IBM Data Science Capstone

Recommending Location for an Indian Restaurant in Chicago Using Data Science

BY VISHNU VARDHAN REDDY

JANUARY 30, 2021

Business Problem

- Chicago is one of the biggest cities in the U.S.A with considerable cultural diversity, income level disparity and demographics.
- Chicago already has a big restaurant industry with over 7,300 restaurants catering to growing base of food lovers.
- With Ivy league and reputed educational institutions set up in the Illinois and Chicago area, such as University of Illinois, University of Chicago etc., many international students, particularly from India are expected to be part of the student community.
- How do we find the best location to open an Indian restaurant leveraging data analytics and machine learning techniques?

Data

- Required Data:
  1. List of Chicago neighborhoods
  2. Geocoding of the neighborhoods
  3. Using Foursquare API – getting venue information for each neighborhood

- The data collected for this project includes:
  I. Community area and per capita income (including hardship index)
  II. Census data by community area to understand the demographics
  III. Most recent crime data (prior twelve-month period) along with description and coordinate information
  IV. Active business licenses and location information

Foursquare API

Using the Foursquare API, corresponding venue names with category and location info is pulled and stored into a dataframe

```
venues = getNearbyVenues(names = df['COMMUNITY NO SPACES'],
                                  latitudes= df['Latitude'],
                                  longitudes= df['Longitude']
                                 )
```

```
len(venues['COMMUNITY NO SPACES'].unique())
```

```
77
```

```
venues.head()
```

	COMMUNITY NO SPACES	Community Latitude	Community Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	RogersPark	42.010531	-87.670748	El Famous Burrito	42.010421	-87.674204	Mexican Restaurant
1	RogersPark	42.010531	-87.670748	Morse Fresh Market	42.008087	-87.667041	Grocery Store
2	RogersPark	42.010531	-87.670748	Taqueria & Restaurant Cd. Hidalgo	42.011634	-87.674484	Mexican Restaurant
3	RogersPark	42.010531	-87.670748	Lifeline Theatre	42.007372	-87.666284	Theater
4	RogersPark	42.010531	-87.670748	Rogers Park Social	42.007360	-87.666265	Bar

Methodology

- One-hot Method:
  - Once the nearby venue information is gathered, we then analyze communities by creating a dataframe with “zero” or “one” value for each venue category for a corresponding community.
  - We then group the communities and find the mean score for each venue category.
- K-Means Clustering:
  - Now, we train the data with the K-means clustering algorithm and find clustered communities.
  - Number of clusters chosen were 3 for this analysis which yielded better classification.

```
df['Cluster Labels'].value_counts()
```

```
2    49
0    27
1     1
Name: Cluster Labels, dtype: int64
```

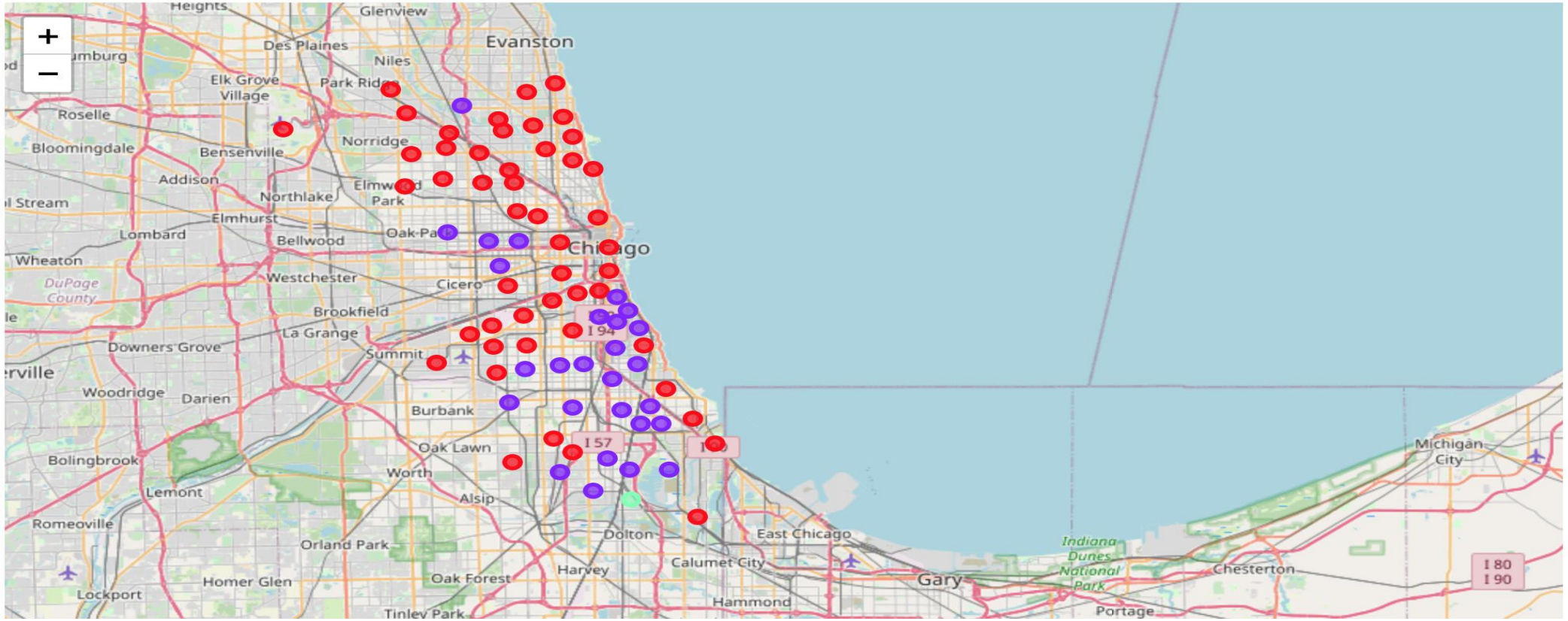
Top 10 Most Common Venues

```
venues_sorted.head()
```

	COMMUNITY NO SPACES	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	AlbanyPark	ATM	Newsstand	New American Restaurant	Neighborhood	Nature Preserve	National Park	Nail Salon	Music Venue	Music Store	Museum
1	ArcherHeights	ATM	Noodle House	Non-Profit	Nightlife Spot	Newsstand	New American Restaurant	Neighborhood	Nature Preserve	Office	National Park
2	ArmourSquare	ATM	Organic Grocery	Optical Shop	Office	Noodle House	Non-Profit	Nightlife Spot	Nightclub	Other Great Outdoors	Newsstand
3	Ashburn	ATM	Optical Shop	Office	Noodle House	Non-Profit	Nightlife Spot	Nightclub	Newsstand	Organic Grocery	New American Restaurant
4	AuburnGresham	ATM	Optical Shop	Office	Noodle House	Non-Profit	Nightlife Spot	Newsstand	New American Restaurant	Organic Grocery	Neighborhood

K-Means Clustering

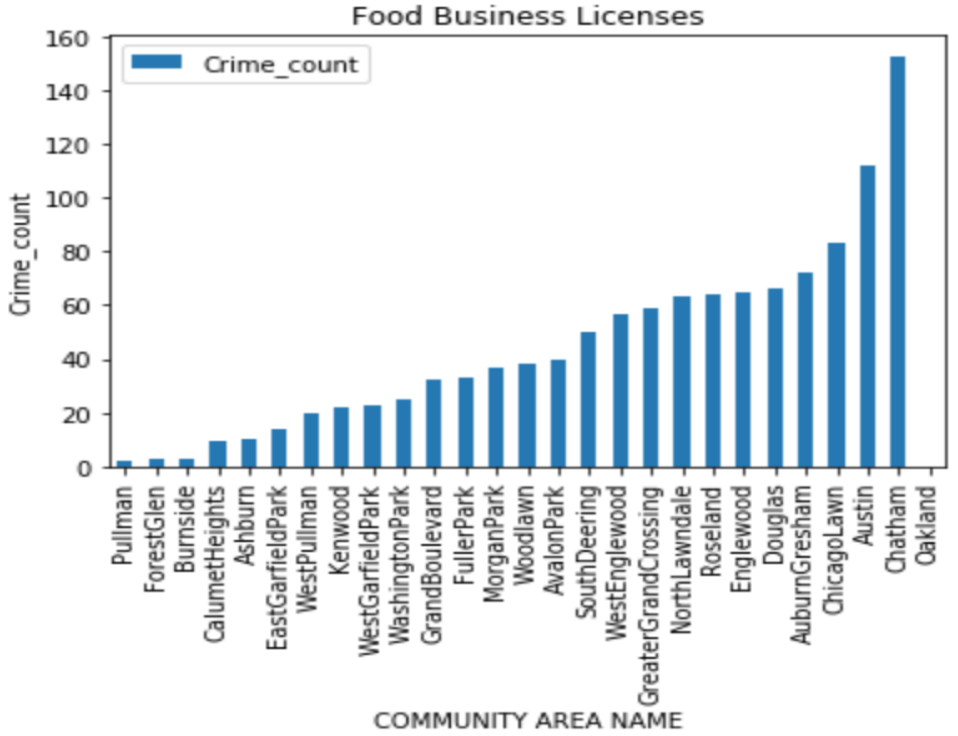
```
map_clusters
```



Visualization

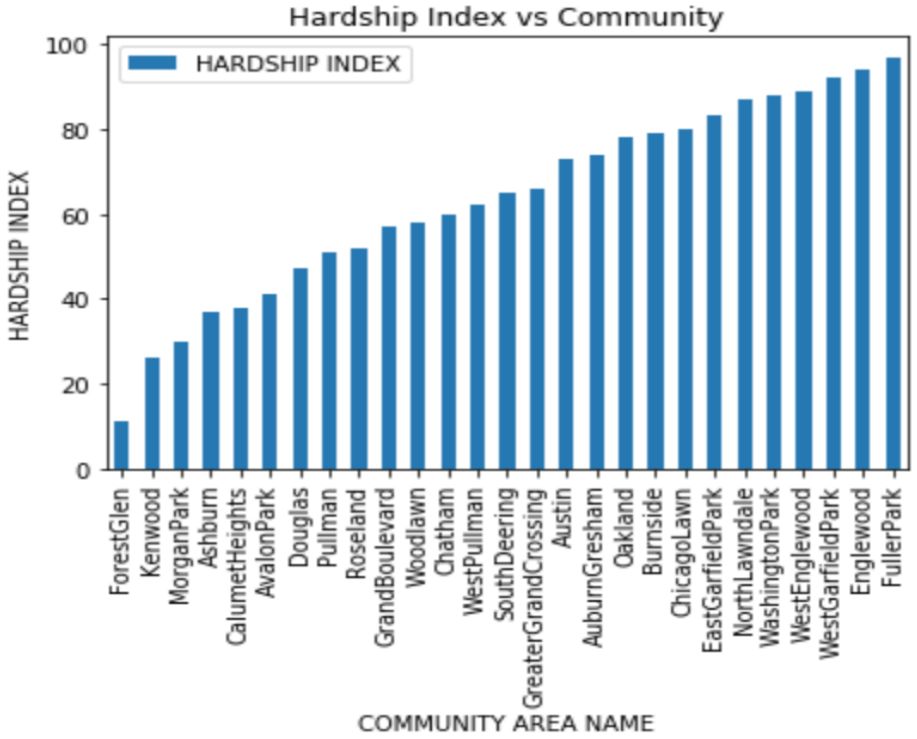
Community Area vs License Count

Community Area vs Crime Count

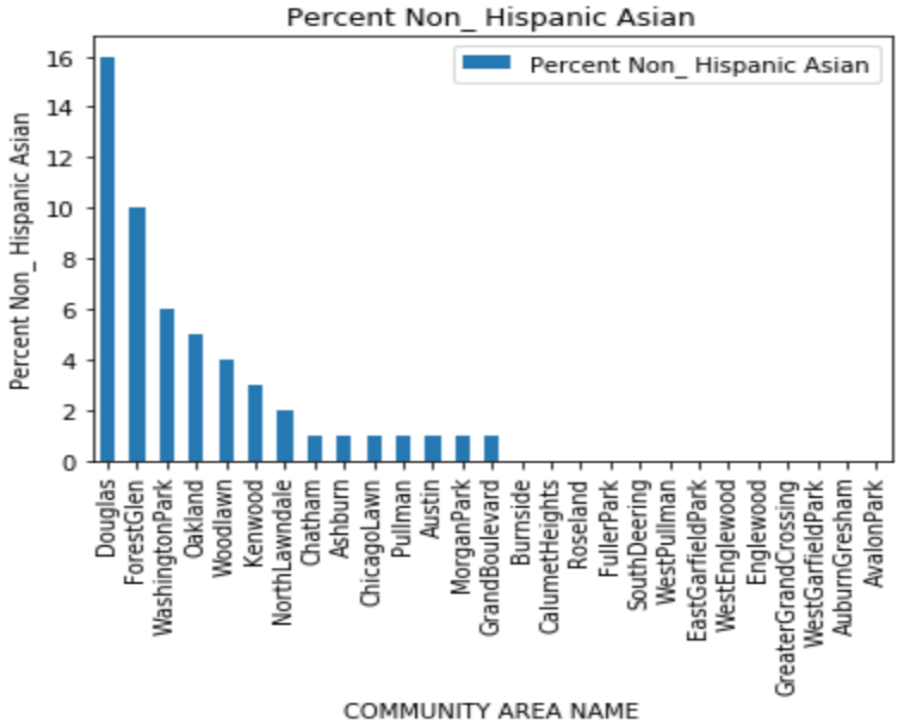


Visualization- Cont.

Community Area vs Hardship Index



Community Area vs Non-Hispanic Asian

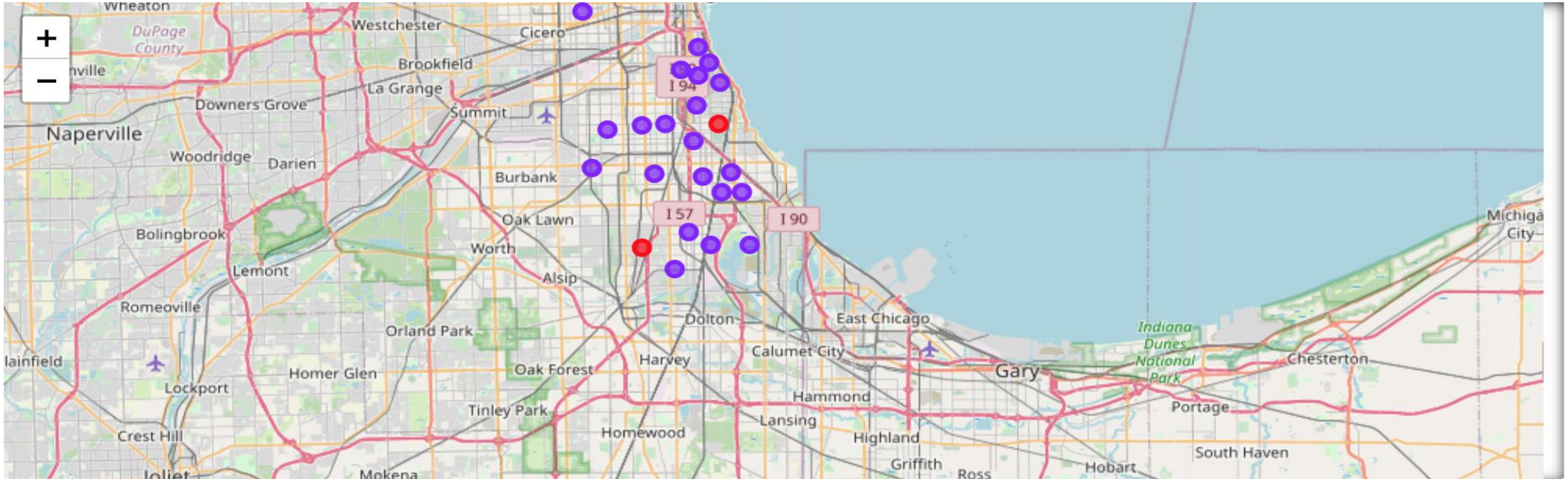


Results

Now, we identify the best communities with

- High business licenses (high business activity attracts higher workforce and customers for food business)
- Low crime count (safer community with lower insurance costs)
- Low hardship index (people have higher spending power)
- High Asian population (better chances of attracting people with similar food tastes and preferences)

Results – Best Two Communities



	COMMUNITY NO SPACES	License_Count	Crime_count	HARDSHIP INDEX	Percent Non_ Hispanic Asian
0	MorganPark	53.0	37.0	30	1
1	Woodlawn	53.0	38.0	58	4

Discussion & Recommendation

- Based on the analysis, it is beneficial to open a new Indian restaurant in either Morgan Park or Woodland communities.
- If we analyze these two communities, we note that these two communities are similarly positioned in the number of active businesses and crime counts in the areas.
- However, the hardship index is better in Morgan Park compared to the Woodlawn community and non-Asian population is higher in Woodlawn community.
- We can offer the client/ end customer with both options and decide a community based on the scale of the restaurant.
- If the client wants to set up a high scale restaurant, Morgan Park is the recommended community. If the plan is to set up a low/mid-scale restaurant, Woodlawn community is a better one.

Conclusion

- In this project, we have been able to understand the needs of the end customer and offer recommendations/ solutions using machine learning and proper data analysis techniques.
- This can be an iterative process with feedback and inputs for improvisation. There may arise a need to increase the cluster count or machine learning techniques depending on additional data.

References

- https://data.cityofchicago.org/Health-Human-Services/Census-Data-Selected-socioeconomic-indicators-in-C/kn9c-c2s2
- http://www.actforchildren.org/wp-content/uploads/2018/01/Census-Data-by-Chicago-Community-Area-2017.pdf
- https://data.cityofchicago.org/Public-Safety/Crimes-One-year-prior-to-present/x2n5-8w5q
- https://data.cityofchicago.org/Community-Economic-Development/Business-Licenses-Current-Active/uupf-x98q
- https://www.chicago.gov/city/en/about/facts.html

THANK YOU