

# Capstone Project - The Battle of Neighborhoods (Week 1)

## Best Neighborhood for retirement in South Florida

Machine learning allows for the creation of computational models capable of identifying patterns in multi-dimensional datasets. This project aims to utilize all Data Science Concepts learned in the IBM Data Science Professional Course i.e define a Business Problem, the data that will be utilized and using that data, analyze the data using Machine Learning tools. In this project, we will go through all the steps to provide a solution that can be leveraged by the business stakeholders to make their decisions.

### 1. Introduction/Business Problem:

#### Introduction:

South Florida, with its year-round warm weather, beaches, golf courses and flat geography, is an attractive option for retirement. First, the weather, older people hate cold and they want to spend their retirement days in a warm and sunny place. And the Sunshine State is the perfect place for this. Florida is also cheaper and has no income tax. This is a major factor for senior citizens living on restricted income. There are more seniors moving to Florida than in some other parts of the USA. Florida is a very popular travel destination. And many people who come to live in South Florida are the ones who have been visiting it for many years. And they already know all the pros and cons of living in Florida.

**Problem:** For someone considering retiring to South Florida, there are dozens of cities, hundreds of neighborhoods available, and it can be a daunting task deciding where to move to.

**The objective** of this project is to help someone who is unfamiliar with South Florida decide where to move to. Crime rate is the first factor to look at, since crime is prevalent in some communities. The program filters all the cities of Florida, to just the cities in South Florida counties, and then it uses crime statistics from FBI to mark the cities as green, yellow, orange and red, green being the safest cities. Next, since moving to a beach city is more desirable for most retirees, the program selects the safe-beach cities.

Finally, the project uses Foursquare location data and regional clustering of venue information to determine what might be the 'best' neighborhood out of the safe cities for retirement.

Through this project, we will find the most suitable location in South Florida for retirement.

## 2. Target Audience

This project is aimed towards retired or retiring people, or old people considering moving to South Florida. The analysis will provide vital information that can be used by the target audience.

## 3. Data Overview:

The data that will be required will from multiple sources which will provide the list of cities in Florida, neighborhoods in these cities (via Wikipedia), the Geographical location of the neighborhoods (via GeoPy Geocoder package) and Venue data (via Foursquare). The Venue data will help find which neighborhood is best suitable for retirement.

### 3.1. Data Acquisition

- List of all the Florida Cities by counties is available at:  
<https://dos.myflorida.com/library-archives/research/florida-information/government/local-resources/citycounty-list/counties/>

This Division manages the State Library and Archives, supports public libraries, directs records management services, and is the designated information resource provider for the state of Florida.

Data is in format that is suitable for analysis, its loaded to Pandas dataframe directly

```
11]: url='https://dos.myflorida.com/library-archives/research/florida-information/government
dffcities=pd.read_html(url, header=0)[0]
dffcities.columns = ['county','city','citytype']
dffcities.head()
```

ut[11]:

	county	city	citytype
0	Alachua	Alachua	city
1	Alachua	Archer	city
2	Alachua	Cross Creek	populated place
3	Alachua	Earleton	populated place
4	Alachua	Gainesville	city

- Crime Data for all Florida cities is available at FBI Website. The FBI collects these data through the Uniform Crime Reporting (UCR) Program. The table provides the volume of violent crime as reported by city and town law enforcement agencies. The link for Florida is: <https://ucr.fbi.gov/crime-in-the->

[u.s/2015/crime-in-the-u.s.-2015/tables/table-8/table-8-state-pieces/table 8 offenses known to law enforcement florida by city 2015.xls](https://ucr.fbi.gov/crime-in-the-u.s.-2015/crime-in-the-u.s.-2015/tables/table-8/table-8-state-pieces/table-8-offenses-known-to-law-enforcement-florida-by-city-2015.xls)

Data is in format that is suitable for analysis, its loaded to Pandas dataframe directly

```
flcrimeurl='https://ucr.fbi.gov/crime-in-the-u.s/2015/crime-in-the-u.s.-2015/tables/table-8/table
header = {
    "User-Agent": "Mozilla/5.0 (X11; Linux x86_64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/50
    "X-Requested-With": "XMLHttpRequest"
}
r = requests.get(flcrimeurl, headers=header)
flcrime = pd.read_html(r.text)
dfflcrime = flcrime[0].dropna(axis=0, thresh=4)
dfflcrime['crimeperk'] = round(1000*dfflcrime['Violentcrime']/dfflcrime['Population'],0)
dfflcrime.head()
df_flcrime = dfflcrime.filter(['City','crimeperk'], axis=1)
df_flcrime.columns = ['city','crimrate']
df_flcrime.head()
```

]:

	city	crimrate
0	Alachua	3.0
1	Altamonte Springs	4.0
2	Apalachicola	0.0
3	Apopka	4.0
4	Arcadia	8.0

- List of best beach cities, as of June 2020, is available from <https://wallethub.com/edu/best-beach-towns-to-live-in/36567/>

Data is in format that is suitable for analysis, its loaded to Pandas dataframe directly

```
urlbeaches = 'https://wallethub.com/edu/best-beach-towns-to-live-in/36567'
header = {
    "User-Agent": "Mozilla/5.0 (X11; Linux x86_64) AppleWebKit/537.36 (KHTML, like G
    "X-Requested-With": "XMLHttpRequest"
}
r = requests.get(urlbeaches, headers=header)
lsbeached = pd.read_html(r.text)
dfbeaches = lsbeached[0].dropna(axis=0, thresh=4)
dfbeaches[['city','state']] = dfbeaches['City'].str.split(', ',expand=True)
dfbeaches.head()
```

3]:

	' Rank	City	Total Score	'Affordability' Rank	'Weather' Rank
0	1	Naples, FL	62.50	30	100
1	2	Lahaina, HI	61.25	68	31
2	3	Boca Raton, FL	60.96	17	5
3	4	Newport Beach, CA	60.01	69	56
4	5	Santa Monica, CA	59.87	122	26

- Venue Data using Foursquare

The places by Foursquare API is a database of more than 105 million places worldwide and is going to be consulted for this project. To explore the cities, we will use the Venue Recommendation which returns a list of recommended venues near a certain location. In order to make the query in the Foursquare API we need the coordinates given in Latitude and Longitude for a given Neighborhood. So. the first thing me we need to do is to get coordinates for each Neighborhood in each of the cities.

To get the coordinates given Neighborhood we use Geopy geocoder. Geopy makes it easy for Python developers to locate the coordinates of addresses, cities, countries, and landmarks across the globe using third-party geocoders and other data sources. The Geopy geocoder documentation can be found in the following link: <https://geopy.readthedocs.io/en/stable/>

```
from geopy.geocoders import Nominatim
from geopy.extra.rate_limiter import RateLimiter

locator = Nominatim(user_agent="myGeocoder")
geocode = RateLimiter(locator.geocode, min_delay_seconds=1)
df_sfcities['location'] = df_sfcities['address'].apply(geocode)
df_sfcities['point'] = df_sfcities['location'].apply(lambda loc: tuple(loc.point))
df_sfcities[['latitude', 'longitude', 'altitude']] = pd.DataFrame(df_sfcities['point'].values, index=df_sfcities.index)
df_sfcities.head()
```

```
]:
```

	county	city	citytype	address	
0	Broward	Coconut Creek	city	Coconut Creek, Florida	(Coconut Creek, Broward
1	Broward	Cooper City	city	Cooper City, Florida	(Cooper City, Broward Co
2	Broward	Coral Springs	city	Coral Springs, Florida	(Coral Springs, Broward
3	Broward	Dania Beach	city	Dania Beach, Florida	(Dania Beach, Broward C