**Group 6:** Rucha Patil, Emmanuel Jonathan Yalla, Pavan Kumar, Surojit Sasmal

**Motivation:**

Retail sector has always fascinated us, this sector is highly influenced by social as well as economic factor which makes this sector very attractive to us.The consumption pattern reflects the true economic growth of a country, in our case for the city of Boston and spending capacity of any individual. This dynamic nature of this sector and vast diversity has encouraged us to deep dive into studying the Boston pattern data and its factors along with the places. There are always limited resources available to retail enterprises to better service their client base. To maximize your potential return, you must pick which physical sites to invest in, how to allocate resources for the demand and where to direct your advertising and marketing efforts.

**Why is it more important than other areas?**

As mentioned above, retail industry has been a buzz of discussion for many investors and shoppers for a while now. People have shifted their shopping preferences to online platforms as it provides them convenience and ability to review options in a matter of clicks. Furthermore, the rate at which big retail store brands have started filing for bankruptcy has substantially increased. Since the covid pandemic, many shoppers have exclusively moved to online transactions which has made this problem even worse, forcing stores to shut down physical locations and move their business operations to completely virtual. Overall, working on this will give us more insight on the trajectory of this industry POIs. We will study the patterns of the top brands that the customers prefer to visit over other smaller brands which in turn, will give us deeper insights about their products, pricing, and their customer servicing capacity. Lastly, we would like to know which ones are thriving market leader POIs to recommend opening of new businesses, changing of location or expansion of current businesses.

**Question to analyze?**

To understand the customer patterns to the Clothing & Grocery stores across the Boston area and see the brands that are the most popular and track their changes during the period from January to August 2022 and find out why those brands are visited over other brands and to provide insights and recommendations on how they can position and service their customers considering several factors.

* Economic downturn this year and seasonal impact that makes us ask which stores are seeing the biggest change in the customer visit patterns each month.
* High and low-income groups areas; for example, Seaport and Dorchester and understand how it impacts consumer traffic at expensive grocery stores such whole foods vs cheap stores such as Dollar tree.
* Use the census block group to understand the visit patterns and gain insights by creating a customer profiling as per the location and get target segments and customer demographics if we have the open census data.
* Factors influencing the consumption such as substantial jump in the gas prices during the middle of this year due to the Russia and Ukraine war and how it impacted the foot traffic to the clothing stores vs grocery stores.
* Distance travelled to those store and why they might prefer those stores even though they are far .
* Dwell time analysis could help us understand the time spent by the customer. This analysis and time can understand how long the customer would like to spend at the store. Also, the store can plan employees accordingly if they weren’t able to meet the user demand.
* Factors influencing the number of revisits for different stores and in different regions. Which type of stores has high percentage of revisits and why? Are they willing to travel longer distance? How much is the dwell time if so?

**Data manipulation & Strategy:**

Compare both the data sets with joining based on the place key, to understand the stores and its foot traffic frequency. Based on that we are going to classify out targeted retail stores by their categories such as the Clothing (“**359**” POI’s) and Groceries (‘**150**’ POI’s) from the places dataset and map it with the patterns dataset to find out the busiest and the highest number of stores across different Boston neighborhoods. We also can use the NAICS code to classify , mainly using the first two digits using the census gov manual , in this case ‘**44**’ and ‘**45**’ for the Retail Industry as given by the government. Additionally, we can understand the time of the visit in each month and find the peak months of consumptions, find the factors influencing this peak and low. We can also understand how much time a consumer is spending at the store and the factors influencing the same. Further, it’s interesting to analyze the factors affecting the number of revisits and for which stores this is common. This could be very important for any kind of business to increase the foot fall/revists to get the customers hooked to the services and products offered. We will be able to stratify into the top retailers and top grocery stores by the number of visitors in those areas and do a competitor analysis to understand the customer preference and recommend the business to give the demanding product mix and their preferences in that area to boost their sales and be the best in that area. To deep dive into other factors as well such as distance travelled from home to store and find out why the customers chose to travel that distance instead of going to nearby neighborhood stores and convert the distance\_from\_home column from meters to miles and find out how much they are willing to travel to those particular clothing and grocery stores despite other economic and geo-political factors and plot a map using longitude and latitude columns and understand the distribution of those store branches across the popular ‘cbg’ groups and give recommendations to open or close the stores basing on the insights received.

Very Good.

**Challenges:**

* To identify and classify the parent POI and independent POI will be difficult to understand as for most of the times there might be situations that the customer could’ve visited other brand, but they all are situated in the same locations which could cause possible report errors in our analysis.
* Sampling bias could be one as the patterns dataset can be biased towards high income individuals or certain geographic regions .
* We cannot answer who are the customer that are going to these POI’s.
* Dwell time can imply couple of interpretations based on the type of the store and may not always indicate that the store is busy with less employees to meet the demand of the users. It might be the case that user is willingly spending time at the store for window shopping or working. It is interesting to see the patterns but difficult to come to conclusions in this case.
* Duplicate customer visit data if an individual has more than one mobile device.

Overall, it is excellent.

Here is what I want you to do:

1. Find all the POIs relevant to the clothing and groceries
   1. Provide me the specific ways in words
      1. Consider the way we can group them whether the POIs are independent (or local) store or the franchise store
   2. Provide me the result
      1. Summary statistics of POIs
         * If possible, group them by the indicator for local shop
      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: Good

Writing quality: Good