The Best Neighborhood in New Jersey for setting up a outpatient physical therapy facility

1. Introduction

1.1 Business Problem

A friend of mine is thinking of moving to New Jersey to open her physical therapy facility. Both of us met for a drink and she discussed her ideas with me. She has every other thing figured out except the best location (optimal neighborhood) to set up the facility. Her criteria for the optimal neighborhood is as follows:

- Less crime (safe areas)
- Cost of rent
- Close to long-term health care facilities e.g. hospitals, nursing homes, home-care centers, rehabs etc.

This report outlines some basic assumptions, data sets, and analysis that can inform our decision when selecting the optimal neighborhood in New Jersey for setting up a physical therapy facility.

1.2 Target audience

The target audience of this report would be anyone who wants to buy or set up a physical therapy business in New Jersey. This report will also be useful for government, hospitals or other health care facilities that are interested in setting up outpatient physical therapy facility.

2.0 Data

Making the best decision for the best neighborhood for a physical therapy facility is not a trivial process because several factors have to be considered such as the closeness to long-term health care facilities, areas with low crime rate and cost of rent. To be able to make the best decision, data is needed. Fortunately, New Jersey has several public databases that describes various aspects of the state and Foursquare API allows free access to some of its venue and location data. We will use five sets of data for our analysis. They are:

1. New Jersey Health care facilities data
2. New Jersey Crime Data
3. New Jersey average rent data

4. Foursquare data

2.1 New Jersey Health care facilities data

We have a list of all registered health care facilities in New Jersey from New Jersey Department of Health website - https://healthapps.state.nj.us/Facilities/fsSearch.aspx. We downloaded this data and place it into a pandas dataframe using python. This data will help us to know where most of the health facilities are located. Our physical therapy facility will be situated close to one of these, since most of the people in this health care facilities are likely to need physical therapy.

```
health = health.drop(['ALPHA_NAME', 'CSZ','TELEPHONE','FAXPHONE','FACEMAIL'], axis=1)
print(health.shape)
health.head()
```

```
(859, 9)
```

7]:

	FACILITY_TYPE	FACID	LICENSED_NAME	ADDRESS	FAC_CITY	COUNTY	Lic_Beds_Slots	LAT	LNG
0	ADULT DAY HEALTH CARE SERVICES	NJ80770	1st Cerebral Palsy of New Jersey	7 SANFORD AVENUE	BELLEVILLE	ESSEX	27	40.341649	-74.462594
1	ADULT DAY HEALTH CARE SERVICES	NJ708112	2nd Home Adult Medical Day Care	100 HAMILTON PLAZA GROUND FLOOR	PATERSON	PASSAIC	120	40.916300	-74.172438
2	ADULT DAY HEALTH CARE SERVICES	NJ308113	2nd Home East Orange	115 EVERGREEN PLACE	EAST ORANGE	ESSEX	150	40.762161	-74.222470
3	ADULT DAY HEALTH CARE SERVICES	NJ308116	2nd Home Newark Operations, LLC	717-727 BROADWAY	NEWARK	ESSEX	240	40.774457	-74.159509
4	ADULT DAY HEALTH CARE SERVICES	NJ308117	2nd Home Orange Operations, LLC	37 NORTH DAY STREET	ORANGE	ESSEX	110	40.773370	-74.228835

From the figure above, there is a total of 859 entries in the health care facility data frame. It has the name of each heath care facility, their type, ID, address, number of bed slots, latitude and longitude. This data frame contains other features but they have been dropped during the process of cleaning the data. We want to see the number of heath care facilities registered in New Jersey, grouped by city (neighborhood). After coding in python to get the neighborhood with the most heath care facilities, we get the data frame below:

In [61]:
```
health1 = health[['FAC_CITY','LAT']]
grouped_city = health1.groupby(['FAC_CITY'],as_index=False).count()
#grouped_city.head(10)
grouped_city.rename(columns={'LAT': 'Num_of_facility'}, inplace=True)
grouped = grouped_city.sort_values(by = 'Num_of_facility', ascending=False).reset_index()
grouped = grouped.drop(['index'],axis = 1)
grouped.head(10)
```

Out[61]:

	FAC_CITY	Num_of_facility
0	EDISON	16
1	TOMS RIVER	16
2	VOORHEES	14
3	NEWARK	13
4	WAYNE	13
5	WEST ORANGE	12
6	CHERRY HILL	12
7	LAKEWOOD	11
8	JERSEY CITY	10
9	BRICK	9

Edison and Toms river have the most heath care facilities in the state but the top 10 are all looking like they have a good amount of heath care facilities.

2.2 New Jersey Crime Data

We got the New Jersey crime data from the New Jersey state police website - https://www.njsp.org/ucr/uniform-crime-reports.shtml. This will help us to select the safest place for our physical therapy facility.

In [68]:
```
body = client_1f9700cb326643a6871f6fb9e313e64b.get_object(Bucket='segmentingandclusteringneighborho-donotdelete-pr-2hamo3a0fhscf
# add missing __iter__ method, so pandas accepts body as file-like object
if not hasattr(body, "__iter__"): body.__iter__ = types.MethodType( __iter__, body )

crime = pd.read_csv(body)
#crime.head(10)
crime = crime.drop(['NJ','Population','Violent','Law enforcement per 1,000','Property','Law enforcement','Crime rate per 1,000',
crime = crime.drop([0])
print('The total number of crime is ',crime['Total crimes'].sum())
crime.head(10)
```

```
The total number of crime is  118109.0
```

Out[68]:

	City	Total crimes
1	River Vale Township	25.0
2	Tenafly	41.0
3	Kinnelon	26.0
4	Bergenfield	77.0
5	Mount Olive Township	83.0
6	Dumont	54.0
7	Sparta Township	59.0
8	New Milford	57.0
9	Chatham Township	33.0
10	Warren Township	55.0

We have a total of 118,109 crime reports in this data frame. We grouped this data into different neighborhood using python code. This data frame gives us a good indication of how safe or dangerous each neighborhood is today.

2.3 New Jersey average rent report

We will use the New Jersey average rent report to select a neighborhood with affordable rent. For the purpose of this analysis, we will be flexible with this criteria. The data was downloaded from the website of RentCafe - https://infogram.com/rentcafe_new-jersey-rentreport-january2020-1h7k23n3m8dv6xr.

In [37]:
```
body = client_1f9700cb326643a6871f6fb9e313e64b.get_object(Bucket='segmentingandclusteringneighborho-donotdelete-pr-2hamo3a0fhscfj',Key='rent_prices.csv')['Body']
# add missing __iter__ method, so pandas accepts body as file-like object
if not hasattr(body, "__iter__"): body.__iter__ = types.MethodType( __iter__, body )

rent = pd.read_csv(body)
#print(rent.shape)
rent['Average rent'] = rent['Average rent'].str.replace(r'\D', '').astype(float)
rent['City'] = rent['City'].str.upper()
rent['City'] = rent['City'].map(lambda x: str(x)[:-4])
rent = rent.sort_values(by = 'Average rent', ascending=True)
rent.head(10)
#rent.dtypes
```

Out[37]:

	City	Average rent	Percentage increase
26	LINDENWOLD	988.0	-0.40%
3	CAMDEN	1013.0	-0.20%
2	BURLINGTON	1054.0	0.10%
51	TRENTON	1108.0	0.30%
22	IRVINGTON	1131.0	0.30%
34	NEWARK	1217.0	0.40%
17	HACKETTSTOWN	1221.0	-0.10%
29	MAPLE SHADE	1228.0	0.20%
44	PLAINFIELD	1232.0	0.90%
41	PATERSON	1260.0	0.40%

From the figure, the LINDENWOLD neighborhood has the lowest rent prices in the state.

2.4 New Jersey Neighborhood map

According to ontheworldmap.com - http://ontheworldmap.com/usa/state/new-jersey/ , there are 565 neighborhoods in New Jersey. The largest by population is Newark with 277,140 residents while the smallest is Tavistock.