**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.

**Submission Requirements**

Write your code in Python (either script or jupyter notebook is acceptable). 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 to determine the **total** quarterly traffic for each store for 1Q18 through 4Q20 (1Q18 == 1/1/2018 through 3/31/2018).
   1. “Traffic” is defined as devices\_store / devices for the time-period in question.
   2. Should the “devices\_store” and “devices” columns be summed or averaged?
   3. The “devices” column represents the panel size (# of cell phones tracked that day) and grows significantly throughout the dataset. Please use the panel size to normalize/adjust the traffic so that the data is comparable across time.
2. For each grocer, which store location has the highest traffic and which store has the lowest traffic each quarter?
3. Calculate the quarterly year-over-year growth (e.g. 1Q19 vs. 1Q18 % growth) in device traffic for each store for 1Q19 through 4Q20. For each grocer, which store had the highest growth?
4. How many stores does each grocer have per state?
   1. Use the Address column (column G) to determine the state of each grocery.
5. For each grocer, which state had the highest growth from 4Q19 to 4Q20 in terms of:
   1. # of stores.
   2. Average # of devices visiting a store.
6. 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?
7. Which grocer has rebounded the best in terms of traffic since the Coronavirus lockdowns (Use April 30, 2020 as the assumed starting date of the rebound period)?
   1. Is there a pattern to the kind of store that rebounded faster vs. other stores?
8. 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?

**Data Dictionary:**

|  |  |
| --- | --- |
| **Attribute** | **Description** |
| devices\_store | number of tracked devices that went into a particular store on a particular date |
| devices | number of devices being tracked across the entire country on a particular date; also known as the “panel size” |
| 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 |