**BUS211A Final Project Initial Report**

Group 4: Ziyu Fan, Shipeng Ji, Gengchen Zhang, Yiqing Wang

**What is your specific area of interest?**

Our team aims to investigate how Boston’s COVID-19 policy has affected visiting patterns of both full-service and limited-service restaurants since the beginning of 2022. Due to two COVID-19 policy changes that went into effect on February 15th, 2022 and July 1st, 2022, respectively, we are particularly interested in the change in visiting patterns, including the visitor count of a restaurant during various time periods, the average amount of time visitors spend in restaurants, the visitors’ home countries and their average distance from home. Our hypothesis is that after a certain period of time as the mask restrictions are lifted, we would observe more active visits to both types of restaurants from a larger variety of origins.

**Policies and effects:**

Regarding the effects of the policies, we will deal with three time periods. The first period is from January 2022 to February 2022. During this period of time, the Department of Public Health guided that all people in Massachusetts (regardless of vaccination status) are required to continue wearing face coverings in certain settings, including transportation and healthcare facilities. During this period, people are required to wear masks in transportation.

From February 2022 to July 2022, DPH advises that a fully vaccinated person should wear a mask or face covering when indoors (and not in their own home) if they have a weakened immune system, if they are at increased risk for severe disease because of your age or an underlying medical condition, or if someone in your household has a weakened immune system and is at increased risk for severe disease or is unvaccinated. Individuals who are not fully vaccinated should continue to wear a face covering or mask when indoors with others to help prevent the spreading of COVID-19. In this case, people are no longer required to wear masks in public transportation and indoor settings. We are expecting more people to visit restaurants regularly since this policy does make the life of people who use public transportation easier.

And finally, from July 1, 2022, the new guidance advises that masks indoors are optional for most individuals, regardless of vaccination status. This would make people that are not willing to take vaccines have chances to go to the restaurant which should increase the number of visits significantly. We, as a team, would like to test different factors that are related to the running of restaurants that may correlate with the updates of the policies.

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

COVID-19 might have brought irreversible changes to the world. Although policies related to COVID-19 have been gradually lifted, we are still uncertain whether the world will return to its pre-pandemic state. As one of the most common consumption activities, visiting patterns of restaurants might act as a representative indicator of how people spend their spare time and thus help us answer the question. By conducting this research, we also hope to provide useful advice to help restaurant owners to adjust their business strategies.

**Why is the cell phone tracking data fit for your interest?**

Cell phone tracking data would be the most accurate source to answer our research question. When we are considering the pattern of restaurants, it is important to know what is the population that visits these places. Since almost everyone carries cell phones nowadays, we would have a good estimate of how many people visited these restaurants in a specific duration of time. Moreover, there are many statistics that would only be available in cell phone data: dwelling time in a specific location, related location, and visitor’s country of origin. These are dimensions hard to obtain for traditional methods, yet they are a detrimental part of our analysis.

When using the places and pattern data from Safegraph, we need to first filter out all the relevant data in the places data frame using “naics\_code” - 722511 for full-service restaurants and 722513 for limited-service restaurants. Using the data filtered we left join with patterns since the patterns data frame has detailed information on visitors, for example, the counts of visitors and visits, the distance from home, and the average time visitors tend to stay at the restaurants. That results in a total of 8395 rows for full-service restaurants and 2367 rows for limited-service restaurants. Then, based on the release and effective date of the three policies, using the date\_range\_start to separate the data frame into three different time ranges - January, February-June, and July-August. In total that would be 6 different data frames. In the end, we do some data cleaning and remove the columns in each data frame that contains only one unique value like city, region, date\_range\_start, and date\_range\_end. At last, we can make some comparisons between each time frame data based on the average result of each month, and hope to find out any correlation between the visitors' pattern and the release of the COVID-19 policy.

**What is the main challenge in answering the questions?**

The first policy we intend to study took effect on February 15, 2022. However, most of the columns in our dataset are mainly recorded on monthly statistics, such as raw\_visitor\_count, and distance\_from\_home; only the column visits\_by\_day records daily data. Therefore, it is difficult to observe changes in the fifteen days after the policy came into effect on February 15. If we decide to observe it 15 days after the policy was published, our second policy, effective July 1, 2022, would have the same problem. Our team plans to focus on the visits\_by\_day from February 15 to March 1, 2022, in order to reduce the error.

Moreover, we want to find out as the mask restriction is gradually lifted, whether this would attract more local US customers and whether people from nearby countries like Canada and Mexico will be willing to come to visit. But in the visitor\_country\_of\_origin column, the countries are not listed in a stationary manner, so it brings challenges in separating the corresponding data with the correct country name. For this problem, we are considering methods like separating the datasets by conditioning whether all visitors of this restaurant are from the U.S. And we would also categorize these rows by their distance from home.

Finally, since we only have 8 months of data, the covid situation isn’t that severe now and the policies released are just slightly different, it might not show much difference in the visitors’ patterns compared to a year ago.

Ok. The questions are not only well-defined, and they address significant matters in real life.

Here is what I want you to do:

1. Find all the POIs relevant to restaurants
2. what’s the difference between 722511 and 722513?

* try to research what’s the exact definition of each code
* could the code be differed by its location? For example, some pizza stores would not offer a delivery (I am just guessing)

1. how do you group POIs within each code?

* otherwise, you only have two groups

1. Provide me the result
2. Summary statistics of POIs for each category
3. 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: Fair