**Seasonal Consumer Coffee Habits in Boston**

Boston has the highest number of coffee shops per citizen in the United States; for every one hundred thousand citizens, there are sixty-one coffee places[[1]](#footnote-1). In addition, according to WalletHub[[2]](#footnote-2), Boston is among the finest twenty places for coffee in the United States. Moreover, coffee shops are a sociocultural phenomenon that people from all over the world can relate to and enjoy. Coffee is a big part of day-to-day life for many people. To illustrate, people grab coffee when they need to go to work, when they have homework to do, when they need a pick me up, when they want to catch up with a friend, and more. Additionally, specialty cafes are popping up in the suburbs; each with its own customer base and main selling points. With that in mind, leveraging consumers’ behavior in the industry of coffee shops leads to fruitful insights that shed light on various social, cultural, and economic factors, making it a highly relevant area of research.

In particular, this research project seeks to analyze consumers’ coffee habits across seasons. More specifically, given that SafeGraph offers data from January 2022 to August 2022, this research project aims to analyze whether consumers in Boston are more likely to visit coffee shops during the winter, spring, or summer. In addition to that, throughout this project, the implementation team plans to provide information regarding the geographic areas that have the highest and lowest number of coffee shops, hours, days, and months, during which visitors frequent coffee shops more, coffee shops with the highest and lowest number of visitors, and so on.

In terms of utilizing the data, the team plans on employing a few different filtering strategies and on choosing the optimal one based on the robustness of the data the strategy produces. The first strategy is to choose a few larger chain coffee shops, such as Starbucks and Dunkin, and filter the remaining coffee shops out of the location\_name or brands columns. This allows the implementation team to choose on which coffee shops the research will be conducted. Under this strategy, the research would be conducted on larger Boston brands, such as Starbucks and Dunkin, and would exclude individual, local coffee shops. Another strategy would be to filter location\_name or brands for businesses that have as their title some of the main keywords used to describe coffee shops. Among such keywords can be the terms “coffee” or “cafe”.

When cleaning up the data for this research project, the following columns will be included: location\_name, brands, street\_address, postal\_code, date\_range\_start, date\_range\_end, raw\_visit\_counts, raw\_visitor\_counts, visits\_by\_day, distance\_from\_home, median\_dwell, popularity\_by\_hour, and popularity\_by\_day. In particular, date\_range\_start and date\_range\_end will be used in conjunction with the visits\_by\_day and raw\_visit\_counts columns to compare visit behaviors in January versus July, capturing the seasonal aspect. In addition, raw\_visitor\_counts will be used to see the amount of unique visitors, as the two visit columns used previously may capture a unique consumer visiting one coffee shop multiple times a day and account for it accordingly. The popularity\_by\_hour and popularity\_by\_day columns will be analyzed to find which brands and locations are the most popular each month and which seasons and months are the most and least popular to visit a coffee shop. Moreover, median\_dwell will allow the implementation team to analyze which coffee shops consumers spend the most amount of time at and whether the season has an impact on this dwell time. Furthermore, distance\_from\_home will be utilized to find out if consumers are less likely to travel farther from home when it is a colder month, such as January. Lastly, postal\_code and street\_address will enable the team to analyze the number of coffee shops in one area and connect this information back to the statistics on the strong prevalence of coffee shops in Boston given earlier. Hence, the usage of SafeGraph cell phone tacking data is of crucial importance to the implementation team and the completion of this research, as it provides detailed information that this project aims to address and analyze.

Several challenges that are both external and internal to the data set are expected to arise when completing this project. An external factor is the impact of COVID-19 policies on consumers’ coffee consumption behavior. In January, there were more COVID-19 related policies and consumers acted more cautiously in relation to masking, social distancing, and attempting to minimize contact with others. This may skew the data to less visits during months when COVID-19 and related policies and behaviors were more prevalent, such as January compared to July. Other challenges that are expected to arise are internal to the data set. One of them comes from the collection of data based on smartphones. On one hand, some individuals do not own a smartphone and so the data cannot capture their behavior. On the other hand, some individuals own more than one smartphone, thus inflating the data. Another internal challenge is mobile ordering. Consumers can mobile order coffee on their Dunkin and Starbucks apps. In this, the consumer simply runs into the location, grabs his/her coffee, and leaves. The concern here is that the data set may not be able to capture these consumers. In addition, another challenge is related to the filtering process of the SafeGraph data set so that it only contains information on coffee shops. The data set cannot be filtered by NAICS code because NAICS code 722515[[3]](#footnote-3) captures not only coffee shops but also all “snack and nonalcoholic beverage bars.” Moreover, filtering the data by keywords, such as “coffee” or “cafe” is not feasible and will not be used because this method would remove key players, including Dunkin and Starbucks, two of the biggest coffee chains in the United States. Moreover, missing values within columns of interest pose another challenge, for example brands and distance\_from\_home. The implementation team is considering a few different routes to deal with this missing data. One option is to exclude the columns from the data altogether, as the sheer number of missing values makes the information ineffective. Another option is to create a boxplot of each column and utilize it to identify and drop the outliers as seen fit.

Once the data is clean, the implementation team will utilize the data set to give more precise information relative to the project. The results and intuitions found in the research will allow the team to draw in-depth conclusions on the relationship between coffee consumption habits and the three seasons, namely, winter, spring, and summer in the year of 2022.

The introduction is fine. One concern is the purpose of the study is too descriptive, so I recommend you add some Covid situations in your question. The questions are not only well-defined, and they address significant matters in real life.

I want to see more strategies on how to filter the coffee shops.

Here is what I want you to do:

1. Find all the POIs relevant to coffee shops
   1. Provide me the specific ways in words
      1. Need to have at least FOUR DIFFERENT strategies
         * Give me a FULL list of keywords if you want to filter based on the name of POIs
   2. Provide me the result
      1. Summary statistics of POIs for each strategy
      2. Find the number of raw visitors for the corresponding strategy
         * Summary statistics of visitors
         * Add time (month) dimension if necessary

Send me the result by 11th. If you want to talk with me, please use:

<https://calendly.com/ymoon-econ/30min_moon>

Motivation: Good

Answer Strategy: Poor

Writing quality: Good

1. Moore, Eric. “U.S. Cities With the Most Coffee Shops.” *Overheard on Conference Calls*, 13 Jan. 2020, https://overheardonconferencecalls.com/business/cities-most-and-least-coffee-shops/. [↑](#footnote-ref-1)
2. Buyinza, Alvin . “Boston Is Still One of the Best Coffee Cities in the US, WalletHub Says.” *Masslive*, 21 Sept. 2022, https://www.masslive.com/news/2022/09/boston-is-one-of-the-best-coffee-cities-in-us-wallethub-says.html. [↑](#footnote-ref-2)
3. “NAICS Code: 722515 Snack and Nonalcoholic Beverage Bars.” *NAICS Association*, https://www.naics.com/naics-code-description/?code=722515. Accessed 3 Nov. 2022. [↑](#footnote-ref-3)