Philadelphia - the city of brotherly love

Selecting the Best Location for a Boutique Coffee Shop in Philadelphia

Jacob Good

March 23, 2021

1. Background

A. Philadelphia is the sixth largest city in the United States with an estimated population of over 1.5 million people. Extending the boundaries to Philadelphia county which are easily accessible by Amtrak and other public transportation, the city has over 6 million residents. There are many large corporations including Comcast, Aramark, and Urban outfitters. These alone account for over $100 billion worth of yearly revenue, with an estimated gross metropolitan product (GMP) just shy of $500 billion. The city also hosts a diverse racial make-up and multiple well respected universities.

B. The most recent median income (2013) for Philadelphia was $36,836, which his relatively lower than the median income for metropolitan areas at $60,482. These incomes range from a high of $93,720 to a low of $14,185. [1]

2. Description of the problem

A. My client is already a proud owner of two coffee shops, two eateries, a bakery, and a large family restaurant. He is looking to expand his business, particularly in the field of boutique coffee shops. However, he's not exactly sure where the new coffee shop should be located. He has already had great success with a coffee shop located near a university, and would like

to expand a bit beyond university clientele. He wants to make sure that his coffee shop is not overcrowded by other chain coffeeshops, i.e. Starbucks, or eateries that would take away his revenue, but this is not a. He is based in Pennsylvania and would like to continue to build his business in that area. He is very interested in the Philadelphia market and would like to know more details about where a new coffeeshop would bring in additional revenue for his business, while also expanding beyond university-centralized locations.

B. This is a preliminary search. There are a few spots in the city that my client has already chosen as possible places to set up a new coffee shop, but he wants to see data that will help guide his decision. This would be initial research, and if the data yields a neighborhood that he is unsure of (extraneous factors like shop rent prices, crime, etc.) further analysis will have to be conducted.
C. I plan to do a few things.
    a. Explore Philadelphia neighborhoods using metrics including but not limited to…
        a. population size
        b. population density
        c. average income
    b. Identify the most common venues in each neighborhood to help make a location decision that could have an impact and immediate influence on the business.

	Zipcode	City	Population	Density	Avg Income	Latitude	Longitude
2	19139	Philadelphia, Pennsylvania	43866	24062.66	61720.0	39.961529	-75.230259
3	19107	Philadelphia, Pennsylvania	12340	22723.74	60179.0	39.951623	-75.158637
4	19102	Philadelphia, Pennsylvania	4396	22186.08	51949.0	39.953423	-75.165384
5	19104	Philadelphia, Pennsylvania	50125	16806.34	46520.0	39.960323	-75.197883
6	19106	Philadelphia, Pennsylvania	8359	12566.59	44776.0	39.950472	-75.147231
7	19127	Philadelphia, Pennsylvania	5465	10756.87	43629.0	40.027929	-75.224083
48	19128	Philadelphia, Pennsylvania	36420	5163.86	43629.0	40.049525	-75.230253
8	19152	Philadelphia, Pennsylvania	31379	10538.35	43490.0	40.061595	-75.046385
10	19111	Philadelphia, Pennsylvania	58874	10679.91	41592.0	40.063318	-75.077631
11	19146	Philadelphia, Pennsylvania	35783	22007.33	39075.0	39.939069	-75.182585

3. Description of the data:

A. *Postal Codes* - I was able to download a file from https://www.geonames.org/postal-codes/US/PA/101/philadelphia.html that listed all the postal codes in Philadelphia City. I was also able to web scrape some population data for each postal code from http://zipatlas.com/us/

pa/philadelphia/zip-code-comparison/population-density.htm. I've already put some of the data together, but will try to include an average salary for each zip code.

B. *Income Data* - http://zipatlas.com/us/pa/philadelphia/zip-code-comparison/population-density.htm. This is the income data that I used to merge with the population density data.

C. *Foursquare API* - I will use the Foursquare API to get the most common venues in Philadelphia, and connect them to the postal code data.

D. *Starbucks Data* - I was able to download a CSV file for the Starbucks store locations from www.kaggle.com. I'll use this information to display on a map of possible competitive stores to cross reference my own data.

4. Data Cleaning

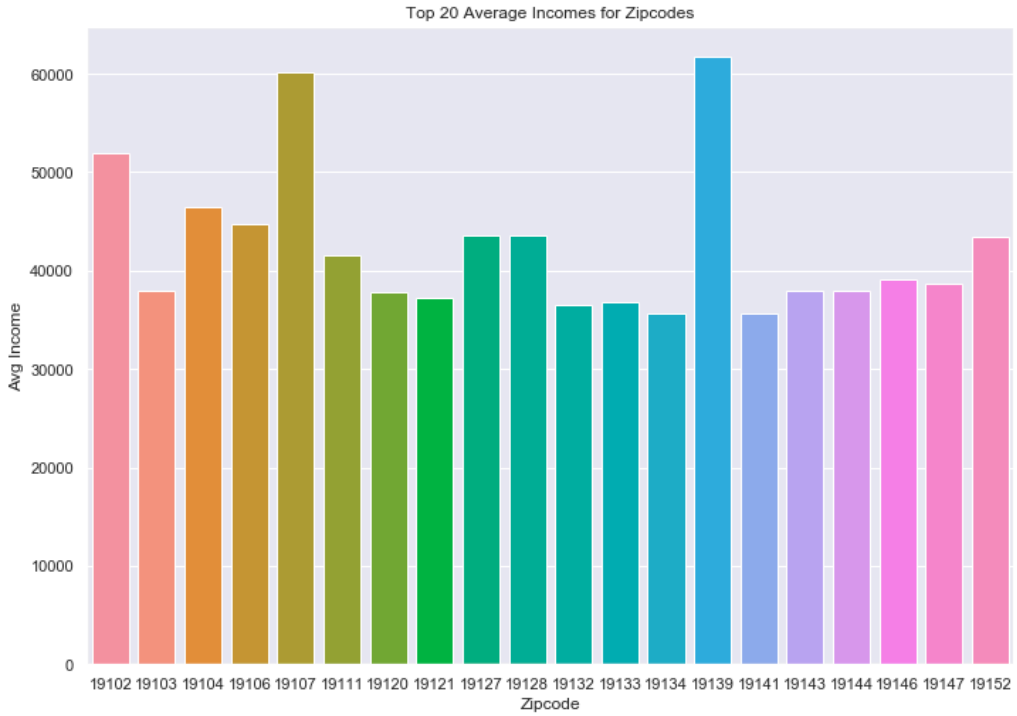
E. Data was downloaded from the above sources and combined into one table. There was difficulty in finding an accurate list of zip codes associated with neighborhood names, and although I could have included these names into the report, I believed that the time-consuming process may have been an influencing factor in choosing a location for a coffee shop. Thus, I did not decide to include the neighborhood names. Finding this information is easy enough after deciding which zip code would be most suitable.

F. All of the datasets were set as objects and each had to be converted to an appropriate float, integer, or string. Upon doing this, some values were missing, and others were mislabeled. I decided to drop the missing values since they would have been outliers.

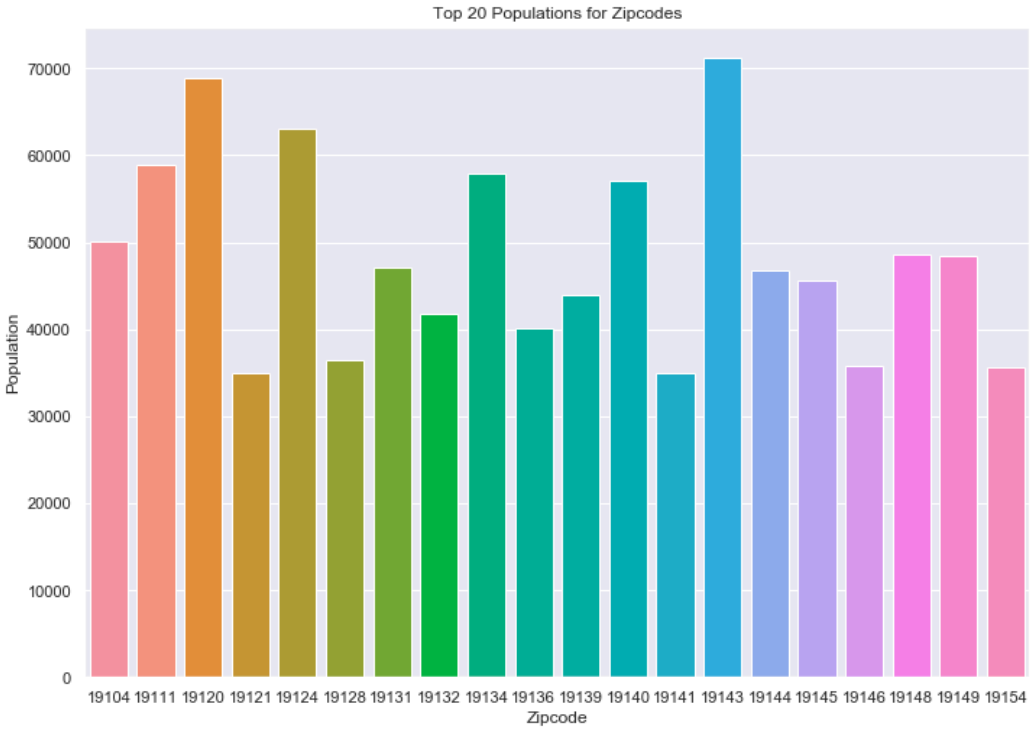
G. Some of the values were also too small to have an impact on location. I decided to drop values that did not have a sufficient population (less than 500) or average income (less than $2000). These would have adversely influenced the dataset.

5. Methodology

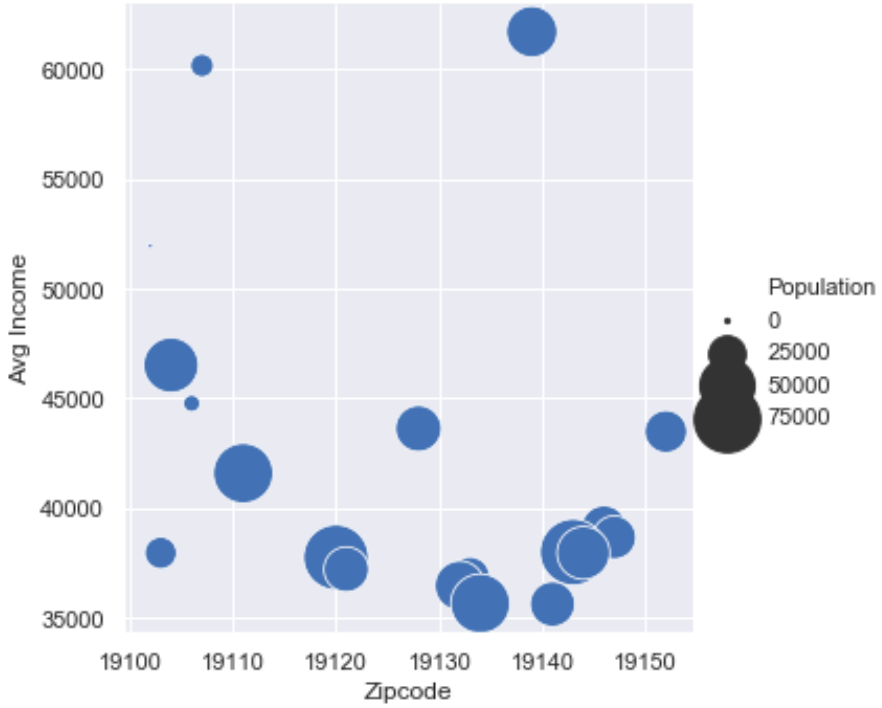
A. A selection for a new coffee shop should be in a location that has a decent amount of extra income. Since the median income in Philadelphia is listed at $36,836, I thought it would be wise to maintain focus on upper incomes in the different Zip Codes. I plotted a bar graph with highest incomes in each Zip Code. *(See graph on following page)*



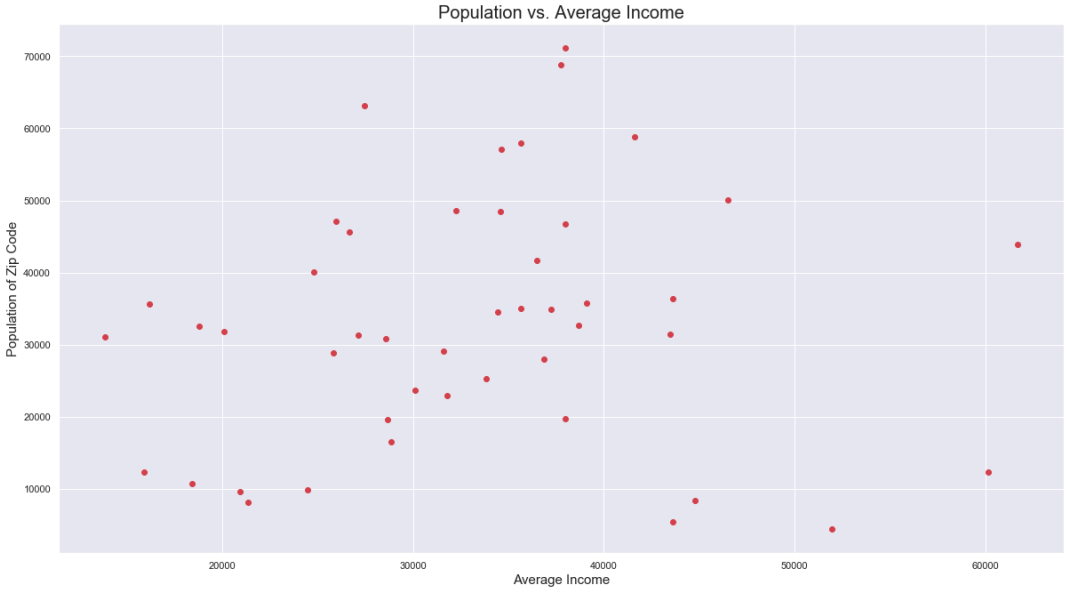
B. Another selector that would be beneficial to know would be the amount of people in a certain area. This would provide a higher number of foot traffic that may come into the store and purchase coffee. I also used a bar graph for this to show the different Zip Codes and most populated areas. I chose not to include the population density, because although density may have been a good factor to use, a high density doesn't mean that the population would be high enough to support a conversion ratio of foot traffic to customers. *(See graph below)*



C. I also used a relationship plot to show the average income with the zip code along with the size of the population within that area. This shows the relationship between the average income and the population that can be easily visualized. The idea is that we want to use the big bubbles that also have a high average income. We'll notice that 4 or maybe 5 zip code provide the appropriate statistics. (Zip Codes, 19104, 19111, 19128, 19139, and 19152)



D. Finally, to add another visualization was a scatter plot that also plotted the average income and the population. *(See graph below)*



E. Finally, I used k-means clustering to identify how similarly the different zip codes are when comparing the most common venues in the area. There were only two clusters that were created, so this machine learning did not have much emphasis on the overall project. It would have been nice to identify an area that was focused on food and/or coffee, but I can't say that this is what happened after examining the results. However, the Center City cluster (high rate of coffee shops) falls into the same cluster as the recommended areas for a new coffee shop location. However, this should not be taken as a definitive reason for selecting that area.

6. Results

Zipcode	Population	Density	Avg Income	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
19104	50125	16806.34	46520.0	Pizza Place	Deli / Bodega	Light Rail Station	Food Truck	Szechuan Restaurant	Office	Beer Garden	Lounge	Food	Bakery
19120	68831	21110.03	37760.0	Korean Restaurant	Shoe Store	Seafood Restaurant	Chinese Restaurant	Women's Store	Kids Store	Mobile Phone Shop	Fast Food Restaurant	Sandwich Place	Supplement Shop
19128	36420	5163.86	43629.0	Convenience Store	Public Art	Flower Shop	Gym	Beer Garden	Yoga Studio	Eastern European Restaurant	Food & Drink Shop	Food	Fast Food Restaurant
19139	43866	24062.66	61720.0	Pharmacy	Metro Station	Food Truck	Caribbean Restaurant	Grocery Store	Shoe Store	Lounge	Breakfast Spot	Fast Food Restaurant	Flower Shop
19143	71169	21569.56	37996.0	Intersection	Chinese Restaurant	Locksmith	American Restaurant	Dessert Shop	Convenience Store	Market	Metro Station	Pharmacy	Park
19152	31379	10538.35	43490.0	Pizza Place	Bar	Pharmacy	Restaurant	Diner	Liquor Store	Baseball Field	Burger Joint	Martial Arts School	Supermarket

The results above are the recommended locations from the data, graphs, and k-means clustering. They all have relatively similar venues, but there should be a focus on either dessert shops or bakeries. These places often work well with coffee shops.

7. Discussion

From the research and data, I wanted to make sure that the place in Philadelphia that we were choosing was in the upper range of the Average Income and the upper range of the Population. From there, I plotted multiple graphs to see if there were any other outliers that may be a good choice. There were a few, but these were all located in the Center City of Philadelphia. This would be a great place, but the cost of upkeep, rent, and competition from other big name brands would be very difficult. Building a reputation in Philadelphia first should be priority number one. If this shop works out, then the high cost in Center City may be a good gamble. It wasn't the focus here.

There are not a lot of coffee shops that were found with the Foursquare API in Philadelphia. Most were concentrated in Center City, but a lot of opportunity lay outside of that area. I narrowed down the selection to 6 Zip Codes above.

Most of the areas are very similar, with the exceptions of 19128 and 19152. These areas have a much lower population density (half as low) as the other possible Zip Code locations. This may be something to discuss in a meeting with the CEO and owner. Ultimately, he will make the decision where to go. I am very satisfied that there will be zero, or close to zero, competing Coffee Shops in any of these locations. I am confident that with this data the CEO can make a more informed decision. As a final note, I purposely left out the Neighborhood names in this data, because it may have been an influencing factor

8. Conclusion

I would recommend Zip Code 19139 because it has the highest density and highest income of all of the candidates locations. With this being said, there should be sufficient income for people to spend time on the weekend to enjoy a coffee at a new coffee shop, or grab a coffee before heading to work. Notably, there is a metro station that would provide a great opportunity to people to get their morning coffee before getting the metro and heading to work. I hope with this information my friend and CEO can make a great decision to expand his business.

In conclusion, this data should suffice for getting to a specific zip code, but the CEO will still need to do more research. I would recommend to continue collecting data and research the following.

- available spaces for rent and price of rent
- walkability - is there easy access to sidewalks?

[1] Philadelphia Wikipedia. https://en.wikipedia.org/wiki/Philadelphia#Economy