Sprint 0: Initial Design

Planning for Sprint 1

Project Group 8: Yao-Fu Yang, Rong Chang and Qiushi Li
Trello board: https://trello.com/b/61ULBOVX/grocery-store-data-support
Git: https://github.ccs.neu.edu/yangyaof/cs5500_sum2020_group8

User stories

the Manager's viewpoint:

As a store manager, I need to know the patterns of customers shopping, so that we can plan operations to reduce customer waiting time.

At a detail level:

1. (P-1) As a store manager, I need to differentiate the **patterns of customer volumes** over the week including **weekdays** and **weekends**, so that I can assign the employee shifts effectively.
2. (P-1) As a store manager, I need to know the typical **customer visiting time and length** on each day, so that I can identify the peak hour and adjust our service and staff schedule.
3. (P-2) As a store manager, I need to know the **customer wait time** ~~for parking and~~ **outside of the store**, so that I can adjust the number of employees helping the waiting line.
4. (P-3) As a store manager, I need to know the change in customer volume and visiting pattern due to **weather** conditions, so that I can adjust employee shifts accordingly.
5. (P-3) As a store manager, I need to know the change in customer volume and visiting pattern during **special event days** such as senior discounts on Tuesday mornings and **holidays**, so that I can delegate employee duties efficiently.

Abstraction and Relationship

The user needs to use the data to plan the operation under COVID-19 scenario (and normal scenario) so that the store can produce the services more efficiently, such as more staff at the needed place, minimize customer waiting time, in order to give customers better shopping experience. To achieve that, the data will allow the user to:

1. ~ observe customer shopping **distribution** during the day and during the week
2. ~ find the **peak** time during the day
3. ~ estimate the **time** each customer spends at different points such as waiting outdoors, food count, groceries, and checkout
4. ~ find the **factors** affecting the amounts of customers, such as weekday/weekend/holiday and weather.

A .csv file of data with sufficient parameters allowing the Manager to try out some operational planning. In the product of a data file, **each row of the csv file represents a shopper's visit**. In addition, the parameters are planned:

- *Non-time related parameters and relations*
  - **Weather condition** (categorical: sunny, cloudy, rainy).
  - **Temperature** (numeric).
  - **Date** (format: month/day/year). Has three dependent parameters:
    - **Holiday**
    - **Day of Week** (Categorical: Mon, Tue, …Sun)
    - **Traffic** (Categorical: weekBefore, dayBefore, dayOf): Either 15%, 40%, or 20% more.
- *Time related parameters and relations (in sequential order)*
  - **Arrival time** (format: h:m:s). The specific time when the shopper arrives at the store and begins waiting to enter the store.
  - **Waiting time** (minutes). Waiting outside of the store, it is defined as (entryTime – arrivalTime). May be 0 when there is no waiting line.
  - **Entry time** (format: h:m:s). The specific time when the shopper enters the store.
  - **Total visit time** (minutes). This is associated with entry time and leave time, it is defined as (leaveTime - entryTime).
  - **Leave time** (format: h:m:s). The time shopper leaves the store.

Implementation planning

We take advantage of the technology already installed in the store, which is able to track how the customers go through the store in the "normal" schedule, as described by the Manager. <u>Given the knowledge of customer shopping patterns in the normal schedule (*normalData*), we are able to simulate the shopping patterns under COVID-19 scenario by borrowing the statistical distributions of associated parameters (*covidData*).</u>

We are going to use Java as our primary language. Initially, external libraries include random and DateTime might be considered for the implementation of simulation. Due to the unexpected situations under COVID-19, the implementation of the data generation should be maintained to allow any changes of the parameters or distribution of parameters. This includes the simulation model, statistical inference and the coding lines. The data is expected to be delivered in a series of versions allowing it to be tested and negotiated.

The planned class diagram is as followed:
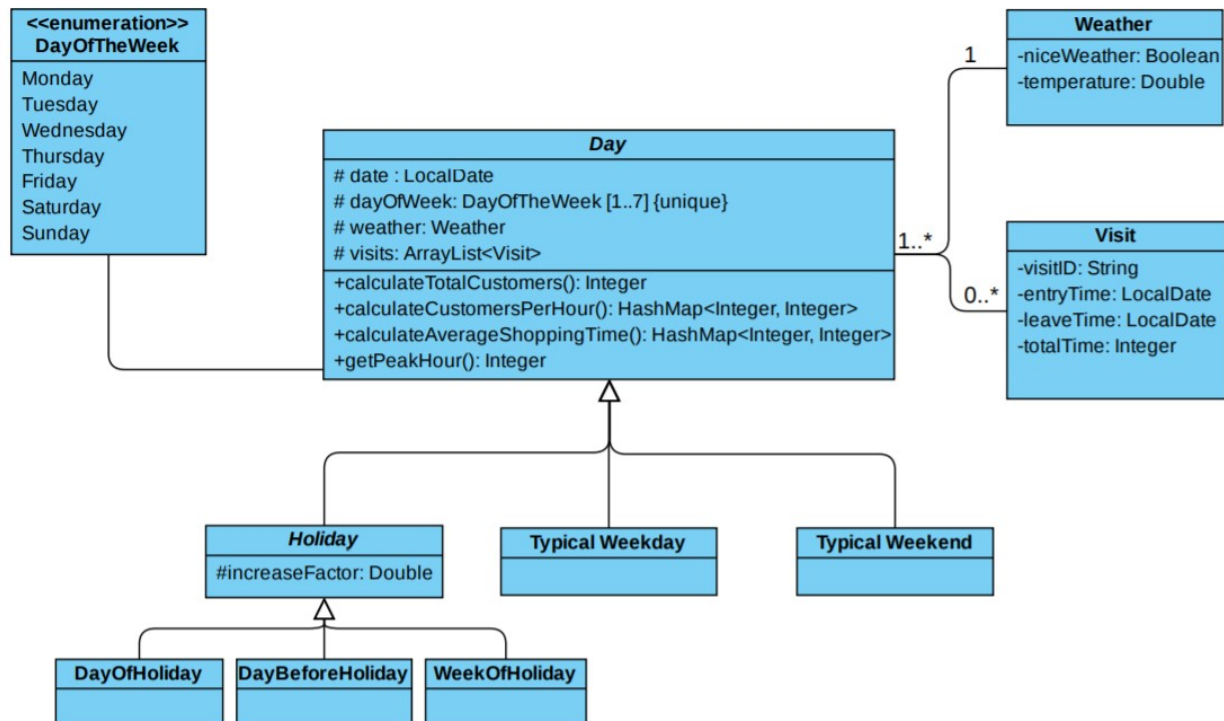
The planned procedure is as followed:

week of 05/18/2020   Prepare the UML class diagram

week of 05/25/2020   Prepare the parameters and the parameter distributions in normalData.
week of 06/01/2020   normalData generation.
week of 06/08/2020   Data testing and listen to user feedback.

week of 06/15/2020   Prepare the additional parameters, make the assumption on the
                      parameter distributions due to the changes of scenario.
week of 06/22/2020   covidData generation
week of 06/29/2020   Data testing and write up for delivery