Suggesting Best Neighborhood for a Business

Jair

June 2020

1. Introduction

1.1 Background

About a quarter million people are moving to Colorado every year. Denver itself was in the top 5 cities with the most inbound growth of the year 2020. Even big business and tech companies are moving to Colorado. One of those being PalantirMany of these people are looking to set up new business and may not know what city they would want to set up their business.

1.2 Problem

Many of these new people coming in want to set up business but are so new to the area that they don't know what area is best.

1.3 Interest

As someone who lives in Colorado, I am very interested in the economics of Colorado and what businesses are most prevalent in each area. I also hope to create a business some day and choosing a location is part of that.

2 Data Acquisition and Cleaning

2.1 Data Sources

The data set I acquired was from simplemaps and includes data sources that include U.S. Postal Service™, U.S. Census Bureau, National Weather Service, American Community Survey, and the IRS.

2.2 Data Cleaning

I didn't need to clean the data too much. I merged the two data sets together on the zip code in order to get a data set with the neighborhood names, zip codes, latitude, and longitude.

I also got data from foursquare in order to get all the venues in each venue and what venue category they were in. It also gives the latitude and longitude of each venue.

The dataset has all the neighborhoods in Denver with their latitude and longitude. Each neighborhood has all the corresponding venues in the area with their latitude and longitude as well.

3 Exploratory Data Analysis

I ran this code to count the number of venues in each neighborhood.

```
denver_venues.groupby('Neighborhood').count()
```

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Athmar Park	10	10	10	10	10	10
Auraria	14	14	14	14	14	14
Baker	3	3	3	3	3	3
Barnum West	3	3	3	3	3	3
Barnum	14	14	14	14	14	14
...	...	...	...	...	...	...
West Colfax	14	14	14	14	14	14
West Highland	22	22	22	22	22	22
Westwood	3	3	3	3	3	3
Whittier	9	9	9	9	9	9
Windsor	5	5	5	5	5	5

77 rows × 6 columns

Also found that there were 162 different unique venue categories.

I hot encoded whether each venue appeared in each neighborhood with either a 1 or a 0. I then find the mean/frequency amount of times each venue category appears in each neighborhood.

```
----Athmar Park ----
                         venue  freq
0           Convenience Store   0.2
1                 Sports Bar   0.1
2                        Park   0.1
3                     Brewery   0.1
4  Construction & Landscaping   0.1


----Auraria ----
        venue  freq
0  Print Shop  0.07
1    Pharmacy  0.07
2        Café  0.07
3  Taco Place  0.07
4         Gym  0.07
```

I then found the top 5 most frequent venues in each neighborhood to print out. The top picture is an example of this.

From this I created a new data set that showed the top 10 most common venues in each neighborhood.

	Neighborhood	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	Athmar Park	Convenience Store	Gym	Pharmacy	Clothing Store	Park	Brewery	Construction & Landscaping	Sports Bar	Home Service	Disc Golf
1	Auraria	Convenience Store	Distillery	Mexican Restaurant	Spa	Pharmacy	Baseball Field	Taco Place	Café	Light Rail Station	Discount Store
2	Baker	Pizza Place	Scenic Lookout	Flower Shop	Wine Bar	Distillery	Fast Food Restaurant	Eye Doctor	Event Space	Dog Run	Doctor's Office
3	Barnum West	Locksmith	Disc Golf	Park	Wine Bar	Dive Bar	Flea Market	Fast Food Restaurant	Eye Doctor	Event Space	Dog Run
4	Barnum	Convenience Store	Distillery	Mexican Restaurant	Spa	Pharmacy	Baseball Field	Taco Place	Café	Light Rail Station	Discount Store
...	...	...	...	...	...	...	...	...	...	...	...
72	West Colfax	Convenience Store	Distillery	Mexican Restaurant	Spa	Pharmacy	Baseball Field	Taco Place	Café	Light Rail Station	Discount Store
73	West Highland	Marijuana Dispensary	Convenience Store	Chinese Restaurant	Breakfast Spot	Bed & Breakfast	Gym / Fitness Center	Grocery Store	Liquor Store	Mexican Restaurant	Event Space
74	Westwood	Locksmith	Disc Golf	Park	Wine Bar	Dive Bar	Flea Market	Fast Food Restaurant	Eye Doctor	Event Space	Dog Run
75	Whittier	Park	Food	French Restaurant	Rental Car Location	Trail	Bar	Dog Run	Grocery Store	Dive Bar	Fast Food Restaurant
76	Windsor	Lake	Bar	Pool	Construction & Landscaping	Cycle Studio	Deli / Bodega	Cosmetics Shop	Flea Market	Fast Food Restaurant	Eye Doctor

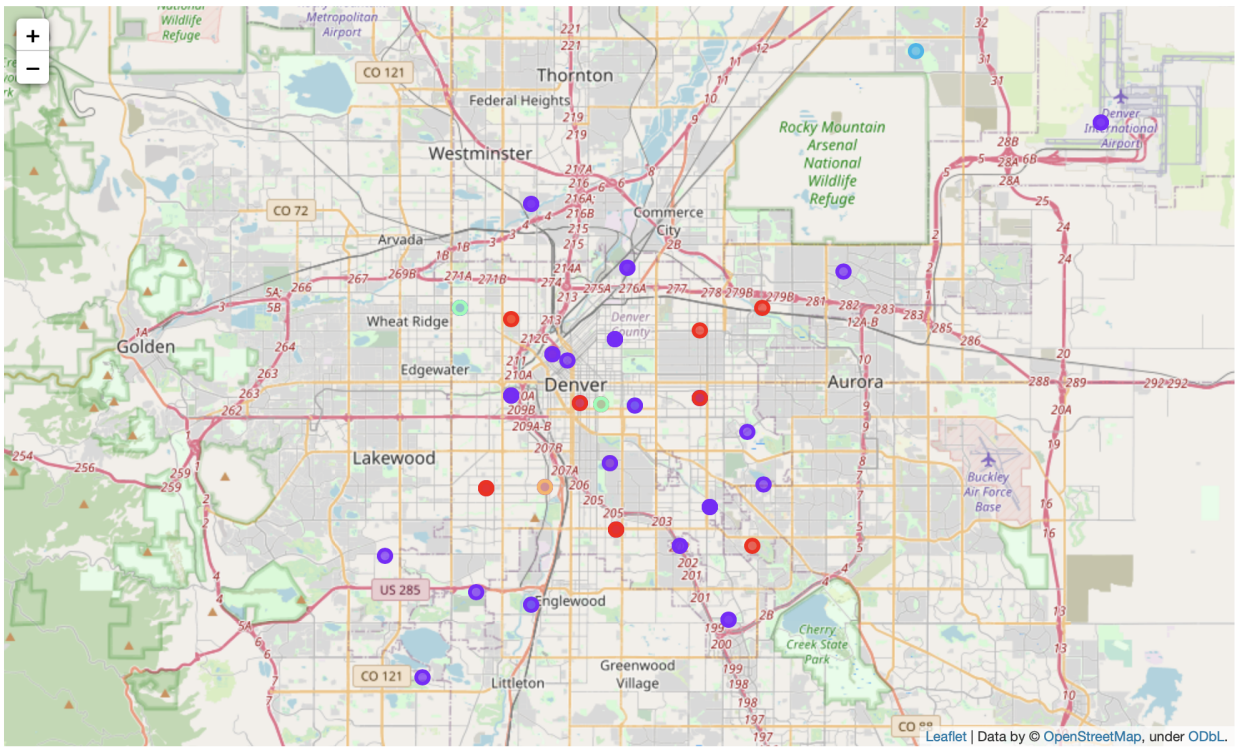
Here is an example of the data set with each venue with the top 10 most common ones.

4. Regression Models

The types of regression models that I user were K Nearest Neighbors and Logistic Regression

In order to find which business are best in each neighborhood or zip code

4.1 Clustering



4.2 Logistic Regression Model

Using the dummy variables that I created, I fitted the dataset with all the venues. What I wanted to predict

```
Alternative Healer
[80216]
Beer Garden
[80216]
Breakfast Spot
[80210]
Burger Joint
[80210]
Coffee Shop
[80210]
Construction & Landscaping
[80210]
Lake
[80210]
Locksmith
[80210]
Marijuana Dispensary
[80210]
Mexican Restaurant
[80210]
Nightclub
[80210]
Other Repair Shop
[80210]
Art Gallery
[80123]
Bar
[80247]
Baseball Stadium
[80236]
Café
[80221]
Cocktail Bar
[80221]
Coffee Shop
[80220]
Food
[80204]
Grocery Store
[80204]
Gym
[80205]
Italian Restaurant
[80205]
Mexican Restaurant
[80205]
Park
[80205]
Pharmacy
[80205]
Convenience Store
[80210]
Dive Bar
[80210]
Flower Shop
[80210]
French Restaurant
[80210]
Gay Bar
[80210]
Grocery Store
[80210]
Gym / Fitness Center
[80210]
Liquor Store
```

was what zip code matched the venues' dummy variables.

Below shows the output of each zipcode with what venue category is more popular depending on which other two venues were closely related to your venue as well.

4. Methodology/ Results

Some of the exploratory data analysis that I completed was learning a little more about the data I had created. I found out a little about each city, the size of the dataframe, and learning more about the data types. I observed some of the duplicate values that were in the data set and was able to handle them accordingly.

Some of the machine learning algorithms that I conducted were clustering and logistic regression.

Some of the things that I noticed when clustering neighborhoods that were similar were that cluster 0 had a lot of different places you could buy alcohol /drugs or just hang out at night like at bars or not night clubs. Cluster 1 had a lot of grocery stories/places you could buy different things. There were quite a lot of retail stores which we don't see in other clusters. We can presume there might be a shopping mall around those. There were many parks and scenic routes so this tells us these areas might have beautiful landscapes and sights compared to the other neighborhoods in other clusters. In Cluster 2, the neighborhood's most popular venues were mostly places like restaurants and coffee shops. These neighborhoods have a lot of dining and food options around them. Cluster 3 was the smallest of the 5. The neighborhoods here had a lot of dining but also had a lot of exercise and healthy venues like gyms and pilates studios. In Cluster 4, it looks like Cluster with a lot of venues that sell alcohol and drugs but there are more restaurants.

I also used logistic regression in order to find the best zipcode to build a certain venue depending on what venue was most likely to be in each zip code. A lot of the neighborhoods have the same zip codes. What I found was that some of the more popular zip codes were 80210 and 80204. 80210 had a lot of dining and groceries stores in it. 80204 was lisited most of the time. It had a lot of restaurants and shops. As well as outdoor places like trails and pools. This zipcode has a lot of venues that are a lot of fun and could be great for kids. 80247 has a lot of bars. This means the demand for them is definitely there so you can't go wrong with building bars in that zip code. 80221 is popular for cafes and building one there would be optimal. If you were looking to set up a physical therapy business or other health related business, 80216 is a popular choice for alternative healers.

5. Discussion

Like we discussed previously, many zipcodes are ideal for different businesses. From the logistic regression model predictions, we see exactly which zipcodes are best for each business. For any business related to food. 80204 is perfect since it has a lot of food related venues that include:

mexican, sushi, sports bar, and sandwich. The night time life here also is good so any business that caters towards night time clients would be ideal. We see that 80204 has a lot of nightclubs and sports bars. 80210 also is a good option for dining especially if you want to make a burger business, coffee shop, or a french restaurant. It has a lot of liquor stores and also marijuana shops. Marijuana being something super popular is a perfect business for this zip code. I would say these two zipcodes are the most popular and really encompass a lot of different areas.

6. Conclusion

Neighborhoods in Denver are a really good place to start businesses. There is a constant flow of new people, new housing being built, and more demand for different people. Creating a business is a big dream many people have. The dream of being your own boss and in charge of your own schedule. It can provide financial freedom to some; however, it can also be a burden and a money dump. That is why it is important to find the most optimal place for your business. A change in a couple of numbers in a zip code can make a big difference for a business in revenue. That's why this report is so important so people can make the right choice.