What is the best Tucson neighborhood for Embassy Tire to expand to next?

Authored by Jon Ingram

Introduction

Embassy Tire & Wheel Company is a growing, family-owned business in the Tucson Metro Area. Starting from a single location in 2004, it is now expanding to a third location set to open in the summer of 2019. The positioning of the three current locations provide a decent range of coverage of the region. As a growing business, they most certainly aren't planned to settle on just three locations. After the third location opens and business starts to pick up, the company is sure to look to expand once again. But where to?

The report seeks to answer the question of "given the population and tire shop distribution of the Tucson Metro Area, where would be the best neighborhood(s) to look to expand into?"

Data

The types of data that are required to execute this idea are mostly population data and location data. Specifically, it requires data like the neighborhoods of Tucson, their populations, population densities, and the locations of all neighborhoods and tire shops in the Tucson metro area.

The data for the populations and population densities of the top 50 populated neighborhoods in Tucson are publically available at the following Statistical Atlas website: https://statisticalatlas.com/place/Arizona/Tucson/Population. The neighborhood names of the 50 most populated neighborhoods, their populations, their rank in population, their population densities, and their rank in population densities will be ripped directly from the website. This will be accomplished using the Beautiful Soup library.

The population data as well as the Statistical Atlas links for each neighborhood will be collected into a single DataFrame for ease of use. This DataFrame will also contain an Area (in square miles) column calculated by dividing Population (number of people) by Population Density (number of people per square mile) for each neighborhood.

The geographical locations of these neighborhoods will be retrieved using OpenStreetMap data obtained from the geocoder library. Locations which are not retrieved using this method will be found manually using the Google search engine.

The acquisition of tire shop location data is a process in several steps. First, the zip codes of major postal area codes in Tucson will be taken manually from the following Statistical Atlas website: https://statisticalatlas.com/place/Arizona/Tucson/Overview. Their geographical coordinates do not

appear in OpenStreetMap so their coordinates will be acquired manually using the Google Search engine. Finally, these locations will be used as the basis for a series of Foursquare API search queries (with keyword 'tire') to collect a comprehensive list of tire shops in the Tucson Metro Area. The IDs, Names, and Location in coordinates will be taken from the results of the queries and be compiled into a pandas DataFrame.

Methodology Part 1: Data Collection

Step 1:

- Import the required libraries (pandas, numpy, requests, and beautifulsoup4)

Step 2:

- Run a simple html script to remove 'Run This Cell' buttons from the display of the Notebook
- Not necessary to do but these buttons tend to be visually distracting with editing

Step 3:

- Collect the initial data from the Statistical Atlas website
    - Retrieve the html data of the website
    - Create a Beautiful Soup object using the html text
    - Define a function to rip the data from the formatted tables on the site
    - Use the function to create a pandas DataFrame for the population counts
    - Repeat the process for the population densities
    - Merge the resulting DataFrames

Step 4:

- Collect neighborhood location data using geocoder
    - Import geocoder library
    - Use geocoder to collect lists of latitudes and longitudes
    - Add found latitudes and longitudes to main DataFrame

Step 5:

- Find rows in DataFrame that are missing coordinates data
- Collect and data manually

- Import data into main DataFrame from temporary DataFrame of manually collected data

The resulting full DataFrame:

Neighborhood	Population	Population Rank	Density	Density Rank	Area	Link	Latitude	Longitude
Casas Adobes	52552.0	1	2437.98	41	21.555550	https://statisticalatlas.com/neighborhood/Ariz...	32.248908	-111.116696
Drexel Heights	28627.0	2	1452.47	46	19.709185	https://statisticalatlas.com/neighborhood/Ariz...	32.141249	-111.028443
Tanque Verde	17718.0	3	541.25	50	32.735335	https://statisticalatlas.com/neighborhood/Ariz...	32.264927	-110.736165
Rita Ranch	16282.0	4	1044.09	47	15.594441	https://statisticalatlas.com/neighborhood/Ariz...	32.103666	-110.769505
Flowing Wells	15667.0	5	4493.78	28	3.486374	https://statisticalatlas.com/neighborhood/Ariz...	32.286127	-111.026304
Sunnyside	15483.0	6	5784.51	20	2.676631	https://statisticalatlas.com/neighborhood/Ariz...	32.221892	-110.926235
Midvale Park	13635.0	7	3617.70	35	3.768969	https://statisticalatlas.com/neighborhood/Ariz...	32.133702	-110.998957
West Flowing Wells	9553.0	8	2080.07	43	4.592634	https://statisticalatlas.com/neighborhood/Ariz...	32.286127	-111.026304
Cherry Avenue	9453.0	9	7575.59	5	1.247824	https://statisticalatlas.com/neighborhood/Ariz...	32.230555	-110.948137
Elvira	8876.0	10	5594.56	21	1.586541	https://statisticalatlas.com/neighborhood/Ariz...	32.126708	-110.977584
Amphi	7461.0	11	7481.24	6	0.997295	https://statisticalatlas.com/neighborhood/Ariz...	32.272222	-110.972222
Dietz	7375.0	12	5319.21	25	1.386484	https://statisticalatlas.com/neighborhood/Ariz...	32.200096	-110.831062
Campus Farm	6943.0	13	6164.02	14	1.126375	https://statisticalatlas.com/neighborhood/Ariz...	32.280435	-110.952443
Garden District	6555.0	14	6592.01	13	0.994386	https://statisticalatlas.com/neighborhood/Ariz...	32.231805	-110.971782
South Harrison	6424.0	15	3705.55	34	1.733616	https://statisticalatlas.com/neighborhood/Ariz...	32.186900	-110.786700
Corbett	6169.0	16	6084.53	15	1.013883	https://statisticalatlas.com/neighborhood/Ariz...	32.213088	-110.919030

Step 6:

- Collect tire shop location data
    - Collect zip code locations
        - Acquire list of zip codes from Statistical Atlas website
        - Search for coordinate data manually
        - Import data into intermediate DataFrame
    - Collect tire shop data using Foursquare API
        - Define Foursquare required variables such as `CLIENT_ID` and `CLIENT_SECRET`
        - Loop through zip code coordinates from intermediate DataFrame and put relevant data in lists
        - Merge data lists to single DataFrame
        - Use Foursquare unique venue IDs to remove all duplicates from trie shop DataFrame
- Acquire Embassy Tire locations from Foursquare data for later use

The first five rows of the DataFrames containing the general and Embassy-specific Foursquare data:

Out[68]:

	ID	Name	Latitude	Longitude
0	5555904d498e2f2c61b12d17	Phil's Fleet Tire Service Inc.	32.214113	-110.944973
1	5afe24bb345cbe002cd224f4	Discount Tire	32.187005	-110.943963
2	4c2d09bd260bc928b25119d3	Discount Tire	32.275145	-110.977893
3	57adfb19498e812df760d199	Jack Furrier Tire & Auto Care	32.261595	-110.960618
4	5334785811d2caf0f822559e	Discount Tire	32.153825	-110.987217

In [69]:
```
embassy_coords = tucson_shop_data.loc[tucson_shop_data['Name'].str.contains('Embassy'),:]
embassy_coords
```

Out[69]:

	ID	Name	Latitude	Longitude
6	4e24a307d16474063225f037	Embassy Tire & Wheel Company	32.272466	-110.988764
62	4f83603ce4b03e850b0a0c6a	Embassy Tire & Wheel Co.	32.133841	-110.968153

Methodology Part 2: Data Analysis

Step 1:

- Import Python's `math` library
- Define important constants
    - `LAT_CONSTANT`: latitude degrees per mile
    - `LONG_CONSTANT`: longitude degrees per mile
- Define important functions
    - `getNeighborhoodRadius`
        - Use Area of neighborhood to return a radius suitiable for check if a tire shop is in the neighborhood or not
        - Radius returned must have a minimum value of 1.5
        - NOTE: Radius is in miles
    - `latDistanceInMiles`
        - Returns difference of latitudes 'a' and 'b' in miles
    - `longDistanceInMiles`
        - Returns difference of longitudes 'a' and 'b' in miles
    - `modifiedEuclidistance`

        - 'modified' because function accepts two pre-calculated differences into the formula
        - Returns euclidean distance of two points using pre-calculated differences
    - `getShopCounts`
        - Given the name of a neighborhood get the search radius and coordinates of that neighborhood
        - Iterate through the list of tire shops
            - Calculate the euclidean distance of each shop with the given neighborhood
            - Increase shop count if shop is inside the search radius of the given neighborhood
        - Return the shop count for that neighborhood

Step 2:

- Use `shop_counts` method to calculate the number of shops in or close to each neighborhood
- Put those values into a new DataFrame

Step 3:

- Build a DataFrame of shop count to 'Shop Count Rank' conversion using the unique values of the shop sount DataFrame
- Add a column to the shop count DataFrame with the correct 'Shop Count Rank' for each neighborhood

The resulting DataFrame of Shop Counts and Shop Count Ranks per neighborhood (first 10 rows):

	Neighborhood	Number of Shops	Shop Count Rank
0	Casas Adobes	0	11
1	Drexel Heights	10	1
2	Tanque Verde	0	11
3	Rita Ranch	4	7
4	Flowing Wells	1	10
5	Sunnyside	4	7
6	Midvale Park	3	8
7	West Flowing Wells	1	10
8	Cherry Avenue	1	10
9	Elvira	3	8
10	Amphi	10	1

Step 4:

- Compile all of the statistical ranks for each neighborhood into a new DataFrame
- Append an 'Optimal' neighborhood row as a metric to measure the quality of each neighborhood
    - Optimally, a neighborhood would have a Population and Density Rank of 1 and a Shop Count of 11
    - This means that there is high population and low tire shop competition (perfect for expansion!)

Step 5:

- Calculate the 'Overall Rank' for each neighborhood
    - Find the euclidean distance of each neighborhood to the Optimal neighborhood
    - Add all distances to the DataFrame of ranks
    - Sort all rows ascending by 'Distance From Optimal'
    - Add the 'Overall Rank' as a column to the final ranks DataFrame

The first 15 rows of the DataFrame containing all Rank data:

Neighborhood	Overall Rank	Population Rank	Density Rank	Shop Count Rank	Distance From Optimal
Cherry Avenue	1	9	5	10	9.000000
Amphi	2	11	6	1	15.000000
Campus Farm	3	13	14	8	17.944358
Garden District	4	14	13	8	17.944358
Sunnyside	5	6	20	7	20.049938
Corbett	6	16	15	10	20.542639
Elvira	7	10	21	8	22.135944
Lakeside Park	8	18	17	9	23.430749
Myers	9	22	18	10	27.037012
Palo Verde	10	17	22	5	27.073973
Dietz	11	12	25	5	27.073973
Flowing Wells	12	5	28	10	27.313001
Julia Keen	13	21	19	6	27.367864
Keeling	14	26	8	2	27.477263
Wakefield	15	27	7	4	27.586228

Results

The analysis of the results was begun by looking at the top ten neighborhoods in terms of their overall rank. Those neighborhoods, as determined by the Ranks DataFrame are 'Cherry Avenue', 'Amphi', 'Campus Farm', 'Garden District', 'Sunnyside', 'Corbett', 'Elvira', 'Lakeside Park', 'Myers', and 'Palo Verde'.

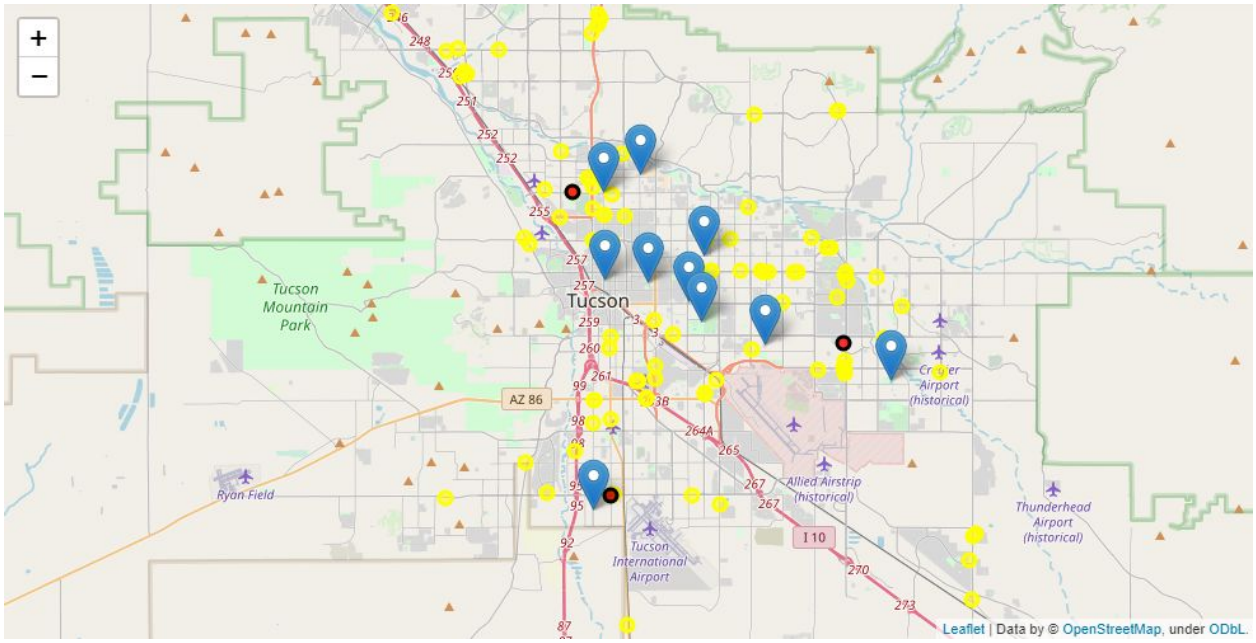
Neat! Looks like Cherry Avenue, Amphi, and Campus Farm are some top contenders. For a better look at the results, though, our top ten neighborhoods were plotted to compare them to the locations of Tucson's tire shops as well as the distance of the neighborhoods to the three existing Embassy Tire & Wheel locations.

With the folium library imported, the locations of the three Embassy Tire locations were compiled using the coordinates from Foursquare acquired before as well as coordinates for the third location found on Google using it's address of 1431 S Kolb.

Finally, using all of the acquired location data, an interactive Leaflet map was created to view the data points visually. The map was centered around a moving services company determined by trial and error to give the best view for all of the data. The map includes three types of markers:

- Yellow Circle Markers representing tire shop locations
- Red and Black Circle Markers representing the Embassy Tire & Wheel locations
- Point Markers representing the ten best neighborhoods for expansion
    - These markers include popup labels with the neighborhood name and that neighborhood's Overall Rank

The Folium Map containing all of the mentioned data:



Discussion

Of our top 10 neighborhoods, some of them have to be ruled out due to their proximity to the current three Embassy Tire & Wheel locations and other shops. For example, using the generated map as a reference, Amphi places 2nd overall as a very ideal candidate but is only about a mile away from the location at Prince. As a small, yet, growing business, opening a new shop a mile away from a current location is a poor idea. So, Amphi must be ruled out as a potential neighborhood for expansion.

It is also important to note that just the population, population density and tire shop count alone may not be enough to give the best objective suggestion in selecting Tucson neighborhoods for expansion. Other factors may include Average Individual Income, Average Household Income, Household Types, and Industries in each area. This is just a non-comprehensive list of additional factors which are potentially important to look into but all of the data in that list, like the Population data, is also available on Statistical Atlas.

With the brevity of the dataset in mind, an interesting observation is the proximity of some of the top ten neighborhoods to the locations at Prince and Valencia (the locations on the Western side of the city). The success of the Prince and Valencia locations is what has allowed the company to continue to expand. While these close neighborhoods (namely: Amphi, Campus Farm, and Elvira) need to be ruled out with regards to new expansions, the fact that such high ranking neighborhoods (which

include neighborhoods ranked 2 and 3) are geographically close to already successful locations suggests that the current algorithm used to generate the list of top neighborhoods can imply future success in expansion into other top neighborhoods.

Conclusion

Using the generated map as a reference, it is determined that Cherry Avenue, Sunnyside, and Corbett are the ideal Tucson neighborhoods to look to expand into. Placed near the center of town, they all rank high statistically (ranked 1, 5, and 6, respectively) and are effectively equidistant from the current three Embassy Tire & Wheel locations. The map also indicates that the center of town is fairly lacking in tire shops in general which would also make this region great for expansion.

For information on the location data, demographics and more on those neighborhoods, the links to their Statistical Atlas websites will be provided below.

Thank you for reading!

Links:

Cherry Avenue:
https://statisticalatlas.com/neighborhood/Arizona/Tucson/Cherry-Avenue/Population
Sunnyside:
https://statisticalatlas.com/neighborhood/Arizona/Tucson/Sunnyside/Population
Corbett:
https://statisticalatlas.com/neighborhood/Arizona/Tucson/Corbett/Population