# Project Report: Workforce Scheduling for a Retail Chain

Group Members: Felipe Calderon, Nikki Gong, Rebecca Peng, Vivian Peng

## 1. Background

[1] Brick-and-mortar retail stores employ about 15% of the American workforce. In retail, variable schedules are the norms where schedules typically change every day and every week with three to seven days' notice of next week's schedule. [2] Researchers have found that matching labor to incoming traffic is a key driver of retail profitability. Most retailers operate under the assumption that stabilizing employees' schedules would hurt the stores' financial performance. [3] Numerous studies have also found that variable schedules have deleterious effects on employees' well-being and thus incur invisible expenses, such as, inefficiency at work.

[3]A few leading retailers have made the attempt to adopt more data-driven approaches for scheduling and been able to capture between 4 and 12 % in cost savings among other facets of store operations. Stable and consistent retail staffing scheduling that is able to meet the store traffic demand is likely to benefit both the employees and employers.

## 2. Problem Statement

Current solutions and softwares used by majority of the retailers produce only generic schedules that fail to account for store-specific factors and workload fluctuations, and they disregard the impacts of scheduling on the service level and staff overall well-being. These solutions have led to undesired results including:

- Overstaffing leads to high labor cost
- Understaffing that would hurt the stores' profitability
- Inconsistent schedules that lower staff satisfaction

In this project, we intend to help solve the above three pressing issues in retail workforce scheduling by developing effective retail workforce scheduling models that will:

- ensure the number of staff will match the store customer traffic
- minimize the total staffing cost
- improve the employees' satisfaction and well-being
- help managers make better staff scheduling decisions based on staffing cost and the impact of staffing on the store's service level which is correlated to the store revenues in most scenarios

## 3. Data Source

The datasets we use come from LC Waikiki, which is a Turkish retailer operating over 150 stores in the apparel sector. It has three types of stores based on their location, sizes, and target population:

- Neighborhood Business District (NBD)
- Shopping Mall (SM)
- Central Business District (CBD)

The dataset we acquired is the store traffic and staff scheduling requirement data of LC Waikiki's shopping mall. [3] Please see below for the dataset we acquired through an academic paper:

| Hourly requirements of SM | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Days/Hours | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 |
| Monday | 3 | 5 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 |
| Tuesday | 3 | 5 | 7 | 7 | 7 | 7 | 7 | 8 | 7 | 7 | 7 | 7 |
| Wednesday | 3 | 5 | 7 | 7 | 7 | 8 | 8 | 8 | 7 | 7 | 7 | 7 |
| Thursday | 3 | 5 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 |
| Friday | 3 | 5 | 7 | 7 | 7 | 8 | 8 | 8 | 7 | 7 | 7 | 7 |
| Saturday | 3 | 6 | 8 | 9 | 10 | 10 | 10 | 10 | 10 | 10 | 9 | 8 |
| Sunday | 3 | 6 | 8 | 10 | 10 | 11 | 11 | 11 | 10 | 10 | 9 | 8 |

Fig. 1 Sample data for the hourly staffing requirement throughout a week for a single store.

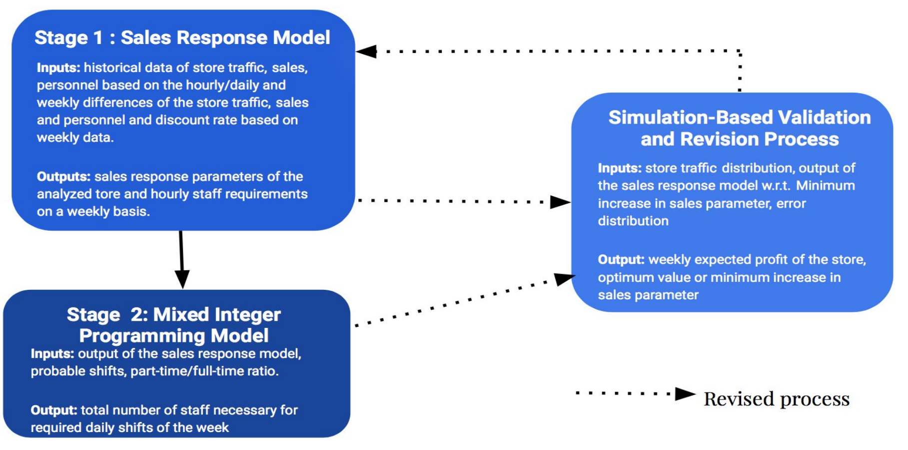
| | Monday | | | Tuesday | | | Wednesday | | | Thursday | | | Friday | | | Saturday | | | Sunday | | | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Dshift | *s* | *b* | Dshift | *s* | *b* | Dshift | *s* | *b* | Dshift | *s* | *b* | Dshift | *s* | *b* | Dshift | *s* | *b* | Dshift | *s* | *b* | Duration |
| Full1 | OFF | – | – | 12:00–22:00 | 10 | 17 | 11:00–22:00 | 11 | 15 | 10:00–19:00 | 9 | 14 | 12:00–22:00 | 10 | 16 | 14:00–22:00 | 8 | 18 | 14:00–18:00 | 4 | 18 | 52 |
| Full2 | 10:00–22:00 | 12 | 14 | 14:00–22:00 | 8 | 18 | OFF | – | – | OFF | – | – | 12:00–16:00 | 4 | 16 | 10:00–21:00 | 11 | 14 | 10:00–22:00 | 12 | 14 | 47 |
| Full3 | OFF | – | – | 10:00–22:00 | 12 | 15 | 10:00–22:00 | 12 | 15 | 14:00–18:00 | 4 | 18 | 14:00–18:00 | 4 | 18 | 10:00–21:00 | 11 | 14 | 10:00–21:00 | 11 | 15 | 54 |
| Full4 | OFF | – | – | 10:00–19:00 | 9 | 14 | 12:00–19:00 | 7 | 16 | 12:00–22:00 | 10 | 17 | 11:00–19:00 | 8 | 15 | 11:00–20:00 | 9 | 15 | 11:00–22:00 | 11 | 15 | 54 |
| Full5 | 14:00–22:00 | 8 | 18 | OFF | – | – | 14:00–22:00 | 8 | 18 | 10:00–22:00 | 12 | 16 | 10:00–22:00 | 12 | 14 | 13:00–22:00 | 9 | 17 | 14:00–18:00 | 4 | 18 | 53 |
| Full6 | 14:00–18:00 | 4 | 18 | 10:00–19:00 | 9 | 14 | 12:00–16:00 | 4 | 16 | 10:00–22:00 | 12 | 16 | 10:00–22:00 | 12 | 14 | 10:00–22:00 | 12 | 15 | OFF | – | – | 53 |
| Full7 | 12:00–22:00 | 10 | 17 | OFF | – | – | 15:00–19:00 | 4 | 19 | OFF | – | – | 12:00–22:00 | 10 | 17 | 11:00–22:00 | 11 | 16 | 13:00–22:00 | 9 | 17 | 44 |
| Full8 | 11:00–22:00 | 11 | 16 | 11:00–22:00 | 11 | 16 | OFF | – | – | 12:00–22:00 | 10 | 17 | 14:00–22:00 | 8 | 18 | 15:00–22:00 | 7 | 19 | 12:00–19:00 | 7 | 16 | 54 |
| Part1 | 11:00–22:00 | 11 | 16 | 14:00–22:00 | 8 | 18 | 14:00–22:00 | 8 | 18 | OFF | – | – | OFF | – | – | OFF | – | – | 13:00–22:00 | 9 | 17 | 36 |
| Part2 | OFF | – | – | OFF | – | – | OFF | – | – | OFF | – | – | 10:00–22:00 | 12 | 15 | 10:00–22:00 | 12 | 15 | 10:00–22:00 | 12 | 14 | 36 |
| Part3 | OFF | – | – | OFF | – | – | 10:00–22:00 | 12 | 14 | 14:00–22:00 | 8 | 18 | OFF | – | – | 13:00–17:00 | 4 | 17 | 10:00–22:00 | 12 | 14 | 36 |
| Part4 | 10:00–22:00 | 12 | 15 | OFF | – | – | OFF | – | – | OFF | – | – | 11:00–22:00 | 11 | 17 | 14:00–18:00 | 4 | 18 | 11:00–20:00 | 9 | 15 | 36 |
| Part5 | OFF | – | – | OFF | – | – | 11:00–22:00 | 11 | 17 | 11:00–22:00 | 11 | 15 | OFF | – | – | OFF | – | – | 12:00–22:00 | 10 | 16 | 32 |
| Part6 | 12:00–22:00 | 10 | 17 | 10:00–22:00 | 12 | 15 | OFF | – | – | OFF | – | – | OFF | – | – | 12:00–22:00 | 10 | 16 | 18:00–22:00 | 4 | 22 | 36 |
| Part7 | OFF | – | – | 11:00–22:00 | 11 | 16 | 10:00–22:00 | 12 | 14 | OFF | – | – | 16:00–20:00 | 4 | 20 | OFF | – | – | 14:00–21:00 | 7 | 18 | 34 |
| Part8 | 10:00–19:00 | 9 | 14 | OFF | – | – | OFF | – | – | 11:00–22:00 | 11 | 15 | OFF | – | – | 12:00–22:00 | 10 | 17 | 12:00–16:00 | 4 | 16 | 34 |

Dshift: detailed shift; *s*: length of the shift; *b*: time of the one-hour break.

Fig. 2 Available detailed shifts to create the schedules. 28 different shifts were found from this sample output in the paper.

## 4. Original Model

Our project is based on an academic paper published by Turkish operations research scholars. According to the paper, the construction of an optimal schedule is divided into two stages. During the first stage, the Sales Response Model forecasts the expected hourly sales based on the store traffic, average price of the product, number of sales staff, etc. Using this model, we are able to find the optimal number of staff that will maximize the expected profit in any given hour. The output from stage 1 is used as the input for the second stage where each employee is assigned to different shifts throughout the week with a mixed integer program. We further develop upon the stage 2 MIP model by adding service level and consistency constraints which will be discussed in more detail in later sections.



## 4.1 Stage 1 Sales Response Model

*Input Parameters*

- $N_t$ , store traffic based on hourly time series data
- $S_t$ , overall store sales revenue during hour t based on hourly time series data
- $l_t$ , number of staff members present during hour t based on hourly time series data
- $\pi_t$ , store sales profit for hour t
- $\alpha$ , sales potential
- $\beta$ , traffic elasticity
- i , average price
- $\gamma$, parameter associated with staff number
- G, average gross margin
- D, average wage rate (per hour)

*Index*

- $t$, index for the hourly observation arranged in chronological order ($t$ = 1 for the first observation, etc.)

$$S_t = \alpha N_t^{\beta} e^{i\gamma / l_t}, \tag{1}$$

Model (1) accounts for the fact that the effect of sales staff on sales will decrease whenever there is a discount. The average price *i* is used to account for the impact of discounting. For example, if the discount is 10% then the *i* = 0.9 instead of 1.0. $\alpha$ is a parameter that reflects sales potential. $S_t$ is expected to increase with store traffic depending on the value of $\alpha$. As customer traffic increases, due to the decrease in interaction time with each customer and the increased waiting time, there are diminishing returns in the purchase yield rate and average sales.

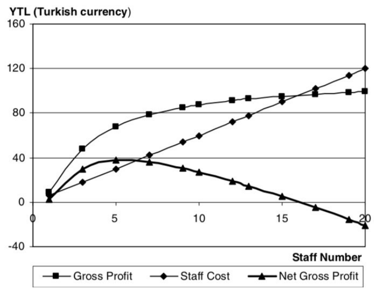



Fig. 3 Net gross profit, gross profit, and staff cost functions as a function of staff number

$$\pi_t = S_t g - l_t d, \tag{2}$$

By using Model (2), the optimum net gross profit $\pi_t$ is found. Model (2) is a net gross profit function with respect to the number of salespeople. This function is found to be concave downward and the $l_t$ value that maximizes this function is calculated.

$$N_t = \tau C_{t-1} + (1-\tau) C_t, \tag{3}$$

Also, many customers who visit a store only purchase items after spending a certain amount of time. This is reflected in Model (3) where $\tau$ represents the percentage of customers who purchase in the hours following their arrival, and $C_t$ represents the number of incoming customers in period t.

$$S_t = \alpha_m N_t^{\beta_m} e^{i_m \gamma_m / l_t} \quad \text{for } m = 0, 1, 2, 3, \tag{4}$$
$$N_t = \tau_m C_{t-1} + (1 - \tau_m) C_t \quad \text{for } m = 0, 1, 2, 3. \tag{5}$$

The sales response model also considers the fact that customer can be divided into different types based on the way they react to different factors, such as the presence or absence of staff assistance, the level of in-store traffic, time limitations, etc. So, the coefficients in Models (1) -(3) $\alpha$, $\beta$, $\gamma$ and i may vary according to the specific customer type. Model (4) and (5) illustrate that hours can be divided into 4 groups according to the different customers. A dummy variable, m, is used: m=0 for weekday office hours, m = 1 for weekdays outside of office hours; m=2 for weekend hours; m=3 for special periods such as New Year or Christmas, during which the customers behave quite differently. For each m, the parameters in Model (1) and Model (3) will be different.

In order to specify the parameters of the Model (1)-(3), weekly sales are converted to daily and hourly sales. The daily and hourly sales corresponding to different days and hours of the week are placed into the corresponding group m. The parameters are determined by minimizing the sum of squared errors between the real and forecasted sales.

The deviations between forecasted ($S_t$) and realized ($R_t$) sales can be expressed as follows:

$$R_t = S_t \cdot e^{\varepsilon_t}, \tag{6}$$

where $\varepsilon_t$ is an error term at time $t$. The error term can be derived from (6) as follows:

$$\varepsilon_t = \ln(R_t / S_t). \tag{7}$$

Our objective is to minimize the sum of squared errors

$$\text{Min} \sum_{t=1 \text{ to } n} (\ln(R_t / S_t))^2, \tag{8}$$

4.2 Stage 2 (Part 1) MIP Optimization Model

*Input Parameters*

- $n^{SL}$, number of possible daily shift lengths
- $n^{DS}$, number of detailed shifts
- OT, opening time of the store
- CT, closing time of the store
- $n^F$, maximum number of full-time staff
- $n^p$, maximum number of part-time staff
- pfrate, allowed ratio of part-time to total staff = 1/19
- FG, allowable gross work hours per week for full-time staff = 40 hours
- FD, allowable work days per week for full-time staff = 5 days
- PG, allowable gross work hours per week for part-time staff = less than 30 hours
- PD, allowable workdays per week for part-time staff = 4 days
- CostF, cost of a full-time staff member to the store (assuming payments on a weekly basis) = \$21K/year / 52 weeks = \$10/hour
- CostP, cost of a part-time staff member to the store = \$10/hour
- hreq(i,h), the number of workers required on day i, at hour h
- dursh(j), duration of shift j (hours)
- workh(s,h) = 1 if a worker who is assigned to shift s, works at hour h; 0 otherwise
- shiftint(s,j) = 1 if detailed shift s has shift length j; 0 otherwise

*Indices*

- i, index for the day of the week; i = 1, …, 7
- j, index for the shift length; j = 1,..., $n^{SL}$
- s, index for the detailed shift type; s = 1, …, $n^{DS}$
- h, index for the hourly time periods during a day; h = OT, ... , CT
- k , index for full-time staff members; k = 1,..., $n^{F}$
- l, index for part-time staff members; l = 1, …, $n^{P}$

*Decision Variables (Pure Integer)*

- staffosh(*s,i*) number of staff assigned to detailed shift s on day i
- assignfull(*i,j,k*) = 1 if full-time staff k is assigned to a shift of length j on day i; 0 otherwise
- assignpart(*i,j,l*) = 1 if part-time staff l is assigned to a shift of length j on day i; 0 otherwise
- full(*k*) = 1 if full-time staff k is assigned to a shift of any length on any day; 0 otherwise
- part(*l*) = 1 if part-time staff is assigned to a shift of any length of any length on any day; 0 otherwise

*MIP Model*

$$\text{minimize } z = \text{CostF} * \sum_k \text{full}(k) + \text{CostP} * \sum_l \text{part}(l)$$

**Objective:** minimize total labor cost of part-time and full-time staffs

Subject to the following constraints:

**Constraints:**

$$\sum_l \text{part}(l) * (1 - \text{pfrate}) \leqslant \sum_k \text{full}(k) * \text{pfrate},$$

1. Define allowed ratio of part-time to full-time staff

$$\text{FD} * \text{full}(k) \geqslant \sum_i \sum_j \text{assignfull}(i,j,k) \quad \forall k,$$

2. Limit full-time staff working days and days off

$$\text{PD} * \text{part}(l) \geqslant \sum_i \sum_i \text{assignpart}(i,j,l) \quad \forall l,$$

3. Limit part-time staff working days and days off

$$\sum_i \sum_j \text{dursh}(j) * \text{assignfull}(i,j,k) \leqslant \text{FG} \quad \forall k,$$

4. Place upper bound on # of working hours/week (full-time staff)

$$\sum_i \sum_j \text{dursh}(j) * \text{assignpart}(i,j,l) \leqslant \text{PG} \quad \forall l,$$

5. Place upper bound on # of working hours/week (part-time staff)

$$\sum_j \text{assignpart}(i,j,l) \leqslant 1 \quad \forall k, i,$$

6. Ensure full-time staff only assigned one shift per day

$$\sum_j \text{assignfull}(i,j,k) \leqslant 1 \quad \forall k, i,$$

7. Ensure part-time staff only assigned to one shift per day

$$\sum_s (\text{stafftosh}(s,i) * \text{workh}(s,h)) \geqslant \text{hreq}(i,h) \quad \forall i, h,$$

8. Guarantee that the hourly requirements resulting from the sales response are realized

$$\sum_k \text{assignfull}(i,j,k) + \sum_l \text{assignpart}(i,j,l) \geqslant \sum_s (\text{stafftosh}(s,i) * \text{shiftint}(s,j)) \quad \forall i, j.$$

9. The detailed shifts are transformed to hourly shifts per day, and the staff is assigned to those shifts

## 5. Revised Model

### 5.1 Model Reconstruction

In the original model, the total weekly labor cost is computed using weekly wage, which means a worker gets paid weekly a fixed amount if the worker works for at least one shift during the week. Because most companies apply hourly wage policy in the United States, we modify the objective function so that it is calculated based on hourly wage, and added the minimum working hours and minimum working days constraints to make the model more realistic.

The other disadvantage of the original model is its output format: the output variable is the assigned shift length index $j$ on day $i$ for employee $k$. Therefore, we cannot directly produce a schedule for an employee given the model output. Given a large-size store, it requires significant time and effort to generate a feasible schedule for the staff. We modified the model output format, so that the model directly outputs the assigned shift schedule index $s$ on day $i$ for employee $k$, from which we obtain the specific details of each employee's schedule. This modification requires adding a few additional variables while the number of constraints stay the same, and most importantly reduces the time to obtain a final schedule significantly.

### 5.2 Probability Constraints

One of the main disadvantages of the original model is that it only takes into account the operational impact of the staffing policy. With an increasing competition and the pressure to create unique experiences in brick-and-mortar stores, retailers depend heavily on the service level they provide to their customers to thrive and survive financially. One indicator they use is the service availability, defined as the number of customers that each staff member can assist. For this retailer, and only as an illustration, we have defined a target of 10. If we use the expected store traffic then we can easily calculate the hourly requirement for staff

$$E(N_t)/min\ staff = 10 \ \rightarrow \ min\ staff \ = \ E(N_t)/10$$

Basically, this is done in the original model, where we obtain an optimal cost of \$8,655. This approximation might be reasonable in scenarios where there is not high variability, however if store traffic fluctuates over a long range of values then this approximation is no longer justifiable. We shall then include our service level as probability constraints.

Probability constraints arise naturally in many different applications and there has been an increasing number of research papers that explain how to solve these problems. We have divided the planning horizon (1 week) in 1-hour increments and within each interval we assume the arrival process follows a homogeneous poisson process. The following figure illustrates the different arrival rates for a typical week.

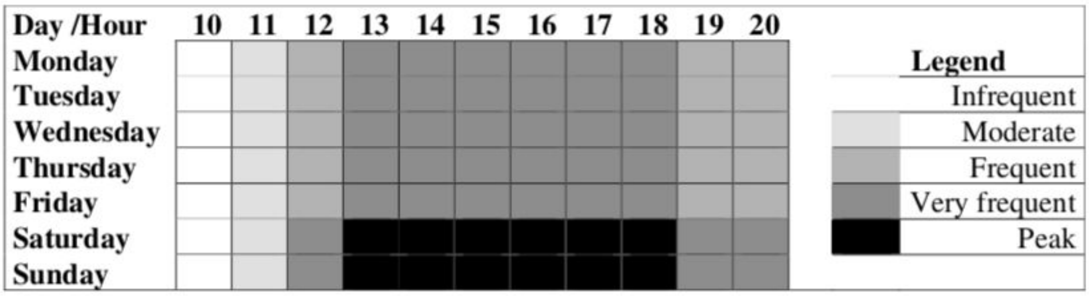



Fig. 4. Example of categorization of customers arrivals. Soure: [2]

There are two different ways to write the probability constraints, i.e. individual probability constraints or joint probability constraints,

Consider first the individual probability constraints. These constraints are written as follows,

$$\mathbb{P}[g_j(x, Z) \leq 0] \geq p_j, \ j \in J, \ p_j \in [0, 1]$$

Dentcheva D. [4] shows that for separable functions, this constraint can equivalently be expressed as a single linear constraint, as shown next

**individual prob. constr.:**

$$P(h_j(x,\xi)\geq \mathbf{0}) \geq p \Leftrightarrow P(g_j(x)\geq \xi_j) \geq p \Leftrightarrow g_j(x) \geq \underset{\text{p-quantile}}{q_p} \quad (j=1,\ldots,m)$$

This implies that $min\ staff \geq p - percentile\ of\ poisson\ distribution$. We have one such constraint for every hour, and we can solve the increased LP with any standard solver such as Gurobi. Figure 5 shows the impact of the parameter $p$ on the total staffing cost. We observe that when the probability of meeting the target service is very low (around 0.5) the cost is basically the minimum cost without any probability constraint. As we increase the service level the cost goes up and it can be as high as \$11,000 and the team grows from 20 people to 27 people. For our simulations we used $p = 0.95$, and the final result is denoted with a red dot. The optimal cost that we obtain is \$10,290 which is about 20% larger that the minimum cost obtained using the expected value.

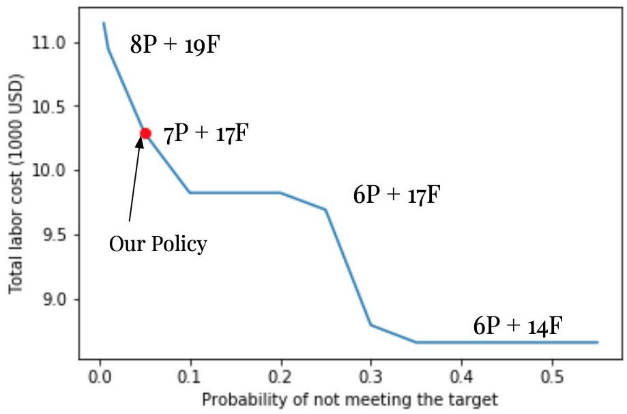



Fig. 5. Impact of the probability threshold in total staffing cost.

Consider now the joint probability constraint,

$$\mathbb{P}[g_j(x,Z) \leq 0,\ j \in J] \geq p$$

This is a much more stringent constraint and it is usually more difficult to solve because there is no percentile for multivariate probability distributions. Nonetheless, it has been shown that this constraint can be replaced by a set of constraints that contain the feasible region with high probability [5]. This method is known as Scenario Approximation, and it's similar in nature to the Sample Average Approximation (SAA) that is used to estimate the expected value in the objective function. With the Scenario Approximation you simulate N values, $Z_i$, from the multivariate probability distributions and add N constraints to the problem of the following form

$$g_j(x,\ Z_i) \leq 0$$

The set of constraints above will produce a feasible solution with probability $\delta$ and $\varepsilon$ is the probability of not meeting the service availability target. Additionally, a lower bound on the number of scenarios required is given by [5].

$$N \geq \frac{2}{\epsilon}\log\left(\frac{1}{\delta}\right) + 2n + \frac{2n}{\epsilon}\log\left(\frac{2}{\epsilon}\right)$$

For a problem of our size we would need at least 86,000 scenarios. Figure 6 shows the result of the three different methods. If we use the expected store traffic you would need a team of 20 people and there is a probability of 40% of not meeting the target service availability, but if you use the individual constraints this probability drops to 6% and the team grows to 24 people. Finally, if you use the joint probability constraint, the probability of not meeting the target is 0.001%, however the team grows to 34 people and the staffing cost goes up to $13,395 (+54%). Clearly, the decision of which constraints to add have a great impact in the optimal cost and the decision should be based on market research, strategic analysis, customer insights or from what the competitors are doing. This constraints show clearly the tradeoff of a better service and the staffing costs.

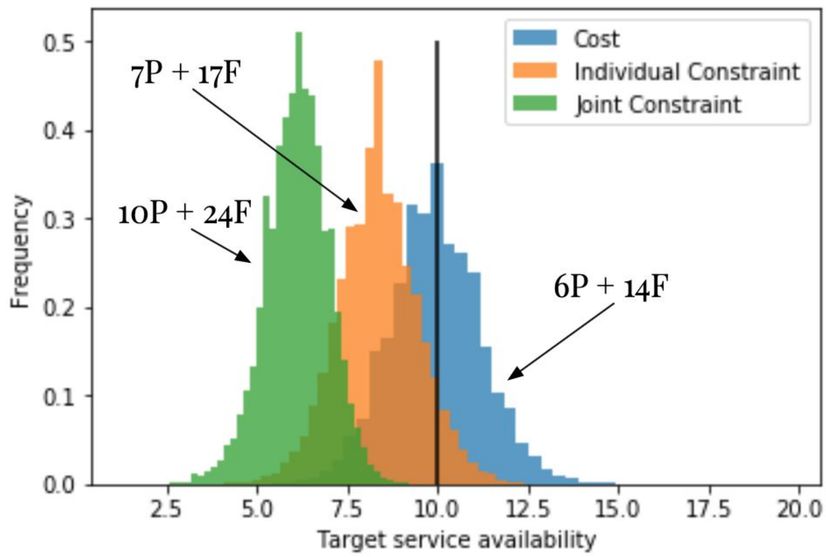



Fig. 6. Comparison of the service availability under three different staffing policies: using expected store traffic (blue), individual probability constraints (orange) and joint constraint (green)

## 6. Conclusion

After gaining a thorough understanding of the existing models we modified their formulations to account for other concerns that might affect the business bottom line, i.e. wellbeing of employees and service level to customers. In spite of the size of the model, we can conclude that mixed integer programming can be used to efficiently create workforce schedules. Using simulation we were able to quantify the tradeoff between the number of staff and the service level so that the manager of the store can follow a staffing policy that allows her to provide a superior service experience to her customers with high probability. That is why we added additional constraints to the stage 2 model to help managers make better decisions after knowing the potential impact.

The main observations of the project are:

- Matching labor to incoming traffic is a key driver of retail store profitability.
- The previous models show the tradeoffs that any retailer face between the cost of staffing, and welfare of its employees, and service level to its customers.
- Consistency in the schedule of an employee would increase his productivity and satisfaction and ultimately help drive the profitability of a company.
- Probability constraints are a very effective way to account for the uncertainty in the parameters and dynamics of the system. They provide a feasible region where we can reliably provide a

service level and deliver on our value proposition.

## References


[1] Lam, S., Vandenbosch, M., Pearce M. *Retail Sales Force Scheduling Based on Store Traffic Forecasting*. Journal of retailing.

[2] Kabak, O., Ulengin, F. et al. *Efficient shift scheduling in the retail sector through two-stage optimization.* European Journal of Operations Research.

[3] Williams, J., Lambert, S. et al. *Stable Scheduling Increases Productivity and Sales.* The Stable Scheduling Study.

[4] Dentcheva, D. *Optimization Models with Probabilistic Constraints*. Stevens Institute of Technology.

[5] Luedtke, J. And Ahmed S. *A Sample Approximation Approach For Optimization With Probabilistic Constraints.* Society for Industrial and Applied Mathematics.

[6] András Prékopa. *On Probabilistic Constrained Programming.* Princeton Symposium on Mathematical Programming.