Why is a Planet Fitness location Is Under-performing?

Scenario

- Planet Fitness (a chain of gyms) asked us why the venue at '10790 Alpharetta Hwy, Roswell' is underperforming.
- Your goal is to investigate the data, find a possible explanation and package the results in a customer-facing deck.

  Expected deliverables:
  - A client facing deck with visualizations including charts and maps.
  - The data files you used for all your calculations (e.g., Excel spreadsheets, Python Code, etc.)

Data Sample (provided in the attached zipped file):

- In the file, you will find data exports of the query venues plus 3 different additional Planet Fitness venues nearby.
- Venues info file with venue metadata (You will find in the "notes" column the underperforming venue)
- 3 data dumps of each venue raw visits (2018-02-01 to 2019-01-31). Each record is equal to a panel visit (i.e., raw visit) and carries 2 "unique" fields that are the product of our proprietary estimation model:
  1. visit_weight — the extrapolated value of the raw visit. Summing visit weights over a time period gives you the total estimated visits for that time period.
  2. customer_weight — the extrapolated value of the visitor ("device_id). Each unique visitor will always have the same weight. Note that there are visitors with weight = 0 (use it as is)
  3. To calculate the estimations/extrapolated values you will need to use these 2 fields.

Something to keep in mind:

- Gyms customer visit frequently (on average more than 1 time per week)
- Generally speaking, people don't like to drive much and prefer venues that are near their home or work.
- Data visualization is very important, you can use any mapping tool (CSV's can be used as input)

Tips on weights:

- visit weight — this is our extrapolated visits counts, it is the output of a model that takes into account over 60 factors (including app type, home and its daily coverage, etc...),
  For example: one row is a visit by device id X that is representative of 42 estimated visits to planet fitness within the given time of visit

  the weight of visits can differ between visits per device (each row has different weights) – because within each time frame our extrapolation model is basing its results on varying coverage and sample sizes we have (for example a certain census block may have 50 panelists in one week and 60 the next week which changes the device weight)

  for over time analysis - use the visit weight

- customer weight - this applies to the entire report period, each device has 1 weight (dedup if you want to use this metric). the way to read this is as follows:

  device X represented 54 estimated unique customers (regardless of how many visits they had) within the entire time frame

  some devices have 0 0 meaning they don't need to be counted (but don't drop them from the over time visits analysis)