**Finding a Chicago community to open a Mexican Restaurant**

Rajesh

09/08/2019

Introduction:

|  |
| --- |
| A friend of mine is a good cook and moving from another city to Chicago with a plan to open a restaurant. He asked my help to find a best community that fits his needs. His requirements are, The place and surroundings should have sizable middle age people and good employment rate. Also since he is going to open a Mexican restaurant there should be a sizable Latino and Hispanic community in the neighborhood but not with many Mexican restaurants around. Also his restaurant price range is going to be Moderate to High so the community income should be decent to afford. |
|  |  |
|  |  |
| Datasets  To Solve the problem, I need the following datasets:  1) As the restaurant is going to carry moderate to high price range dishes , I need the Communities in Chicago, their employment rates and per capita income  2) As the need is to open in a community with decent employment rates and sizeable Latino community and middle age groups, I need a dataset with demographic information with age.  3) I need to query the foursquare API to address the final need, make sure there are not many Mexican restaurants arounds and their price ranges to get an idea of the community  After careful research, Here are a few datasets I have found that helps me find the best place that fits his needs:  Census data to help with first dataset need above:  <https://datahub.cmap.illinois.gov/dataset/2010-census-data-summarized-to-chicago-community-areas>  Dataset Attributes:   |  |  | | --- | --- | | **Attribute** | **Description** | | Geog | Unique ID for each community | | GeogKey | Community Name | | Total Population | Population Count, 2010 Census | | Not Hispanic or Latino but White alone | Population Count, Not Hispanic or Latino but White alone | | Not Hispanic or Latino but Black or African American alone | Population Count, Not Hispanic or Latino but Black or African American alone | | Not Hispanic or Latino but American Indian and Alaska Native alone | Population Count, Not Hispanic or Latino but American Indian and Alaska Native alone | | Not Hispanic or Latino but Asian alone | Population Count, Not Hispanic or Latino but Asian alone | | Not Hispanic or Latino but Native Hawaiian and Other Pacific Islander alone | Population Count, Not Hispanic or Latino but Native Hawaiian and Other Pacific Islander alone | | Not Hispanic or Latino but Some Other Race alone | Population Count, Not Hispanic or Latino but Some Other Race alone | | Not Hispanic or Latino but Two or More Races | Population Count, Not Hispanic or Latino but Two or More Races | | Hispanic or Latino | Population Count, Hispanic or Latino | | Male: to 5 yrs | Population Count: Male: to 5 yrs | | Male: 5 to 9 yrs | Population Count: Male: 5 to 9 yrs | | Male: 10 to 14 yrs | Population Count: Male: 10 to 14 yrs | | Male: 15 to 17 yrs | Population Count: Male: 15 to 17 yrs | | Male: 18 and 19 yrs | Population Count: Male: 18 and 19 yrs | | Male: 20 yrs | Population Count: Male: 20 yrs | | Male: 21 yrs | Population Count: Male: 21 yrs | | Male: 22 to 24 yrs | Population Count: Male: 22 to 24 yrs | | Male: 25 to 29 yrs | Population Count: Male: 25 to 29 yrs | | Male: 30 to 34 yrs | Population Count: Male: 30 to 34 yrs | | Male: 35 to 39 yrs | Population Count: Male: 35 to 39 yrs | | Male: 40 to 44 yrs | Population Count: Male: 40 to 44 yrs | | Male: 45 to 49 yrs | Population Count: Male: 45 to 49 yrs | | Male: 50 to 54 yrs | Population Count: Male: 50 to 54 yrs | | Male: 55 to 59 yrs | Population Count: Male: 55 to 59 yrs | | Male: 60 to 61 yrs | Population Count: Male: 60 to 61 yrs | | Male: 62 to 64 yrs | Population Count: Male: 62 to 64 yrs | | Male: 65 to 66 yrs | Population Count: Male: 65 to 66 yrs | | Male: 67 to 69 yrs | Population Count: Male: 67 to 69 yrs | | Male: 70 to 74 yrs | Population Count: Male: 70 to 74 yrs | | Male: 75 to 79 yrs | Population Count: Male: 75 to 79 yrs | | Male: 80 to 84 yrs | Population Count: Male: 80 to 84 yrs | | Male: 85 yrs and over | Population Count: Male: 85 yrs and over | | FeMale: Under 5 yrs | Population Count: FeMale: Under 5 yrs | | FeMale: 5 to 9 yrs | Population Count: FeMale: 5 to 9 yrs | | FeMale: 10 to 14 yrs | Population Count: FeMale: 10 to 14 yrs | | FeMale: 15 to 17 yrs | Population Count: FeMale: 15 to 17 yrs | | FeMale: 18 to 19 yrs | Population Count: FeMale: 18 to 19 yrs | | FeMale: 20 yrs | Population Count: FeMale: 20 yrs | | FeMale: 21 yrs | Population Count: FeMale: 21 yrs | | FeMale: 22 to 24 yrs | Population Count: FeMale: 22 to 24 yrs | | FeMale: 25 to 29 yrs | Population Count: FeMale: 25 to 29 yrs | | FeMale: 30 to 34 yrs | Population Count: FeMale: 30 to 34 yrs | | FeMale: 35 to 39 yrs | Population Count: FeMale: 35 to 39 yrs | | FeMale: 40 to 44 yrs | Population Count: FeMale: 40 to 44 yrs | | FeMale: 45 to 49 yrs | Population Count: FeMale: 45 to 49 yrs | | FeMale: 50 to 54 yrs | Population Count: FeMale: 50 to 54 yrs | | FeMale: 55 to 59 yrs | Population Count: FeMale: 55 to 59 yrs | | FeMale: 60 to 61 yrs | Population Count: FeMale: 60 to 61 yrs | | FeMale: 62 to 64 yrs | Population Count: FeMale: 62 to 64 yrs | | FeMale: 65 to 66 yrs | Population Count: FeMale: 65 to 66 yrs | | FeMale: 67 to 69 yrs | Population Count: FeMale: 67 to 69 yrs | | FeMale: 70 to 74 yrs | Population Count: FeMale: 70 to 74 yrs | | FeMale: 75 to 79 yrs | Population Count: FeMale: 75 to 79 yrs | | FeMale: 80 to 84 yrs | Population Count: FeMale: 80 to 84 yrs | | FeMale: 85 yrs and over | Population Count: FeMale: 85 yrs and over | | Median Age | Population Count: Median Age | | Total Households | Population Count: Total Households |   In this dataset, we are particularly interested in the “GeogKey” for Community Name, “Hispanic or Latino” for the community population.  Socio-economic data of Chicago communities to help with first dataset need above:  <https://data.cityofchicago.org/Health-Human-Services/Census-Data-Selected-socioeconomic-indicators-in-C/kn9c-c2s2>  <https://github.com/LRT-YASH/TestDSRepo/blob/master/Demographics_Chicago_2008_2012%20(1).csv>  Dataset Attributes:   |  |  | | --- | --- | | **Attribute** | **Description** | | Community Area Number | Unique ID for each community | | COMMUNITY AREA NAME | Community Name | | PERCENT OF HOUSING CROWDED | PERCENT OF HOUSING CROWDED | | PERCENT HOUSEHOLDS BELOW POVERTY | PERCENT HOUSEHOLDS BELOW POVERTY | | PERCENT AGED 16+ UNEMPLOYED | PERCENT AGED 16+ UNEMPLOYED | | PERCENT AGED 25+ WITHOUT HIGH SCHOOL DIPLOMA | PERCENT AGED 25+ WITHOUT HIGH SCHOOL DIPLOMA | | PERCENT AGED UNDER 18 OR OVER 64 | PERCENT AGED UNDER 18 OR OVER 64 | | PER CAPITA INCOME | PER CAPITA INCOME | | HARDSHIP INDEX | HARDSHIP INDEX |   In this dataset, we are particularly interested in the “GeogKey” for Community Name, “Hispanic or Latino” for the community population.  Chicago communities Latitude and Longitude information:  <https://github.com/LRT-YASH/TestDSRepo/blob/master/Chicago_Lat_Long.csv>  Finally Foursquare API for Mexican restaurants in each community.  After combining all the datasets and removing the attributes(features) that are not needed, left over data frame looked like this:   |  | | --- | | Community Name | | % Aged Under 65 and above 15 | | Per Capita Income | | Total Population | | Hispanic or Latino | | Median Age | | Total Households | | Male: 15 to 70 yrs | | Female: 15 to 70 yrs | | Not Hispanic or Latino | | Hispanic or Latino % of Total | | Latitude | | Longitude | | Income Rank | | H&L\_Pop%\_Rank | | Age Rank | | Income\_&\_Hispanic\_Pop\_Combined\_Rank | | Count of Mexican Restaurants | |  |

### Plan of execution:

Step 1: Import the two datasets as CSV, remove the attributes that are not needed, merge some attributes as per need and changes cases/types

Step 2: Create addition attributes needed for analyzing the data

Step 3: Connect to foursquare and create a list of all the Mexican restaurants and their price tier by community

Step 4: Combine the foursquare data and the community data from the first three steps

Step 5: Segment the data into multiple groups/clusters and find the mean to see which cluster is evenly balanced between Income, Hispanic & Latino community and age group.

Model:

I need find out the communities that fit the requirement. There is no prediction involved which rules out the prediction algorithms like regressions. Between Classification and clustering, since the problem doesn’t have a dependent variable in the datasets, classification is also ruled out which leaves us with clustering. Among the clustering choices I would like to try the K-Means technique on the data set to find the cluster that fits the needs.

### **Results**:

Communities before and after clustering:

|  |  |
| --- | --- |
|  |  |

One cluster that attracts the attention is **cluster 3** which has the mean of income rank and Hispanic Population rank closer to each other and also with a decent median age which in my view indicates enough income levels to support the moderate price range restaurant.

