CS 5500 Sprint 0

Project Group 2: Audrey Jo, Carlo Mutuc, Evan Simpson

Trello Board and GitHub

Trello: https://trello.com/b/z7zJbAuZ/kanban

- Invite link to Trello if you couldn't access the link above:
    - https://trello.com/invite/b/z7zJbAuZ/ab83d84a7fe375145aee368ff47731c4/kanban

GitHub: https://github.ccs.neu.edu/simpsone/5500Project

Inputs

The inputs for this project are:

- Descriptions of shopper behavior in the project description
- List of holidays and their dates obtained from the holidays module (https://www.geeksforgeeks.org/python-holidays-library/)
- Weather data from Sand Point weather station for Seattle, Washington (https://www.ncdc.noaa.gov/cdo-web/datatools/findstation)

The main source of data for creating the csv of shopper behavior would be the descriptions of shopper behavior in the project description. Some examples of provided information are:

- average shopper traffic on each day,
- average time spent in the store,
- peak times of shopper traffic in the store,
- shopper behavior based on age,
- Increase in shopper traffic based on weather.

A sample of the input data is shown below:

	Average shopper traffic	Min in store	Average in store	Max in store	10-12pm (seniors in store)	12-1pm	5-6:30pm
Monday	800	6 min	25 min	75 min		10 min	20 min
Tuesday	1000	6 min	25 min	75 min	45-60 min	10 min	20 min
Wednesday	1200	6 min	25 min	75 min		10 min	20 min
Thursday	900	6 min	25 min	75 min		10 min	20 min
Friday	2500	6 min	25 min	75 min		10 min	20 min
Saturday	4000	6 min	60 min	75 min			
Sunday	5000	6 min	60 min	75 min			

Sunny weekend	Normal traffic * .2	6 min					
Holiday	Normal traffic * .2	6 min	25 min	75 min			
Day before holiday	Normal traffic * 1.4	6 min	25 min	75 min			
Week leading up to holiday	Normal traffic * 1.15	6 min	25 min	75 min			

We would also use the holiday data to account for the increase/decrease in store traffic accordingly for days before and on holiday.

The weather information would be used to account for the increase in store traffic on sunny weekends.

Outputs

The output is a csv with data about how long each shopper spends in the store. The columns for this data would be visitID, date, dayOfWeek, timeIn, timeOut, timeSpent, sunny, holiday, and senior. Each row will represent a shopper visiting the store.

visitID	date	dayOfWeek	timeIn	timeOut	timeSpent	sunny	holiday	senior
0001	05/27/19	Wednesday	10:15 AM	10:45 AM	30 mins	true	false	true
0002	05/27/19	Wednesday	10:55 AM	11:55 AM	60 mins	true	false	false

User Stories

As a technically literate person, I can generate a data file for the grocery store manager so that he can run some analysis.

As a technical user, I can modify the parameters to generate a different version of the data for the grocery store manager to run some analysis.

As a technical user, I can modify average shopper traffic to simulate changes in traffic going through the grocery store.

As a technical user, I can modify the range of dates of shoppers to be generated to drill down on shopper behavior.

As a technical user, I can modify the amount of time shoppers spend in store on average.

As a technical user, I can modify the number of seniors coming in on different days.

As a technical user, I can modify the number of people coming in during lunch and dinner time rush.

As a store owner, I want to know the effects of advertising programs as it relates to changes instore traffic and product sales to see how successful it is.

As a store owner, I want to know the status of our shoppers so I can plan events and discounts (senior discounts, samples, etc).

As a store owner, I want to know how long shoppers are spending in the store compared to profits earned or amount of products sold.

As a stock manager, I want to know what parts of the store customers are spending their time in so I can update the stocking staff appropriately.

As a staff manager, I want to know how busy the cashier lines are so I can open the appropriate amount of cashiers.

As a store owner, I want to know how much store traffic increases during lunch, dinner, and weekends so that I can hire additional staff members appropriately.

As a store owner, I want to know how much store traffic increases during holiday seasons so that I can hire additional staff members appropriately.

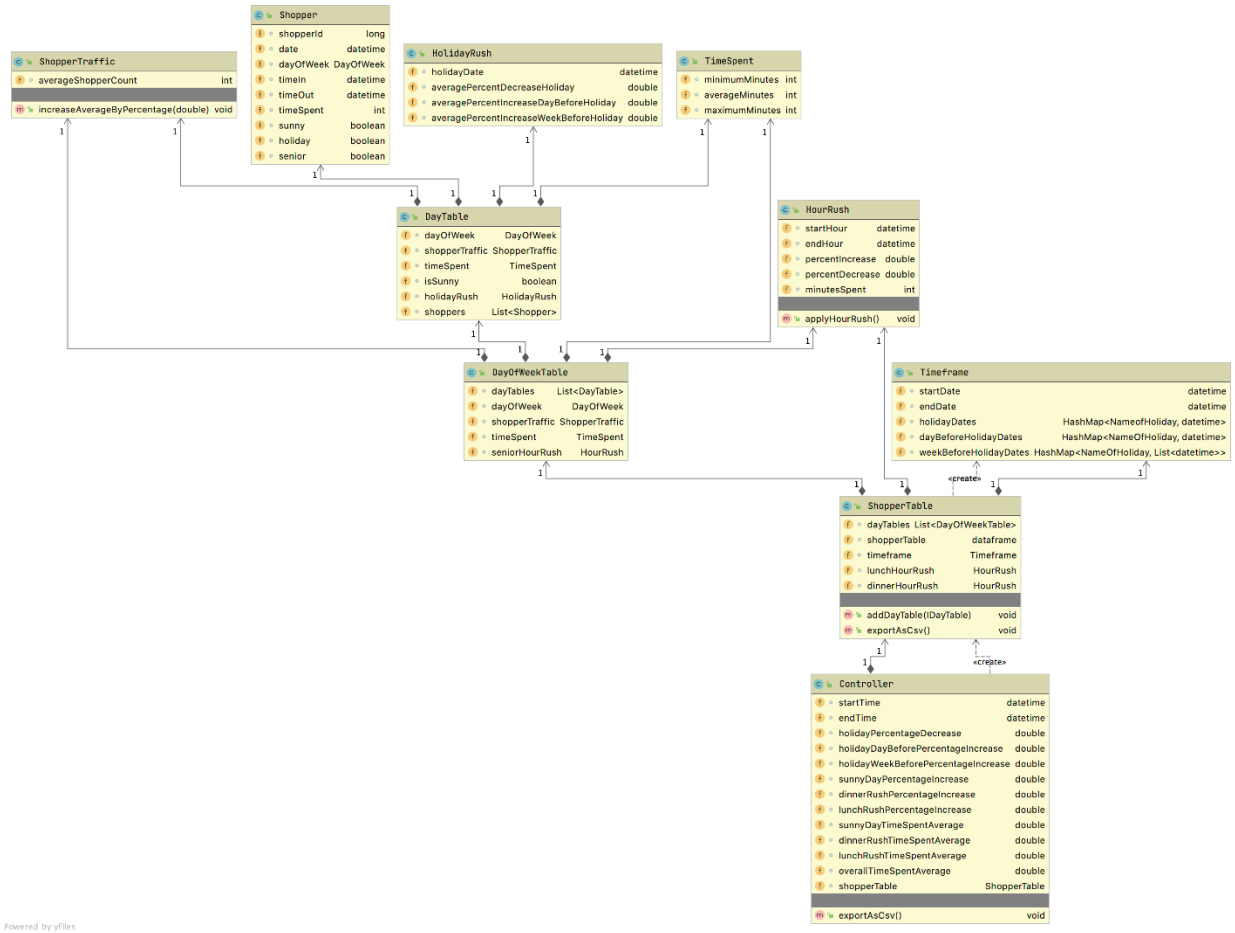
As a store owner, I want to know which products are being sold the most so I can plan buying and budgeting inventory needs.

As a staff manager, I want to know how many people are at each department at different times so I can schedule the amount of staff there appropriately.

Programming language and external libraries

Our team decided to use Python in the PyCharm IDE. We will be using pandas for generating the csv. We will be using numpy and faker to generate fake data from the given trend data (i.e. average shopper traffic, average time spent shopping, etc.). The holidays module will be used to get the actual dates for holidays.

Major abstractions and relationships



Here is our proposed UML for the project. We have a Shopper class that encapsulates a shopper that has visited the grocery store and represents a single row for the output csv. Shoppers are contained in the DayTable class, which represents a specific date. The DayTable has a reference to the HolidayRush class. This represents the increase or decrease in traffic to the store based on where it falls in relation to a holiday. It will be used to apply changes in traffic to specific dates. In addition, we have a DayOfWeekTable which is a different day of the week. Since each day has an average amount of shoppers and time spent, this can hold that information, which will then be referenced by each specific DayTable. There is also a concept of an HourRush, which represents the increase in shopper traffic based on dinner or lunch time which will be used to modify the shoppers coming in at that time and the time they spent in the store. TimeFrame contains the date to generate data for and will find/hold information for all the holidays, day before holidays, and week leading up to a holiday. The ShopperTable will contain the data in a dataframe to be converted into csv. It has a reference to DayOfWeekTable so that it can get the shoppers contained further down the hierarchy. The Controller class will be what handles the different parameters that our technical person can set to modify the data that is generated.