Sips & Flips

*A Business Proposal and Geographic Analysis*

Alec Biehl IBM Applied Data Science Capstone December 2020

# Section 1: Problem Statement

Contemporary young urban professionals, colloquially referred to as *yuppies*, lie at the crux of two major societal shifts in the United States, both of which are partially rooted in the push towards more sustainable travel (i.e. less car dependency):

1. Migration towards urban centers to experience the plethora of opportunities tied to economic investment, as well as the allure of “social connectivity” in comparison to the relative isolation of low-density suburbia; and
2. Telecommuting, or the ability to work remotely away from a central office location, that is the result of a growing flexibility offered by employers to meet policy goals of traffic congestion mitigation, in addition to the growing attraction and feasibility of sustained freelance work, independent consulting, and various other “lone wolf” professions that offer the freedom of daily mobility.

Concurrently, although “work from home” is certainly an option that has exploded due to the Covid-19 pandemic, there is sufficient evidence that a psychological distance between work and home life is good for mental health, especially for individuals who are apartment or condominium dwellers.

This is where *Sips & Flips* enters the picture. As a hybrid between a café and bookstore, it would be nostalgic for the degree-holders who might recall those not-yet-distant college years and the many hours spent reading or writing in libraries, yet offer a more relaxing atmosphere than what is typically associated with such institutions. Furthermore, *Sips & Flips* would differ from the setup between Starbucks and Barnes & Noble as follows. First, rather than a collection of tables adjacent to one another, the business would offer Wi-Fi-equipped “work nooks” to nurture productivity with minimized distractions. Second, to foster a sense of community among recurring patrons, books will be made available for rental (as is the case with libraries) to encourage reading & discussions during short “mental reset” breaks. Third, and perhaps the easiest outcome to achieve, the coffee will be better than that offered by the insinuated competition…

Accordingly, it is critical to identify an advantageous location for this entrepreneurial endeavor, which would ideally be within a city characterized by a high influx of young urban professionals and a sizable university graduate student population (since this latter market segment is essentially the yuppie equivalent in academia). Based on recent web rankings by Niche[[1]](#footnote-1) and studyportals[[2]](#footnote-2), the U.S. metro area that appears the highest on both lists is Boston, Massachusetts (as well as its neighbor, Cambridge); therefore, it is selected for the case study analysis comprising this report.

# Section 2: Data Description

To determine where the first *Sips & Flips* should open within Boston, two sets of data are required at minimum: (a) the “neighborhood composition” of the city to be matched with geospatial attributes and (b) the current set of cafés/coffee shops and bookstores. The former will be used as input into the Foursquare location-based API that will be used to extract the latter. In addition to the *absolute count* of these institutions within each neighborhood, the *relative count* in comparison to other institution types among the Top 100 reviewed locations will allow for the clustering of neighborhood based on general activity type. In other words, we would want to identify places that should be avoided either since the number of competitors to *Sips & Flips* is too high or because the presence of similar institutions is minimal, which could indicate low demand.

Wikipedia can be used to extract a list of neighborhoods for both cities:

* Boston: <https://en.wikipedia.org/wiki/Neighborhoods_in_Boston>
* Cambridge: <https://en.wikipedia.org/wiki/Cambridge,_Massachusetts#Neighborhoods>

While these delineations typically reflect socio-political and cultural heritage, population data is often aggregated using official U.S. Census boundaries, thus making it difficult to merge these geographies with socio-demographic attributes such as composition by race/ethnicity, education level, and household income. As an alternative approach so that the **2019 American Community Survey estimates** could be integrated into the clustering analysis, one could use the following website to extract all zip codes covering Boston and Cambridge: <https://www.unitedstateszipcodes.org/>.

On a final note, a two-step clustering approach will be implemented to identify relatively homogeneous groups of neighborhoods. To expound, this method first employs (a) Ward hierarchical clustering to identify the optimal number of clusters; next, the ascertained cluster centers are input as parameters for (b) K-means partitioning to refine the solution. According to [this paper](https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=1540560&casa_token=d38L5YvsjD0AAAAA:ZapffVCAhny2Bi3EARJTvNWkwh-SZmQRHm584zuXroyhZvuvCylHTcUCBJG-OTBDiVbvNKJ7pOQ) by Bernard Chen et al (2005), this hybrid approach circumvents the shortcomings that using each one in isolation entails.

1. <https://www.niche.com/places-to-live/search/best-cities-for-young-professionals/> [↑](#footnote-ref-1)
2. <https://www.mastersportal.com/articles/1839/top-10-student-cities-in-the-us-and-canada-perfect-for-a-graduate-degree.html> [↑](#footnote-ref-2)