Finding the ideal location for Maggie & Maggie's Beach Library and Bar

Introduction

Two of my friends (both called Maggie) want to set up a business on a beach that is both a bar and a library, allowing customers to unwind on the sand with a book and a drink. The main requirements are:

- The bar must be located on a beach, in the USA
- States suggested are Florida and North Carolina
- Stakeholders have told me they will not consider setting the business up in are California, Alabama, Iowa, Kansas, Montana or Nebraska

The business problem as a one-liner:

"Where is the best location to set up Maggie and Maggie's Beach Library and Bar?"

My questions for the stakeholders, and the answers given:

1. **Can the business be located on a riverside, lakeside, or coastal only?**
   A: We would prefer a coastal beach, but ideally want somewhere with a lot of through traffic of tourists and locals
2. **Are there any states which are personal preferences?**
   A: Florida and North Carolina
3. **What other businesses or things do you want nearby?**
   A: Touristy-type businesses
4. **What other businesses or things do you want to be away from?**
   A: Other businesses serving drinks or those that are child-centric (eg schools or play parks)
5. **What size of city or town would you prefer?**
   A: No preference, but as stated above somewhere with a lot of foot traffic

Data: How Foursquare and other data will be used to help solve the problem

Data from Foursquare:

The business problem is a matter of location, therefore data from Foursquare will be useful in the following ways:

- Identifying locations where there is a beach. This will be the foundation of the data, as the location for the business must be a beach on a coast.
- Locating other businesses in the local area which could have a favourable or adverse impact on the business.

Data from other sources:

As well as the proximity of other local businesses, other data will be essential in producing a sound location proposal for Maggie & Maggie's Beach Bar and Library:

- **Population centres in the USA** (acquired from https://simplemaps.com/data/us-cities): the business will need to have a customer base. This dataset includes the latitude and longitude of the town / city, so it can therefore be narrowed down to areas that have a Foursquare beach location.

Methodology

1. Background: A brief statement on the notebook covering the purpose of the project. This section also shows questions put to the stakeholders in order to gain steer in regards to locations they would prefer / avoid. These requirements then feed into the variables used within the Python script.

2. Prepare the development environment: Importing of the various packages and libraries the Python development environment needs in order to produce the output. The variables are also created at this stage, based on information provided by the stakeholders, and assumptions.

Variables based on stakeholder are:

- *Excluded States*. US States the stakeholders have expressly stated they will not consider. These are:
    - California
    - Nebraska
    - Alabama
    - Iowa
    - Kansas
    - Montana
- *Preferred States*: US States the stakeholders would greatly prefer:
    - Florida
    - North Carolina

Variables based on assumptions are:

- Minimum population of a city to be considered. Set to 100 people.
- Maximum distance of a city from a beach. Set to 2000 meters.
- Local area distance. This is used to ascertain the population of the area surrounding the city with a beach. Set to 10 miles
- Minimum threshold for a local area to be considered viable. Set to 100,000 people.
- List of words to identify other businesses that would be good to have nearby:
    - 'Bar', 'Gift', 'Shop', 'Tour', 'Ice Cream' etc
- List of words to identify other businesses to be avoided:
    - 'School', 'Kid', 'Child', 'Play', 'Nursery' etc
- Range within which other businesses will be considered: Set to 1,000 meters

The US Cities Dataset is also imported at this stage. It is cleaned of rows where required information is not populated (such as city population is missing), preventing useless rows from being included in the analysis. Unneeded columns are also removed.

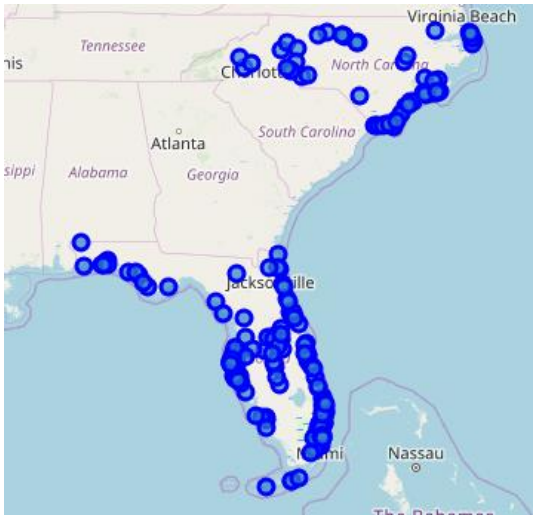
3. Clean the Data: Coastal Locations: The dataset is now refined, concentrating only on cities in Florida and North Carolina. The data is then passed through Foursquare, which provides location data based on latitude / longitude co-ordinates. These co-ordinates are present in the USA Cities dataset. Using the keyword of beach, the dataset is now further refined to cities in Florida and North Carolina which have or are close to a beach.

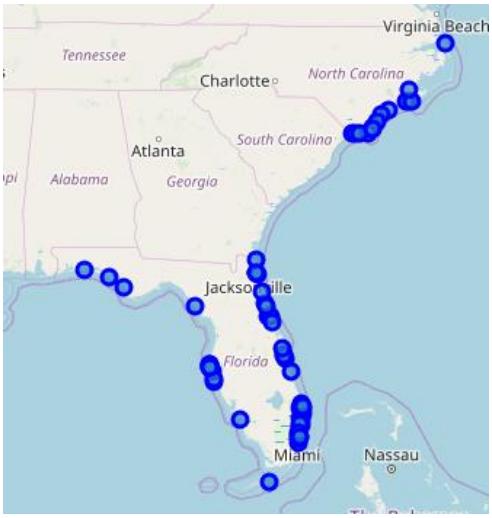
However, upon inspection of the plot there are many locations which are inland. Therefore, a further filter is applied to find locations with names matching the following list:

- Beach
- Coast

- Shore
- Sand
- Sands

After filtering, the dataset is optimized to coastal locations only. The number of locations is reduced from 224 to 64.



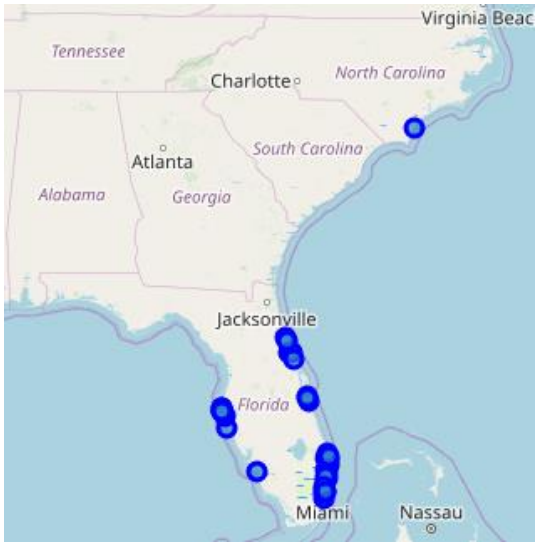


4. Enhance the Data: Local Area Populations: Each city remaining has latitude and longitude positions as per the USA Cities dataset. Using the local area setting, a range of latitude and longitude is determined for each city, covering an area up to 10 miles North, South East and West of the beach.

For each of these areas, the sum of the local population is then generated from the USA Cities dataset. This gives an indication of how busy the local area will be and potential foot traffic.

The logic for how to convert between differences of latitude/longitude and miles was obtained here: https://gis.stackexchange.com/questions/142326/calculating-longitude-length-in-miles

By this point, the field of prospective locations had been narrowed down to 38.



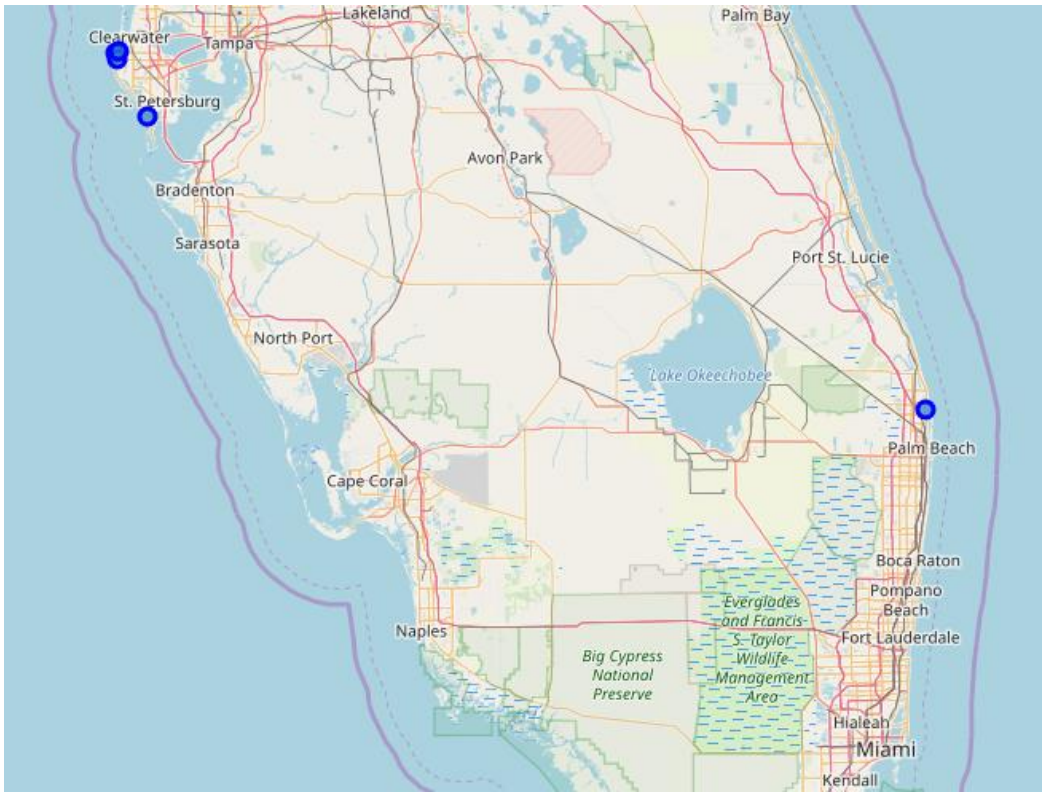
5. Enhance the Data: Other Local Businesses: The final aspect for adding insight to the analysis was including information about local businesses which would be advantageous or disadvantageous to have in the local area.

This was achieved by using the Foursquare data to identify instances of the lists of good and bad keywords, further narrowing the data.

The count of good matches minus bad matches gives a local business score.

6. Final Recommendation: Top 5

The top five recommended beach locations are based on the local area population divided by the business score. Therefore, the lower the number, the better, based on the logic of better business conditions as a ratio to the local population.



Results

The top recommendation for Maggie & Maggie's Beach Library and Bar is the area of Belleair Beach, Florida. Four of the top 5 recommendations for the business location are within this local area. The area is populous with a large area of beach, in a climate conducive for relaxation.

Discussion

Whilst Belleair beach has been recommended as the best location to start this business, this is based on assumptions I have entered for location distances and populations. If the stakeholders disagree with this recommendation, then we can re-visit the variables defined above, so I can revise the output of this process.

Further information which may be useful to include, but was difficult to do so, includes economy and crime statistics. However, this information was difficult to find at national level. Although state-level data may have been useful, it would have been difficult to incorporate for multiple states, as different methodologies may have been used, covering different time periods etc.

A further point to note is the data has been narrowed down to Florida and North Carolina, as per the stakeholder's preferences, and is also advantageous to me as reduces data being processed on the IBM and Foursquare systems. However, to enhance this report further, I recommend removing the restriction of favoured states only.

In addition, I also question the stakeholders' refusal to consider California as an option, as a general assumption would be that it would be a good state to consider for such a project.

Conclusion

Belleair beach looks like a sensible recommendation based on the data available. However, given this is based on data for only two states, and the multitude of other factors which contribute to the decision to locate a business, such as local economic performance and state laws to name just two, I'd also recommend the stakeholders consult other information before reaching a decision.