Brewery Site Evaluation

for Montgomery County, PA

July 24, 2019

Coursera - IBM Data Science Capstone project
“Battle of the Neighborhoods”

Overview

When a business is proposed, the location is often very important and an analysis is typically initiated to do a site selection and evaluate the pro's and con's of various locations.

In the same way, this project was to do a site selection evaluation for a brewery in central Montgomery county, PA.

Project objective:

Identify good locations for a new brewery in the philadelphia suburbs, centered on Montgomery county using free information from various data sources.

How was this done

Sample Targets

Identify existing breweries in the target area and acquire rating information of the brewery location (not the beer).

Gather Local Venues

Retrieve venue location near the sample breweries and perform clustering analysis to identify criteria to evaluate the sample location ratings

Find New Locations

Using the clustering analysis, find new locations in the target location that satisfy the ‘good’ criteria while avoiding the ‘bad’.

Data Acquisition

Data Sources

- Location data from Foursquare
  - Brewery venues identified based on CategoryId search exposing 61 target locations
  - Cleansing and dropped locations based on criteria and distance from target
  - Brewery ratings and number of ratings
  - Venues in close proximity identified by category revealing 251 additional locations

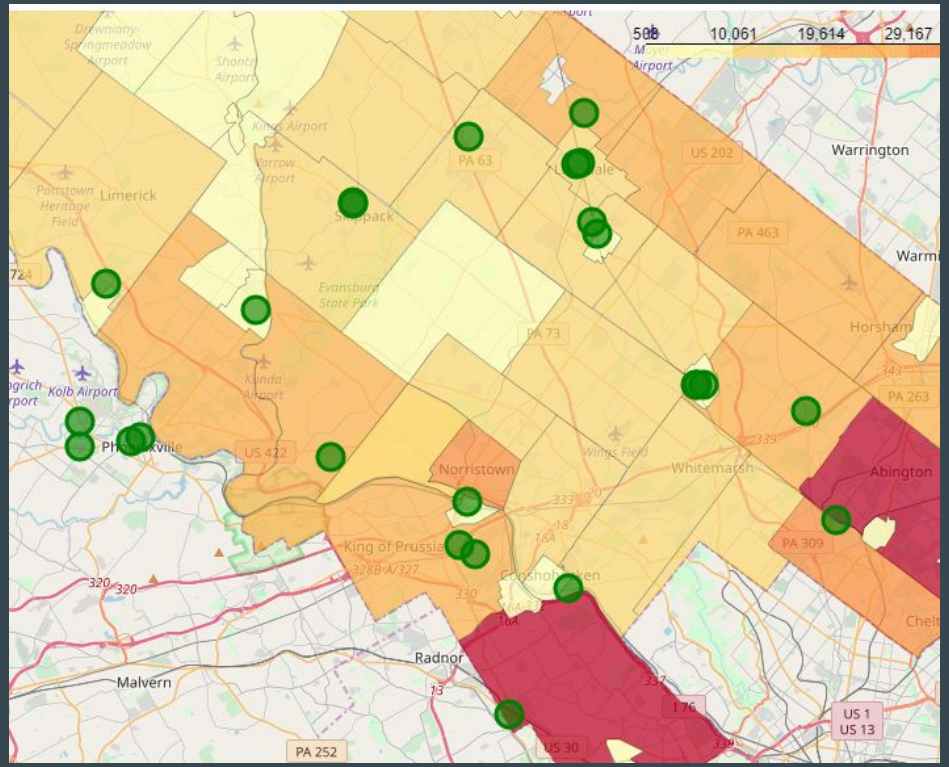
- Pennsylvania spatial data for township borders
- The Penn State Data Center Municipal level data for population density
- Nominatim and Geolocator API for location translation to latitude/longitude

- Google maps and search
  - To validate venues and breweries as operational venues and not fake data.

Methodology and Exploration

Existing Breweries Identified

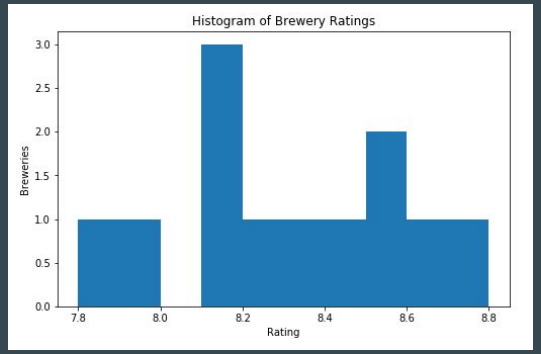
- 24 Target locations identified
- Locations correlate to population density and transportation corridors
- Large central area lacking any target breweries

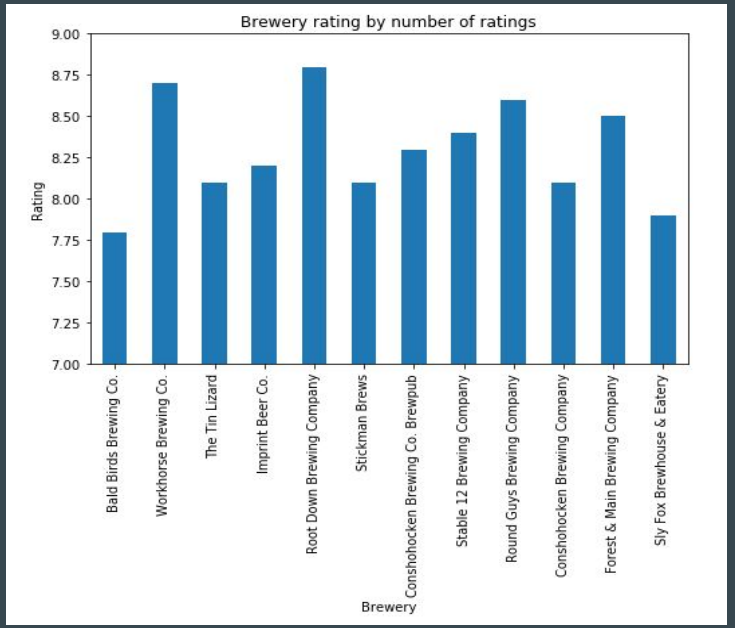


Existing Breweries Rating Data

- Ratings roughly followed Gaussian distribution expectation

- Ratings of breweries were not correlated to number of ratings (this was good)

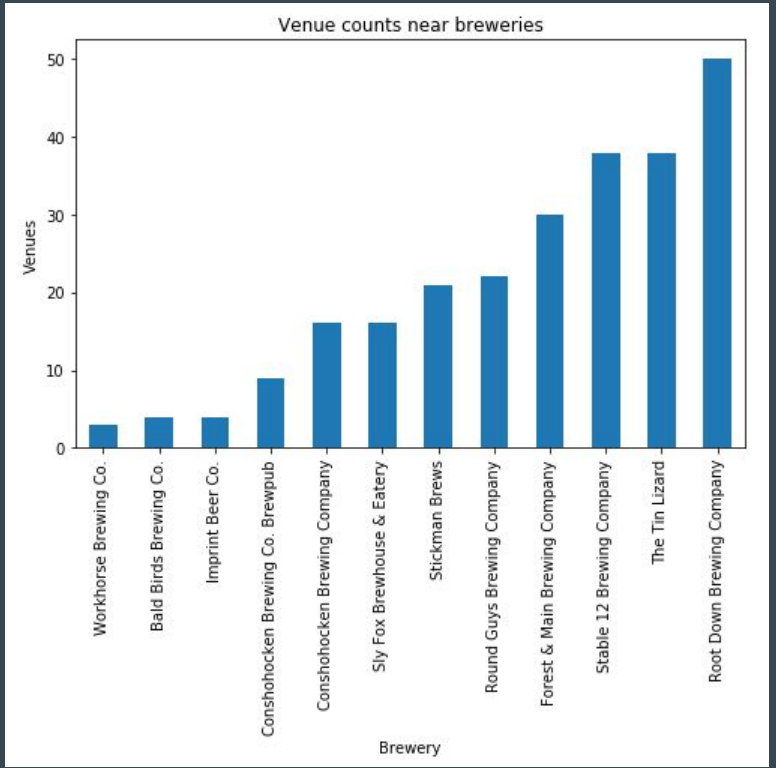




Existing Breweries Rating Data

- Ratings were also not distributed according to the number of nearby venues

- Number of nearby venues varied by a large range.



K-Means Clustering Analysis

	Cluster Labels	Brewery	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
0	0	Bald Birds Brewing Co.	Other Great Outdoors	Business Service	Coffee Shop
1	1	Conshohocken Brewing Co. Brewpub	Pizza Place	Gym	Ice Cream Shop
2	1	Conshohocken Brewing Company	Italian Restaurant	Athletics & Sports	Yoga Studio
3	1	Forest & Main Brewing Company	Pizza Place	American Restaurant	Bakery
4	0	Imprint Beer Co.	Pharmacy	General Entertainment	Business Service
5	1	Root Down Brewing Company	American Restaurant	Pub	Pizza Place
6	1	Round Guys Brewing Company	Pizza Place	Ice Cream Shop	Bakery
7	1	Sly Fox Brewhouse & Eatery	Pizza Place	Fast Food Restaurant	Bank
8	1	Stable 12 Brewing Company	American Restaurant	Pub	Pizza Place
9	1	Stickman Brews	Pizza Place	Fast Food Restaurant	Discount Store
10	1	The Tin Lizard	Pizza Place	Indian Restaurant	Café
11	2	Workhorse Brewing Co.	Breakfast Spot	Food Truck	Wine Shop

Clustering targeted 3 categories (high, medium, low)

Breweries grouped into those 3 categories

Conclusions

Results

Given the limited dataset, the results showed a correlation. The criteria for the best and the worst did not overlap.

Worst Rated Breweries Proximity to:

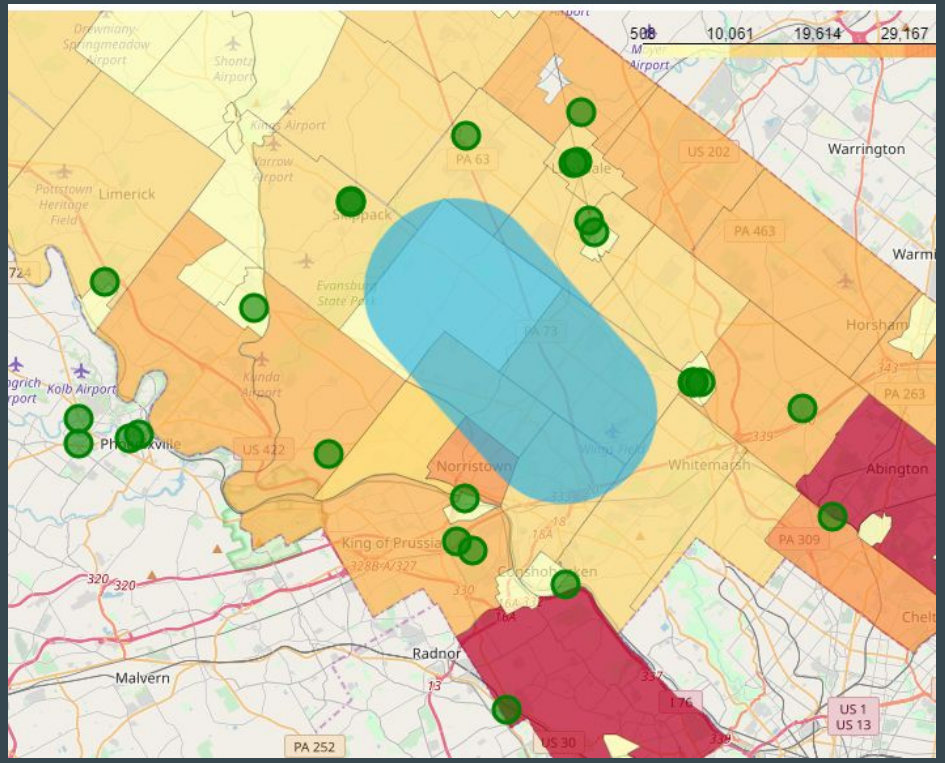
1. Fast Food (Wendy's, Dairy Queen, Subway, ...)
2. Business Services (Cleaners, corporate offices)

Best Rated Breweries Proximity to:

1. American Restaurants (Black Lab Bistro, Bistro on Bridge, Great American Pub, ...)
2. Pubs (Molly Maguire's, Junction House, The Foodery, PJ Ryan's Pub, ...)

The Desired Target Area

- The central area of Montgomery County was lacking any breweries
- This was the target area designated to search
- This area is represented by a blue oval on the map



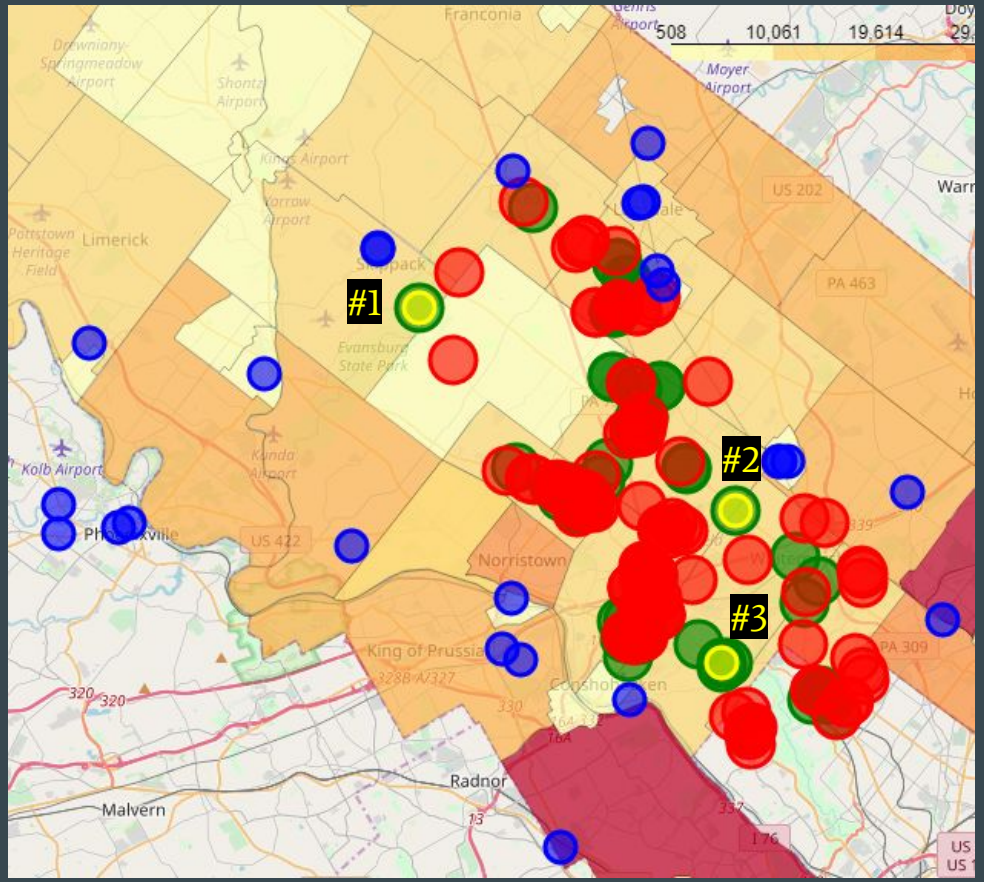
The Suggested Target sites

Using the criteria of good and bad sites:

- Blue sites are existing Breweries
- Green sites are “good” criteria
- Red sites are “bad” criteria

The yellow sites satisfy:

1. Low proximity to other breweries
2. Low proximity to ‘bad’ sites
3. High proximity to ‘good’ sites.



1. Near the Skippack Golf Club in Evansburg
2. Near the Phil’s Tavern in Blue Bell
3. Near Brittingham’s Pub in Lafayette Hill.

Discussion

Discussion

- Free data is has lots of errors and fake information
- Extensive manual clean up was needed
- Free data is limited in terms of value. With better data, the results and options for evaluation would have improved substantially.
- Other data source besides Foursquare should have been used for venue information, ratings and proximity. This would not be free though.