Analysis of Chicago city for selecting optimal location for a cocktail bar

1. Introduction

Chicago, officially the City of Chicago, is the most populous city in the U.S. state of Illinois, and the third-most-populous city in the United States. With an estimated population of 2.7 million, it is also the most populous city in the Midwestern United States. Chicago is an international hub for finance, culture, commerce, industry, education, technology, telecommunications, and transportation. Chicago has one of the world's largest and most diversified economies, with more than four million employees and generating an annual gross regional product (GRP) of over $609 billion. The city is an efficient economic powerhouse, home to more than 400 major corporate headquarters, including 36 in the Fortune 500. Among the most diverse economies in the nation, Chicago is a key player in every sector from risk management innovation to manufacturing to information technology to health services. It has attracted many different players into the market. It is a global hub of business and commerce. This also means that the market is highly competitive. As it is a highly developed city, the cost of doing business is also extremely high. Thus, any new business venture or expansion needs to be analyzed carefully. The insights derived from the analysis will give a good understanding of the business environment which would help in strategically targeting the market. This will reduce risk as well as reap generous returns on the investment.

2. Business problem

The city of Chicago is famous for its flashy restaurants, bars and nightclubs at flamboyant neighborhoods. Cocktails bars have always been a go-to spot for most of the residents after a hard week at work to blow off steam. Starting a cocktail bar presents a really promising business opportunity but you have to distinguish yourself from the competition to expect long-term success.

Opening a cocktail bar would require a large capital, however, strategically selecting the best location for the venture would definitely result in better returns. My client is willing to open the cocktail bar in those community areas of Chicago whose residents have a high standard of living, the crime rates are comparatively low and the competition is less. My client is the owner of several medium sized businesses all across the globe and has ample of financial resources. He has hired me to present a comprehensive analysis of the city of Chicago and present to him a list of three neighborhoods which have great potential for opening a cocktail bar.

3. Data requirement and description

For the analysis I am leveraging four separate datasets which would provide me information about the coordinates of community areas in Chicago, the socioeconomic factors of the neighborhoods, crime rates of each neighborhood and fetch nearby points of interests using foursquare API.

Dataset 1: I am scraping a list of all 236 community areas of Chicago from Wikipedia(https://en.wikipedia.org/wiki/List_of_neighborhoods_in_Chicago) and then using geocoder I stored the latitude and longitude coordinates of each area in a pandas data frame.

	Community area	Latitude	Longitude
0	Albany Park	41.968290	-87.723380
1	Riverdale	41.654410	-87.602250
2	Edgewater	41.980460	-87.668340
3	Archer Heights	41.811540	-87.725560
4	Armour Square	41.834580	-87.631890
5	Ashburn	41.747850	-87.709950
6	Ashburn	41.941674	-88.198809
7	Auburn Gresham	41.743190	-87.655040
8	Avalon Park	41.745070	-87.588160
9	Avondale	41.939250	-87.711250
10	Irving Park	41.948461	-87.723240

Dataset 2: The second dataset is the socioeconomic factors of Chicago data which is available on the Chicago city data portal. Since the dataset is extremely large(> 1.5gb), I am using a snippet of the the dataset available at https://ibm.box.com/shared/static/05c3415cbfbtfnr2fx4atenb2sd361ze.csv. I am only using the hardship index as a parameter for my analysis.

	Community area	Latitude	Longitude	COMMUNITY_AREA_NUMBER	hardship index
0	Albany Park	41.96829	-87.72338	14.0	0.536082
4	Riverdale	41.65441	-87.60225	54.0	1.000000
8	Edgewater	41.98046	-87.66834	77.0	0.185567
13	Archer Heights	41.81154	-87.72556	57.0	0.680412
14	Armour Square	41.83458	-87.63189	34.0	0.835052

Dataset 3: For the crime data in Chicago, I am leveraging the Chicago city crime dataset on the portal. Again, since the dataset is extremely large, I will be using a snippet of the dataset available at https://ibm.box.com/shared/static/svflyugsr9zbqy5bmowgswqemfpm1x7f.csv.

	Community area	Latitude	Longitude	COMMUNITY_AREA_NUMBER	hardship index	Total crimes
0	Albany Park	41.96829	-87.72338	14.0	0.536082	5
1	Riverdale	41.65441	-87.60225	54.0	1.000000	2
2	Edgewater	41.98046	-87.66834	77.0	0.185567	2
3	Ashburn	41.74785	-87.70995	70.0	0.371134	8
4	Auburn Gresham	41.74319	-87.65504	71.0	0.752577	14

Dataset 4: For the final dataset I am using the foursquare API to fetch the details of all the bars at a distance of 10 km around all community areas. The data fetched consists of coordinates and type of the venue.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Albany Park	41.96829	-87.72338	Luxe Hookah Lounge	41.968836	-87.727922	Hookah Bar
1	Albany Park	41.96829	-87.72338	Rotana Cafe	41.968806	-87.728004	Hookah Bar
2	Albany Park	41.96829	-87.72338	240 Lounge	41.968283	-87.727558	Dive Bar
3	Albany Park	41.96829	-87.72338	Alpha Delta Dyke	41.961101	-87.717582	Fraternity House
4	Albany Park	41.96829	-87.72338	Albany Place	41.964036	-87.727688	American Restaurant