Dance Studio Placement in Atlanta, Georgia

An IBM Data Science Capstone Project

Kyle Hutchings*

April 10, 2020

*Contact at: kyle.sh.hutchings@gmail.com

Contents

Abstract

A dance studio is a business which provides the service of dance lessons, most often for school-age children. They have unique challenges which can be mitigated by carefully considering where to locate one. This report details the use of data science techniques and freely available data to answer the question, "What is the best place to start a new dance studio?"

More will be included in the abstract section as the progress is made

1 Introduction

1.1 Background

Dance studios are places where children, teenagers, and even adults can go to learn dance from qualified instructors. There are various kinds of dance studios that specialize in certain forms of dance like a ballet specific studio. Other studios may offer training in a variety of dance types and techniques.

Irrespective of the kind of dance focused on in the studio, a dance studio may choose to enter into dance competitions. Dance competitions are a means for studios to push their students to excel, they create a common goal and sense of loyalty among dancers, and winning competitions can boost a studio's reputation. However, because of costs, long practice hours, and logistics associated with competitions, studios may choose to forgo competitions and instead focus on recreational dance.

1.2 The Business Problem

Though the teaching and learning of dance is an athletic, artistic, and fulfilling endeavor, a dance studio is, at its heart, a business. Every business has unique challenges and needs, and dance studios are no exception. The needs of dance studios include, and are not limited to, competition with other studios for students, student retention, and finding qualified teachers. These needs and challenges necessitate that a potential studio director carefully choose the location for their studio. A good location can help to minimize some of the common issues that studios face.

While some needs like a good student base may be helped by obvious factors like being near public schools, other needs may be influenced by some non-obvious forces. A real issue studios face is student attrition. A somewhat hidden factor in the retention of students is the commitment of parents. Parents need to be on board in order to insure students stick around, because they are likely paying for the lessons and providing transportation. Some locations may provide benefits for parents that will decrease the chances of students being pulled out of dance lessons. Among these bonus features could be close proximity to schools to minimize drive time for parents and as well as being located near shops or a grocery store so parents can run errands while their child is at dance.

A more complete list of issues and "bonus" features, explanations as to why they matter, as well as implications for a data science analysis, has been compiled in Table 1. This list was devised after an interview with a former dance studio Assistant-Director, so it is based on relevant, and real-world experience about what can make a significant impact on the success of dance studios. With all of these factors to consider, it may not be immediately obvious where the best place to start a new studio is. This report seeks to remove some of the difficulty in determining the best location for a new dance studio using data science techniques. Put succinctly, this analysis will seek to answer the question, "What is the best area to start a new dance studio?"

Issue/Bonus	Explanation	Model Implication
Competition with other studios	Two similar studios located too near each other could start competing for students which is obviously undesirable. It's best to be farther away from other studios if possible.	Maximize distance from existing studios
Student base	A dance studio needs a steady base of students. Most dance studios accept only students that are school-aged. Locating in an area near public schools insures there are plenty of potential students. It is also important to note that low-income families may not have the disposable income to sign their children up for dance lessons. Locating in an area with a higher average salaries and lower cost of living may increase student base.	Proximity to public schools; located in a more affluent area
Finding teachers	Finding qualified dance teachers can be made much easier by locating near a college or university. Often, such schools have dance majors or students who formerly danced in studios.	Proximity to colleges and universities
Parent convenience	Parents can have a lot of competing demands to balance. Locating a dance studio near public schools can help eliminate the need for a parent to pick their child up from school and drive across town to dance lessons. Locating the studio near to shops or a grocery store can allow parents to minimize the number of trips they make. They can get groceries or go out to eat while their child is at dance, and not have to worry about being late to pick their kid up.	Proximity to public schools, grocery stores.
Affordable performance venue	Recitals and performances give students a goal to work toward and help parents see the fruits of their financial and time investment. A recital can also help the studio raise funds, but that is only the case if the performance venue is relatively nearby and not too expensive to use.	Proximity to dedicated performance venues; colleges and universities

Table 1: A list of needs and bonus features, explanations about them, and implications they have on a data science model.

1.3 Interested Parties

This report is targeted to any person or persons interested in starting a dance studio. This report may be appealing to those interested in starting a dance studio because it does a lot of the heavy lifting involved in deciding what a good location might be. Namely, it avoids areas with existing dance studios, locates in areas with a good potential student base, and locates in areas where there might be more teachers available to hire. Obviously, personal preferences, financial limitations, and other factors will come into play when considering where to start a dance studio, and this report does not address all factors. Instead, the aim is to supply a potential studio owner with targeted regions that will support the basic needs of a new studio, and maybe even provide a slight edge over their competition.

The results of this report may be more optimal for those seeking to run a recreational dance studio, as competition studios may have additional needs that are not be considered here. Given that the author lives in Atlanta, Georgia, this report will focus on the greater Atlanta metropolitan area. Even with the given limitations, it still stands to reason that any current or potential studio directors in any area may also glean insights from this report.

2 Methodology

2.1 Prioritization of Features

The aim of this report is to find areas for dance studios to be located that will be mutually beneficial for parents, students, the studio owner, and studio staff. Based on the features discussed in Table 1, potential areas will be identified based on the following constraints (in order of priority):

1. A minimum 2 miles away from other existing dance studios
2. Near as many public schools as possible
3. Close to a college or university
4. Near a grocery store
5. Near a performing arts venue

There is reasoning behind this order of priority, and it was informed by interviewing a former dance studio assistant director. First and foremost it's important to have paying customers, i.e. students. Keeping some distance from existing studios will ensure a good reputation for your studio as well as minimal competition for students. Being near lots of public schools will ensure a good general student base as well. Second, a studio needs qualified teachers to function even at a small scale. Close proximity to a university or college is important not only to have potential teachers, but also to make it possible for them to commute. Many college students rely on public transportation, so being within a reasonable walking distance or bus ride of a college could be important to make it reasonable

for college students to consider working at your studio. Lastly, being near a grocery store or performing arts venue are nice bonuses, but nowhere near as vital as having students and teachers at your studio, so they are given lowest priority.

The sources and processing of the data for all of the various features are discussed in Section 3.

2.2 Preliminary Analysis – Maps

As has been mentioned before, the aim of this analysis is to find some broad areas within which a dance studio could locate to avoid some common problems all studios face. As data surrounding different features and venues are gathered, it will be helpful to generate maps to get a feel for how things look. This will give a broad idea of what areas might be good, and what areas might be best avoided.

2.3 Fine Analysis – DBSCAN Clustering

After getting an intuitive feel for what areas might be good by making maps, it will be most helpful to determine more discrete areas which could serve as candidate regions for the opening of a dance studio. For this reason, the Density-based spatial clustering of applications with noise (DBSCAN) algorithm will be employed. DBSCAN will be an effective choice here because it will divide the schools, colleges, grocery stores, and performing arts venues into clusters based on how close they are to each other.

In order to avoid being near existing studios, after gathering and cleaning the data (see Section 3), it will make sense to eliminate any venues within 2 miles of the dance studios. That way, DBSCAN will automatically make clusters which not only exclude existing studios, but also only include relevant locations which are 2 miles or more away from existing studios.

After running the DBSCAN algorithm, the various clusters will be sorted based on which of them have more of the most important features. The central point of the clusters will be calculated and the zip code that the central point lies in will be found. Any studio which locates in this zip code, therefore, will be surrounded by schools, colleges, grocery stores, and performing arts venues.

3 Data

3.1 Sources

To answer the question, "What is the best location for a new dance studio in Atlanta, Georgia?", first, necessary data must be collected. Due to the nature of the problem, location data including features from Table 1 will be the main kind of data utilized – namely public schools, colleges, grocery stores, and performing arts venues. In particular, in order to facilitate the use of a clustering machine learning approach, latitude and longitude data will be required for each of the features. Another important piece of information for each

feature will be its zip code, because a zip code is more easily usable and searchable by an interested party than geographic coordinates.

The Foursquare API was used to find data on existing dance studios, performance venues, and grocery stores. Foursquare is able to provide geographic data and other information like latitude, longitude, the name of the venue, etc. For example, as Figure 1 shows, a list of existing dance studios was easily generated with data provided by FourSquare.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
1050	Downtown Atlanta - Grant Park	33.744222	-84.393238	Rhythma Dance Studio	33.742962	-84.406353	Dance Studio
1264	Cabbagetown - East Atlanta Village - Ormewood ...	33.736990	-84.350905	Dance 411 Studios	33.741120	-84.349610	Dance Studio
1350	Cabbagetown - East Atlanta Village - Ormewood ...	33.736990	-84.350905	The Beam: Moving In The Spirit	33.741769	-84.361998	Dance Studio
1599	Morningside/Lenox Park - Piedmont Heights - Le...	33.821216	-84.368706	Academy Ballroom	33.830702	-84.363622	Dance Studio
2139	Hapeville	33.672086	-84.371992	MoveWithAPurpose Dance	33.681068	-84.381351	Dance Studio

Figure 1: A data frame containing data gathered from the FourSquare API.

The majority of the data in this report started with information scraped from the following webpages:

- The Keen Team – Real Estate website containing a list of zip codes and neighborhood names for Atlanta, Georgia. No other pre-made list was found elsewhere.
- Education Bug – List of public schools in Georgia by county
- Wikipedia – List of colleges and universities in Georgia

After scraping the information from the web pages using the Beautiful Soup package, it was necessary to add in latitude, longitude, and zip code. For example, the web scrape of the List of Colleges and Universities in Georgia (U.S. State) web page provided a list of the names of colleges in the state of Georgia (see Figure 2. After that, the geographic coordinates were found by geocoding the names using the Geopy package in Python, and zip codes were found using the US Zip Code package in Python. The results of this can be seen in Figure 3

3.2 Cleaning

Data from the Foursquare API contains information about all venues in a given locality. Foursquare has done a very thorough job categorizing venues. For this situation, a large number of unnecessary categories were present that were dropped so only categories relevant to features in Table 1 were included.

Knowing that many schools in the United States might share the same name, schools were fed to the Geopy geocoder using their name along with "Georgia, USA". Even with these measures taken, some of the schools were geocoded incorrectly as other schools in the

	College Name
0	Georgia Institute of Technology
1	University of Georgia
2	Augusta University
3	Georgia State University
4	Georgia Southern University
5	Kennesaw State University

Figure 2: The first few entries of a data frame containing a list of colleges scraped from Wikipedia.

	College Name	Latitude	Longitude	Zip Code
0	Georgia Institute of Technology	33.776033	-84.398841	30363
1	University of Georgia	33.940428	-83.373049	30606
2	Augusta University	33.469335	-81.988562	30912
3	Georgia State University	33.754794	-84.387896	30303
4	Georgia Southern University	32.421438	-81.784505	30458
5	Kennesaw State University	34.038832	-84.582745	30144

Figure 3: The first few entries of a data frame containing the geographic coordinates and zip codes of the colleges.

USA and around the world. These schools that were assigned an incorrect location were weeded out.