

# Predicting suitable commercial business type and location in New Castle County, Delaware

*Debasish Chaulia*

*June 12<sup>th</sup>, 2020*

## Introduction and Business Problem

A large scale business owner, who has a wide scale presence in many cities in USA, wants to start a new business in New Castle county in Delaware, USA. He is already exposed to various types of business setups (eg grocery stores, restaurants, fast food centers etc), but it is his first time he is planning to setup some business in New Castle, Delaware. Even though he has some business ideas floating in his mind, he thinks now it's the time to balance it with a little reality. Does those ideas have the potential to succeed? Well he realizes the need to run his business ideas through a validation process before he go any further and take a final call.

Based on the report from U.S. Bureau of Labor Statistics (BLS), it shows that approximately 20% of **new businesses** fail during the first two years of being open, 45% during the first five years, and 65% during the first 10 years. Only 25% of **new businesses** make it to 15 years or more. Bad location chosen and incorrect business types are the most common reasons of such a failure, in spite of having sufficient financing.

In order for the business to be successful, it must solve a problem, fulfill a need or offer something the market wants. There are a number of ways we can identify this need, may be research, focus groups, and even trial and error etc.

Given the fact there are multiple options of new businesses as well as many choices in terms of location, he wants to have better understanding and clear information/knowledge about the most suitable business options in New Castle county and that too which specific location/neighborhood, so that it maximizes his ROI (return on investment) and profit. Location selection of the place of the business is equally important along with the type of business for the smooth operation of the business.

Once the solution is figured out regarding the suitable location and type of business, he may possibly need to extend his thinking to find out whether his business can deliver something other companies don't (or deliver the same thing, only faster and cheaper).

## Target Audience

This solution of this problem can benefit bigger audience (ie other similar business owners) who are also intending/planning to open similar new businesses (even startups also) in the city of New Castle, Delaware. Because it is very important and critical for the business owners to know the pros and cons of the type of business they are investing into. And on top of that, its very crucial selecting the correct location so as to target maximum possible customers for his / her business.

## **Data**

### **Data sources**

I will source dataset from USPS(United States postal service) website having all USA county wise geographical details in it. I will do analysis on the dataset to determine which of the fields can be used for our problem. The dataset has information related to all counties, zip codes, state codes, neighborhoods and coordinates.

### **Data cleaning/wrangling and feature selection**

Data downloaded has information for all the counties across USA. So I will filter out the dataset based on state code 'DE' for Delaware. And after that I will select only those records that are for New Castle county. I will then utilize the features("County", "Neighborhood", "Latitude", "Longitude" ) of the selected records to proceed further.

Based on the neighborhood location information available, I will then be using(latitude and longitude) of it in making foursquare API calls to get/explore more information and facts about that particular neighborhood . I will then reformat the foursquare returned info by filtering/refining the information based on the need to approach towards the solution.

For eg based on the neighborhood (say Newark) , I will make foursquare API call to get all the nearby venues (say in a radius of 5 m) . And then use this info to find what are the types of venues (say restaurant, grocery stores, gymnasium etc) that are most popular in that neighborhood and based on those types we can filter out any specific venue type which fits closely his business needs.

Fourquare API calls will also be used to get/retrieve information related to the most trending and happening venues in any particular neighborhood, so that cross analysis can be done about what are already popular venues in that neighborhood and what are the trending venues in that neighborhood.

## **Methodology**

I am using segmentation, aggregation and clustering methods to do detail analysis on the neighborhoods. I am using free version of USPS ZIP code database available at "<https://www.unitedstateszipcodes.org/zip-code-database/>".

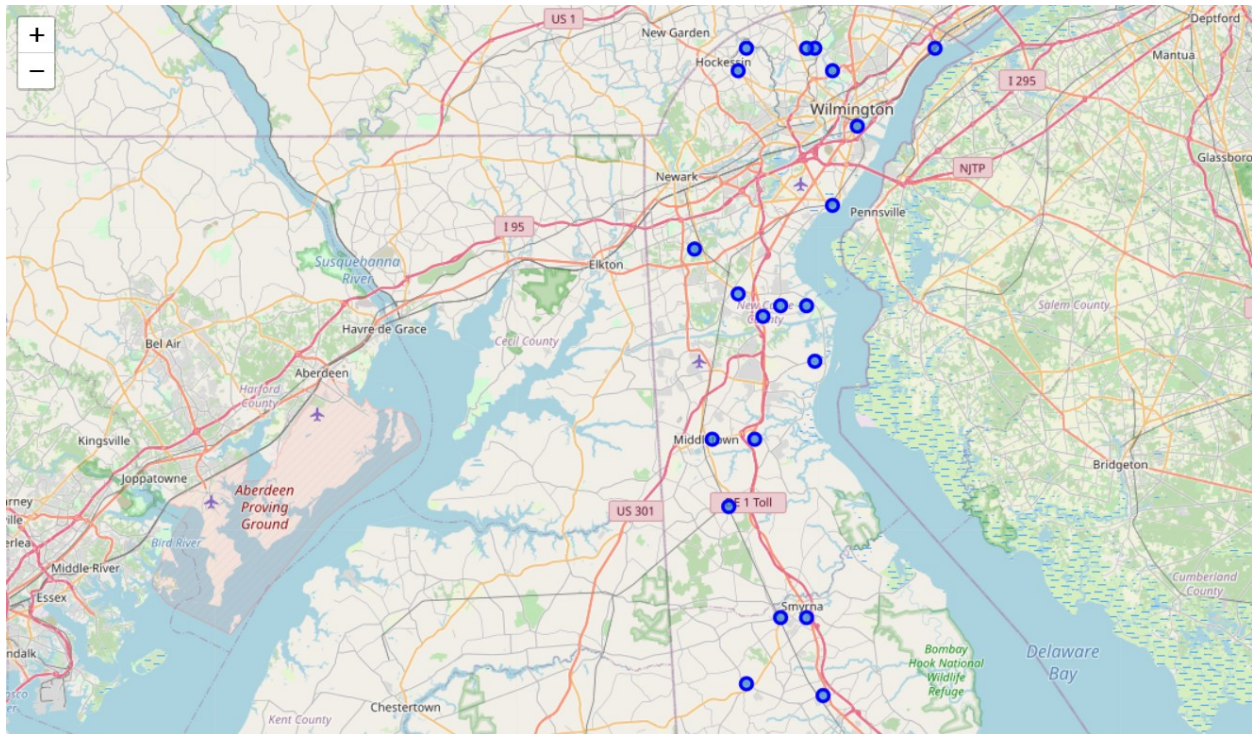
### **Features engineering :**

The dataset has all the below features and we will be focusing only on the highlighted fields based on our needs :

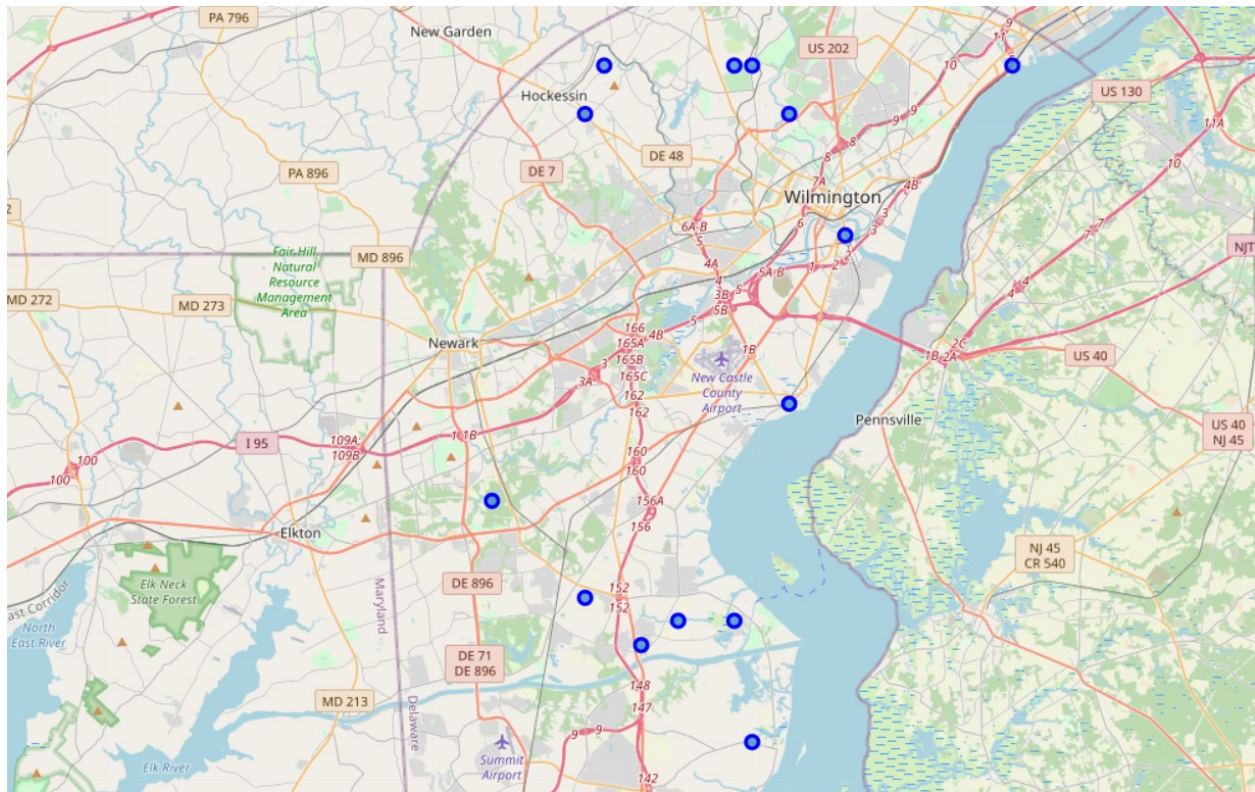
- a) Zip-code
- b) Type
- c) Decommissioned

- d) Primary city
- e) Acceptable-cities
- f) Unacceptable-cities
- g) State
- h) County
- i) Time-zone
- j) Area-codes
- k) World-region
- l) Country
- m) Latitude
- n) Longitude
- o) Irs\_estimated\_population\_2015

Since our problem deals with the state of Delaware (abbreviated as DE), I will extract records for Delaware state alone and store in a file for further processing. I will also perform exploratory data analysis on the dataset to see the main characteristics or summarization of the dataset, and then check if there are any missing values or null values in the dataset . I am then renaming the columns giving them more meaningful names based on the problem I am working on. Next I am plotting a graph using folium to visualize all Delaware neighborhoods in the map.



From the above extracted file, I will then extract only the records for New Castle counties alone.  
(Note : pls note that in “New Castle county” there is one neighborhood called “New Castle” !!.)  
Next I am plotting neighborhoods of New Castle county alone.

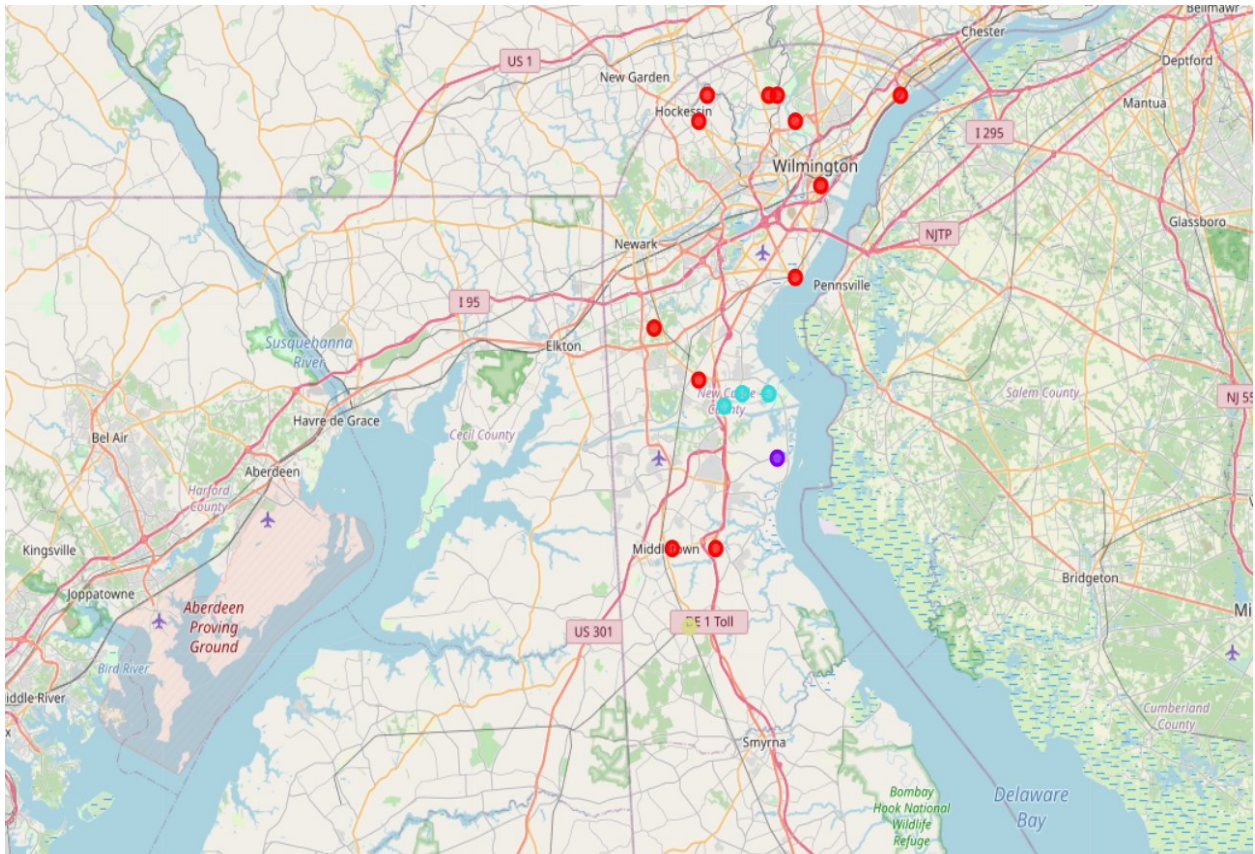


We will then use foursquare API calls to get/explore all the venues of each of the neighborhoods of New Castle county. We are limiting our exploration by max 100 venues within 5 miles. Result obtained from the call will then be used to find out what are the types of venues (say restaurant, grocery stores, gymnasium etc) available for a neighborhood and then storing it into a dataframe. We are then sorting the venues based on mean frequency (ie descending order or most popular first) of the venue types and based on those types we can filter out any specific venue type which fits closely his business needs.



## **Results**

Based on the 17 neighborhoods in New Castle county, we grouped these neighborhoods into 4 clusters based on the most common venues for the neighborhood.



Here for illustration purpose only first 5 most common venues are being displayed but the clustering has been done based on all the 10 most common venues. (pls refer the jupyter notebook for the details).

Below is the table depicting dataframe that includes the cluster as well as the top 10 venues for each neighborhood.

(For instance for the Newark neighborhood, the most common venue is “Pizza Place”, 2<sup>nd</sup> most “Pharmacy”, 3<sup>rd</sup> most “American Restaurant” and so on).

	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	New Castle County	Bear	39.58	-75.68	0	Pizza Place	Pharmacy	Intersection	Convenience Store	Discount Store
14	New Castle County	Winterthur	39.80	-75.60	0	American Restaurant	Golf Course	Coffee Shop	Gym / Fitness Center	Bagel Shop
11	New Castle County	Rockland	39.78	-75.57	0	Park	Pharmacy	Pizza Place	Seafood Restaurant	Gym / Fitness Center
9	New Castle County	Odessa	39.45	-75.66	0	Pharmacy	Donut Shop	Video Store	Construction & Landscaping	Chinese Restaurant
15	New Castle County	Yorklyn	39.80	-75.67	0	Park	American Restaurant	Café	Sandwich Place	Gym / Fitness Center
7	New Castle County	Montchanin	39.80	-75.59	0	American Restaurant	Coffee Shop	Ice Cream Shop	Pizza Place	Pharmacy
8	New Castle County	New Castle	39.66	-75.57	0	Donut Shop	Pharmacy	American Restaurant	Ice Cream Shop	Discount Store
4	New Castle County	Hockessin	39.78	-75.68	0	Pizza Place	American Restaurant	Park	Sandwich Place	Bar
2	New Castle County	Claymont	39.80	-75.45	0	Pizza Place	Pharmacy	Convenience Store	Discount Store	Sandwich Place
1	New Castle County	Newark	39.62	-75.73	0	Pizza Place	Pharmacy	Ice Cream Shop	Sandwich Place	Convenience Store
6	New Castle County	Middletown	39.45	-75.71	0	Restaurant	American Restaurant	Pharmacy	Sandwich Place	Fast Food Restaurant
16	New Castle County	Wilmington	39.73	-75.54	0	Pharmacy	Italian Restaurant	Discount Store	Brewery	Park
10	New Castle County	Port Penn	39.52	-75.59	1	Pub	IT Services	Event Space	Nature Preserve	Beach
5	New Castle County	Kirkwood	39.57	-75.63	2	Pizza Place	Seafood Restaurant	Convenience Store	Bakery	Event Space
3	New Castle County	Delaware City	39.57	-75.60	2	Seafood Restaurant	State / Provincial Park	Boat or Ferry	Bakery	Café
12	New Castle County	Saint Georges	39.56	-75.65	2	Pizza Place	Construction & Landscaping	Theme Park	Deli / Bodega	Convenience Store
13	New Castle County	Townsend	39.39	-75.69	3	Liquor Store	State / Provincial Park	Park	Music Venue	Food

### 1<sup>st</sup> Cluster :

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Bear	Pizza Place	Pharmacy	Intersection	Convenience Store	Discount Store
1	Newark	Pizza Place	Pharmacy	Ice Cream Shop	Sandwich Place	Convenience Store
2	Claymont	Pizza Place	Pharmacy	Convenience Store	Discount Store	Sandwich Place
4	Hockessin	Pizza Place	American Restaurant	Park	Sandwich Place	Bar
6	Middletown	Restaurant	American Restaurant	Pharmacy	Sandwich Place	Fast Food Restaurant
7	Montchanin	American Restaurant	Coffee Shop	Ice Cream Shop	Pizza Place	Pharmacy
8	New Castle	Donut Shop	Pharmacy	American Restaurant	Ice Cream Shop	Discount Store
9	Odessa	Pharmacy	Donut Shop	Video Store	Construction & Landscaping	Chinese Restaurant
11	Rockland	Park	Pharmacy	Pizza Place	Seafood Restaurant	Gym / Fitness Center
14	Winterthur	American Restaurant	Golf Course	Coffee Shop	Gym / Fitness Center	Bagel Shop
15	Yorklyn	Park	American Restaurant	Café	Sandwich Place	Gym / Fitness Center
16	Wilmington	Pharmacy	Italian Restaurant	Discount Store	Brewery	Park

### 2<sup>nd</sup> Cluster

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
10	Port Penn	Pub	IT Services	Event Space	Nature Preserve	Beach

### 3<sup>rd</sup> cluster

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
3	Delaware City	Seafood Restaurant	State / Provincial Park	Boat or Ferry	Bakery	Café
5	Kirkwood	Pizza Place	Seafood Restaurant	Convenience Store	Bakery	Event Space
12	Saint Georges	Pizza Place	Construction & Landscaping	Theme Park	Deli / Bodega	Convenience Store

#### 4<sup>th</sup> cluster

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
13	Townsend	Liquor Store	State / Provincial Park	Park	Music Venue	Food

### **Discussions**

Based on the results obtained, one of the cluster 1 neighborhoods would probably a good option for opening new businesses as it shares lot of similarities in terms of most common venues with other neighborhoods in the same cluster 1. Also since all these neighborhoods in the same cluster share so much similarities in terms of most common venues, businesses can even potentially target customers residing in the vicinity of other neighborhoods also within same cluster. So that should also definitely businesses get/expect more additional customers.

Now we should also take into consider the school district of the neighborhood before deciding on the neighborhood, because that is one of the most important factor people take into consideration who are planning to migrate from one place to another place. So better the school district more the people and more the customer base for your business. For this I referred several relevant websites like <https://www.schooldigger.com>, <https://www.usnews.com/education/best-high-schools>, <https://www.niche.com/k12/search/best-school-districts/c/new-castle-county-de/> etc.. and “**Appoquinimink School District**” is best school district in New Castle county, Delaware and the neighborhood for this school district is “**Middletown**”.

Now lets do some analysis on the type of business to be chosen. If you see in “Middletown”, you will find 4 of the 5 most common venues are eatery items like (Restaurant, American Restaurant, Sandwich place and Fast food restaurant). So is it not already overcrowded with restaurants ? Opening another restaurant in this neighborhood might not be a good decision as it might be already saturated with restaurants. So,what about other venues types like Gymnasium, gas station ? If there are so many restaurants in that neighborhood, it is high likely people should also need /look for some mechanism and way to stay fit and maintain their health thru exercises and so on. Also gas station might also work well. But again choosing this option might be little risky for the businessman as he is opting to deliver something other companies don't. Other option might be to go with the type which is already a most common trend in that neighborhood (ie some kind of restaurant) and in that case risk is relatively less. But competition would be relatively high for the second option and he would also need to keep in mind (deliver the same thing, only faster and cheaper and better quality).

### **Conclusion**



Conclusion section where you conclude the report.

After studying the results, we can possibly propose the business owner below two options.

1) opening a Gymnasium or gas station in Middletown, which means he is opting to take more risk by selecting to deliver something other companies don't in that area. But it is also quite obvious that returns on such investments is expected to be more (high risk-reward ratio)

2) opening a type of business which is already a most common trend in Middletown neighborhood (ie some kind of restaurant) and in that case risk is relatively less (low risk-reward ratio). But he also has to keep in mind about the fierce competition he would need to face from other competitors doing same type of businesses and need to focus on delivering the same thing, but faster, cheaper and better quality he may possibly need to extend his thinking to find out whether his business can deliver something other companies don't (or deliver the same thing, only faster and cheaper).

## **References**

Websites referred –

<https://www.schooldigger.com>

<https://www.usnews.com/education/best-high-schools>, <https://www.niche.com/k12/search/best-school-districts/c/new-castle-county-de/>

<https://www.unitedstateszipcodes.org/zip-code-database/>