**Background**

The deal team wants to understand the gourmet grocery space a bit better, specifically focusing on a handful of companies (Whole Foods, Sprouts, Natural Grocer, Fresh Thyme, Fresh Market, and Earth Fare). You are given access to a geolocation dataset which tracks cell phone pings when customers enter a gourmet grocery store. Please use this data to answer the following questions and to make a sole recommendation of which grocery store brand is your top pick for an investment (assume all of them have the same revenue, financials, and purchase price).

Data frequency for the geolocation data is daily from 9/10/2015 to 11/16/2020. The demographics file is a snapshot as of October 2020.

**Excel Functionality and Data Normalization**

Write your code in Python (either script or jupyter notebook is acceptable) but answer the questions below and show your output in Excel. Submit both your code and excel output prior to the case study deadline.

**Evaluation and Questions**

1. Aggregate the cell phone device data by quarters (i.e. 1Q2016 = 1/1/2016 to 3/31/2016).
   1. Should the “devices\_store”, “devices”, and “traffic” columns be summed or averaged?

These columns should be averaged so that we know how’s the footfall on a daily basis spread overs quarters. Also, the same number of cellphone tracking is done across all the stores on a particular day and in the dataset some of the stores don’t have data for all days of the quarter so it would inflate the numbers for those who have better data collection.

Apart from that traffic column is a measure of one-day only and shouldn’t be summed.

* 1. The “devices” column represents the panel size (# of cell phones tracked over time) and grows significantly throughout the dataset. Please use the panel size to normalize/adjust the traffic so that the data is comparable across time.

Took the quarterly average and then created norm\_traffic using devices\_store and devices.

1. For each grocer, which store has the highest traffic (and how many devices visited) and which store has the lowest traffic each quarter?



1. For each grocer, which store had the highest year-over-year growth (i.e. 1Q2017 vs. 1Q2016 growth) in device traffic?



1. Which grocer has rebounded the best in terms of traffic since the Coronavirus lockdowns (Use April 30, 2020 as the assumed starting date)?
   1. Is there a pattern to the kind of store that rebounded faster vs. other stores?

Earthfare rebounded the best



If comparing just the stores, I noticed



Sprouts Farmers Market showing great growth and it looks like people have thronged towards this store.

1. How many stores does each grocer have per state?
   1. Use the Address column (column G) to determine the state of each grocery.

See attached excel

1. For each grocer, which state had the highest growth from 4Q2019 to 4Q2020 in terms of:
   1. # of stores.
   2. Average # of devices visiting a store.

See attached excel

1. Join the aggregated data with the demographics file.
   1. What state has the highest average income per grocer?
   2. What state has the youngest average visitor age per grocer?

See attached excel

1. Based on the above and other creative ways to analyze the data (do not use external sources beyond the geolocation & demographics data we provided), which grocer would you recommend Advent International invest in?

I would recommend Sprout Farmers maket as the grocer to invest in as it has the 2md largest # stores in the country (wider presence= more footfalls) and it has also high average income / grocer.

Also, almost all of its stores have reported substantial % change in their businesses from Q4 2019 to Q4 2020 indicating a changed business model that’s bringing more footfalls.

One of its stores reported the highest % change in traffic since the beginning of coronavirus thus indicating a special interest among grocery customers.

1. Save your outputs into Excel.

**Data Dictionary:**

|  |  |
| --- | --- |
| **Attribute** | **Description** |
| devices\_store | sum of tracked devices that went into the store |
| devices | sum of total tracked devices |
| traffic | devices\_store / devices for a store over a period (one day) |
| Id | grocery store id (used to link device counts and demographics) |
| Address | grocery store address |
| Income\_avg | average income in USD of visitors to the store |
| Age | average age of visitors to the store |