**Mid-Term Draft Projects**

Dhananjay Kumar [Dhananjay.kumar@spsmail.cuny.edu](mailto:Dhananjay.kumar@spsmail.cuny.edu)

Data 698 -04 - Final Semester

CUNY – MS Data Science 2019

Under Guidance of Professor Haiyuan Wang - [haiyuan.wang@sps.cuny.edu](mailto:haiyuan.wang@sps.cuny.edu)

**Title: Traffic based Retail store staffing optimization**

**Introduction:**

Retail store staffing optimization is one of the most efficient ways of saving cost and increasing productivity. It is one of the most crucial aspect for any organization working in retail industry and has direct impact on labor-related expenses as it constitute one of the biggest chunk of operating cost. Currently the retail sector is under severe pressure from online ecommerce player, thus saving labor related cost is one of the most prominent goal for any retailer working in the current economic scenario.

Since, retail is an old industry, many retailers have indeed found statistical and mathematical solution to resolve or optimize this problem. One of the most prevalent ways of optimizing labor force is the use of Linear Programming where we set an objective and then find its minima or maxima, depending on the goal.

The above solution based on Linear Programming does not work for every labor scenario as we know that in the real world the retailers have to consider many constraints, where it is quite possible that a feasible solution does not exist and hence they have to come up with near feasible solution. Some of this can be achieved using Constraint Programming. But given the complexities of the environment that a typical retail store worked, there are factors, which are quite dynamic, and changes very frequently. One such factor is Traffic Data. The goal of this Project is to come up with a working model for a retail store where we are able to optimize retail labor workforce schedule considering customer traffic data.

**Initial Set Up:**

There is a 100-store retail chain in North America. Lets call this retail chain as “Dhananjay Book Shop” or DBS. In DBS there are various Employees working to serve the customer. The senior management of DBS has decided for an exhaustive In store workforce optimization. The CEO of DBS wants to keep the in store Employees in proportion to the customer traffic attracted by the respective stores. For this to happen, each store is equipped with traffic counters, which records the customer.

The first challenge is to keep the workforce in accordance with the customer traffic attracted by Store.

The second challenge would be to keep the traditional scheduling constraint into the equation. We need to build a model that would be a traditional scheduling model but will be dynamic enough to respond to changing customer traffic pattern.

**Data Required**:

We would need three sets of data:

1. Employee: this dataset would tell us number of employees working in a store at given time. The dataset would help us in determining optimum conversion.
2. Customer Traffic: This dataset will give us the count of customer traffic at any point of time. We will aggregate this dataset at hourly level to observe the customer traffic pattern.
3. Customer Sales Transaction: This dataset would have information of all customer sales transaction with timestamp. The timestamp would help in understanding customer traffic and actual sales.

**Data Source:**

The data for this project comes from following source:

Sales transactions (download .csv file)

The Sales Jan 2009 file contains some “sanitized” sales transactions during the month of January. Below is a sample of a report built in just a couple of minutes using the Blank Canvas app. These 997 transactions are easily summarized and filtered by transaction date, payment type, country, city, and geography. Note that these records are already geocoded, so you can use the existing latitude/longitude in the file.

Link of the website: <https://support.spatialkey.com/spatialkey-sample-csv-data/>

Based on above data we need to create two more dataset:

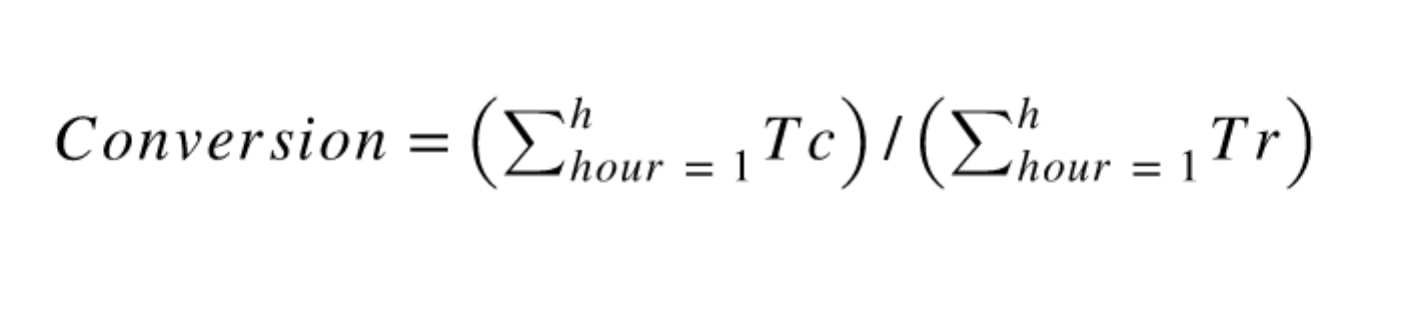
1. Customer Traffic: This is the data which will tell us how much traffic does a retail store attracts.
2. Employee: This dataset will tell us how many Employees are present at any given time within a store

The link of the Jupyter Notebook, which has source code for data retrieval and data generation, is given below:

<https://github.com/dhnanjay/CUNY/blob/master/FinalCapstoneProject/Data.ipynb>

**Implementation:**

Before we begin, we must define a mechanism to determine the correct number of employees need at a given hour in store i.e. We need to define a function which measures the effectiveness of employees present in store. This function could be "Conversion". Conversion is defined as sum of total transaction count divided by Sum of Total traffic out at every hour of the operating hours of a retail store:



Where:

Tc => Transaction Count at a given hour

Tr => Customer Traffic at a given hour

h => Operating hours which is between 9:00 AM to 9:00 PM which comes out to be 12, hence h is a range of 1 to 12

The conversion above defined will be measured against a parameter. Using this parameter we will analyze conversion and understand how it varies. This parameter would be defined as Traffic to Employee Ration or TER. TER is nothing but a ratio, which is defined as Average No. Of Total Customer Traffic divided by Total No. of Employees present in Store at give hour:

*TER = Tr/Emp*

Where:

Emp => Employee present in Store

We need to find a Goldilocks’ ratio where both Conversion and TER is high. This will help us in determining how many Employees are needed to attend Customers in order to achieve high Conversion. A TER of 1 would mean Average One Employee for every customer, which seems little unreasonable, as it would most likely raise the labor cost.

We might not be able to choose very low TER in order to have a high conversion as this would mean that we need large number of Employees in store which would increase the overall labor related cost and every enterprise has limited budget for operating cost including labor.

Lets Define Constraints now for generating schedule:

1. A Retail Store would have at least a Store Manager (SM) and Assistant Store Manager (ASM)
2. The store will have multiple Temporary workers (TEMP) based on customer traffic attracted by store.
3. Both SM and ASM will not work more than 5 days in a week .
4. TEMP cannot work more than 6 days a week.
5. The minimum shift for any TEMP employee will be 3 hours and maximum will be 10 hours per day.
6. The minimum shift for SM and ASM will be 7 hours and maximum will be 9 hours per day.
7. The maximum number of total shift hours in a given day cannot exceed 10 hours for TEMP.
8. The operating hours for retail for store would be 9:00 AM to 9:00 PM, 7 days a week.
9. Total number of Employees in a given week cannot be more than 5.

In real world scenario lot of things are subjected to financial constraints including Labor Scheduling. There could be case where the model suggests weekly total hours that might not be feasible based on the financial budget. Hence we need to add another constraint where we can force the algorithm to find solution near to the targeted total hours. To do that we need to develop a simple back propagation model, which will generate new schedule keeping traffic pattern into consideration and then matching the total hours inputted into it.

For finding a schedule, we will require a solver. In this case we will be using CBC solver that comes with package PuLP. For constraints, we will be using a Python Library called "PySchedule"

For complete working of the above, please refer to the below link to see working Jupyter Notebook based on the above mentioned principles:

<https://github.com/dhnanjay/CUNY/blob/master/FinalCapstoneProject/Traffic%20Based%20Automated%20Scheduling.ipynb>