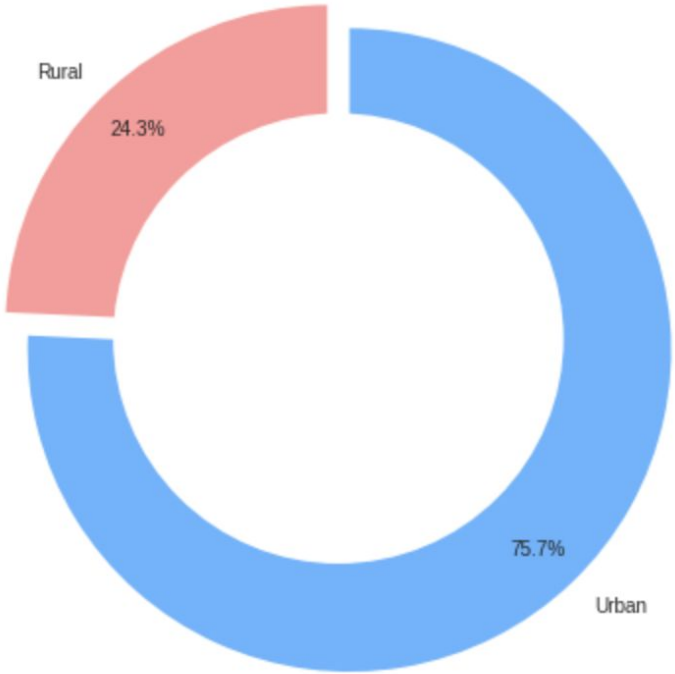


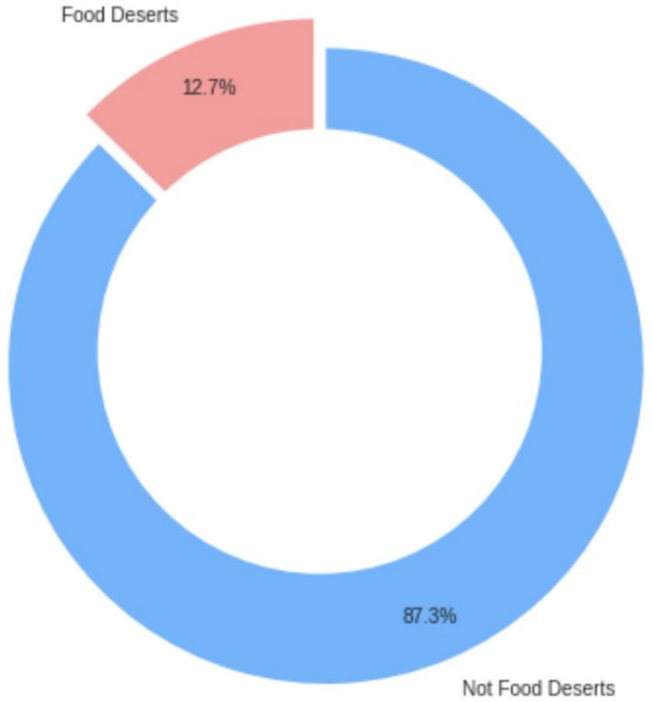
UC Berkeley School of Information: MIDS Capstone Project "The Food Desert Predictor" (Putman, Iqbal, Acconciamessa 2017)

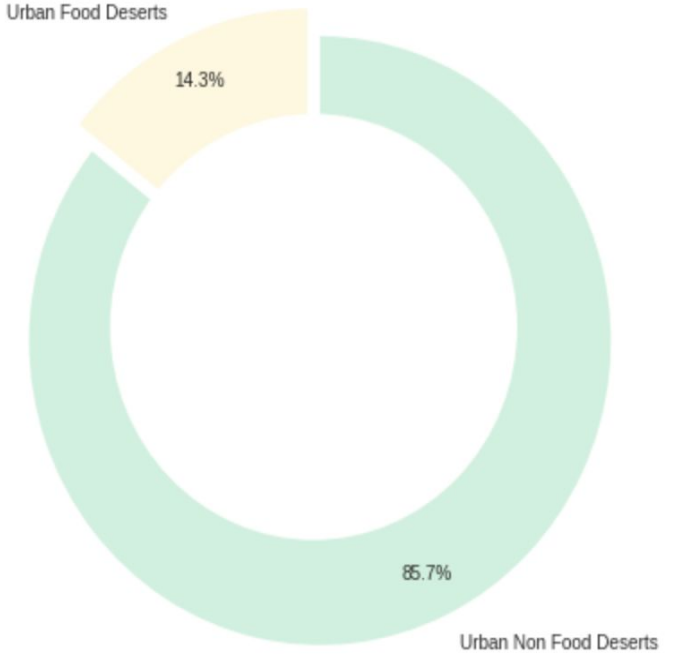
Data

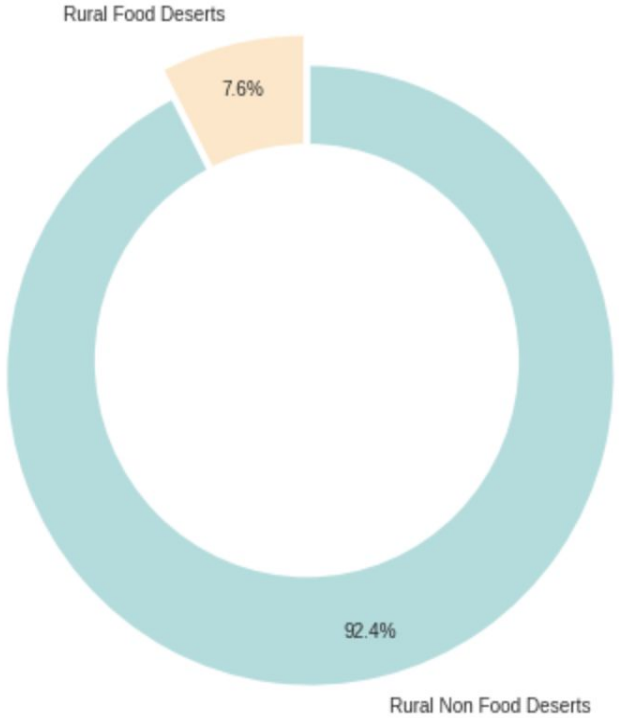
- Food Deserts in the U.S. Dataset on Kaggle
- Data is pulled from the USDA Food Access Research Atlas, and contains information on supermarket access at various distances
- Combines Food Access data with other fields such as age, race, rural/urban, income, etc.
- Each row represents a census tract
    - Total number of census tracts: 72,864

Exploratory Data Analysis



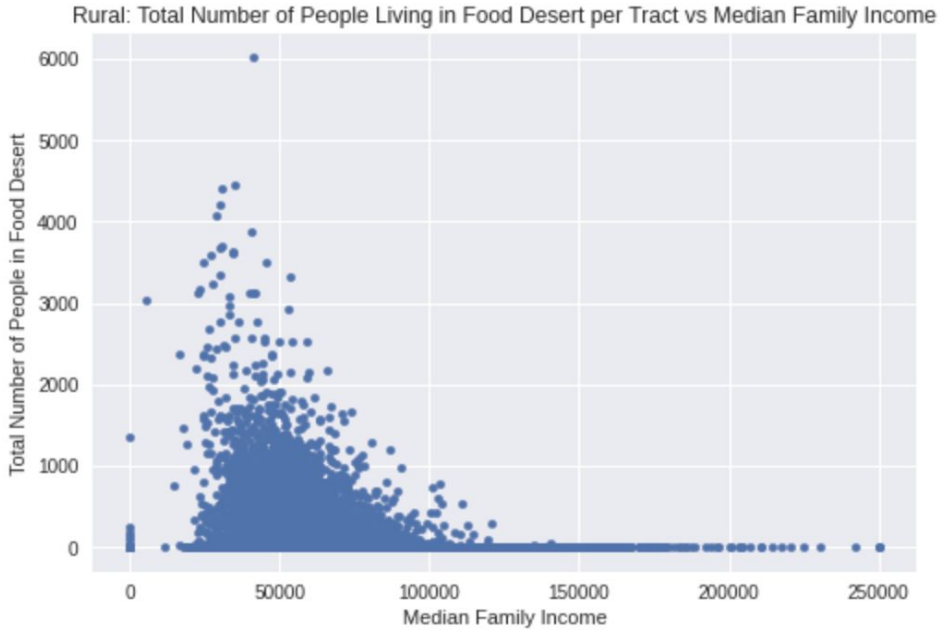


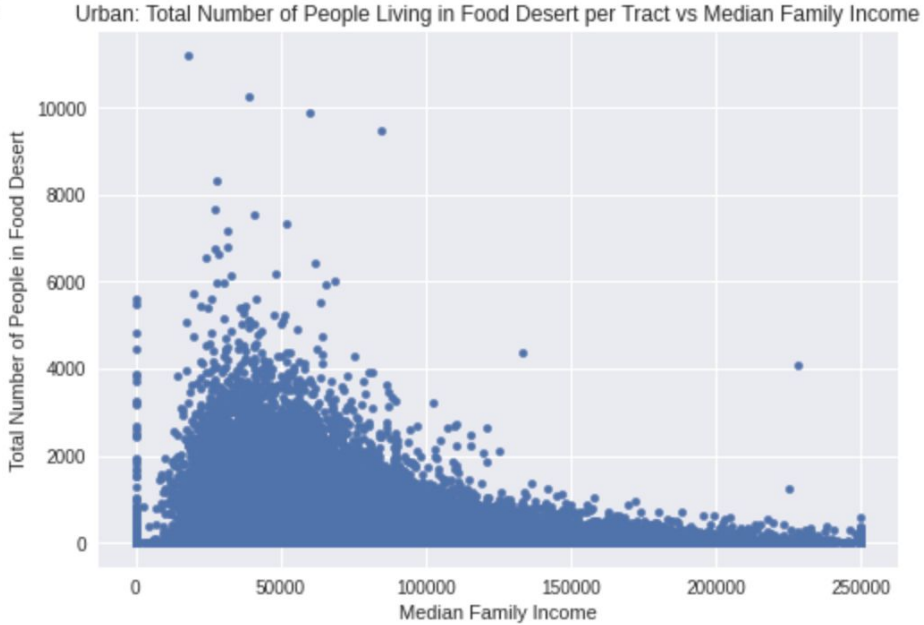




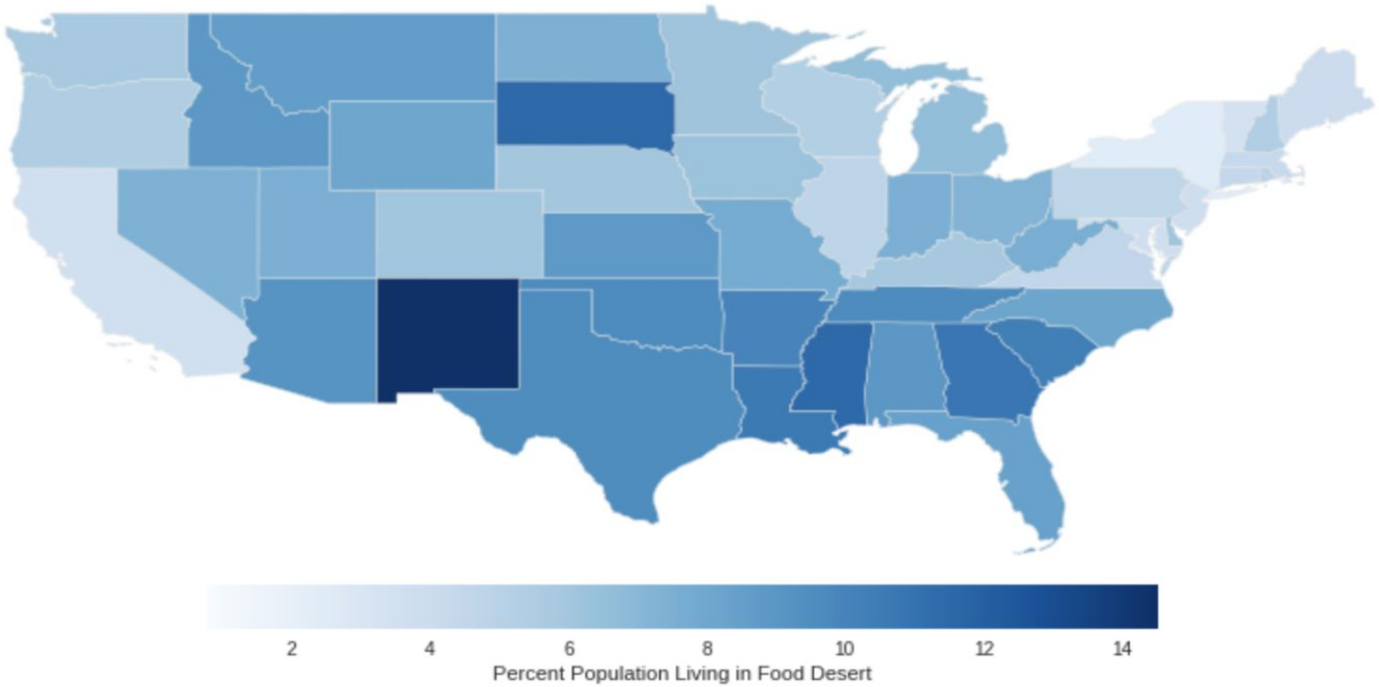
- Given that you are in an urban area, you are almost twice as likely to live in a food desert than if you lived in a rural area

Relationship between Median Income and Number of People Living in Food Desert

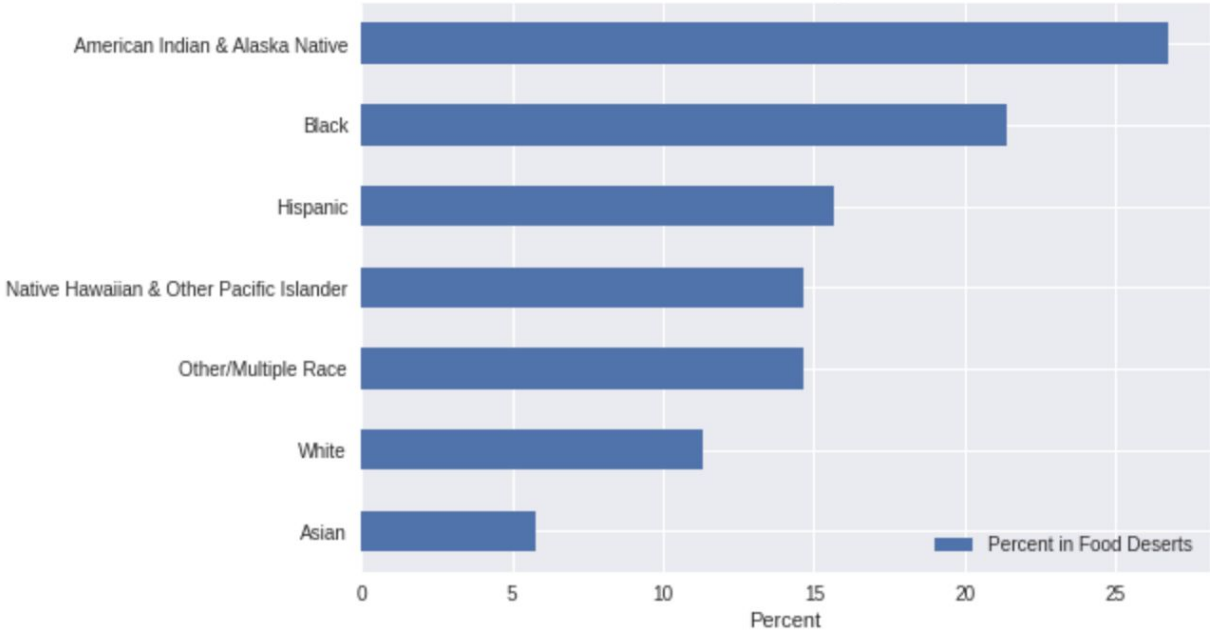




Percent of State Populations Living in Food Deserts



Percent of Each Race Living in Food Deserts



- In general, higher percentages of minority groups live in food deserts

Methods

Dimensionality Reduction:

- PCA
- t-SNE
- Autoencoder

Clustering Methods

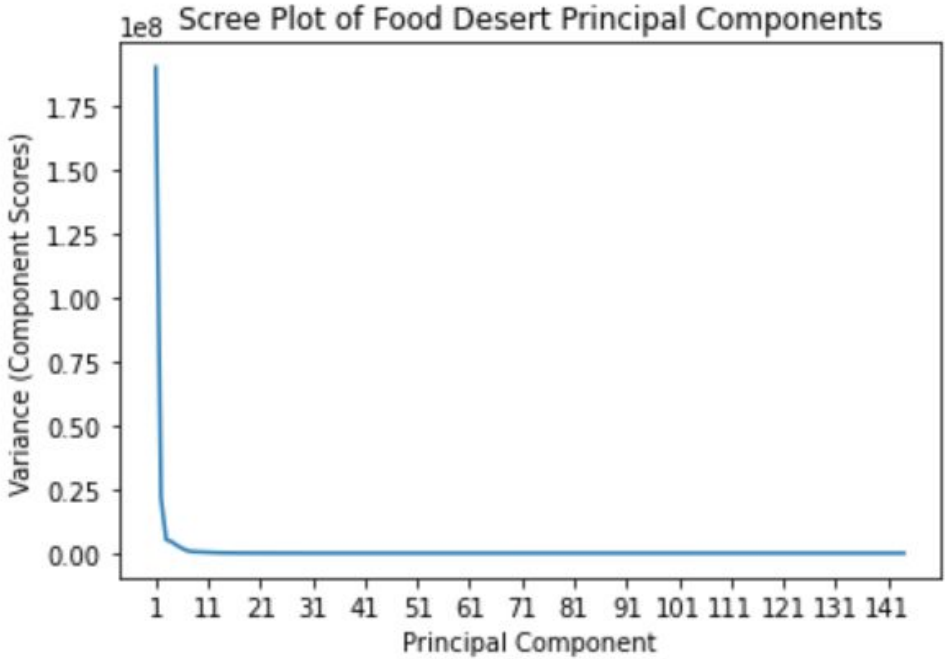
- k-means clustering

Methods - Principal Component Analysis (PCA)

- Filter for only food deserts where feature LILATracts_1And10 = 1
- Center all features for PCA
- Take the singular value decomposition (SVD) of the dataframe as a matrix
- Get dot product of centered features and first 2 columns of vt to get our first 2 principal components, reducing our dimensions **from 144 to 2**

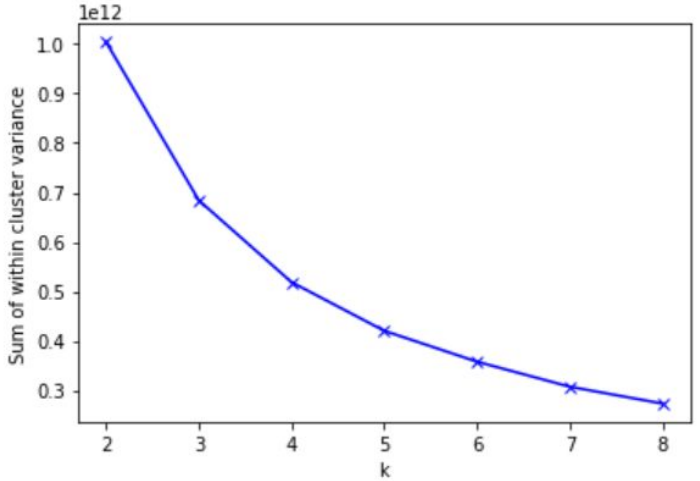
PCA Setup

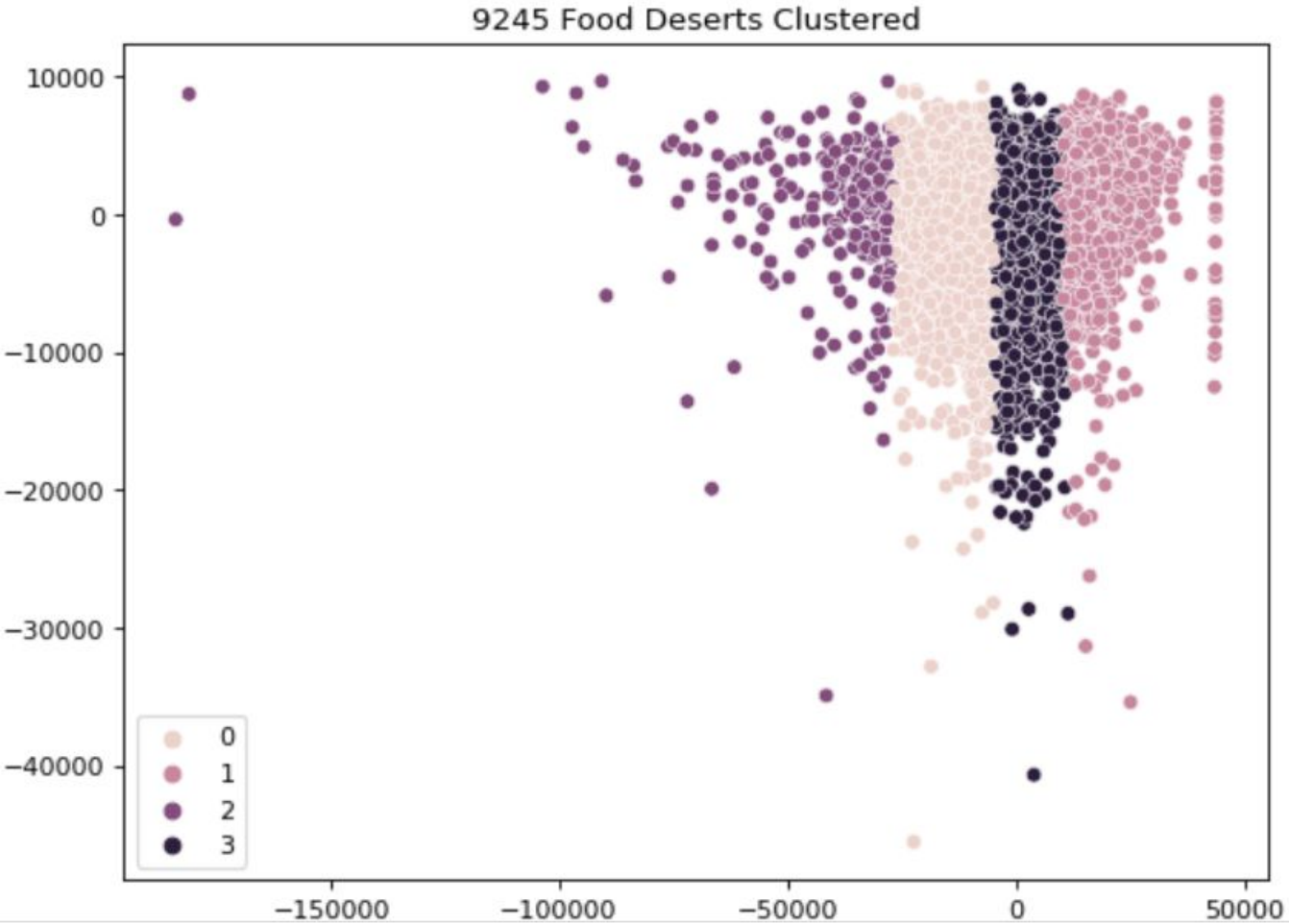
- Check what proportion of variance is accounted for by 2 dimensions (2 largest PCs)
  - 91.75%, most variance accounted for



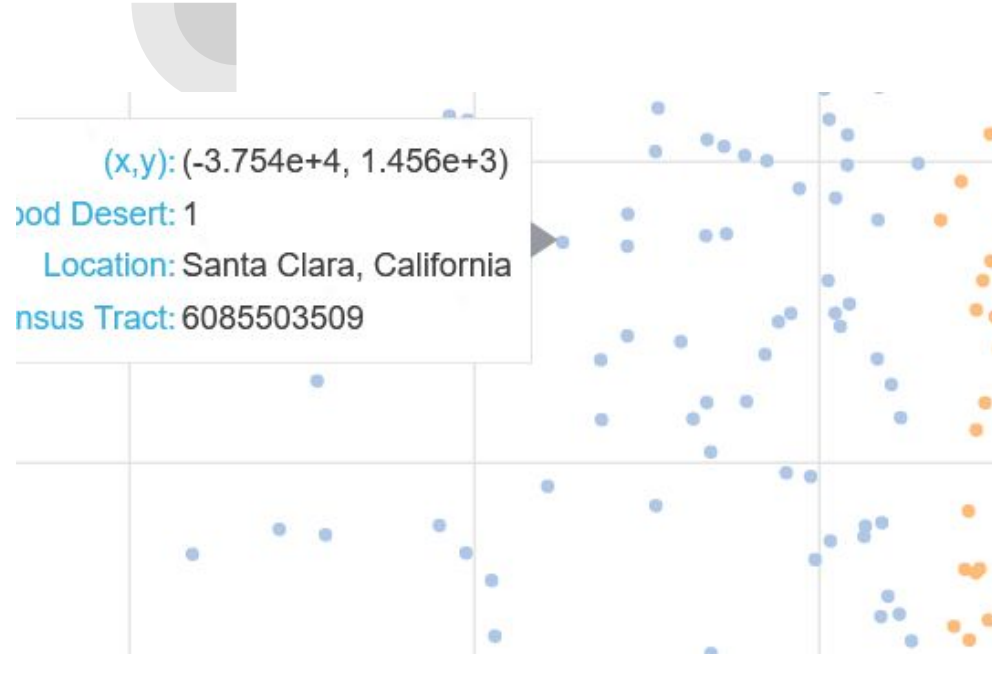
Clustering

- Using k-means
- Decided on k=4 using elbow method
- PCA data points mapped onto a scatter plot with respective clusters as hues using Seaborn (static plot) and Bokeh (interactive hover plot)
    - Where X = PC1 and Y = PC2





Tooltip version available using Bokeh



PCA - Clusters and geography

t-SNE

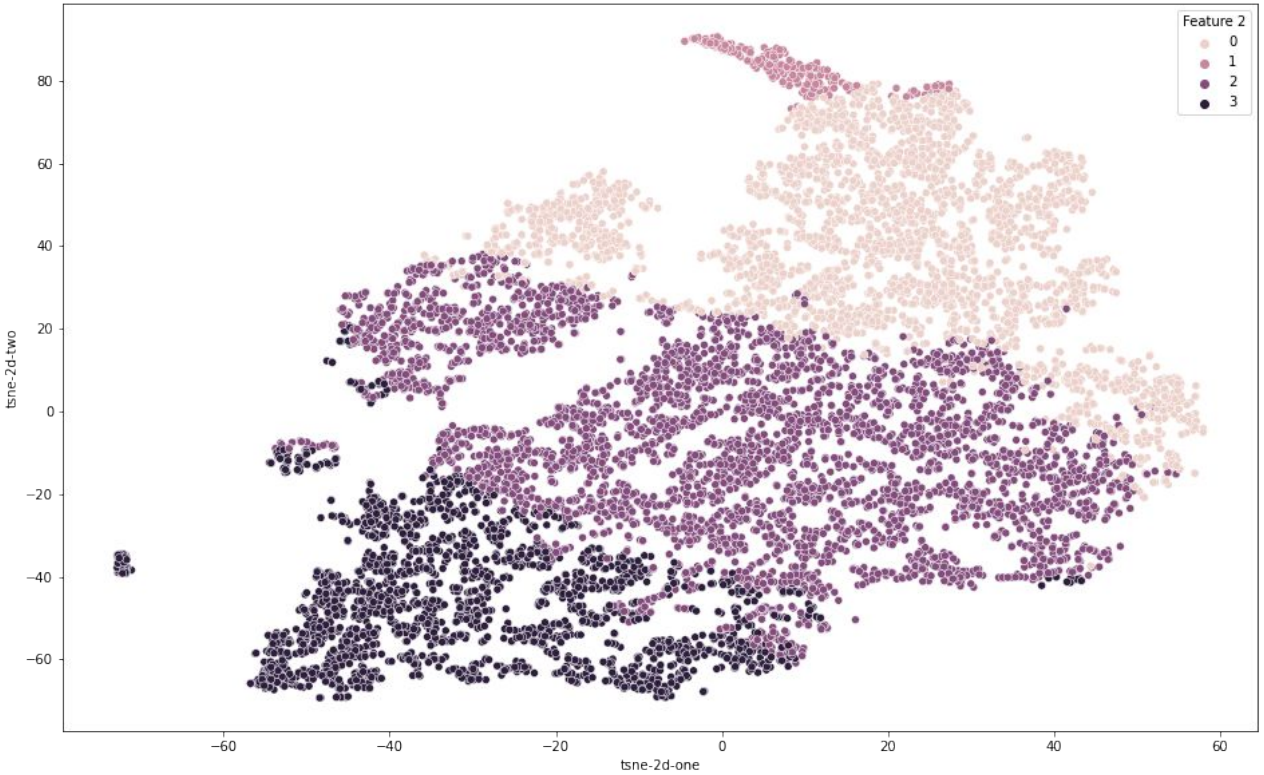
- Dimensionality reduction from high dimensionality to **2 Dimensions**
- Using similar setup and clustering methods as mentioned in PCA

  - Filter for Food Deserts

  - kMeans, k = 4

  - perplexity of 50

  - mapped onto scatter plot where X = t-SNE-1 and Y = t-SNE-2


Feature 2
0
1
2
3
tsne-2d-two
tsne-2d-one
80
60
40
20
0
−20
−40
−60
−60
−40
−20
0
20
40
60

Auto-encoder

- Conversion
  - Process of converting data into a format required for a number of information processing needs
  - Process of applying a specific code, such as letters, symbols and numbers, to data for conversion into an equivalent cipher

Auto-encoder

- One-hot encoding
  - Categorical values (Boolean values)
    - ex: Urban
    - Flag for food desert when considering low accessibility at 1 and 10 miles
- Normalization
  - Continuous values
    - ex: Population, housing unit

Real World Applications

- Clusters and demographic similarity
- Clusters and geographic similarity (same county)
- Clusters and urban/rural split
- Helps identify locations where food inequities are the highest

Conclusion

- Nearby census tracts are generally clustered together (Regional/Statewide trends)
- Lower income areas have more issues with food access and food deserts
- Unclear/inconclusive evidence of clustering together via race
- Clustering of tracts may have a rural/urban component