Customer Analytics for Social Good:

Machine Learning to identify "Sister Cities" for Benchmarking

Consultant: Krishnesh Pujari | **Professor:** Dr. Kathy Weaver
Client: Montgomery County – Office of the County Executive – CountyStat Team

UNIVERSITY OF MARYLAND 18 56 COLLEGE OF INFORMATION STUDIES

Introduction

- Montgomery County is a recognized leader in open data and innovation
- County seeks to benchmark itself against like jurisdictions to assess its performance and identify opportunities for improvement

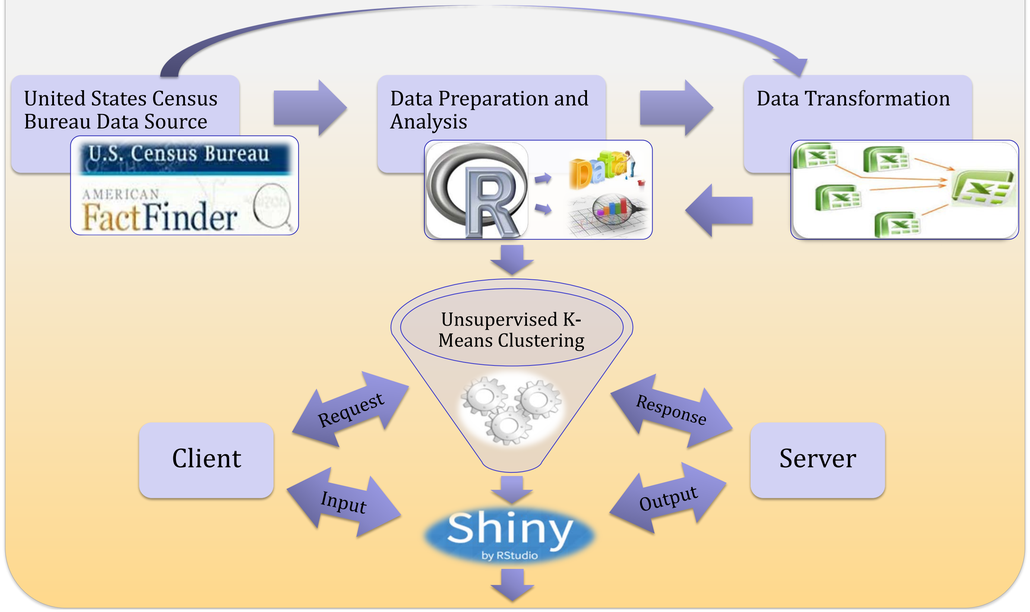
Problem Statement

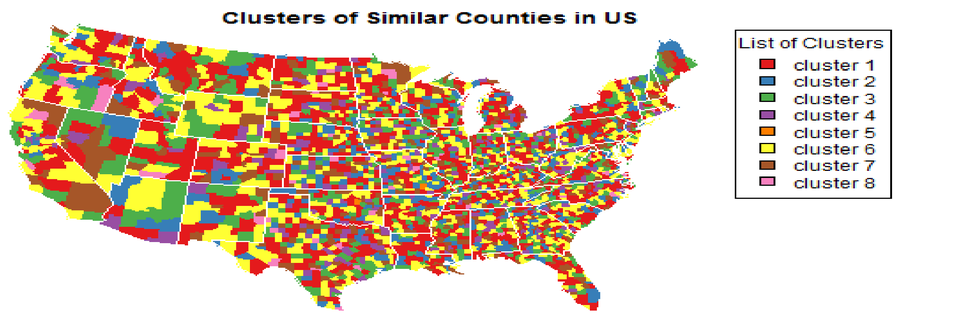
- Currently, the CountyStat team identifies similar counties purely based on subjective determination
- CountyStat team wishes to automate the process using unsupervised machine learning algorithm

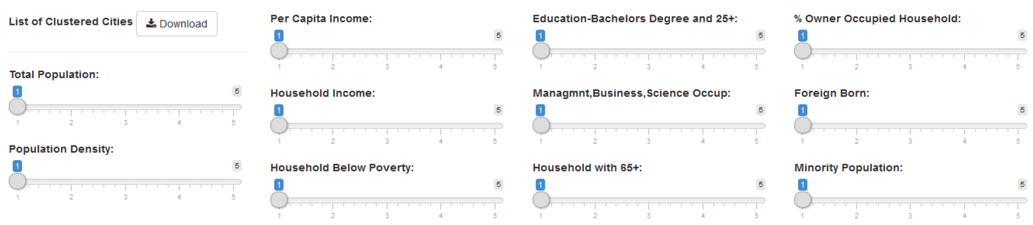
Goals

- Need to implement clustering algorithm that identifies similar jurisdictions
- Create a tool that can be used by any County Government to benchmark themselves against their peers

Project Methodology







Accomplishments

- Automated benchmarking process by identifying clusters of similar counties
- Developed Interactive tool that can be used by any US County Government
- Hands on experience working on R programming, Shiny App, and Excel

Challenges

- Working with multi-dimensional data to identify prime factors
- Feature engineer the data discard masking variables for better results
- Build interactive visualization using Rstudio's Shiny App Framework

Future Scope

- Optimize the algorithm to factor data and identify similar counties across different industry sectors
- Integrate data preparation task with the tool