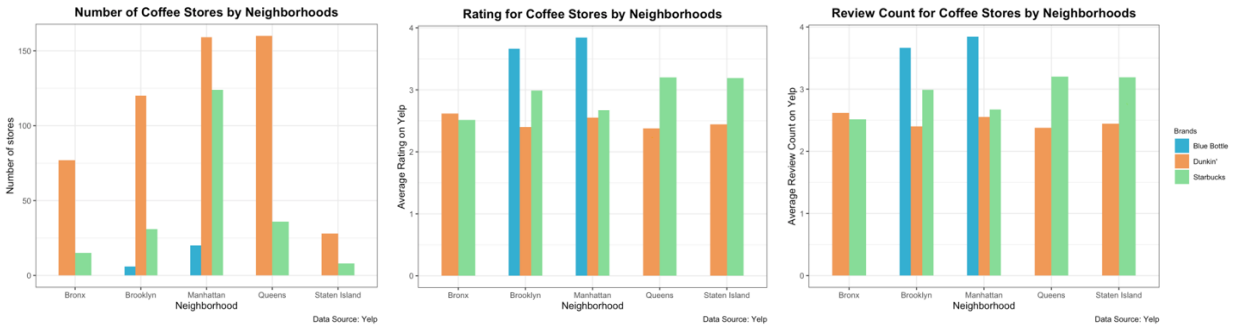
Data Analysis

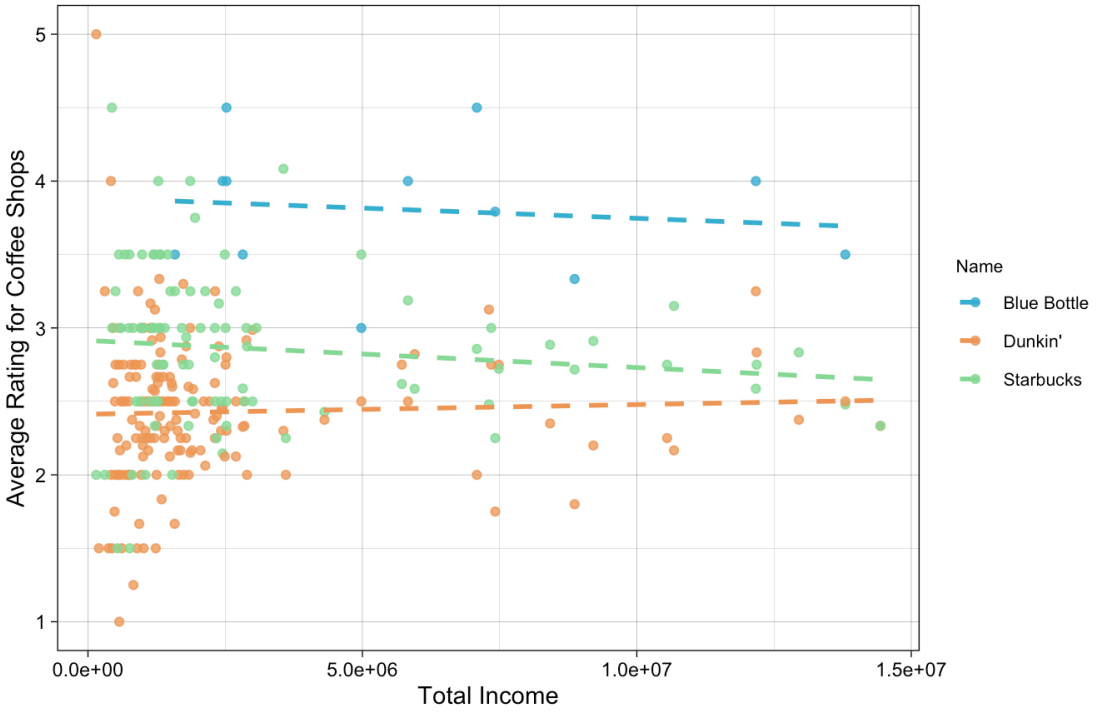
In fact, lots of restaurants also service Starbucks, but this study will ignore the rating for these stores as they are not essential coffee shop chains. After interacting with Yelp API, available information for all coffee shops in New York City has been downloaded, and this study focuses on representative coffee chains, like Starbucks, Dunkin', and Blue Bottle. There are 214 Starbucks and 544 Dunkin' with Yelp rating and review counts in New York City, and this study will also add Blue Bottle Coffee into consideration for the first stage of the analysis.

On average, Starbucks has 2.82 out of 5 for rating and 49 review counts, while Dunkin' has 2.47 out of 5 for rating and 14 review counts. Compared to these two, Blue Bottle Coffee has the best rating and the most review counts on average, which are 3.80 and 73.4 respectively. It seems that although Blue Bottle Coffee is short on numbers, this coffee chain effectively wins consumers' hearts by its high quality and popularity. Overall Blue Bottle Coffee, followed by Starbucks, has the upper hand in the rating and leads the number of review counts in almost all boroughs where they have stores, except in Bronx where prefers Dunkin' over Starbucks. Linear regression results demonstrate interesting findings as ratings for these three coffee chains are statistically significant. If there is no review, a Starbucks store will have a rating of 2.79 out of 5 on average, and a Blue Bottle coffee store will have a rating of 3.62. A Dunkin' store will have a rating of 2.59 on average if there is no review for the store. While each review will drag down the rating for Dunkin' stores by a significant 0.008, each review for Blue Bottle Coffee will boost its rating on Yelp by a significant 0.002.



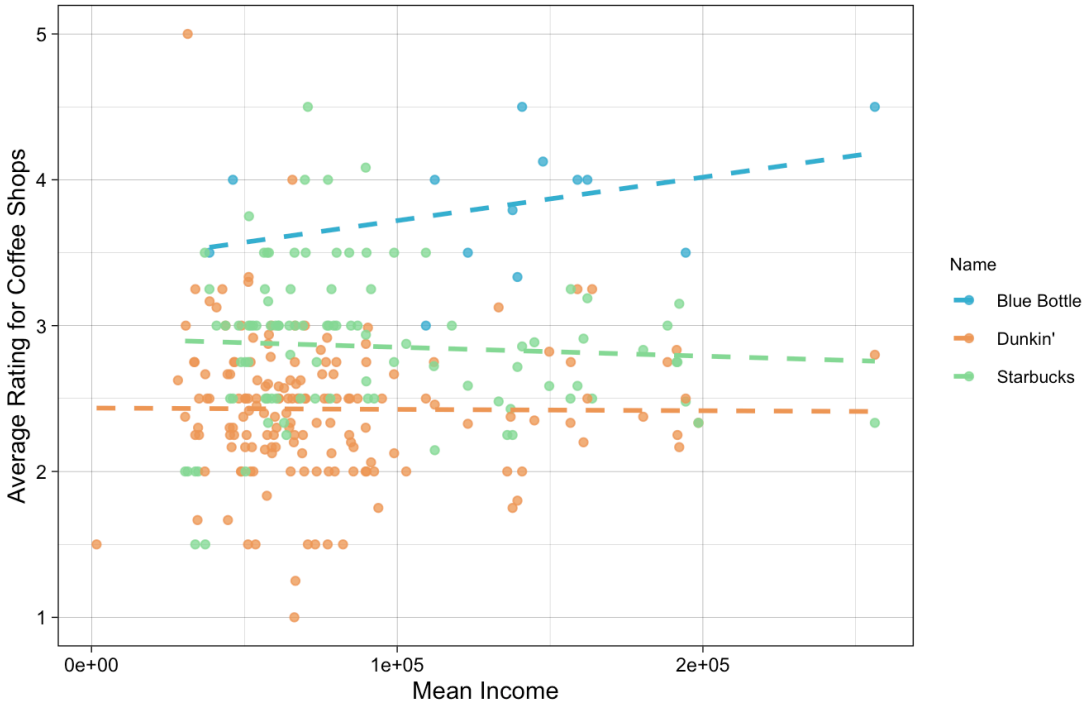
The research further combines the information about coffee brands in New York City from Yelp and the demographic information in the surrounding neighborhood by zip codes. From an overall snapshot of the distribution of shops, Blue Bottle Coffee tend to locate in neighborhoods with relatively high income, while Dunkin' and Starbucks are less biased in the selection of their location. Overall, the total income, the average income, and the population of the surrounding neighborhood do not significantly affect the rating for Starbucks and Dunkin'. Although the influence is not statistically significant, neighborhoods with higher average income shows preference for Blue Bottle who is a niche player in the market.

Total Income vs Average Rating for Coffee

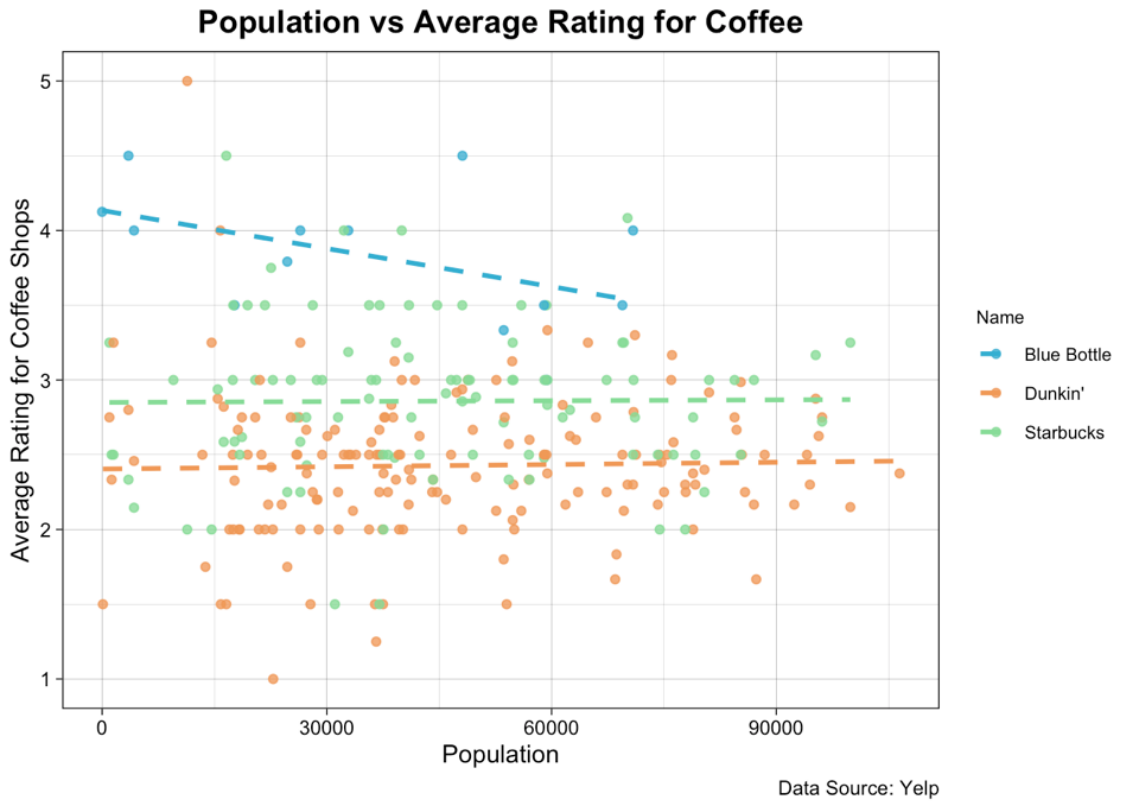


Data Source: Yelp

Mean Income vs Average Rating for Coffee



Data Source: Yelp



There is another dataset on Kaggle that collects the number of Starbucks stores by counties and the demographic characteristics in the counties. The result shows that New York, Suffolk, Nassau, Queens, and Brooklyn are the top four counties with the most Starbucks. After plotting the scatter plots between different parameters, there is no obvious correlation between the number of Starbuck stores and other variables, except the density of shops. The correlation matrix shows that, besides the density of shops, there is a strong correlation between the number of Starbucks stores and the population in the county, as well as the density of population. The result from linear regression demonstrates that there is a strong correlation between the population and the number of shops in the county. Although the correlation is statistically significant, the amount of increase that the population brings to an increase in the number of shops is very small. Similarly, one dollar increase in the average income of a county only results in a small increase in the number of Starbucks stores in that county. However, one increase in the

density of the population will generate around 561 more Starbucks stores in one county, and the incitement force is statistically significant. Meanwhile, a country will have almost two Starbucks stores fewer than other counties if the the median age in the county is one year older than other counties, but this discrepancy is not statistically significant.

Besides population, population density, income, and median age of the population, what are other factors that might have the potential to influence the number of Starbucks stores in a neighborhood? To answer this question, this study used data for health code and violation from the Department of Health and Mental Hygiene (DOHMH) for New York City restaurant inspection results. After filtering out cleaned data for Starbucks and Dunkin', the study will focus only on five boroughs of NYC. There are more Dunkin' in all boroughs except Manhattan than Starbucks, and Dunkin' has more health violations in total and higher inspection scores in total in all boroughs except Manhattan than Starbucks. Looking separately, one more Starbucks store generates 8.75 more health code violations, while one more Dunkin' store results in only 7.55 more health code violations. The research further includes the interaction between the coffee brand and the total violations. The regression result shows that one more Starbucks store creates statistically significant 7.55 more health code violations, and open a new Starbucks store has 1.19 more violations on average than one more Dunkin' store. Applying supervised machine learning techniques, linear discriminant analysis shows 70.62% accuracy to predict whether a coffee shop is Starbucks or Dunkin' given the number of checks, health code violations, inspection scores, and the borough.

Having said all the progress has been made in the research, for the next step in research, sentiment analysis for Starbucks stores on Yelp will be conducted, along with topic modeling in the visualization of word clouds. I have emailed the owner of the dataset that gathered Yelp

reviews for businesses in NYC and waited for their response to continue the study. As there are not enough number of Blue Bottle Coffee, this study only include the brand for a general comparison with Starbucks and Dunkin'. However, future study should definitely consider adding Blue Bottle Coffee as the representative for the Third Wave coffee chain that is actively expending.