Select Business Location in Denver, Colorado

Bingxin Qi

Jan 11, 2021

1. Introduction

1.1 Background

Founded in 1858, the city and county of Denver is located on the banks of the south Platte River close to the Rocky Mountains. As the most populous state capital in the Mountain states, Denver began to thrive due to its mining industry, and continued to grow and expand due to its development of technological and manufacturing industries. Most importantly, after the World War II, Denver gained its recognition as commercial hub because of the I-25 highway construction, military contracts, and so forth.

Previously, since Denver has been relying heavily on gas industry, it had experienced recession in economy. Through strategic investments on various field such as aviation bioscience, health, education, and so forth. Consequently, over the past ten years, the city and county of Denver, especially downtown area has experienced a revive of economy. For example, approximately 4.4 million square feet of office space, over 10,000 residential units, and 3,700 hotel rooms has been built, and this results in a total of about $6.6 investment in the center city. Therefore, the examination of history of Denver and its development revealed the significant role of diversification on economy, which is also the focus of this project.

1.2 Statement of the Problem

The project will contribute to: (1) better analyze the diversification of location in Denver by wrangling geographical data; (2) cluster and segment location based on their geographical distribution and venue categories; and (3) use machine learning algorithm and geographical statistics to identify appropriate locations for setting up a new local business.

1.3. Interest

The business owners who want to start their new branch in Colorado would be interested in this report because it strategically analyze ideal locations for them to maximize profit based on population distribution and categorization of kinds of business in Colorado.

2. Data Acquisition and Cleaning

2.1 Data Sources

The longitudes and latitudes of location data were found here. This website provided us with zip code along with location information across the U.S. at neighborhood levels. This dataset contains 43,191 records for all 50 states in the U.S., and I will use central point of Denver to build the bounding box and select the neighborhoods surrounding Denver metropolitan area since this study focuses on analyzing locations in Denver, Colorado. Foursquare API will be used to get the venue information such as business categories and locations.

2.2 Data Pre-processing

Data downloaded and scraped from various sources were then combined into one table. *Figure 1* shows a preview of information of regions near Denver. Using the latitudes and longitudes of neighborhoods, the venue data was pooled and joined into one dataframe with neighborhood information. After a closer examination of the data, duplicates of venue information was dropped. This results in a total of 3650 venues with 7 relevant location features. No additional data pre-processing procedures and steps were conducted because missing values are not present in this dataset.

The minimum bounding box method is commonly used to selected geographical features based on the distribution of vectors. Therefore, in order to select neighboring regions based on the Denver metropolitan data, a minimum bounding box is created based on the centroid o Denver and a radius or creating the maximum extend using WGS 84 coordinates. The minimum

bounding box was created to select regions included for further analyses. First, coordinates of Denver was set as the center point. Second, range of longitudes and latitudes of surrounding regions were set to 0.5. Therefore, here's a preview of the subset of Colorado data selected based on this minimum bounding box.

```
[ ] # select regions near Denver
    # here are the denver coordinates
    dlat = 39.7
    dlon = -104.99
    # create a bounding box of +/- 1 degree
    df_co2 = df_co[(df_co['Latitude']>dlat-0.5)&(df_co['Latitude']<dlat+0.5)&(df_co['Longitude']>dlon-0.5)&(df_co['Longitude']<dlon+0.5) ]
```

```
df_co2.head()
```

	Zip	City	State	Latitude	Longitude
5	80281	Denver	CO	39.7439	-104.9876
10	80233	Denver	CO	39.9030	-104.9544
18	80001	Arvada	CO	39.5220	-105.2239
25	80419	Golden	CO	39.5220	-105.2239
32	80247	Denver	CO	39.6936	-104.8766

Figure 1. A preview of surrounding regions of Denver and location information

2.3 Feature Selection

With considerations of the overall purposes of this study, the zip code, city information, longitude, latitude were selected for further analyses. In addition, venue category information was also included because the key stakeholders (i.e., the business owners) would be interested to explore what kinds of business were clustered or grouped in a specific area. By having this information, business owners can make informed decisions about whether to put investment in that regions. The venue data is obtained using Foursquare API. Nevertheless, due to the adjacency o all the selected regions, it is highly likely that there are some duplicates within the venues information pulled. To avoid this, I have dropped all the duplicates.

3. Exploratory Data Analysis

3.1 Distribution of Venues

To directly observe the overall distributions of locations of venues, a map (Figure 2.) provides us with visualizations of relevant information, which also serves as a preliminary presentation of venue location distributions across neighborhoods around Denver metropolitan area in Colorado. Through this map, we can observe that most venues were located in downtown Denver with most points overlapped in that area.

Moreover, we can observe in this map that the venues are more concentrated around Denver central downtown areas including the union station, northwest downtown, and cherry creek neighborhood in southwest.

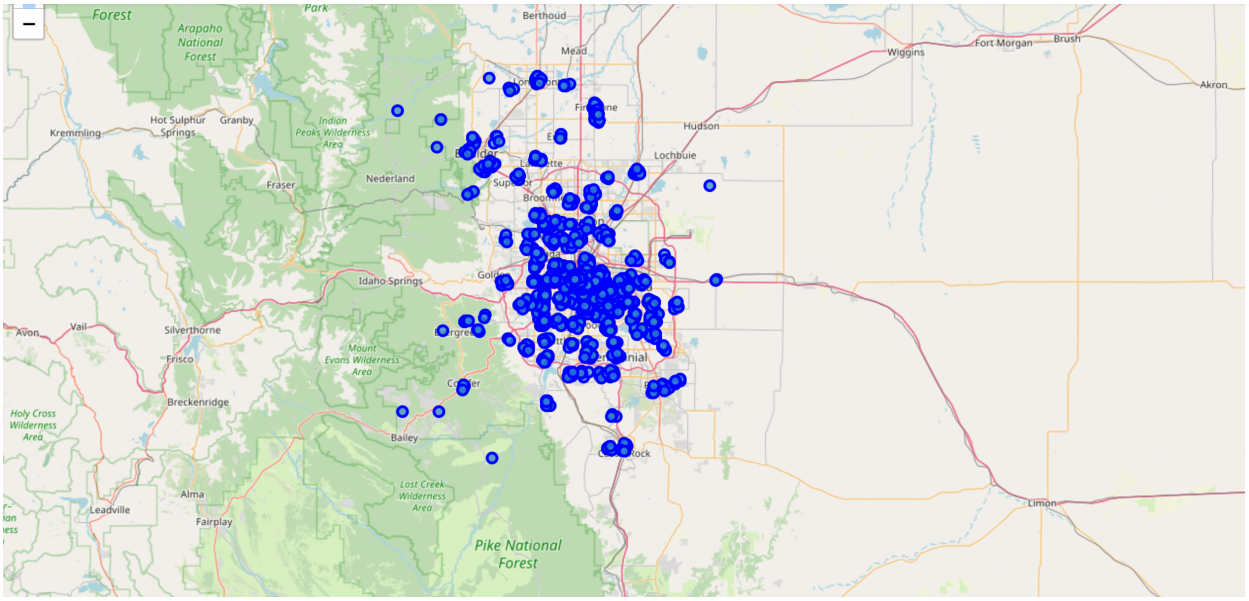

Figure 2. Distribution of data points.

In the next step, I aggregate the data based on location to explore how many venues are selected based on each neighborhood.

Arvada	77	77	77	77	77	77
Aurora	267	267	267	267	267	267
Bailey	1	1	1	1	1	1
Boulder	93	93	93	93	93	93
Brighton	42	42	42	42	42	42
Broomfield	35	35	35	35	35	35
Buffalo Creek	1	1	1	1	1	1
Castle Rock	42	42	42	42	42	42
Commerce City	9	9	9	9	9	9
Conifer	13	13	13	13	13	13
Dacono	13	13	13	13	13	13
Denver	2376	2376	2376	2376	2376	2376
Dupont	6	6	6	6	6	6
Eldorado Springs	4	4	4	4	4	4
Englewood	89	89	89	89	89	89
Erie	5	5	5	5	5	5
Evergreen	1	1	1	1	1	1
Firestone	11	11	11	11	11	11
Frederick	8	8	8	8	8	8
Golden	37	37	37	37	37	37
Henderson	10	10	10	10	10	10
Hygiene	1	1	1	1	1	1
Idledale	2	2	2	2	2	2
Indian Hills	3	3	3	3	3	3
Kittredge	8	8	8	8	8	8
Lafayette	37	37	37	37	37	37
Littleton	181	181	181	181	181	181

Lochbuie	1	1	1	1	1	1
Lone Tree	12	12	12	12	12	12
Longmont	66	66	66	66	66	66
Louisville	12	12	12	12	12	12
Morrison	2	2	2	2	2	2
Parker	31	31	31	31	31	31
Pine	1	1	1	1	1	1
Sherrelwood	21	21	21	21	21	21
Thornton	4	4	4	4	4	4
Ward	1	1	1	1	1	1
Watkins	3	3	3	3	3	3
Westminster	101	101	101	101	101	101
Wheat Ridge	23	23	23	23	23	23

Figure 4. Data aggregation results by neighborhoods.

Then, in the following step, I explore some of the most common venues in each neighborhoods.

	City	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Arvada	Park	Fast Food Restaurant	Mexican Restaurant	Pizza Place	Pharmacy	Construction & Landscaping	Convenience Store	Big Box Store	Discount Store	Home Service
1	Aurora	Convenience Store	Mexican Restaurant	Fast Food Restaurant	Sandwich Place	Pizza Place	Coffee Shop	Park	Chinese Restaurant	Liquor Store	Pharmacy
2	Bailey	Bakery	Yoga Studio	Food Truck	Exhibit	Eye Doctor	Fabric Shop	Farm	Farmers Market	Fast Food Restaurant	Filipino Restaurant
3	Boulder	Trail	Park	Bus Stop	Café	Gym	Gym / Fitness Center	Marijuana Dispensary	Shopping Mall	Sandwich Place	Chinese Restaurant
4	Brighton	Fast Food Restaurant	Mexican Restaurant	Hotel	Breakfast Spot	Gas Station	Liquor Store	Sandwich Place	Salon / Barbershop	Pizza Place	Coffee Shop

Figure 5. Sample result of data aggregation results by venues in all neighborhoods.

3.2 K-mean clustering

A k-mean clustering analysis was conducted using the data mentioned above in order to explore some patterns within this dataset. Due to the data distribution, my initial assumption is that there could be multiple clusters. Therefore, I set clusters k = 5. Then, I utilize the kmeans tool provided my sklearn to run the clustering analysis. Here's the result:

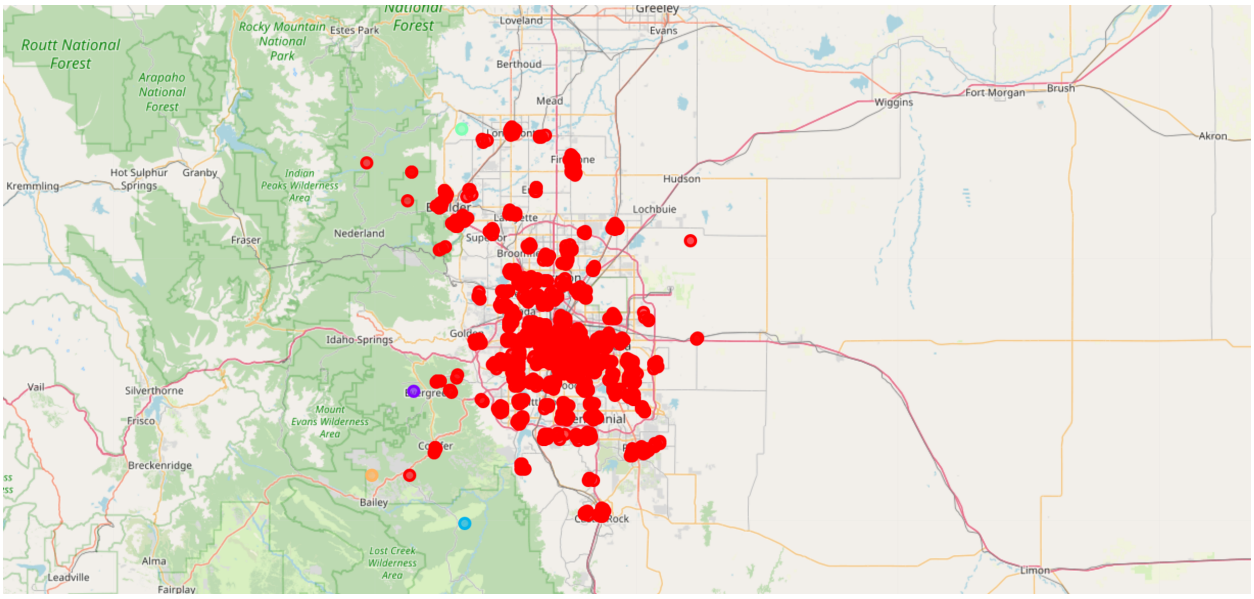

Figure 6. Clustering results.

The data shows that most of the venues are clustered as one single group mostly within the Denver region. There are few outliers in the Rocky mountain regions. This shows that it is better to set up a business within the Denver region since there are more diversity, high economic activity, and potentially higher demand due to high population as well.

In order to confirm this idea. I have also calculated the geographical mean center of all the venues. The geographical mean center is derived based on the distribution of all venues' location information. Here's the venues' mean center.

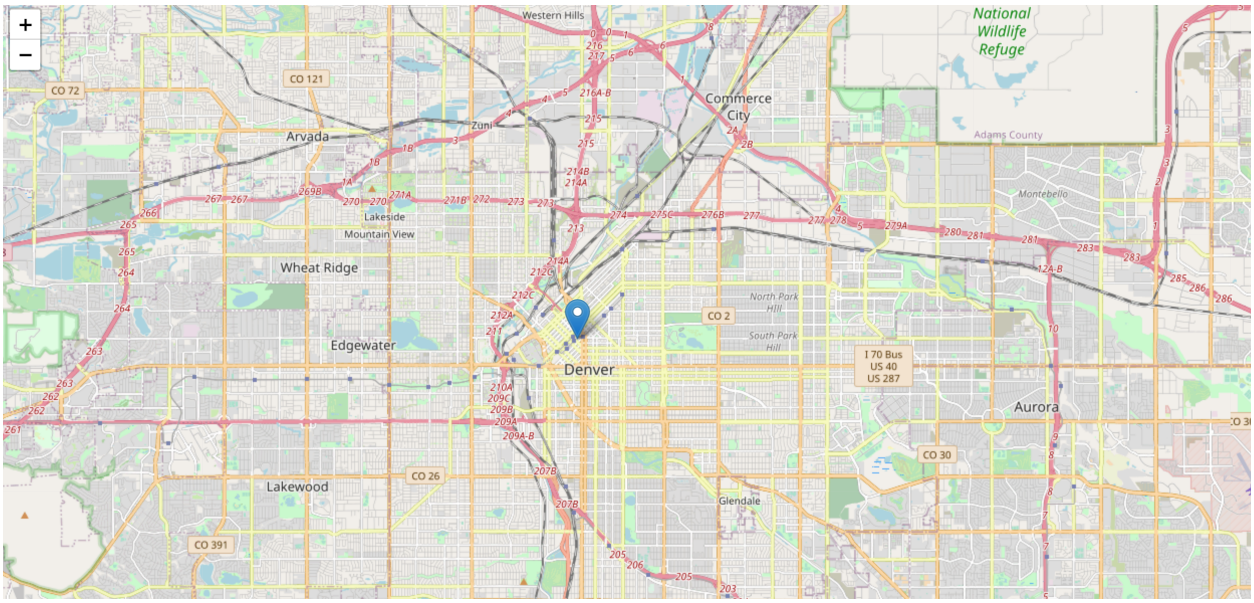

Figure 7. Geographical mean center.

The mean center of this dataset is located in the downtown region of Denver.

4. Discussion

Based on the clustering and mean center analysis, I am confident to recommend this person for starting the new business in the downtown Denver area (as known as LODO). First of all, due to the better transportation system (as seen on the map), it is very likely for this place to attract a lot of potential customers. Second, based on the clustering analysis, there is a very few small clusters located in the Rocky Mountains. Therefore, if the business is located in the mountains, it may not have enough customers and suffer from seasonal variations. Third, since LODO is the center for all venues, it can benefit from the vibrant economic activity and potentially from competition as well.

5. Conclusion

In this study, I have analyzed the distribution of business in Denver and its neighboring regions. Based on the venues data collected from Foursquare, I highly recommend that the new business should be established in the downtown region instead of in the Rocky Mountains.