

# Stop & Shop Visits Trend

**Scenario** - Analyze visitation patterns at [Stop & Shop @ 2335 Dixwell Ave Hamden CT](#)

**Data Sample** ([download here](#))

1. **Venue Visits** - raw visits (March 1st, 2019 - August 31st, 2019).  
Each record represents a device visit to the venue.
2. **Devices Daily Activity** - device (user) activeness per day; a record per device per day in the sample (i.e. for the devices appearing in the **Venue Visits** file).  
Coverage varies from 0 [no coverage at all that day] to 1 [full coverage]

**Required Delivery** - share **Google Drive** folder with the following items:

- A customer-facing **Google Slides** deck which outlines your findings and insights you would add to them. Note: You will be evaluated on the packaging (look and feel).
- Python file with the script you used to analyze the data and generate the results (preferably **Google Colab** notebook, but any other type works i.e. jupyter notebook/py file)
- **Google Spreadsheet** file you used for visualizing the results and preparing the graphs, in case you used one

## **Guidelines**

- Use Python to calculate the numbers
- NO NEED to generate the visualizations/graphs using Python (only if it's easier for you)
- Use Google Spreadsheets to generate the graphs once you have the numbers on hand (or any other visualization tool if easier for you).
- Tip: You can use [Kepler.gl](#) in order to generate heatmap if needed

## Analysis

- **How does the user's activity look like?** Analyze the user's activeness distribution (number of active days) and present it.

User is defined as active on a given day if his coverage  $\geq 0.75$

For example, looking at the below record, we will count this day (7/1/2019) as active for this user as coverage is  $\geq 0.75$

date	device_id	coverage
07/01/2019	1101	0.962326389

**Note:** you are not allowed to use Pandas/Dataframes **in this question only** (you **CAN** use pandas in the rest of the assignment). You should use Python standard Data Structures (lists/dictionaries etc.) to solve this part.

- **Visitation Patterns** - Placer's core metric is the foot-traffic estimation (i.e. estimated number of visits)

Analyze the visits and present the visitation patterns in the best way it represents the data provided as you see it. Notes:

- The graphs should include visits of **active users only**: visits where the associated user (i.e. device\_id) was active for more than **90** days during the entire period (using what you've done in the first question!)
- **Use visit\_weight to present any visits data.** visit\_weight is the extrapolated value of the raw visit, so if you want to get to the estimated number of visits to the location, you would sum on this field (as opposed to simply counting the records).
- **How to define a 'Loyal Customer'?**
  - Analyze the distance of the customer's home location to the venue (think what is the best way to present this data)
  - How would you define a 'loyal customer'? Think of thresholds that make a customer loyal (for example: visits frequency, distance from home or any other feature). Explain why you chose this threshold and present the loyal/casual customers breakdown
  - **Bonus** - Using [Kepler.gl](https://kepler.gl) generate 2 heatmaps:
    - a. Customers home location
    - b. Only loyal customers home location
      - Make sure you include a pin for venue location as well

*Tip: You will need to calculate additional columns such as 'Home distance from Stop & Shop. Use the following coordinates as Stop & Shop's location: 41.3723654,-72.9141964*