Discover the pattern of Starbucks in city and neighborhoods of US

Jackie Yong Leong Shong
May 4, 2019

1. Introduction

Can we deduce how Starbucks choose their target city or neighborhood? Can we spot the pattern behind their existing store locations?

In the United States (US), fast food becomes one of the cuisine cultures, and start growing globally since last century. McDonald, KFC, Starbucks, Burger King and others make one feels no different when situated at any cities. Starts firstly as wagon located just beside highway, fast-food culture already intruding in majority urban neighborhoods. Therefore, location becomes the most crucial strategy to open one fast-food store.

This study is aimed to discover the pattern of **city and neighborhood** which has high popularity of fast-food chain, specifically, **Starbucks**. The pattern study includes the **demographic, economic and venues / Point Of interest (POI)** for this time. Target of audience will be any fast-food entrepreneurs. The discovered pattern might help them to justify the popularity of one location before open a new store nearby.

2. Data

Based on the stated problem, involved dataset including:

- Existing Starbucks stores data in US cities
- City data in US city
- Neighborhood data in targeted city that has high density of Starbucks
- Venues data in chosen neighborhood

2.1. Existing Starbucks stores data in US cities (2017)

First, the existing Starbucks store data in US[1] is collected, the data will help us to narrow down to one target city. Please note that **city** here is similar to the concepts of **place**, in which population ranging from 10,000 to over 1,000,000.

The United States Census Bureau defines a place as a concentration of population which has a name, is locally recognized, and is not part of any other place. A place typically has a residential nucleus and a closely spaced street pattern, and it frequently includes commercial property and other urban land uses. A place may be an incorporated place (a self-governing city, town, or village) or it may be a census-designated place (CDP)[2].

For example, New York City has 5 boroughs (Manhattan, Brooklyn, Staten Island, The Bronx and Queens), and we will study these places respectively, rather than view them as an entity.

In 2017, Starbucks has 13608 franchises in US, in which also includes Coffee House Holdings, Evolution Fresh and Teavana. There are 3567 US cities (places) that has at least one Starbucks in that time. Among these places, **New York's Manhattan** itself has the highest number of Starbucks: 233 stores. Manhattan is the fondest place of Starbucks in East Coast, but others top city candidates mainly distribute along West Coast.

	address	city	state_ab	country	postcode	long	lat
0	5600 Debarr Rd Ste 9	anchorage	AK	US	995042300	-149.78	61.21
1	1725 Abbott Rd	anchorage	AK	US	995073444	-149.84	61.14
2	1501 Huffman Rd	anchorage	AK	US	995153596	-149.85	61.11
3	320 W. 100th Ave, 100, Southgate Shopping Ctr ...	anchorage	AK	US	99515	-149.89	61.13
4	1005 E Dimond Blvd	anchorage	AK	US	995152050	-149.86	61.14

Figure 1. Starbucks store data in US 2017 (partial)

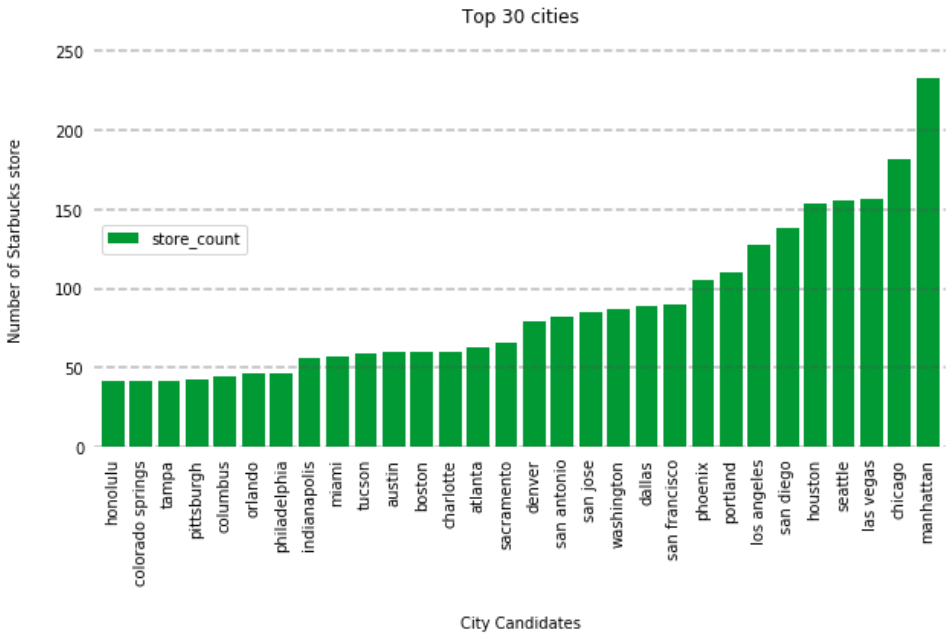

Figure 2. City candidates have at least one Starbucks in US 2017 (top 30)

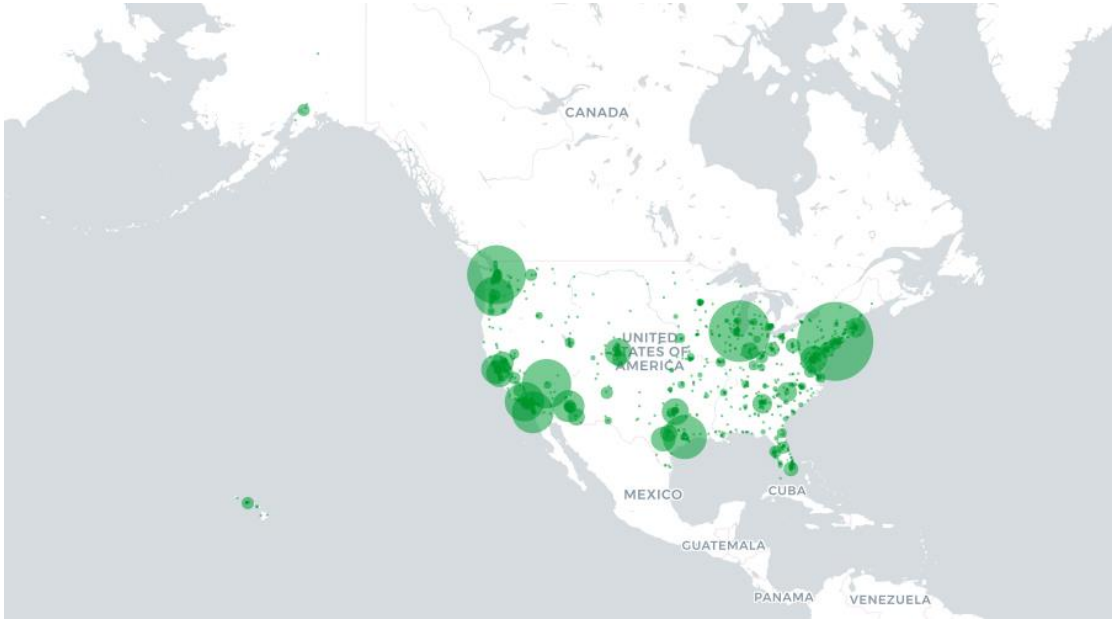

Figure 3. Starbucks stores location and number in US 2017 (top 1500)

2.2. City data in US (2017)

City data (all level population) is collected[3,4], features include city, state, population, gender, age median, marriage conditions, mortgage, high school education, household income and family income. Due to the data limitation, we only investigate the pattern of Starbucks stores among 2960 places (around 5.4% data does not match with Starbucks stores data).

	state_ab	city	land_area	water_area	population	male_population	female_population	population_density	male_age_median	female_age_median
0	AK	anchorage	1.102774e+09	4.936872e+08	150561.0	77433.0	73128.0	1.365294e-04	33.083330	34.000000
1	AK	copper center	6.395523e+10	1.214892e+09	2554.0	1298.0	1256.0	3.993418e-08	36.000000	35.333330
2	AK	craig	3.358007e+08	8.439239e+07	2513.0	1326.0	1187.0	7.483606e-06	41.583330	39.583330
3	AK	delta junction	1.729676e+10	3.763714e+08	4794.0	2742.0	2052.0	2.771618e-07	35.583330	35.833330
4	AK	dillingham	4.096520e+08	7.194810e+06	2335.0	1144.0	1191.0	5.699960e-06	33.750000	32.250000
5	AK	eagle river	5.092349e+08	1.859309e+06	3727.0	1780.0	1947.0	7.318823e-06	39.333330	39.666670

Figure 4. Cities / places data in US 2017 (partial)

2.3. Neighborhood data in Manhattan (2017)

2.3.1. Starbucks stores in Manhattan

Beside city level, correlation of features in a more zoom in scale, which is neighborhood is also needed to study. From 2.1 it is concluded that Manhattan has the highest number of Starbucks store.

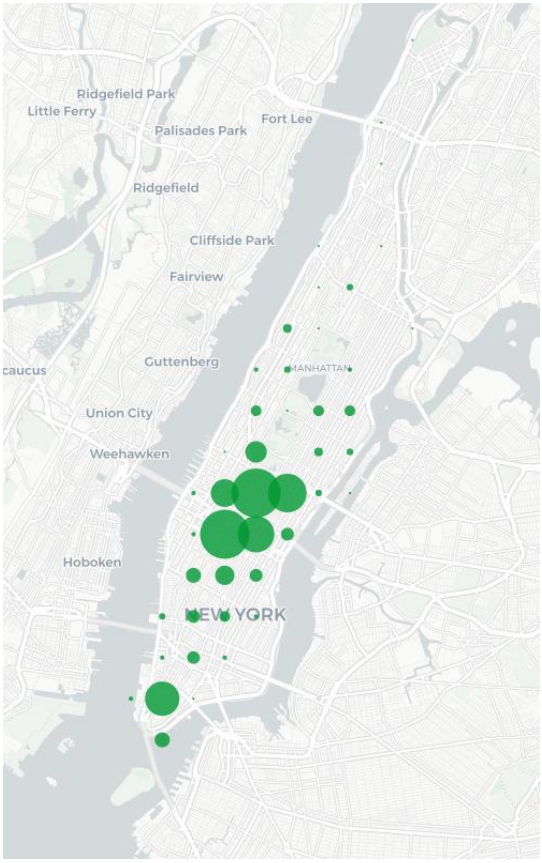

Figure 5. Starbucks stores location and number in Manhattan 2017

From figure above, we could discover that majority of Starbucks in Manhattan distributed around **Midtown Manhattan**, which is composed of **Midtown, Chelsea, Clinton, Gramercy Park & Murray Hill**. Next, we will investigate which neighborhood has high popularity of Starbucks.

2.3.2. Prepare Manhattan neighborhood list and postcode table

Since majority census data normally recorded in terms of 'place' (does not includes neighborhood level), neighborhood is identify according to postcode of stores[5], in which result in 11 groups of neighborhood: 1) East Harlem; 2) Central Harlem; 3) Inwood and Washington Heights; 4) Morningside and Hamilton Heights; 5) Lower East Side; 6) Greenwich Village and Soho; 7) Upper West Side; 8) Upper East Side; 9) Financial District; 10) Gramercy Park and Murray Hill; 11) Chelsea, Clinton and Midtown. The last group has the highest number of Starbucks store (68).

	neighborhood	postcode
0	central harlem	10026
2	central harlem	10030
3	central harlem	10037
4	central harlem	10039
43	chelsea & clinton & midtown	10103
44	chelsea & clinton & midtown	10105
45	chelsea & clinton & midtown	10107
9	chelsea & clinton & midtown	10020
10	chelsea & clinton & midtown	10036
7	chelsea & clinton & midtown	10018

Figure 6. Manhattan neighborhoods and postcode table (partial)

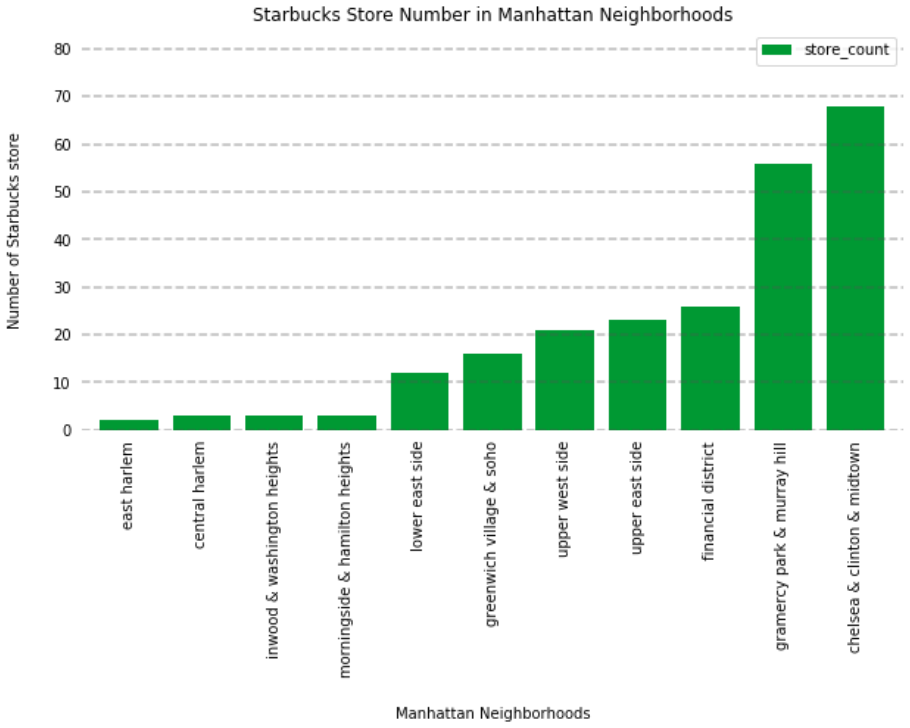

Figure 7. Starbucks stores number in Manhattan neighborhoods (postcode) 2017

2.3.3. Import neighborhood data

Neighborhood data in Manhattan is collected[6], features include demographics, housing market, land-use and development, neighborhood service and rentals.

	neighborhood	store_count	Born in New York State	Disabled population	Foreign-born population	Population	Population aged 65+	Households with children under 18 years old	Single-person households	Percent Asian	Percent black	Percent Hispanic	Percent white	Racial diversity index
0	east harlem	2	0.570	0.136	0.255	128316	0.128	0.250	0.449	0.083	0.264	0.519	0.119	0.64
1	central harlem	3	0.533	0.087	0.252	147442	0.114	0.264	0.389	0.041	0.530	0.247	0.149	0.63
2	inwood & washington heights	3	0.370	0.083	0.440	219998	0.149	0.234	0.310	0.024	0.083	0.669	0.202	0.50
3	morningside & hamilton heights	3	0.424	0.059	0.313	136017	0.131	0.211	0.350	0.102	0.224	0.365	0.285	0.73

Figure 8. Manhattan neighborhoods data in US 2017 (partial)

2.3.4. Import venues (POI) data around neighborhoods

The location of the existing Starbucks store is calculated and achieve a mean location for each neighborhood. Venues data around these locations are collected via Foursquare API[7]. There are overall 197 venue categories, 1099 venues in total, in which every neighborhood has 100 venue data.

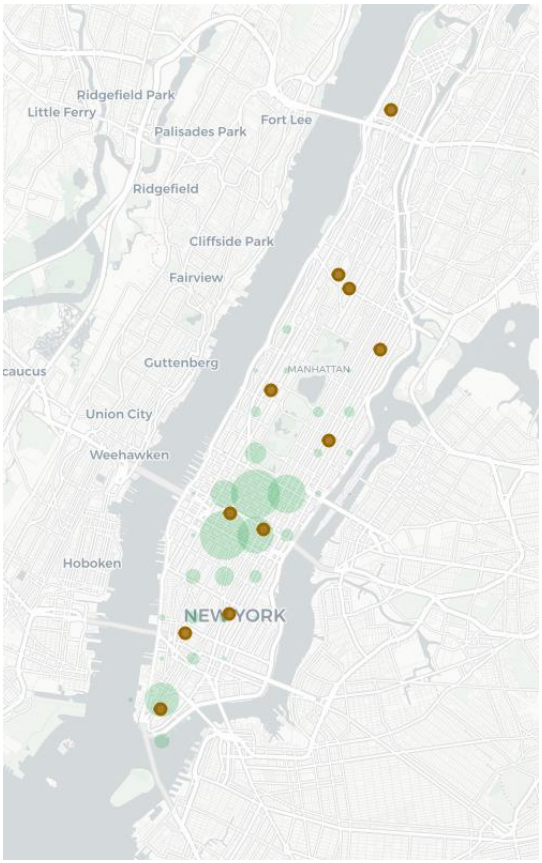

Figure 9. Suggested locations to achieve venue data in each neighborhood

	neighborhood	store_count	Accessories Store	American Restaurant	Arepa Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Auditorium	BBQ Joint	Bagel Shop	Bakery	Bar	Beer Bar
0	east harlem	2	0	0	0	0	1	0	0	0	0	0	7	1	2
1	central harlem	3	0	3	0	2	1	2	0	0	0	1	2	0	2
2	inwood & washington heights	3	1	1	1	0	0	0	1	0	0	0	5	3	0
3	morningside & hamilton heights	3	0	4	0	2	1	2	1	0	1	0	2	0	1
4	lower east side	12	0	0	1	1	0	1	1	0	1	1	1	1	0
5	greenwich village & soho	16	0	3	0	0	1	1	0	0	0	1	1	0	1

Figure 10. Venue category data in Manhattan neighborhoods (partial)

	neighborhood	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
0	central harlem	Jazz Club	French Restaurant	Southern / Soul Food Restaurant	Wine Shop	Seafood Restaurant	Coffee Shop	Italian Restaurant	Park	American Restaurant	Theater
1	chelsea & clinton & midtown	Theater	Hotel	Yoga Studio	Japanese Restaurant	Coffee Shop	Sushi Restaurant	Bakery	Pizza Place	Plaza	Burger Joint
2	east harlem	Mexican Restaurant	Bakery	Pizza Place	Latin American Restaurant	Café	Thai Restaurant	Deli / Bodega	Park	Italian Restaurant	Spa
3	financial district	Coffee Shop	Hotel	Pizza Place	Café	Steakhouse	Wine Shop	Sandwich Place	Cocktail Bar	Park	Gym

Figure 11. Venue category grouped and sorted data in Manhattan neighborhoods (partial)

3. Methodology

In this project we will focus on cities in US that have high numbers of Starbucks, we will study what features will influence the most on the number of Starbucks stores in these city candidates, both in city level and neighborhood level.

In city level, we will first conduct the study to investigate what demographic and economic feature (such as population, income, education, gender ratio, debt, mortgage and etc.) is related the most to number of Starbucks store within US cities that has at least 100,000 population, the study is conducted via correlation matrix [4.1]. Next, we will discover what is the city we should focus in study the correlation of higher popularity by Lorenz Curve, Gini and Nakamoto Coefficient [4.2]. After identifying the targeted city, we will explore the crucial features that highly impact on the density of Starbucks in one city by correlation matrix again [4.3].

Next, we will focus on the neighborhood scale. The characteristics of chosen neighborhoods in selected city will be studied further in terms of demographic, economical and POI composition [4.4, 4.5]. Clustering will be attempted to study the pattern of nearby POI composition [4.6].

4. Analysis, Result & Discussion

4.1. Pattern in city level

4.1.1. Starbucks pattern in all city candidates in terms of demographic and economic

In city level of analysis conducted by correlation, it first discovers that **population** (includes male, female and entire population) is the primary positive factor, means higher population to some extent will influence number of Starbucks in one place. However remaining features exhibit vague relationship with store number, next we target on cities has higher popularity of store.

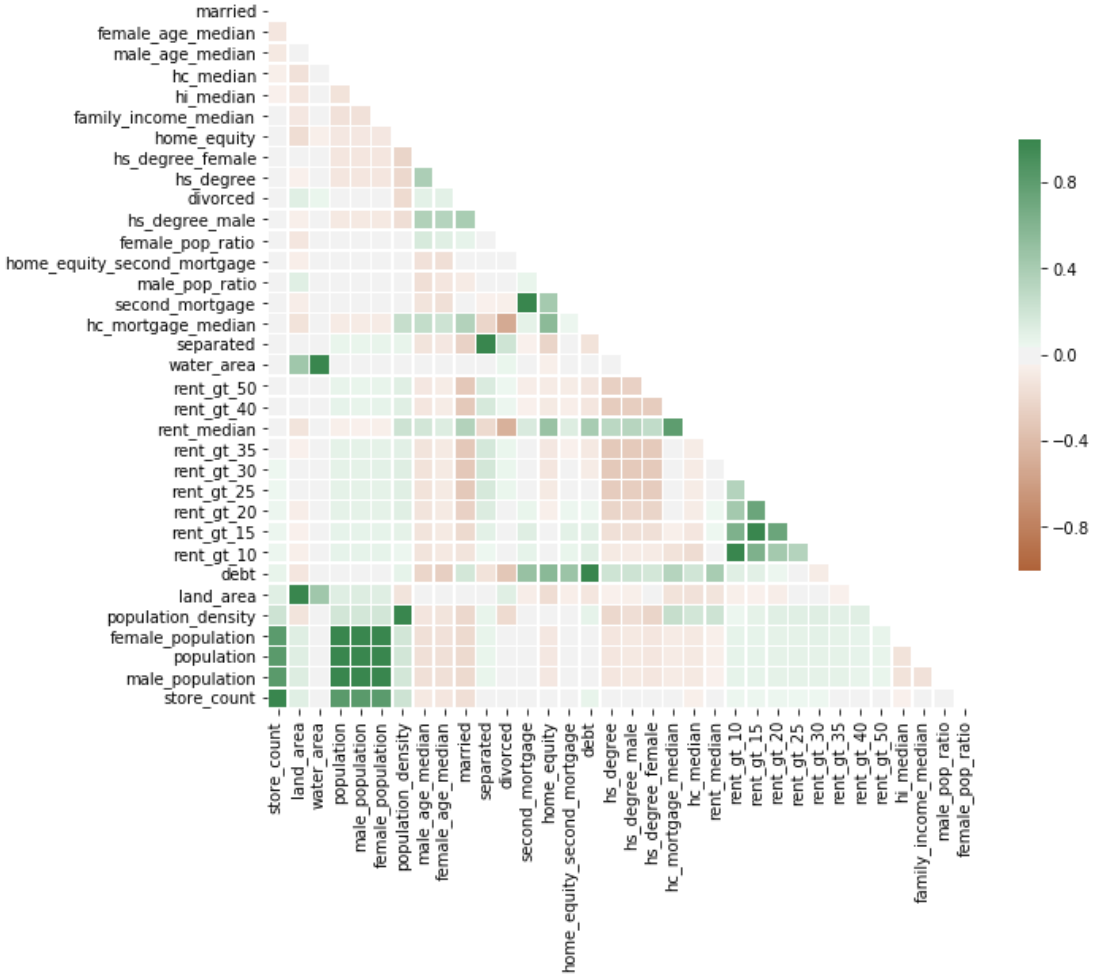

Figure 12. Starbucks correlation matrix of all city candidates in terms of demographic and economic

4.1.2. Starbucks pattern in top 9 city candidates in terms of demographic and economic

There are more than 3000+ cities have at least one Starbucks store. From plot, we discovered that a classic "Lorenz Curve" is exhibited on the bar chart. Usually Lorenz Curve is used to study income inequality, and it does exist too in the distribution of Starbucks store in US. Next, we will study

"Gini Coefficient" and "Nakamoto Coefficient" to calculate the inequality and plot it in Lorenz Curve to study the overall distribution[8].

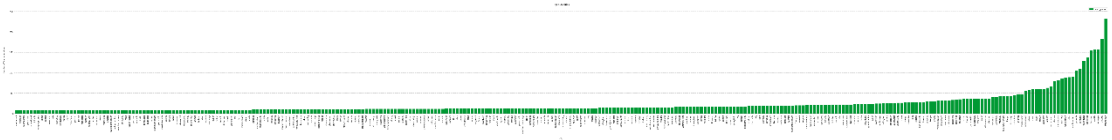

Figure 13. City candidates have at least one Starbucks in US 2017 (top 300)

The less Gini Coefficient, the less inequality. After calculation Gini Coefficient of 0.6 is obtained, it could indicate the inequality distribution of store in cities of US is exist (also can be known as less decentralization). The Nakamoto Coefficient of 295 indicates that it only would require 295 entities (cities) occupy 51% of store distribution.

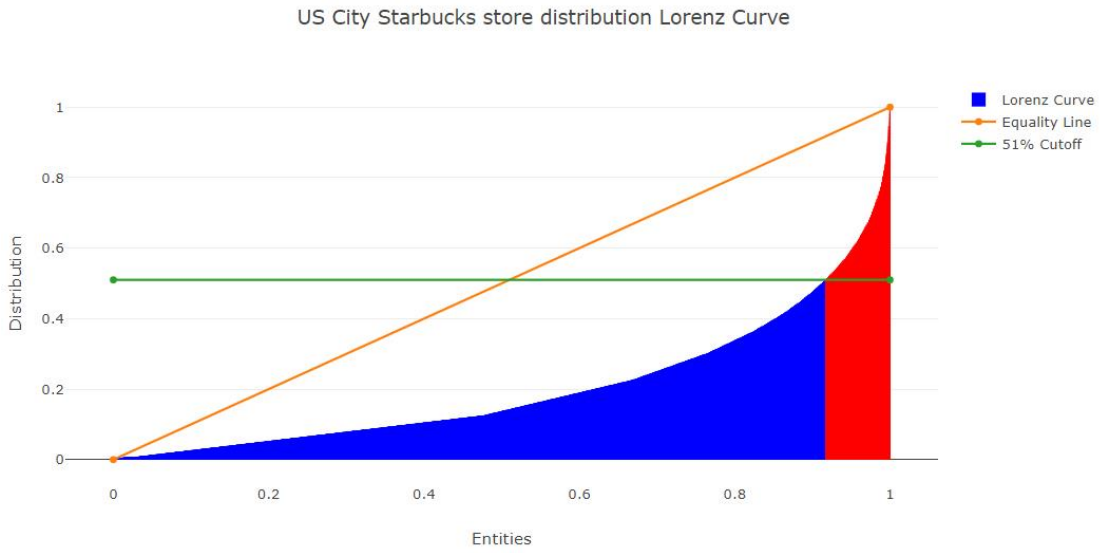

Figure 14. Starbucks store distribution of US 2017 in Lorenz Curve

3567*(1 - 0.9172974) ≈ 295

However, since this topic is to focus on cities with **high popularity**, so we would like to increase 0.51 to 0.90. Therefore, the number of cities we should focus will be:

3567*(1 - 0.997477) ≈ 9

This indicates that the top 9 cities (only 0.25% of 3567 cities) already occupy almost 10% of entire store distribution in US. From Figure 1, you might also notice that these 9 cities have all at least 100 stores respectively.

City candidates has high popularity of Starbucks is narrowed down to 9 by investigating Nakamoto Coefficient, there are Manhattan (NY), Chicago (IL), Las Vegas (NV), Seattle (WA), Houston (TX), San Diego (CA), Los Angeles (CA), Portland (OR) and Phoenix (AZ). These cities all have at least 100 Starbucks stores.

	city	state_ab	store_count
0	manhattan	NY	233.0
1	chicago	IL	182.0
2	las vegas	NV	157.0
3	seattle	WA	156.0
4	houston	TX	154.0
5	san diego	CA	138.0
6	los angeles	CA	128.0
7	portland	OR	110.0
8	phoenix	AZ	105.0

Figure 15. Top 9 US cities list (partial)

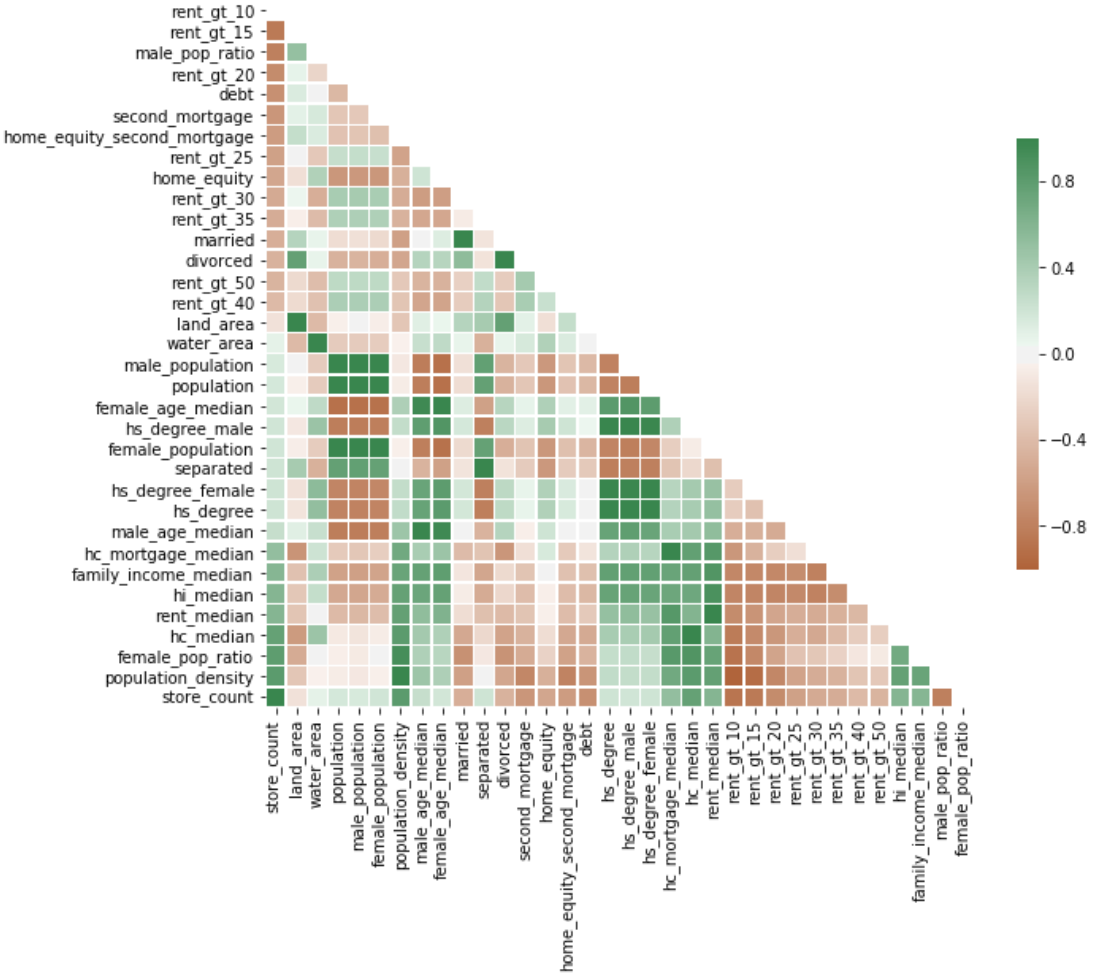

Figure 16. Starbucks correlation matrix of top 9 city candidates in terms of demographic and economic

From correlation study we can discover that **population density** is the most crucial feature this time. Higher population per land area, higher possibility of Starbucks density. Other positive features also include **female population ratio and monthly owner costs**. Interestingly, female population is the second positive impact on Starbucks store. Although it is hard to say which influence the other first,

direct or indirect, Starbucks's business decision maker could think about business promotion targeted on this specific population. For the later features it could be easily understood, higher residents' cost and income at some points depicts higher expenses ability and commercial activity, so stores' number will be higher around that area.

On the contrary, features such as **the empirical distribution value that an individual's rent will be greater than 10-20% of their household income and male population ratio** becomes the negative correlation features. When these features exhibit lower value, the number of Starbucks in that city can be predicted has higher popularity.

The zero impact features are **land and water area, population (includes male and female population), percentage of male with at least high school degree, and female age**.

4.2. Pattern in neighborhood level

From city level data, we found Manhattan is the highest popularity of Starbucks place, so it is chosen to conduct pattern analysis in neighborhood level.

4.2.1. Starbucks pattern in Manhattan neighborhoods in terms of demographic, housing market, neighborhood service and renters

From correlation conducted among these 11 groups of neighborhoods, we discover that **rental vacancy rate** is the primary positive demographic factor in Manhattan neighborhoods. Rental vacancy rates measure the percentage of rental homes that are empty and available for rent out of all the rental stock in an area. Although high vacancy rate means tenants are in a more favorable position, but for this case it might be resulted by high rental price, in which indicates higher housing price of one area will result more Starbucks store. Other positive features include **single-person households, severely rent-burdened households and sales volume (condominium and all property types)**. So, we could also say number of Starbucks in Manhattan could has positive relationship with the number of residents who remains single and spends more on rental. Higher transactions of all residential properties also result in higher number of Starbucks in Manhattan neighborhoods.

Figure 17. Starbucks correlation matrix 11 Manhattan neighborhoods in terms of demographic and economic

On the contrary, features such as **households with children under 18 years old** becomes the negative correlation features in this scale. The percentage of households living with children under 18 also could indicate the birth rate. So lower birth rate or less teenage in one area might result in higher Starbucks store number.

The least impact features includes **income diversity ratio, population and population density, foreign-born population, accessibility of residential units within 1/2 mile of a subway station, serious crime rate, higher-cost refinance loans, the average price changes in repeated sales of the same 5+ family properties, number of properties and units receiving a benefit from the Mitchell-Lama program[i], number of units in properties receiving benefit from HUD Project-based Rental Assistance Program[ii], sales volume of 1 family building and number of properties**

[i] The Mitchell-Lama Housing Program is a non-subsidy governmental housing guarantee in the state of New York. It was sponsored by New York State Senator MacNeil Mitchell and Assemblyman Alfred Lama. The program's publicly stated purpose was the development and building of affordable housing, both rental and co-operatively owned, for middle-income residents. https://en.wikipedia.org/wiki/Mitchell-Lama_Housing_Program

[ii] Project-based rental assistance provides critical affordable housing stock to low-income families across

receiving benefit from a HUD Financing or Insurance Program.

4.2.2. Starbucks pattern in Manhattan neighborhoods in terms of venues composition

From correlation study of Starbucks store number and nearby venue categories' number, we can discover that the number of **South American Restaurant** is related the most to Starbucks store in Manhattan neighborhoods. Other positive features include **Japanese Restaurant, Museum, Hotel and other art and music performing space**.

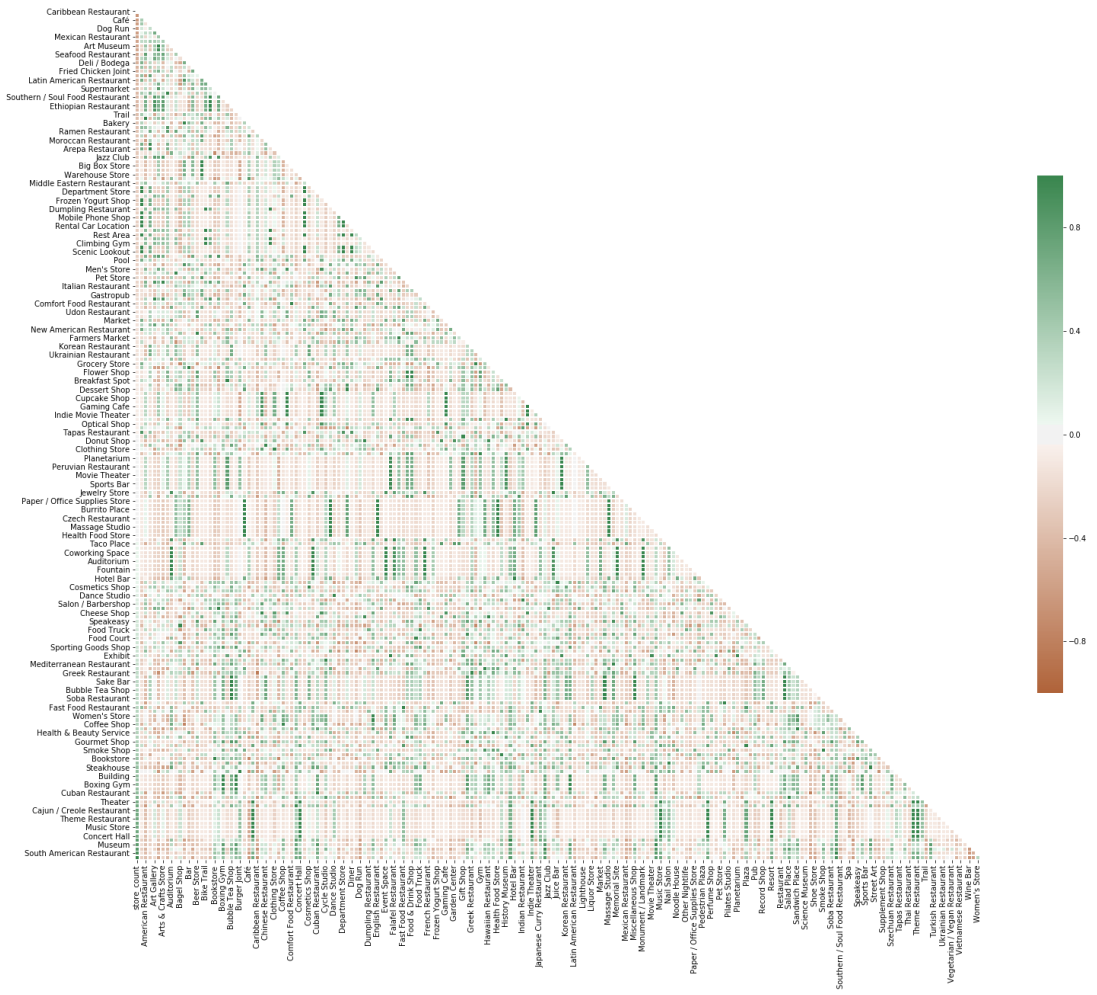

Figure 18. Starbucks correlation matrix 11 Manhattan neighborhoods in terms of venue category

Only one venue could be merely viewed as negative correlation this time, **Caribbean Restaurant**. Moreover, the least impact venue categories (only list range within -0.05 to 0.05%) includes **Gift Shop, Donut Shop, Chinese Restaurant, Clothing Store, Gym, Planetarium, Science Museum, Peruvian Restaurant, Used Bookstore, Movie Theater, Souvenir Shop, Sports Bar, Kids Store, Jewelry Store, Vegetarian Restaurant, Stationery Supplies Store, Pilates Studio, Burrito Place, Noodle House, Czech Restaurant, Turkish Restaurant, Massage Studio, English Restaurant,**

---

the country. This type of rental assistance allows tenants to live in an affordable unit and pay rent based upon their income. https://www.nhlp.org/resource-center/project-based-rental-assistance/

Health Food Store and Falafel Restaurant.

4.2.3. Clustering Manhattan neighborhoods in terms of venues composition

The common venue categories in each neighborhoods cluster at some point will also reveal the characteristics of neighborhoods.

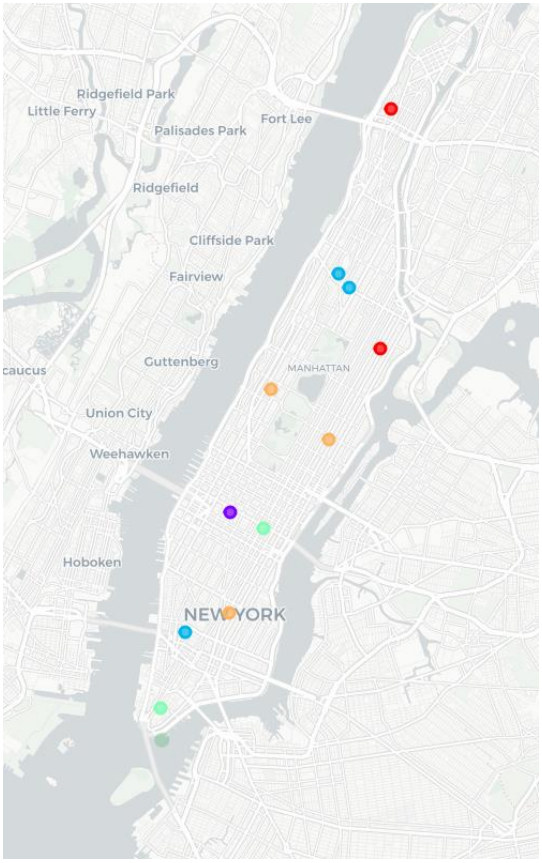

Figure 19. Neighborhoods clustering in each neighborhood based on venue categories

- Cluster 0 includes **East Harlem, Inwood & Washington Heights**. In these neighborhoods, Starbucks store's nearby common venue categories includes **Pizza Place, Bakery, Cafe, Latin American and Mexican Restaurant**.
- Cluster 1 includes **Chelsea, Clinton & Midtown**. In these neighborhoods, Starbucks store's nearby common venue categories includes **Theater, Hotel, Yoga, Japanese Restaurant and Coffee Shops** too.
- Cluster 2 includes **Central Harlem, Greenwich, Soho, Morningside & Hamilton Heights**. In these neighborhoods, Starbucks store's nearby common venue categories includes **Coffee Shop, Italian, French, American and Seafood Restaurant**.
- Cluster 3 includes **Financial District, Gramercy Park & Murray Hill**. In these neighborhoods, Starbucks store's nearby common venue categories includes **Coffee Shop, Hotel, Steakhouse, Sandwich place, Park and Gym**.
- Lastly, Cluster 4 includes **Lower East, Upper East & Upper West**. In these neighborhoods, Starbucks store's nearby common venue categories includes **Coffee Shop, Gym and high diversity of restaurant**.

5. Conclusion

The aim of this study is to discover the pattern of Starbucks in city and neighborhood level by studying the correlation of the number of Starbucks store and relevant demographic, economic and venues data. Clustering of venues was then performed in order to discover cluster of neighborhoods according to number of venues categories number.

In city level, relevant fast-food entrepreneurs could target city / place by the pattern discovered, such as the stated positive and negative correlation factors above of one city. In neighborhood level, stakeholder could target firstly on neighborhood clusters according to venues categories. Next, they could focus on specific neighborhood based on correlation factors in terms of demographic, housing market, neighborhood service and renters.

---

[1] https://www.kaggle.com/starbucks/store-locations

[2] https://en.wikipedia.org/wiki/Place_(United_States_Census_Bureau)

[3] https://www.kaggle.com/goldenoakresearch/us-acs-mortgage-equity-loans-rent-statistics

[4] https://www.kaggle.com/washimahmed/usa-latlong-for-state-abbreviations

[5] https://www.health.ny.gov/statistics/cancer/registry/appendix/neighborhoods.htm

[6] http://furmancenter.org/neighborhoods/

[7] https://developer.foursquare.com/

[8] https://news.earn.com/quantifying-decentralization-e39db233c28e