Which City has the Best Coffee?

Calli Morrison

March 8, 2020

Introduction

Background

Coffee has changed completely in its form, use, and taste over the past century. Since the introduction of coffee to the common household, it has changed from being a commercially manufactured powder to a craft beverage. As the drink has become increasingly popular over the past several decades, the coffee market has grown exponentially in the United States. Consumers have spent billions of dollars annually on coffee and that number is growing. The experience of coffee has become more than just the purchase. Over the past several years, coffee has become an experience rather than simply a product. Aesthetic coffee shops have popped up across the United States offering free wifi, unique desk spaces, and an art design in your porcelain mug.

Problem

In 2018, WalletHub conducted a survey to find the best city for coffee in America. The survey found that Seattle ranked first and New York came in as a close second. Seattle is known as the birthplace of second wave coffee (essentially, a movement made popular by Starbucks) and has given a stage to third and fourth wave movements. New York, as the largest city in the United States, hosts a wide variety of coffee shops, including several iconic craft coffee brands.

The corporate giant Starbucks will be the focus of this analysis project. As a stakeholder in the Starbucks corporation, staying competitive among coffee brands and ahead of the competition is imperative. Public perception of the store experience is essential to the company's success. To get in a glimpse into this insight, I will compare store locations between the two

“hottest” coffee cities' downtown areas. As the results will show the engagement and perception of these stores, stakeholders would be able to highlight exceptional and underperforming stores and make strategic improvements as appropriate.

Data Acquisition

To provide this analysis, a variety of data will be needed. This data includes basic information (name, coordinates, etc.) of Starbucks locations within both downtown areas, engagement information regarding how many people are interacting with the store and some measurement to show what this experience is like, and a numeric comparison on the similar scale so that it is possible to compare all stores.

The Foursquare website is a user-based information source for geographic areas. The Foursquare API provides all of the above listed information. Using the “venue search” endpoint with the “coffee shop” category will provide the coffee shop names and coordinates for each city. The “venue details” will provide the engagement stats, which include the amount of coffee shop “likes” and the amount of “tips” provided from Foursquare users. It will also provide a rating for each coffee shop based on a ten-point scale.

When using the GET request, the information to be imported into a JSON file. At this point, I will be able to convert the information into a pandas dataframe that will be easy to further analyze. I will keep the dataframes separate for each city.

Methodology

After importing the necessary libraries and calling the Foursquare API with the venue search endpoint, I was then able to convert the provided JSON file into a pandas dataframe. I filtered and cleaned the columns to only include the name, foursquare ID, and coordinates of all

the coffee shops in the Downtown Seattle area. After taking an exploratory look at the total number of coffee shops, I filtered my dataset to only include the Starbucks stores in the area.

At this point, I used the venue details endpoint to pull in the price tier, likes amount, tip count, and rating for each of the Downtown Seattle Starbucks locations into the same dataframe with the name, ID, and coordinates as pictured below:

	name	id	lat	lng	price_tier	likes	tip_count	rating
0	Starbucks	4adf37bdf964a520677821e3	47.605054	-122.330211	1	24	5	7.3
0	Starbucks	5c7b51173fffb40025ffd274	47.610703	-122.333494	1	2	0	7.2
0	Starbucks	5a0390ff51950e0aa487d16e	47.608268	-122.338383	1	2	0	7.1
0	Starbucks	5050c269e4b020e9ae019bf2	47.605453	-122.330250	1	5	2	6.9
0	Starbucks	5293a854498e45b5f668644b	47.605286	-122.334370	1	22	2	6.8

Then, this process was repeated for the Downtown New York area:

	name	id	lat	lng	price_tier	likes	tip_count	rating
0	Starbucks	4ad76597f964a5200e0a21e3	40.712204	-74.008052	1	281	46	7.5
0	Starbucks	5bb75d4f9fca56002c7e218b	40.714817	-74.008242	1	1	0	6.8
0	Starbucks	5050b1ade4b05606213f5466	40.711142	-74.004796	1	6	2	6.6
0	Starbucks	4a84eb0af964a520e1fd1fe3	40.711598	-74.006726	1	301	64	6.5
0	Starbucks	59831458e679bc7ea5d6f899	40.712128	-74.010818	1	15	2	6.8
0	Starbucks	59b94fdb2ebb346f98fabd4a	40.710967	-74.000763	1	5	3	6.5

At the end of processing, I had two matching datasets for the Downtown Seattle and Downtown New York areas ready for comparison.

To find the needed information on store engagement, I compared the number of likes and tip counts between the cities. As the population of New York is much larger than Seattle, I normalized the Seattle data to reflect the difference in population. Then, I compared averages of these numbers.

To compare the ratings for each area, I simply compared averages from each dataset. As the ratings were on the same scale, it was a very simple comparison. After presenting the numbers, I put the Starbucks datasets on a Stamen Toner Folium map and showed a visual comparison of the Downtown Seattle Starbucks locations:

and the Downtown New York Starbucks locations:

Results & Discussion

For engagement rates, I examined the average of each city. When I first averaged the areas' number of likes and tips, New York's average totaled 101.5 and Seattle's totaled 11. After dividing the population of New York by Seattle, I found that New York's population is 12 times bigger than Seattle's. Therefore, I multiplied Seattle's results by 12 to more accurately represent the engagement numbers in light of the population difference. The average for Seattle's like count was then increased to 132. I disregarding the tip count data as it was missing for several locations.

For ratings, I examined the highest, lowest, and average score of each city. The highest rated store in Seattle is 7.3, while the highest in New York is 7.5. The lowest score in Seattle was 6.8, while the lowest in New York was 6.5 for the downtown locations. The average rating was a better reflection of overall experience, with Seattle's stores having an average of 7.06 and New York at an average of 6.78. The data resulted in Downtown Seattle's locations having a slighter better engagement rate and overall rating score than Downtown New York's Starbucks locations.

Conclusion

By using the Foursquare data to calculate the average likes, tip_counts, and rating, we found that the averages between the two, once normalized, were slightly higher for Seattle. Overall, the comparison between the cities was fairly similar and it should be noted that further data is needed to determine a full analysis of what internal and external variables may be influencing these results to be able to address them appropriately.

Several considerations exist for further analysis. First, while stakeholders may use the results of this project to locate their best and worst rated stores, this information may be relative to the store placement, management, staffing, time of year, and other variables that commonly influence store performance and public perception. Secondly, looking at this data across a time

period may provide valuable insight into data patterns and highlight external variables that are influencing store performance and perception. Next, while a commonly known brand was chosen, it may be interesting to compare engagement and ratings with other large coffee brands and local craft coffee shops. This analysis could be expanded to look at other areas of the coffee market, such as coffee roasters. Last, Foursquare's data is mostly user-generated, which is susceptible to its user base. Comparing Foursquare's data with other available data sources may provide important insight. In conclusion, as a Starbucks stakeholder, it may be important to have analytic insights such as these when determining business strategy for each of these areas.