Denver Brewery Project

COURSERA / IBM DATA SCIENCE CAPSTONE

Introduction

- Finding an ideal location for a new business is critical decision that may determine whether the business will succeed

- Target Audience
  - Entrepreneur looking to open a new brewery in Denver, Colorado

- The Problem
  - How to determine the location of a new brewery. An ideal neighborhood for this new business would have a large number of potential customers and few direct competitors

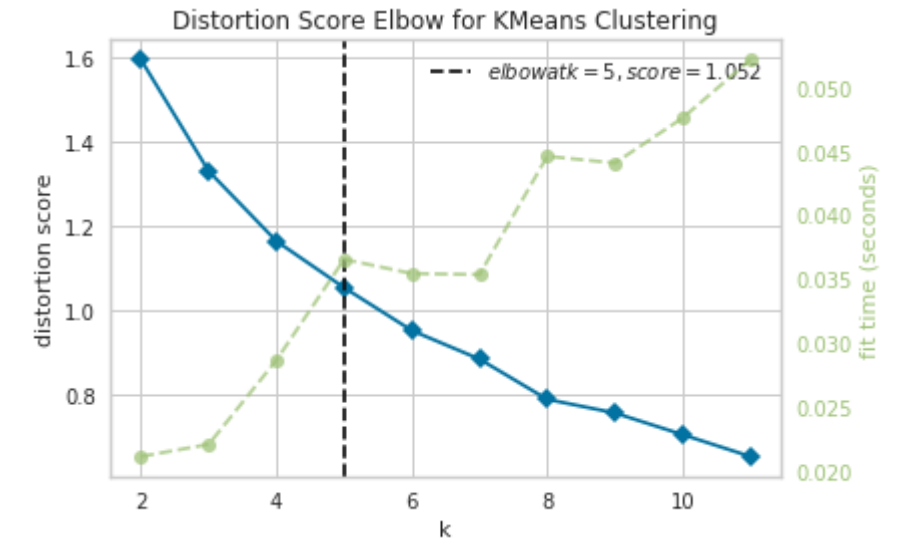
Approach

Extract venue data of existing breweries, bars, and other related business in neighborhoods around Denver. Use machine learning (ML) to identify the relation between various extracted features. Make recommendations based on observed ML model.

- Data
  - FourSquare – location data includes details of "venues" - businesses, restaurants, parks, etc in relation to a specified geolocation
  - Neighborhood and geolocation data was retrieved from Wikipedia and the Python Geocoder package

Methodology

1. Group data by neighborhood and normalized number of venues by category.
2. K-means clustering.
3. Visualize the clusters in a map using Folium.



K- means: Finding a value for K

A good value of K must be found in order to have meaningful data.

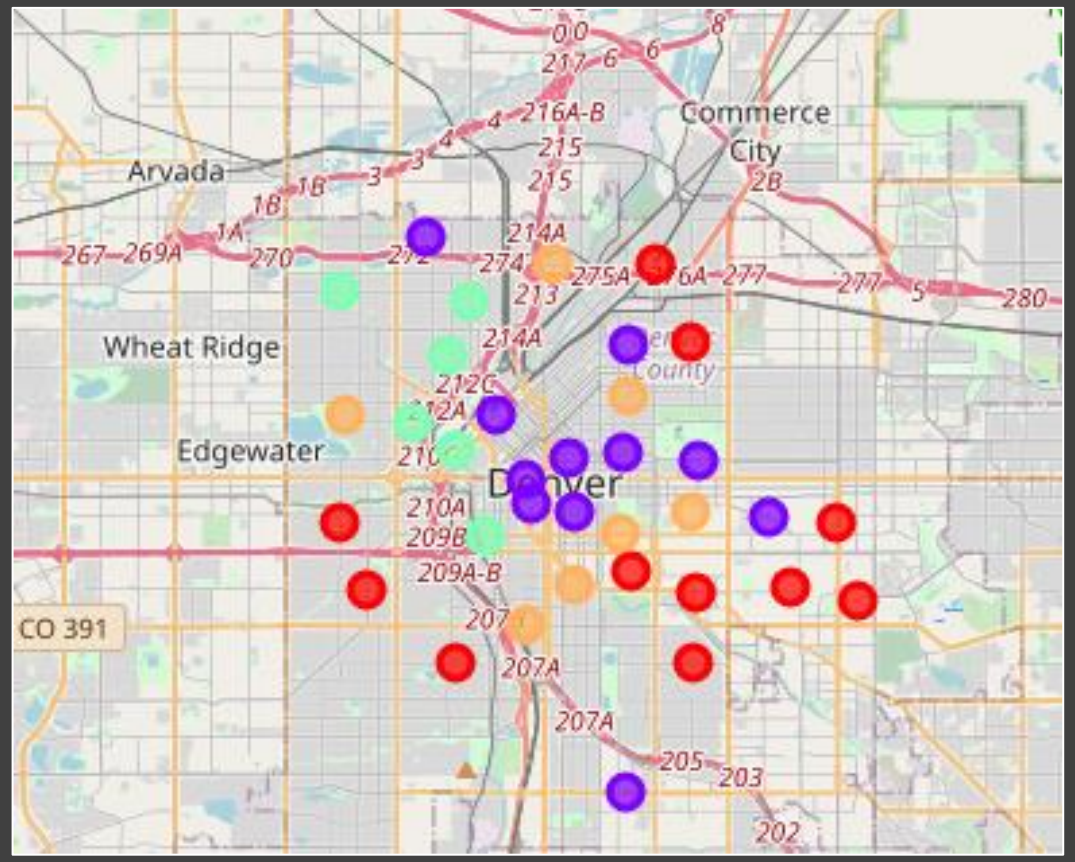
The Elbow method was used to determine the optimal value of K for the datasets.

Geographical Clusters

Cluster 4 identifies the types of neighborhoods we were most interested in: low number of direct competitors (existing breweries) and a medium to high number of related businesses.

Cluster 4 consists of seven neighborhoods: Auraria, Ruby Hill, Sloan Lake, Elyria-Swansea, Chaffee Park, Cole, and West Colfax.

The neighborhoods identified in this cluster will be prioritized for further research in identifying our ideal brewery location.



Conclusion

From the clustering analysis done in this report, it is likely that a neighborhood within Cluster 4 would be the most suitable location for a new brewery. A next step would be to take the neighborhoods of Cluster 4 and find more data on the area, possibly economic data, and perform another cluster analysis to break down the cluster even further. The findings of this project will help the relevant stakeholders to identify and capitalize on the opportunities of locations with high potential.