SIMULATION AND OPTIMIZATION OF OUTBOUND OPERATIONS IN AN E-GROCER'S WAREHOUSE

By

Gustavo Costa

B.S., Electrical Engineering
Universidade Estadual de Campinas, 2010
Ecole Centrale Paris, 2011

Submitted to the MIT Sloan School of Management and the Department of Mechanical Engineering in Partial Fulfillment of the Requirements for the Degrees of

Master of Business Administration and
Master of Science in Mechanical Engineering

in conjunction with the Leaders for Global Operations Program

at the

Massachusetts Institute of Technology

June 2018

© 2018 Gustavo Costa. All rights reserved.

Signature of Author................................Signature redacted

MIT Sloan School of Management
Department of Mechanical Engineering
May 11, 2018

Signature redacted

Certified by................................Signature redacted

Kamal Youcef-Toumi, Thesis Supervisor
Professor of Mechanical Engineering

Signature redacted

Certified by................................Signature redacted

Yanchong (Karen) Zheng, Thesis Supervisor
Associate Professor, MIT Sloan School of Management

Signature redacted

Accepted by................................Signature redacted

Rohan Abeyaratne, Mechanical Engineering Chair
Graduate Program Committee, Department of Mechanical Engineering

Signature redacted

Accepted by................................Signature redacted

Maura Herson, Program Director of MBA Program
MIT Sloan School of Management

The author hereby grants to MIT permission to reproduce and to distribute publicly paper and electronic copies of this thesis document in whole or in part in any medium now known or hereafter created.
SIMULATION AND OPTIMIZATION OF OUTBOUND OPERATIONS IN AN E–GROCER’S WAREHOUSE

By

Gustavo Costa

Submitted to the MIT Sloan School of Management and the Department of Mechanical Engineering on May 11, 2018 in Partial Fulfillment of the Requirements for the Degrees of Master of Business Administration and Master of Science in Mechanical Engineering

ABSTRACT

A second wave of online grocery retailers (“e–grocers”) emerged in late 2000’s after dot.com companies, such as Webvan, Streamline and Homegrocer, failed to break into the grocery market despite investing heavily on automated warehouses and operations. The key success factors for e–grocers identified by literature after studying first–wave players is their ability to increase asset utilization, while adopting a cost–efficient delivery and service level. This thesis focused on improving utilization of a typical e–grocer warehouse with upstream order sortation by alleviating its mechanical throughput bottleneck: the outbound processes.

Batching mechanisms are used by warehouses to consolidate orders with a certain delivery time in the same picking and routing optimization batches. Increasing number of batches eliminates WIP (“work–in–progress”) inventory and increases throughput by improving area utilization, with the trade–off of higher picking and delivery costs, due to sub–optimal picking paths and delivery routes. This thesis proposes a method to determine the optimal number of batches minimizing fixed and delivery costs per order for different warehouse capacities.

The proposed optimal number of batches and a set of lean principles were then applied to a case study to simulate, through a 3D discrete–time event simulation package, a new design and process for outbound operations. Statistical distributions were used for bags arrival rates and process cycle times. The design also leveraged material handling equipment and automation solutions for the shipment label application and bags sortation processes, reducing manual labor and distances walked by workers.

Thesis Supervisor: Kamal Youcef–Toumi
Title: Professor of Mechanical Engineering

Thesis Supervisor: Yanchong (Karen) Zheng
Title: Associate Professor, MIT Sloan School of Management
ACKNOWLEDGEMENTS

The completion of this thesis would not have been possible without the support of multiple individuals that helped me along the way. I would like to take this opportunity to thank:

My thesis advisors, Kamal Youcef–Toumi and Karen Zheng, for their support and advice while writing this document.

My industry supervisor, Jeff Raines, and his team, for sponsoring this project and providing valuable data, input and thought partnership.

The staff from the MIT Leaders for Global Operations program, who helped me navigate multiple deliverables and deadlines over the course of this project.

My classmates at MIT, for their ideas and uncountable adventures shared over these two years.

My former advisor Donald Rosenfield, for his valuable guidance, input and mentorship.

Finally, a special thanks to my parents and my siblings, for their unconditional love and support over the years.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ABSTRACT</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>ACKNOWLEDGEMENTS</td>
<td>5</td>
</tr>
<tr>
<td>I</td>
<td>SECTION I: INTRODUCTION</td>
<td>11</td>
</tr>
<tr>
<td>1.1</td>
<td>Purpose and Thesis Overview</td>
<td>11</td>
</tr>
<tr>
<td>1.2</td>
<td>Brief History of e-Grocery</td>
<td>12</td>
</tr>
<tr>
<td>1.3</td>
<td>Key Lessons Learned from First Wave e-Grocers</td>
<td>13</td>
</tr>
<tr>
<td>1.4</td>
<td>Operations Strategy Choices for e-Grocers</td>
<td>14</td>
</tr>
<tr>
<td>1.5</td>
<td>Process Overview in a Typical e-Grocer's Warehouse</td>
<td>16</td>
</tr>
<tr>
<td>1.5.1</td>
<td>Inbound Operations</td>
<td>17</td>
</tr>
<tr>
<td>1.5.2</td>
<td>Picking Operations</td>
<td>17</td>
</tr>
<tr>
<td>II</td>
<td>SECTION II: IMPROVING OUTBOUND OPERATIONS</td>
<td>19</td>
</tr>
<tr>
<td>2.1</td>
<td>Overview of Outbound Processes</td>
<td>19</td>
</tr>
<tr>
<td>2.2</td>
<td>Issues with Current Processes</td>
<td>21</td>
</tr>
<tr>
<td>2.3</td>
<td>Optimizing Number of Batches</td>
<td>23</td>
</tr>
<tr>
<td>2.3.1</td>
<td>Warehouse Scalability Study</td>
<td>24</td>
</tr>
<tr>
<td>2.3.1.1</td>
<td>Inventory Build-Up Simulation</td>
<td>24</td>
</tr>
<tr>
<td>2.3.1.2</td>
<td>Impact on Fixed Costs per Order</td>
<td>25</td>
</tr>
<tr>
<td>2.3.2</td>
<td>Delivery Costs Study</td>
<td>28</td>
</tr>
<tr>
<td>2.3.2.1</td>
<td>Overview of Delivery Operations</td>
<td>29</td>
</tr>
<tr>
<td>2.3.2.2</td>
<td>Middle Mile Cost Study</td>
<td>29</td>
</tr>
<tr>
<td>2.3.2.3</td>
<td>Routing Optimization</td>
<td>31</td>
</tr>
<tr>
<td>2.3.2.4</td>
<td>Last Mile Cost Study</td>
<td>33</td>
</tr>
<tr>
<td>2.3.3</td>
<td>Total Cost per Delivery (CPD)</td>
<td>38</td>
</tr>
<tr>
<td>2.3.4</td>
<td>Optimal Number of Batches</td>
<td>39</td>
</tr>
<tr>
<td>2.3.5</td>
<td>Potential Impact on Customer Experience and Variable Costs</td>
<td>40</td>
</tr>
<tr>
<td>2.4</td>
<td>Layout and Automation Improvements</td>
<td>42</td>
</tr>
<tr>
<td>2.4.1</td>
<td>Layout Changes</td>
<td>42</td>
</tr>
<tr>
<td>2.4.2</td>
<td>Automation</td>
<td>45</td>
</tr>
<tr>
<td>III</td>
<td>SECTION III: SIMULATING RECOMMENDED OUTBOUND OPERATIONS</td>
<td>49</td>
</tr>
</tbody>
</table>
3.1 Discrete Event Simulation Overview ................................................. 49
3.2 Overview of Simulation Model ...................................................... 50
3.3 Simulation Results ............................................................................. 54
SECTION IV: CONCLUSIONS & RECOMMENDATIONS .......................... 56
  4.1 Summary of Findings ....................................................................... 56
  4.2 Recommendations and Next Steps .................................................. 57
REFERENCES ............................................................................................. 58
LIST OF FIGURES

Figure 1 – Overview of Processes in a Typical e-Grocer Warehouse .................. 16
Figure 2 – Outbound Process Flow ................................................................. 19
Figure 3 – High Level Representation of Outbound Processes Timeline ................. 21
Figure 4 – Impact of Changes in Number of Batches ........................................... 23
Figure 5 – Illustration of WIP Build-Up Simulation Results By Batch ..................... 24
Figure 6 – Comparison of Buffer WIP Simulation with Real Data ......................... 25
Figure 7 – WIP Inventory Area Scalability for Different Number of Batches .......... 26
Figure 8 – Typical Intra-Week Demand Pattern for an e-Grocer ......................... 27
Figure 9 – FCPO for Different Number of Batches ............................................ 28
Figure 10 – High Level Topology of a Hub and Spoke Model ............................. 29
Figure 11 – Middle-Mile Delivery Costs .......................................................... 30
Figure 12 – Histogram of Orders per Number of Bags (Based on Real Data) ............ 34
Figure 13 – DPR Sensitivity by Vehicle Capacity and Delivery Window ................ 35
Figure 14 – DPR Sensitivity (Orders Volume) .................................................. 37
Figure 15 – Last-Mile Delivery Costs ............................................................... 38
Figure 16 – Total CPD for Different Number of Batches .................................... 39
Figure 17 – Optimal Number of Batches Curves .............................................. 40
Figure 18 – Hypothetical Outbound Layout ...................................................... 43
Figure 19 – Proposed Layout ........................................................................... 44
Figure 20 – Illustration of an Auto-Sorter ........................................................... 46
Figure 21 – Sketch of Traditional Auto-Labelers ............................................... 47
Figure 22 – Sketch of Swing Tamp Label Applicator (ID Technology’s 252N model) 48
Figure 23 – Screenshots of Base Case Simulation ............................................. 52
Figure 24 – Screenshots of Proposed State Simulation .................................... 53
Figure 25 – WIP Build-Up Comparison .............................................................. 54

LIST OF TABLES

Table 1 – List of e-Grocers Characterized by Their Operations Strategy .......... 15
Table 2 – Example of Assumptions for Simulation Arrival Rates and Cycle Times ... 51
Table 3 – Additional Simulation Assumptions .................................................. 52
Table 4 – Simulated Improvements in Variable Costs ......................................... 55
SECTION I: INTRODUCTION

1.1 Purpose and Thesis Overview

The primary goal of this thesis is to study and improve the design of outbound operations in the warehouses of online grocers ("e-grocers"). Outbound operations have become critical to these players as the increase in volumes shifted route sorting and loading processes from third-party companies to the e-grocers’ own warehouses, dramatically increasing the footprint required and aggravating the mechanical throughput bottleneck. This document identifies improvement opportunities that would allow e-grocers to alleviate the outbound bottleneck in their network, leading to a direct increase in asset utilization and profitability.

Given the high interdependency between outbound and delivery operations, delivery costs and routing performance will also be studied with the goal of determining how they would be impacted by changes in the outbound processes. In addition, this document identifies opportunities to reduce outbound variable costs through layout changes and automation, which could minimize cycle times, walking distances and redundant processes.

In order to achieve the objectives highlighted above, this document is divided into four key sections. Section I will provide a brief history of the e-grocery industry and highlight lessons drawn from case studies of first-wave players that have attempted to enter this market in the past. In addition, it will introduce a framework to categorize different e-grocers according to choices made in their operations strategy, before providing an overview of the processes conducted in a typical e-grocer’s warehouse.

Section II will focus on outbound operations, highlighting some of the key issues observed in the current processes, before investigating two sets of improvement opportunities: optimizing the number of batches and exploring layout and automation changes. When describing how to optimize the number of batches, the document will explore how warehouse scalability changes with number of batches and then dive deep on delivery costs, studying how delivery performance could be impacted by higher number of batches.
Section III will introduce the discrete–time simulation model that can be used to estimate the reduction in variable costs enabled by changes in layout and automation solutions investigated in Section II. Finally, Section IV will close the document with a summary of key findings and a set of conclusions and recommended next steps for e-grocers when designing and operating the outbound area in their warehouses.

1.2 Brief History of e–Grocery

As the largest single U.S. retail category and with only 2% of online shopping penetration as of 2015, grocery remains an attractive opportunity for retailers migrating to digital offerings (Nowak et al. 2016). Over the last 5 to 10 years, traditional brick and mortar stores, online retailers and start–ups have launched grocery delivery and pick–up solutions with the goal of tapping into this opportunity. Wal–Mart, Tesco, Amazon, Ocado, Peapod and Instacart are just a few examples of companies that have recently started offering digital grocery shopping. This foray into e–grocery is not a new phenomenon. Between late 1990's and early 2000's, dot.com internet grocery retailers (Webvan, Streamline, Homegrocer, Shoplink) have unsuccessfully attempted to break into the market before running out of money (Tanskanen, Yrjölä, and Holmström 2002).

These dot.com companies, known as the “first wave” e–grocers, and their business models have been vastly studied in the literature and many articles have been written with the goal of understanding why they failed and what are the fundamental differences between groceries and other retail categories that make the former harder to digitize. Most authors have identified operations and logistics challenges as the key factors behind the failure of first wave players. These factors are particularly important for the online model when compared to brick and mortar, because order picking, sorting and delivery are all additional activities that e–grocers have to conduct on behalf of the customer (Himmelstein 1999; Ring and Tigert 2001). In addition, grocery demand variation posed an extra challenge to achieve stable workload and efficiently utilize assets across the e–grocers’ network, which led to over–investment and unprofitable operations (Kämäräinen et al. 2001).
1.3 Key Lessons Learned from First Wave e-Grocers

At a more granular level, researchers identified six key challenges faced by previous e-grocery models: over-investment in automation, expensive home deliveries, weak negotiation and purchasing power with suppliers, low ordering frequency / uneven demand and lack of services (Kämäräinen and Punakivi 2002).

Webvan, for example, invested approximately $1.2 billion in its two-year lifespan before going bankrupt in 2001, with most of the investment going to highly automated warehouses with an average cost of $25 million (Barker 2009). As a comparison, competitors' less automated warehouses had an average cost between $4 million and $6 million (Guglielmo 2000). The challenge of having expensive automated e-grocery warehouses is that they are often underutilized. E-grocery customers like to have their orders delivered in very specific days and time windows (e.g., Monday mornings) and warehouses need to be sized to support that demand. However, during the rest of the time, the infrastructure is idle. This high demand variation resulted in a low capacity utilization of 30% in Webvan's warehouses (Sandoval 2002), dramatically increasing fixed costs and hurting business profitability.

On the delivery side, low customer density, over-reliance on the attended delivery model and narrow delivery windows were the key drivers behind high costs faced by first wave players. With the goal of increasing customer satisfaction and stimulating adoption, Webvan offered delivery time windows of 30 minutes, with only attended delivery. This resulted in high number of trucks required to meet customers' demand and extremely expensive delivery operations.

Based on the lessons learned from first wave players, new e-grocery retailers should focus on creating flexible layout designs for their warehouses, allowing them to operate with an optimal product flow, while tailoring their delivery strategies to the demand level that they are currently facing. Changes in layout and policy adjustments (such as batch sizes) could then be continuously made as demand ramps up. This efficient process and product flow, without unnecessary stops, should be regarded as a more important target than rapidly increasing the level of automation in operations (Kämäräinen et al. 2001).
1.4 Operations Strategy Choices for e-Grocers

Both first and second wave e-grocers can be categorized by choices made in terms of focus, picking location and delivery experience.

- **Focus**: e-grocers can be either **pure play** online retailers or **diversified**, offering both brick and mortar and online shopping options to customers

- **Picking Location**: e-grocers can opt to pick all orders from currently existing brick-and-mortar locations or build dedicated warehouses to fulfill online orders

- **Delivery Experience**: e-grocers can deliver all orders to customers’ doors or offer an order pick-up option; For door delivery, they can offer attended and/or unattended delivery

Table 1 lists major first and second wave players according to their choices in terms of focus, picking location and delivery experience. We can see that some players decided to offer multiple choices to customers in each of these categories. Peapod, for example, offers both door delivery and store pickup in Stop & Shop locations (which are owned by the same parent organization, Ahold Delhaize). AmazonFresh opted for door delivery, but have started to experiment with pick-up experience in its Fresh Pick-Up locations in Seattle.

Both Peapod and Wal-Mart leverage an already established physical network to not only deliver from their stores, but also pick and sort orders from the same buildings. Instacart completely foregoes physical locations (and the investments associated with them) and fulfill orders from brick and mortar stores operated by partners (e.g., Star Market, Costco, CVS). Ocado and AmazonFresh adopt a different picking model, fulfilling orders from dedicated warehouses.
Table 1 — List of e–Grocers Characterized by Their Operations Strategy

<table>
<thead>
<tr>
<th>Focus</th>
<th>Picking Location</th>
<th>Delivery Experience</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wal–Mart</td>
<td>Diversified</td>
<td>From own stores</td>
</tr>
<tr>
<td>Peapod</td>
<td>Diversified (owned by brick and mortar Ahold Delhaize)</td>
<td>From own stores &amp; dedicated warehouses</td>
</tr>
<tr>
<td>Instacart</td>
<td>Pure play</td>
<td>From 3rd party stores</td>
</tr>
<tr>
<td>AmazonFresh</td>
<td>Pure play</td>
<td>From dedicated warehouses</td>
</tr>
<tr>
<td>Ocado (UK)</td>
<td>Pure play</td>
<td>From dedicated warehouses</td>
</tr>
<tr>
<td>Webvan</td>
<td>Pure play</td>
<td>From dedicated warehouses</td>
</tr>
<tr>
<td>Streamline</td>
<td>Pure play</td>
<td>From dedicated warehouses</td>
</tr>
</tbody>
</table>

Although investments on dedicated warehouses increase fixed costs, there are many operational benefits in choosing this model. One of the benefits is the ability to reduce demand variability through the risk pool effect and minimize spoilage costs. A second benefit is the increase in stowing and picking efficiency. A dedicated warehouse can be specially designed to increase picking speed, whereas a brick and mortar store also has to focus on properly arranging and displaying products to shoppers. In addition, brick and mortar shoppers can get in the way of online order pickers, negatively impacting picking efficiency. Literature benchmarks define a picking rate range between 100 and 450 lines per hour per worker, with 100 lines per hour corresponding to store picking model with no automation and 450 lines per hour corresponding to a highly automated warehouse (Anders 1999; Kämäräinen et al. 2001).

Given the high demand variation, achieving high asset utilization rate is critical to efficiently operate a dedicated e–grocery warehouse. High utilization can be achieved through retail levers or process improvements. An example of retail lever is offering discount in less popular days and delivery times in order to balance the intra–week and intra–day demand. Peapod, for example, currently offers a discount of up to $5 if customers choose to have their orders delivered during less popular days (e.g., Tuesday or
Wednesday vs. Monday or Sunday) and time (e.g., early afternoon vs. early morning or evenings).

In this thesis, we will study in more detail the fulfillment processes of e-grocers with dedicated warehouses. We will identify opportunities to improve utilization through operational levers, such as changes in policy and layout, while assessing the impact on other cost drivers of the e-grocery model, such as delivery costs.

1.5 Process Overview in a Typical e-Grocer’s Warehouse

The activities in a dedicated e-grocer’s warehouse can be broken down in three macro-processes: Inbound, Picking and Outbound. Figure 1 shows a high level representation of these processes. Inbound encompasses receiving cases of SKUs (Stock Keeping Units) and stowing them into bins. Picking encompasses going through the warehouse’s bins and placing the SKUs ordered by customers into bags (also called packages throughout this document). Outbound encompasses applying the shipment label on bags, taping, route sorting and loading.

![Figure 1 - Overview of Processes in a Typical e-Grocer Warehouse](image)

Some e-grocers consider picking as part of the outbound macro-process, but for the purposes of this thesis, outbound will refer to all processes happening after an order has been picked. From an organizational standpoint, workers are cross-trained to perform any of the activities highlighted above, with one manager, responsible for staffing decisions and problem solving, allocated to each macro-process.
We will briefly discuss the inbound and picking macro-processes in the sections below. The thesis will focus on optimizing the outbound processes, which will be covered in detail in section II.

1.5.1 Inbound Operations

Inbound processes start with trucks arriving at the receiving dock and delivering product pallets. These pallets wait in the receiving area until it is their time to be stowed. A FIFO (first in, first out) approach is used to determine the stowing order.

Workers transfer SKUs from pallets to stow carts and walk around storage bins looking for stowing space. For each SKU stowed, workers scan its barcode, enter expiration date, find a bin with enough space, scan the bin's barcode and place the SKU in the bin. Most warehouses adopt a random stowing method, in which workers can place a certain SKU in any bin with available space.

In e-grocery warehouses, certain restrictions are applied on the random stow method. The most obvious one is the temperature zone (e.g., ambient, chilled, frozen and tropical) restriction. Another one is related to the inventory turnover and fragility of certain SKUs. As an example, heavy and bulky SKUs (e.g., soda packages) have to be stowed in the beginning of the pick path to ensure that, during the picking process, they will be placed on the bottom of the bag and will not damage fragile SKUs (e.g., bread), which have to be stowed towards the end of the pick path.

1.5.2 Picking Operations

The picking processes start with the preparation of the pick cart. Workers place a Quick Response (QR) code sticker on each paper bag, which is then transferred to a pick cart with capacity to hold up to 10 bags. Workers also place an insulator and chilled gel packs inside bags that will store chilled SKUs. At this stage, pick carts can contain bags from multiple orders and bags from a same order can be distributed across different pick carts. An optimization algorithm determines which bags should be consolidated in which cart in order to maximize pick rates.
After being assigned to a pick cart, workers use a scanner to identify the bin location of the first SKU to be picked. For every SKU picked, the worker scans the bin, the SKU and the destination bag. The scanner then directs the operator to the next bin location. After all SKUs from a cart are picked, the worker leaves the cart next to the shipment label application stations (the first process of the outbound area) before re-initiating the picking process with an empty pick cart.

In order to avoid congestion due to excessive number of picking carts and balance the workload in the warehouse, e-grocers can start picking an order up to 12–14 hours before its delivery time. In addition, e-grocers usually batch a certain number of orders for picking, in order to maximize cart utilization and picking rates.
SECTION II: IMPROVING OUTBOUND OPERATIONS

2.1 Overview of Outbound Processes

From a mechanical throughput standpoint, outbound operations are the bottleneck in e-grocers' warehouses. Stowing and picking capacities can theoretically be increased by adding workers to picking and stowing processes. However, the number of shipment label application and taping stations and, more importantly, the space in the floor required to store WIP (Work in Progress) inventory can't be easily increased. They all need to be correctly sized to support operations at maximum capacity. Figure 2 shows a high level representation of the process flow in an e-grocer's outbound area.

![Outbound Process Flow](image)

*Shipment Label Application:* After picking process is complete, workers assigned to shipment label application stations transfer bags from pick carts to the station's table and scan the bags' barcodes. Shipment label containing customer name, address, delivery node and delivery window is printed and applied on the bag.

*Taping:* After applying the shipment label, workers arrange the SKUs inside the bag and close it by first folding and then applying an adhesive tape on top. The taped
bag is then inducted to a conveyor. A belted conveyor is used to allow for round SKUs to be placed on the bottom of the bag. If a roller conveyor was used instead, the bag could become unstable and fall.

- **Buffer Area Induct:** After reaching the WIP area referred to as “buffer”, bags are pre-sorted to a rack, according to delivery node, batch and delivery window. Bags wait in buffer until route generation is run for their batch, which usually happens approximately 3 hours before the start of the earliest delivery window assigned to that batch.

- **Route Label Application:** After route generation, workers re-scan bags, attach a route label and transfer them to a second conveyor that connects the buffer to a sorting area.

- **Sorting:** When bags reach the second WIP area referred to as “sorting area”, they are manually diverted according to delivery node and window and then sorted to route racks.

- **Auditing and Vehicle Loading:** When all bags from a delivery node and window have been sorted, workers audit each route rack and load them into trucks.

As highlighted in Figure 2, the entire process repeats for each batch, which is defined based on the delivery windows offered by the e-grocer. As an example, bags delivered from 5–7am, 6–8am and 8–10am could all be considered part of batch 1, while bags delivered from 8–10am and 10–12am could be part of batch 2. Bags belonging the same batch start to be picked at the same time, are grouped together for routing optimization and have the same critical pulling time (i.e., the latest time a bag needs to be picked in order to be delivered on time, which is usually defined as 3 to 4 hours before the start of the earliest delivery window time). In addition, bags from the same batch are consolidated for middle mile transportation, when route racks are transferred from warehouses to different “delivery nodes”, which are smaller facilities established closer to the customers to avoid excessive displacement between last mile delivery vehicles and the warehouse. Middle mile, delivery nodes and last mile logistics are explained in more detail in section 2.3.2.

If there is available capacity for a given delivery window, customers can place orders up to a certain time before the batch’s critical pulling time. In the example given
above, if a customer would like to have his order delivered from 8–10am and if the order cut-off time is defined as 3 hours before the batch’s critical pulling time, he would have until 10pm of the previous day to place the order (assuming that the first batch’s critical pulling time is 1am).

In addition, since picking can start up to 14 hours before a batch’s earliest delivery window, the e-grocer might pick multiple batches at the same time in order to balance the picking load. Figure 3 shows a high level timeline of the outbound processes for a given batch.

![High Level Representation of Outbound Processes Timeline](image)

E-grocers often design a new warehouse based on a given number of batches, delivery windows and expected demand pattern when operating at full capacity. These are the main drivers when sizing buffer areas, conveyors, number of route racks, number of shipment label application stations etc.

### 2.2 Issues with Current Processes

A fundamental difference in e-grocers’ shopping experience when compared to traditional e-commerce is the ability of customers to choose a delivery window, which is important for groceries because of the perishable and fragile nature of products. The batching mechanism introduced in the previous section allows for this experience, without dramatically increasing delivery costs or decreasing pick rates. It is a mechanism to consolidate, in the same picking path and delivery route, a group of orders that need to be delivered at a certain time, with the trade-off of higher WIP inventory that needs to be accommodated in the warehouses.

The buffer and sorting WIP areas highlighted in Figure 2 represent one of the main challenges for outbound operations. In a typical warehouse, these areas usually take a
disproportional square footage, which could be utilized for more value-added operations in the distribution network, such as cross-docking or storage capacity increase. In addition, since large buildings are hardly found in city centers, high WIP inventory usually deters e-grocers from placing warehouses closer to regions with high population density, limiting their ability to reduce delivery costs. Finally, since capital expenditures on new warehouses are directly related to square footage, larger buffer and sorting areas drive higher depreciation costs, negatively impacting business profitability.

Excessive number of touches is another fundamental issue with current operations. Each bag is scanned twice and has two labels assigned during outbound. The redundant touch happens because delivery routes are not known until after all bags from a batch were picked and stored in the buffer area. Although delivery addresses are known at customer cut-off and total number of bags has already been estimated, operations wait to run routing optimization in case a new bag is added to the order (which might happen when SKUs don’t fit into a bag or when customers add an additional SKU to their orders, for example). Waiting to run routing until all shipment labels for the entire batch are applied allows for higher accuracy in number of bags passed to the routing algorithm, but leads to higher variable cost due to non-value adding time spent re-scanning and re-labeling bags during route label application. This redundancy also results in higher capex spent in additional scanners and labelers.

Inefficiencies also permeate the shipment label application and taping processes. Workers have to navigate an area cluttered with pick carts to retrieve bags and initiate their activities. Transferring bags from pick carts to shipment label application stations and then to conveyors takes on average ~8s, a cycle time that is aggravated by the obstacles and number of carts left by picking workers in the area. In addition, handling the barcode scanner every time a new bag is retrieved from pick cart adds a step that could be avoided if a separate station specialized only on shipment label application. Finally, high variation was observed in the taping process due to the need of checking and repacking certain bags. This is a non-value added activity resulting from low quality picking. In order to increase pick rate metrics, workers try to pick as fast as possible and do not always place SKUs in the most organized way inside bags. As a result, taping workers have to remove SKUs and repack them.
2.3 Optimizing Number of Batches

While a batching mechanism is desired to enable scheduled delivery at a reasonable cost, there is an opportunity to optimize number of batches for a given warehouse size. Adding batches allows for a reduction of WIP bags and square footage, increasing the mechanical throughput of operations without changing fixed costs.

Higher number of batches also has a positive impact on customer experience, since it allows for additional order cut-off’s. As an example, if orders with delivery window 20:00 – 22:00 belong to the same batch as orders delivered from 18:00 – 20:00, they need to be placed by 10:00 (the cut-off for the entire batch), given the time constraint imposed by the earlier delivery window. If delivery window 20:00 – 22:00 was assigned to its own batch, customers would have until 12:00 to place orders for that window.

However, since higher number of batches leads to increase in delivery costs, the entire system needs to be better studied in order to determine an optimal number of batches minimizing total cost. Figure 4 highlights the key impacts on the system of increases in the number of batches. In the following sections, each of these effects will be quantified and the results of these studies will be combined with the goal of identifying an optimal number of batches that warehouses should adopt.

![Figure 4 - Impact of Changes in Number of Batches](image-url)
2.3.1 Warehouse Scalability Study

Since fixed costs are highly driven by capital investments on the warehouse and the investments are highly proportional to square footage, the warehouse scalability needs to be studied in order to calculate the fixed costs impact of an increase in number of batches.

2.3.1.1 Inventory Build-Up Simulation

With the goal of studying warehouse scalability under different number of batches, a simple model can be built in MS Excel to simulate the inventory build-up in the buffer and sorting areas shown in Figure 2.

Using the average of process cycle-times, average intra-week and intra-day demand patterns and timing policies (e.g., cut-off time, delivery vehicle arrival time etc.) that are currently in place in an e-grocer’s warehouse, we can estimate the inventory build-up profile in a typical week and identify the maximum number of bags that each area will need to hold in order to support the simulated capacity for a given number of batches. The average cycle-times can be obtained through time studies for each outbound process. The intra-day demand patterns are obtained by averaging the percentage of the daily volume requested to be delivered at a certain delivery window, while the intra-week demand patterns are obtained by averaging the percentage of weekly volumes usually ordered in each day of the week.

![Buffer Area by Batch (N of Bags)](image1)

![Sorting Area by Batch (N of Bags)](image2)

*Figure 5 – Illustration of WIP Build-Up Simulation Results By Batch*

An example of output of the MS Excel model for a scenario with four batches and a simulation period of one week (starting on Monday) is shown in Figure 5. Each line represents a different batch while the horizontal and vertical axis represent time and the
number of bags accumulating in each area, respectively. The quantities for each batch can be added up to calculate total WIP inventory in either the buffer or sorting area at any given time. We see that the highest WIP inventory occurs on Monday during the first two morning batches, when customer demand is at its peak.

We validated the model by comparing the results with the real WIP build-up on a typical week of a given warehouse. The comparison can be seen in Figure 6, where the dotted line represents the estimated build-up and the solid line represents the real data. Both lines are highly correlated (r-squared of 93%).

![Figure 6 - Comparison of Buffer WIP Simulation with Real Data](image)

Since warehouses are designed to support the peak demand, the WIP inventory area required to support operations for a given capacity can be easily estimated by identifying the peak in total number of bags in the inventory build-up charts and dividing it by the storage density (i.e., the number of bags that can be accommodated in a given area). As an example, if a shelf with area footprint of \( S_{shelf} \) sq. ft. is used to store 10 bags, total square footage required in a given WIP area is given by equation (1).

\[
\text{Area Required} = \frac{\text{Max Buildup}}{\text{Storage Density}} = \text{Max Buildup} \cdot \frac{S_{shelf}}{10}
\]

(1)

2.3.1.2 Impact on Fixed Costs per Order

Eight different scenarios with one through eight batches (\( n=1\ldots8 \)) were used to simulate the area required for WIP inventory for different warehouse capacities. As mentioned previously, the intra-day and intra-week demand patterns were obtained by
averaging real data from current operations (e.g., within a given day, on average 30% of orders could be placed at delivery window 1; within a given week, 20% of orders could be placed on Sunday). In addition, it was considered that the absolute value of demand during the weekday with highest volume scales with the warehouse’s capacity that is being simulated. Therefore, demand and warehouse capacity were considered to scale proportionally.

Maximum of eight batches was considered to match the number of delivery windows usually offered by e-grocers. Figure 7 shows how area requirement varies with warehouse capacity for different number of batches. The chart was normalized by dividing the $x$ and $y$ value of each point by the maximum value from each axis (the same methodology was applied across all charts in this document).

The curves follow the linear model as shown in equation (2), where $S$ is the area requirement and $K$ is daily capacity (given in number of bags per day). $\alpha_n$ and $\beta_n$ are the slope and intercept for the curve corresponding to number of batches $n$ and their units are given by [area/(bags/day)] and [area].

$$S(K) = \alpha_n \cdot K + \beta_n$$

**Figure 7 – WIP Inventory Area Scalability for Different Number of Batches**
As seen in Figure 7, operations become more scalable as number of batches increases, requiring less WIP area for the same capacity (or increasing the capacity for the same area). In other words, $\alpha_n$ decreases with number of batches. Assuming that we know the annual fixed costs ($FC^*$) of a warehouse with number of batches $n^*$, we can estimate the capacity $K_n$ for any number of batches $n$ through equation (3).

$$K_n = \frac{S_{n^*}}{\alpha_n}$$

(3)

Annual fixed costs per order (FCPO) for number of batches $n$ can then be determined by dividing $FC^*$ by $K_n$ and adjusting for number of days in one year, average number of bags per order $n_{bags}$ and a dimension-less demand factor $\sigma$. The demand factor adjusts for the demand variability within a week, since the warehouse usually does not operate at maximum capacity every day of the week. A typical demand pattern over a month is shown in Figure 8.

![Typical Intra-Week Demand Pattern for an e-Grocer](image)

*Figure 8 – Typical Intra-Week Demand Pattern for an e-Grocer*

Equation (4) can then be used to calculate FCPO for each number of batches $n$, after following the methodology described above. The numerator represents total annual fixed costs while the denominator is the total number of orders that a warehouse has in one year when operating at full capacity with the intra-week and intra-day demand pattern described previously. The resulting FCPOs can be plotted against $n$, resulting in the chart shown in Figure 9.
\[
\text{Fixed Costs per Order} = FCPO_n = \frac{FC^*}{365 \cdot K_n \cdot \sigma} \div n_{\text{bags}}
\] (4)

**Figure 9 – FCPO for Different Number of Batches**

From Figure 9, we see that FCPO reduces as number of batches increases. This is in line with what was expected, since a higher number of batches reduces the space requirement in the warehouse, leading to lower investments and, therefore, lower depreciation costs, one of the key components of warehouses’ fixed costs. Similarly, if an already built warehouse is able to increase its capacity by increasing the number of batches, total fixed costs can be divided by a higher number of orders, thus reducing total FCPO. The curve shown in Figure 9 can be generalized to a power function that varies with the number of batches \( n \), as seen in equation (5). The R-squared of the fit is 0.95.

\[
FCPO(n) = \frac{1.05}{n^{0.48}}
\] (5)

### 2.3.2 Delivery Costs Study

After studying the benefits of increasing number of batches \( n \), we now need to assess the impact that a larger \( n \) will have on delivery costs. This chapter provides an overview of delivery operations, before studying in detail the middle and last mile costs. For the last
mile analysis, we will test the performance of a heuristic optimization algorithm used to create e-grocers' delivery routes.

2.3.2.1 Overview of Delivery Operations

A hub and spoke model is commonly used to deliver grocery bags from warehouses to customers. With the goal of reducing real estate costs, e-grocers establish their main warehouses in the suburbs, outside of the urban areas where their market is concentrated. In order to avoid excessive displacement of delivery drivers between urban areas and the warehouse, smaller facilities, called “delivery nodes”, are established closer to customers.

Figure 10 is a high-level representation of the hub and spoke model. Middle mile (MM) refers to the distance between warehouses (where products are stored, picked and orders are sorted) and delivery nodes (where orders, already sorted to delivery routes, wait for last mile delivery). The area covered by each delivery node is defined so that the farthest customer can be reached within the delivery window promised by the e-grocer.

![Figure 10](image)

*Figure 10 – High Level Topology of a Hub and Spoke Model*

2.3.2.2 Middle Mile Cost Study

Middle mile distances are covered by line-haul trucks with capacity of 200+ bags. When a line-haul truck reaches the delivery node, route racks with customers’ orders are unloaded and replaced by empty route racks, which are then brought back to the
warehouse. One or more trucks can be used for a given batch and delivery node, depending on the demand volume.

Total middle mile costs are driven by the price charged by third party transportation companies for each trip made by the line-haul truck. Therefore, costs can be estimated for a batch and capacity scenario by calculating number of trucks required to transport all bags to delivery nodes. If we know truck capacity $K_T$, cost of each trip $C_T$, warehouse daily capacity $K$ and demand factor $\sigma_{i,d,k}$ (fraction of the capacity $K$ that is seen at node $i$, week day $d$ and batch $k$), middle mile cost per delivery (CPD) for a given number of batches $n$ can be estimated by equation (6).

$$\text{MM Cost per Delivery (n)} = \frac{\sum_{i=1}^{\#\text{nodes}} \sum_{d=1}^{7} \sum_{k=1}^{n} \max\left(\frac{K \cdot \sigma_{i,d,k}}{K_T}, 1\right) \cdot C_T}{\sum_{i=1}^{\#\text{nodes}} \sum_{d=1}^{7} \sum_{k=1}^{n} K \cdot \sigma_{i,d,k}}$$

By applying equation (6) to different scenarios of capacity $K$ and batches $n$, we can plot the charts seen in Figure 11. Middle mile CPD decreases exponentially with daily capacity $K$ and becomes practically the same regardless of number of batches. At higher capacities, economies of scale of combining volume in batches for middle mile transportation become less relevant, since multiple line-haul trucks will be needed regardless of number of batches.

![Figure 11 - Middle-Mile Delivery Costs](image)

After studying middle mile delivery costs, we now need to understand the impact that a higher number of batches has on last mile, one of the major cost components for e-
Since last mile delivery routes are mainly driven by optimization algorithms, we will first provide an overview of vehicle routing problems in section 2.3.2.3, before studying the performance of a specific type of routing algorithm used by e—grocers in section 2.3.2.4. We will assess the impact of different vehicle capacities and delivery windows on the overall solution and test how number of batches affect route performance.

**2.3.2.3 Routing Optimization**

Last mile routes are determined by a routing algorithm with the objective function of minimizing number of routes with capacity $C_v$ that leaves a certain delivery node, while delivering all orders within a time window chosen by customers. Before assessing the performance of the last mile optimization conducted by the e—grocer whose operations were analyzed previously, this section first provides a brief literature review of routing problems, one of the most studied optimization problems in the operations research field.

- **The Vehicle Routing Problem (VRP)**

In the classic VRP, first formulated by Dantzig and Ramser (1959), a capacity—constrained vehicle must visit a set of dispersed customers located at $P_i$ that have a certain demand $Q_i$. All vehicles start from a terminal point $P_0$ and distances between each location are expressed by $d_{ij}$. The goal is to minimize total cost of travel (quantified by distance or time), by varying the set of binary variables $x_{ij}$. The variables $x_{ij}$ represent that points $P_i$ and $P_j$ are connected. Therefore, the sum of $x_{ij}$ over all $i$’s must be equal to 1 to ensure that $P_i$ is connected to another point in the route. This problem was first called the “Truck Dispatching Problem” and is considered a sub—type of the traveling salesman problem (TSP).

VRP’s are NP—hard problems and the literature proposing solutions to solve them is vast. Both exact and heuristic algorithms have been proposed, with exact algorithms usually solving simpler sub—sets of the problem (i.e., no distance constraints, no empty routes allowed) (Toth and Tramontani 2008). A wide range of metaheuristic methods have also been proposed, leading to more generic solutions that avoid local optiums. Gendreau et al. (2008) provides a good classification of all types of metaheuristics that have been used to solve VRPs.
- **The Capacitated Vehicle Routing Problem with Time Windows (CVRPTW)**

  The CVRPTW is a type of VRP with one additional constraint dictating that all
delivery vehicles must complete their routes within a certain time. Customers cannot be
served until the beginning of the delivery window and vehicles are not allowed to arrive
after the delivery window has ended. In the case of e-grocers, delivery windows usually
range from 1 to 3 hours. Shorter windows penalize the objective function, while wider
windows negatively impact customer experience and the cold chain (e.g., refrigerated
products might go bad if they wait in non-acclimatized delivery vehicles for too long).

  Traditional routing problems usually minimize total time or distance traveled by
delivery vehicles. However, since e-grocers compensate delivery drivers by number of
shifts (or time blocks), their objective function is to minimize total number of routes.
Similar to VRPs, literature has proposed exact, heuristic and metaheuristic algorithms to
solve CVRPTWs. El-Sherbeny (2008) lists different methods available under each type.
Exact methods are known to perform poorly and, even with the current computing power,
can take hours or days to find the optimal solutions for relatively small problems. E-
grocers, like most companies, usually adopt heuristic methods to solve CVRPTWs.

  A typical heuristic algorithm can be divided in two steps: route-building and
route-improving heuristics. During route building, a set of routes are initialized based on
the customer addresses and constraints at hand. Solomon (1987) proposes a way to
initialize routes by subsequently inserting the customer nearest to the warehouse to a given
route, until a constraint (such as capacity or delivery window) is reached. At that time, a
new route would be initialized. A variation of this method would be to randomly add a
customer to a route and add subsequent customers based on how close they are to the last
added customer.

  After all customers have been added to a route using the route-building heuristics,
the next step of the optimization algorithm is to refine all routes through route-improving
heuristics. A key concept of these heuristics is the notion of a solution neighborhood. As
described by El-Sherbeny (2008), the neighborhood of a solution \( S \) is a set \( N(S) \) of solutions
that can be generated with one modification of \( S \). Route-improving algorithms are based
on assessing whether there is a new solution in \( N(S) \) that performs better than \( S \) and then
adopt this new solution in the next iteration. When no better solution can be found, an
optimum (which can be local or global) is reached. There are multiple ways that $S$ can be modified to determine $N(S)$. These modifications, also known as neighborhood-operator, could involve moving one customer from one route to another, exchanging customers between two routes, changing a segment from one route to a segment from another etc.

The algorithms used by e-grocers are based on the heuristic method described above. During the route-improving step, the algorithm checks for the neighbor solution in every iteration and look for the one with the lowest number of routes. In the last mile optimization, e-grocers monitor the performance of their algorithms by tracking the average number of deliveries made by each route (called in this document as DPR).

### 2.3.2.4 Last Mile Cost Study

One of the assumptions behind last mile is that as more orders are routed together (i.e., the lower the number of batches), the “Number of Deliveries per Route” (DPR) increases, reducing total number of routes. As mentioned previously, last mile drivers are compensated according to the number of routes that they run. Therefore, a higher DPR results in lower last mile costs.

In this section, we will determine how the DPR metric changes under different constraints and scenarios, by studying the heuristic routing optimization algorithm commonly used by e-grocers to solve the CVRPTW. In the first experiment, we will assess how vehicle capacity and window duration impact the DPR metric. In the second experiment, we will assess how DPR varies with total number of orders routed together. The second experiment will help determine the impact that batching has on last mile costs.

**Experiment Set-Up**

Certain parameters and assumptions were considered for the last mile experiments detailed in this section. In order to run the experiments, random grocery orders were generated by sampling from a set of addresses from a region covered by a specific e-grocer. A uniform sampling method was used, with each address being given the same drawing probability. In addition, repetition was not allowed (if an address was drawn, it would not be drawn a second time for the same order stream).
An additional parameter assumed for each order was its number of bags. In order to determine an adequate probability function to draw this number from, a histogram was created with real data observed across the orders of the e-grocer whose algorithm is studied in this document. The resulting histogram was approximated by a lognormal probability distribution with location 1.213 and scale 0.4598 (Figure 12).

![Histogram of Orders per Number of Bags](image)

*Figure 12 – Histogram of Orders per Number of Bags (Based on Real Data)*

Since heuristic algorithms can run for an indefinite amount of time, we gave each optimization 20 minutes to finish running. In other words, whatever solution was achieved at the end of 20 minutes was considered as optimal and was used to calculate the DPR. In addition, we considered that each delivery stop would take six minutes to account for the time that delivery drivers take to park their vehicles and hand the bags to the final customer.

**Experiment 1: Vehicle Capacity (Cv) and Window Duration Sensitivity**

In this first experiment, we varied vehicle capacity (Cv, defined as the number of bags that can be physically fitted into a vehicle) and delivery window duration constraints, while fixing the number of orders simulated in each run at 300. 20 replicates were made for each experiment (e.g., for each pair of capacity Cv and delivery window, 20 sets of 300 orders were randomly generated, simulated and averaged).

Figure 13 shows the results obtained with this experiment. Each point is the average of the 20 DPR’s from each pair of Cv and delivery window. As previously
mentioned, DPR (or “Deliveries per Route”) is the number of stops that each vehicle makes during its route in order to deliver one order, which may consist of one or multiple bags (depending on the number drawn from the probability distribution shown in Figure 12, when the set of simulated orders was generated). The three delivery windows simulated (2, 3 and 4 hours) are represented by different colors, while the different vehicle capacities are listed on the x-axis.

As seen in Figure 13, DPR initially increases linearly with Cv for all delivery windows when Cv is lower than 20. At Cv = 20, DPR for the 2-hour delivery window starts showing a concave behavior and saturates at ~5. However, DPR continues to increase linearly for other delivery windows until reaching their own saturating DPRs. This analysis shows that there is an optimal capacity for a given delivery window and vice versa. As an example, if an e-grocer offers a 2-hour delivery window, delivery vehicles should be able to fit 20–25 bags in order to achieve the saturating DPR. Vehicles with lower capacities would run out of bags before reaching the 2-hour limit chosen by customers, while vehicles with higher capacities would not be able to deliver all bags in two hours.

![Vehicle Capacity Sensitivity](image)

**Figure 13 – DPR Sensitivity by Vehicle Capacity and Delivery Window**

The analysis also shows that the highest theoretical DPR depends on the vehicle capacity and delivery window pair. In other words, if an e-grocer would like to increase last mile efficiency, it needs to simultaneously increase vehicle capacity (defined as the number of bags that a vehicle is capable to transport) and duration of the delivery window. As an example, if e-grocers incentivize customers to choose a delivery window of four
hours, it should allocate SUVs or mini-trucks (capable of fitting more than 25 grocery bags) to deliver those orders in order to achieve the maximum DPR of $\approx 14$ (or 3.5 deliveries per hour, dividing 14 by the route duration of four hours). Conversely, if the e-grocer decides to offer only 2-hour delivery windows, it can adopt a delivery fleet with smaller vehicles (with smaller capacity) to achieve the maximum DPR of $\approx 5$ (or 2.5 deliveries per hour, dividing 5 by the route duration of two hours).

While a wider delivery window can improve the theoretical delivery efficiency (a 40% increase in deliveries per hour, if the window is increased from 2 to 4 hours), it negatively impacts the customer experience for attended deliveries, in which customers need to wait to receive the bags in person. A potential solution to mitigate this impact is to offer delivery bins or lockers to customers in order to incentivize unattended deliveries, for which the wider delivery windows do not have a big impact on customer experience.

- **Experiment 2: Order Volume Sensitivity**

In this experiment, vehicle capacity and delivery windows are fixed at 24 bags and 2 hours, respectively. These numbers were chosen because the average sedan vehicle currently used for last mile delivery fits approximately 24 bags and, as seen in experiment 1, this capacity is compatible with the 2-hour delivery window. The number of orders for each simulation are varied uniformly from 5 to 400, in increments of 5. The ceiling of 400 was chosen because it is compatible with the maximum number of orders seen by the e-grocer’s delivery node during a peak delivery window and week day. The floor of 5 was chosen to capture the performance of the algorithm at minimum order levels.

150 different simulations are conducted for four different delivery nodes, whose service regions were defined based on the e-grocer’s current operations. Figure 14 shows the results for each simulation and indicates that DPR increases with order volume, until a threshold of 50–100 orders is reached (depending on delivery node), when DPR asymptotically approaches a saturating value. This analysis shows that after warehouses reach a certain scale, routing sub-sets of orders separately (which is equivalent to increasing number of batches for routing, since both actions result in lower number of orders routed together) has little impact on DPR and, therefore, on delivery costs.
The difference observed in the saturating DPR values and the order volume thresholds across delivery nodes can be explained by their disparity in terms of customer density and the coverage area. A node with higher customer density (for example, a highly populated downtown area) tends to show higher asymptotic DPR and a lower order volume threshold. The outliers observed in the upper left quartile in Figure 14 can be attributed to simulations in which customers happened to be located close enough to each other (and potentially with a low enough number of bags per order), leading to an unusually higher DPR value. However, as the number of orders increases, these favorable conditions become more and more unusual, and the DPRs tend to converge to an average number.

![Order Volume Sensitivity](image)

**Figure 14 — DPR Sensitivity (Orders Volume)**

From Figure 14, the \( DPR(n_{\text{orders}}) \) curves can be approximated by the asymptotic exponential function shown in (7). The function has parameters \( b_0, b_1 \) and \( b_3 \), which are geography—specific and dependent on the delivery node analyzed.

\[
DPR(n_{\text{orders}}) = b_0 - b_1 \cdot e^{-b_3 \cdot n_{\text{orders}}} \tag{7}
\]

With the relationship shown in equation (7), we can now estimate last mile cost per delivery for a given number of batches \( n \), similar to what was done for middle mile. Assuming cost of each route \( C_R \), warehouse daily capacity \( K \) and demand factor \( \sigma_{i,d,k} \)
(fraction of the capacity $K$ that is seen at node $i$, week day $d$ and batch $k$), last mile cost per delivery can be estimated using equation (8).

$$LM\ Cost\ per\ Delivery\ (n) = \frac{\sum_{i=1}^{\#\ nodes} \sum_{d=1}^{7} \sum_{k=1}^{B} K_i \cdot \sigma_i \cdot \sigma_{d} \cdot \sigma_{k} \cdot \frac{K}{DPR(K, \sigma_i \cdot \sigma_{d} \cdot \sigma_{k})} \cdot C_R}{\sum_{i=1}^{\#\ nodes} \sum_{d=1}^{7} \sum_{k=1}^{B} K_i \cdot \sigma_i \cdot \sigma_{d} \cdot \sigma_{k}} \tag{8}$$

By applying equation (8) to different scenarios of capacity $K$ and batches $n$, we can plot the charts seen in Figure 15. We identify a similar behavior to what was seen in the chart for middle mile costs (Figure 11). LM CPD is significantly higher in scenarios with higher number of batches and low daily capacity, but stabilizes at around the same value when the capacity increases. In the case of last mile, this happens because the solutions calculated by the optimization algorithm at lower capacities (and, therefore at lower order volume) are less efficient than solutions calculated at higher capacities (and higher order volume). The inefficiency is driven by the more constrained set of delivery addresses, which lowers the customer density across the delivery area covered by the e-grocer.

![Last-Mile Delivery Costs by Volume](image)

*Figure 15 – Last–Mile Delivery Costs*

### 2.3.3 Total Cost per Delivery (CPD)

After studying middle and last mile delivery cost drivers, we can combine both results to determine how total CPD varies with the number of batches $n$. This can be done by essentially adding expressions (6) and (8) (which is equivalent to adding the curves seen in Figure 11 and Figure 15) and turning the x–axis variable into number of batches $n$, instead of having them represented by the different lines.
Figure 16 shows the resulting analysis, with $n$ on the x-axis and normalized total cost per delivery (CPD) on the y-axis. Each line represents a different capacity expressed in daily bags volume. We see that for lower capacities the slope of the CPD curve is higher, meaning that delivery costs in smaller warehouses are more penalized when there is an increase in number of batches. In larger warehouses, CPD barely changes when number of batches increases.

These results can now be combined with the fixed costs per order (FCPO) analysis in order to determine an optimal number of batches for a warehouse with daily capacity $K$.

### 2.3.4 Optimal Number of Batches

The results obtained from the scalability and delivery costs studies can be combined to determine the number of batches $n$ that optimizes total cost. By combining the curves from Figure 16 and Figure 9, we can plot the total cost curves seen in Figure 17, where each curve represents a different daily capacity (expressed in number of bags per day). The x-axis is the number of batches and the y-axis is the total fixed costs and delivery costs per order under each scenario.

This analysis shows that warehouses should be designed to support an optimal number of batches in the long term, when operating at full capacity. This number is dependent on the design capacity of the warehouses and increases as they become larger.
However, the layout design should be flexible to support a lower number of batches during the ramp-up, when order volume has not yet attained the levels to offset higher middle and last mile costs. For very large warehouses, operations could decrease the batch sizes (or increase \( n \)) to essentially operate at a constant flow model, leveling the production across the entire building.

\[
\text{Fixed Cost per Order} + \text{Costs per Delivery by Number of Batches and Capacity}
\]

This analysis assumed a delivery topology with four nodes between the warehouse and the final customers. If an e-grocer is able to successfully locate its warehouse close to the city, so that drivers would be able to reach all customers within the promised delivery window, the delivery nodes would no longer be necessary. In this scenario, additional batches would become even more desirable as the minima of the curves from Figure 17 would be shifted to the right.

### 2.3.5 Potential Impact on Customer Experience and Variable Costs

In addition to the impact on fixed and delivery costs per order, modifying the number of batches also leads to changes in customer experience and variable costs. In the customer experience side, customers might have additional time to place their orders. In the variable costs side, pick rates could be reduced due to congestion, while sorting and loading efficiency could be increased with the reduction of outbound area.
As mentioned in section 2.1, when a low number of batches is adopted, orders from multiple delivery windows are batched together. As an example, orders with delivery window 7–9 pm might belong to the same batch as orders from delivery window 9–11 pm. Since operations timings are dependent on the batch (and not on the window), orders from both of these delivery windows will end up having the same cut-off time (for example, 12pm). In other words, customers choosing delivery windows 7–9pm or 9–11pm would both have to place their orders by 12pm, even though, theoretically, the second delivery window could have two extra hours to shop. If number of batches is increased, each delivery window could be assigned to its own batch and customers choosing the later window would have more time to shop. In this model, the click to delivery time is reduced, leading to an improvement of overall customer experience.

A potential negative impact of increasing the number of batches, while maintaining the number of batches picked ahead at two (which is the case for most e-grocers’ warehouses), is the impact on pick rate. Since the number of batches picked ahead is kept at two, the time available for picking can be defined as the time elapsed between the shipment of batch $n$ and the critical pulling time of batch $n + 2$ (see Figure 3 for outbound timelines). When the warehouses operate at a high number of batches, the shortest time available for picking is lower than when the warehouses operate with a low number of batches, since the batches become more frequent.

After analyzing the operations of an e-grocer warehouse currently operating with four batches, we identified that the shortest time available for picking would decrease from eight to four hours, if the number of batches is doubled to eight (the optimal number according to the previous analysis). One might think that the shorter window would double the number of carts in pick paths, leading to a significant increase in congestion. However, a better metric to look at when assessing this impact is the “number of carts/hour” (dividing the volume of each batch by the hours available for picking). After analyzing this metric, we observed a 16% increase in number of carts per hour.

The second potential impact on variable costs is related to the distances walked by workers during outbound. As seen previously, the buffer and sorting areas are significantly reduced when number of batches are increased. Since many of the activities in outbound involve walking and are still manually executed (e.g., sorting bags into route racks, loading
route racks into middle mile trucks), a reduction in the WIP areas can lead to more efficient operations. This benefit is hard to quantify without piloting or simulating the recommended operations. In section III, we will use a simulation software to quantify the reduction in variable costs of a warehouse operating with higher number of batches.

2.4 Layout and Automation Improvements

After studying how policy changes, such as optimizing the number of batches, could improve utilization in e-grocers’ warehouses by reducing and potentially eliminating the WIP areas shown in Figure 2, we will now shift focus to study improvements in layout and automation solutions that could reduce outbound variable costs.

Many of these improvements would only be possible after the number of batches is optimized. Only then, for example, the area required for operations would be small enough to support a more compact layout. Similarly, only with the improved constant flow enabled by more batches, the warehouse would be able to increase utilization of automation assets and make them more economically viable.

2.4.1 Layout Changes

A hypothetical layout for current operations of an e-grocer can be found in Figure 18. Two key issues can be seen in this layout. The first one is the large distance between the sorting area and the loading docks. After sorting is completed, workers need to walk back to the docks, while manually pushing the route racks into the middle mile delivery trucks. This creates a spike in labor required every time that trucks need to be dispatched.

The additional walking distance can be eliminated if the sorting area is placed in the space currently occupied by the buffer. Bags could have their routes assigned during shipment label application and be taken directly from taping to sorting, bypassing the buffer area. An increase in number of batches could allow this to happen by enabling routing optimization to be run at customer cut-off (when the e-grocer already has all orders for a given batch) and shortly after picking starts.
A second issue with the layout in Figure 18 is related to the shipment label application stations. Picking carts accumulate next to the stations, cluttering the area and making it hard for workers to retrieve bags and initiate the outbound processes.

Figure 19 shows a proposed layout, after eliminating the buffer area and moving the sorting operations closer to the loading docks. Loading efficiency increases with this layout, since the distance walked by workers to load route racks into trucks is much shorter. In addition, the overall sorting area decreases with the higher number of batches, reducing distances walked between conveyors and route racks.

The proposed layout also includes feeding conveyors before taping stations. With this change, picking workers empty their pick carts by transferring bags to small sections of conveyors placed before taping. Taping workers are then able to pack and tape bags as they arrive, without needing to handle carts. The feeding conveyors must be correctly sized to support accumulation and ensure no overflow.
Determining the optimal size of the feeding conveyors can be treated as a classic queueing theory problem that depends on the parameters of the Poisson distribution assumed for arrival and service rates. Considering that bags inter-arrival times and taping station cycle times follow a general probability distribution with mean $\lambda$ and variance $\gamma^2$, this process can be modeled as a $G/G/N$ queue. The length of the queue observed when utilization $\rho$ is at its peak (e.g., Monday mornings in the case of e-grocery) can then be used to calculate the required length for each of the feeding conveyors.

Although there are no simple and analytical formulas to calculate the length of a $G/G/N$ queue, equation (9) can be used as an approximation. $N$ represents the number of servers (in this case, taping stations), $C_A$ and $C_S$ are the coefficient of variation for inter-arrival and service times, respectively.

$$L = \rho^{\frac{1}{2}(N+1)} \frac{C_A^2 + C_S^2}{2}$$  \hspace{1cm} (9)
In the context of the conveyor sizing problem, $C_A$ can be determined by performing time studies to measure the time when each bag arrives for taping. These timings can then be used to determine the probability distribution function for inter-arrival times and $C_A$ can be directly calculated by dividing the standard deviation of the underlying distribution function by its mean. $C_S$ can be estimated in a very similar way, by performing time studies on the cycle time of the taping operation, instead.

After analyzing potential layout changes for e-grocers’ outbound operations, we will now propose two potential automation solutions that could be explored to further reduce variable costs.

2.4.2 Automation

As mentioned in section 1.3, automation should always be pursued with caution by e-grocers, since high demand variability usually leads to low asset utilization, hurting the profitability of highly automated warehouses. However, process automation is still a powerful way to reduce variable costs and increase the razor thin margins in the e-grocery business. The constant flow of bags enabled by a higher number of batches and elimination of the buffer area previously described in this document could allow automation of at least two outbound processes: sorting and shipment label application.

- **Sorting**

  As shown in the proposed layout (Figure 19), when bags reach the sorting area, they are pre-sorted according to delivery nodes and batches (since the model assumes that two batches would be picked at the same time, with one batch of pick ahead). In a manual operation, three workers would be stationed in each of the spurs, and would manually divert a bag according to the batch and delivery node designated on its shipment label. Since bags would continuously arrive throughout the day, workers would have to be stationed at all times, driving high labor costs.

  Auto-diverts can be used to automatically pre-sort bags to their appropriate section in the sorting area. Scanners or image recognition software can read the batch and delivery node from the shipment label and an automated system can activate the mechanical sorting mechanism that will divert the bags to the correct location.
There are multiple types of sorters available in the market, such as deflector arms, pushers, pop-up sorters and tilt-trays. Companies must choose which one is more appropriate to their operations by considering multiple factors, including packaging, item diversity and predictability (Rogers 2011). In the case of e-grocers, this is particularly important given their wide range of SKUs, which includes not only bulky and heavy items (such as soda bottles), but also light and fragile ones (such as fruits and vegetables).

The fragility of SKUs and the instability of grocery bags when a round item is placed on the bottom are two limiting factors when choosing automated sorting solutions for e-grocers. Deflector arms and pushers, for example, could easily damage fragile products and destabilize bags. A type of sorter identified in the industry that could overcome this challenge is based on an Activated Roller Belt (ARB) system (Figure 20).

**Figure 20 – Illustration of an Auto-Sorter**

An ARB system consists of a conveyor section with rollers extending through the thickness of the belt. These rollers can be selectively activated and deactivated by lowering the conveyor belt onto and raising the conveyor belt above the roller bearing surfaces (Fourney, Matthew 2012). When the rollers are activated, the items being conveyed can be diverted to a direction perpendicular to the regular flow. This system is particularly interesting for e-grocers because it does not involve pushing the grocery bags' faces and, therefore, exerting additional forces on them. Instead, it leverages the bag surface already
in contact with the conveyor and change the movement direction by simply generating friction forces in a different angle. This method mitigates the risk of destabilizing the bag and damaging fragile products inside it.

- **Shipment Label Application**

  The second potential automation solution consists of automatic shipment label application stations. Auto-labelers are commonly used in the industry and most e-tailers adopt a version of this equipment in their outbound operations. A sketch of one potential design for traditional machines can be seen in Figure 21. As we can see, auto-labelers are usually designed for rigid, corrugated boxes and apply the shipment label by exerting pressure through a steel robotic arm, which essentially stamp the label on top of boxes.

  ![Figure 21 - Sketch of Traditional Auto-Labelers](image)

  Similar to what was observed for the auto-sorters, the fragility of SKUs and the packaging solution adopted by e-grocers make it difficult for them to use the same design as traditional e-tailers. The stamping force of the robotic arm can easily damage grocery products. In addition, with the goal of preserving the brick-and-mortar shopping experience, e-grocers adopt packaging that resembles the brown paper grocery bag, which can’t have shipment labels applied on top.

  One way to overcome this challenge is to adjust the design so that labels are applied to the lead end of the bag using a swing-tamp label applicator. As shown in the sketch in Figure 22 (extracted from ID Technology’s 252N model product sheet), this design consists of a swing robotic arm that rotates to a 90 degrees angle after the bag has been scanned and shipment label has been printed. The robotic arm (also called tamp pad) that makes direct contact with the bag during application can be made of a flexible material to ensure that products will not be damaged and that bags will not be destabilized.
Figure 22 – Sketch of Swing Tamp Label Applicator (ID Technology’s 252N model)

Swing-tamp label applicators available in the market are able to match the manual process cycle times, achieving a rate of approximately 12 bags per minute (or a cycle time of ~5s).
SECTION III: SIMULATING RECOMMENDED OUTBOUND OPERATIONS

In order to test solutions investigated in section II and quantify potential savings, we can simulate operations under the suggested configuration, while assuming random behavior for arrival rates and process cycle times. Simulation can also be leveraged to ensure that conveyor sections (such as the feeding conveyors introduced before taping stations) are properly sized and will not overflow when the warehouse is operating at peak capacity. In this section, we provide an overview of discrete event simulation, detail the simulation models built for e-grocers’ outbound operations and discuss their results.

3.1 Discrete Event Simulation Overview

In a discrete event simulation, the system of interest changes value or state when a certain event occurs at discrete points in time, rather than continuously. During the time span from one event to another, no change in the system is assumed to occur. This is different from continuous simulation, in which the system is continuously tracked at every instant in time, regardless of an event occurring or not. Discrete-event systems are frequently used to model manufacturing plants, inventory systems, distribution systems, communications networks, transportation networks, health-care delivery systems, as well as many other environments measuring their performance in terms of delay, number of units waiting, throughput, and resource utilization (Fishman 2001).

Discrete event simulation is a powerful tool to analyze and study a given system from both a theoretical and empirical standpoint. Literature has provided an extensive list of purposes that a simulation model serves. As listed by Fishman (2001), for example, discrete—event simulation:

- Enables an investigator to organize her/his theoretical beliefs and empirical observations about a system and to deduce the logical implications of this organization
- Provides a framework for testing the desirability of system modifications
- Is easier to manipulate than the physical system
- Expedites the speed with which an analysis can be accomplished
- Leads to improved system understanding
- Brings into perspective the need for detail and relevance
• Permits control over more sources of variation than direct study of a system allows
• Is generally less costly than direct study of the system

Given the versatility of simulation and the increase of computer computing power of recent years, multiple software providers have emerged allowing individuals and organizations to easily model their systems. From a computer language standpoint, most simulation platforms support dynamical data structures, generation of pseudorandom numbers and a variety of probability distributions, conventional arithmetic calculations, 2-D or 3-D animations and CAD (computer-assisted design) drawings. Most of the platforms also allow users to insert codes in other computer languages, such as Visual Basic, C and C++.

Open source software includes java-based platforms (ESMO-J, Ptolemy II), C++ based platforms (PowerDEVS, SystemC) as well as platforms based on Python (SimPy). These platforms usually have limited capabilities to graphically display the simulation through either 2-D or 3-D animations. Their development user interfaces are also less intuitive and user-friendly when compared to most commercial platforms. Despite these limitations, they provide the core functionalities required for discrete-time simulation and are a good alternative in case a commercial platform is not available.

Commercial software includes Rockwell’s Arena, Matlab’s SimEvents, AnyLogic, SIMUL8 and FlexSim, among many others. Most of the commercial platforms provide similar functionalities, with some focusing on a specific industry or providing better visualization capabilities. In the following sections, FlexSim is leveraged to model both the status-quo and the proposed operations for a typical e-grocers’ warehouse. FlexSim provides superior 3-D visualization capabilities and supports a wide range of material handling equipment (MHE) in its simulations, including conveyors, automatic guided vehicles (AGVs), automated storage and retrieval systems (ASRS) and human operators. It is also able to support the probability distribution functions assumed for the arrival rates and process cycle times.

3.2 Overview of Simulation Model

In order to assess potential improvements enabled by initiatives described in Section II, two simulation models can be built. The first one replicates operations using the layout seen in Figure 18 and keeps the current number of batches (Base Case). The second model changes the layout and operations to align with what is shown in Figure 19 and also doubles the number of batches from four to eight (Proposed State). As mentioned previously, the hypothesis is that
the more compact layout will improve variable costs by reducing walking distances. The base case model is built to allow for a relative comparison with the proposed state, ensuring that any simplifications inherent to the simulation methodology is cancelled out.

As listed in Table 2, an exponential distribution function is assumed for the inter-arrival times of grocery bags, with the value varying with the day and time being simulated. The inter-arrival times are directly related to demand for a given delivery window and when its bags are being processed. In peak times, the inter-arrival time is lower (i.e., bags arrive more frequently).

When the process simulated shows a high variation, a probability distribution function is used to model its cycle time. For the e-grocer simulation, lognormal distribution is assumed for the shipment label application and packing cycle times. Lognormal distribution shows the best fit with real data collected with time studies. In addition, a random variable which is log-normally distributed takes only positive real values, which is desirable when modelling cycle times, whose values are never negative. If a normal distribution is used instead, adjustments need to be made to eliminate values corresponding to the normal distribution’s left tail.

<table>
<thead>
<tr>
<th>Demand / Arrival Rates</th>
<th>Probability Distribution</th>
<th>Average Value per Bag</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logistic</td>
<td>Inter-arrival time varying depending on day and time</td>
<td>Store pickup in 1 hr window</td>
<td></td>
</tr>
<tr>
<td>Shipments Label Application</td>
<td>Lognormal</td>
<td>~12 seconds</td>
<td>~1.3 seconds</td>
</tr>
<tr>
<td>Packing</td>
<td>Lognormal</td>
<td>~15 seconds</td>
<td>~ 8 seconds</td>
</tr>
<tr>
<td>Route Label Application</td>
<td>Constant</td>
<td>~6 seconds</td>
<td>N/A</td>
</tr>
<tr>
<td>Sorting</td>
<td>N/A</td>
<td>Dependent on Walking Distance</td>
<td>N/A</td>
</tr>
<tr>
<td>Auditing &amp; Vehicle Loading</td>
<td>Constant</td>
<td>~1.25 seconds</td>
<td>N/A</td>
</tr>
</tbody>
</table>

Table 2 – Example of Assumptions for Simulation Arrival Rates and Cycle Times

The sorting time is highly dependent on the walking distances between the conveyor and the bags’ route racks. Simulation software like FlexSim allows to simulate material
handling by either workers or conveyors, which is helpful when determining cycle times involving bag transportation or bag flow. Table 3 lists additional assumptions from the discrete-event simulation model, such as walking and conveyor speed, which allow to control the underlying factors impacting worker walking and waiting times.

<table>
<thead>
<tr>
<th>Assumption</th>
<th>Value</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Workers Walking Speed (No Load)</td>
<td>1.50</td>
<td>m/s</td>
</tr>
<tr>
<td>Workers Walking Speed (Loaded)</td>
<td>1.00</td>
<td>m/s</td>
</tr>
<tr>
<td>Workers Acceleration / Deceleration</td>
<td>1.00</td>
<td>m/s²</td>
</tr>
<tr>
<td>Conveyor Speed (Packing to Buffer Area)</td>
<td>1.00</td>
<td>m/s</td>
</tr>
<tr>
<td>Conveyor Speed (Buffer to Sorting Area)</td>
<td>2.20</td>
<td>m/s</td>
</tr>
</tbody>
</table>

*Table 3 – Additional Simulation Assumptions*

1. Shipment Label Application and Packing
2. Buffer Induct
3. Route Label Application
4. Sorting

*Figure 23 – Screenshots of Base Case Simulation*
Screenshots of the base case simulation model is shown in Figure 23. Workers apply shipment label and pack bags (1), which are then conveyed to the buffer area and inducted into racks (2). After routing, workers remove bags from buffer racks, apply route labels and induct them to a second conveyor (3). Bags are then conveyed to the sorting area (4) to be placed to their appropriate route racks by the sorting team.

![1. Packing](image1)

![2. Shipment Label Application](image2)

![3. Sorting](image3)

*Figure 24 – Screenshots of Proposed State Simulation*

Figure 24 shows the simulation screenshots of the proposed state. The packing stations now include feeding conveyors and workers no longer need to handle pick carts (1). Bags are then conveyed to an automated shipment label application station (2), as described in section 2.4.2.

A manual station can be added in parallel to the automated one, ensuring enough capacity when the warehouse operates at peak demand, while not hurting asset utilization. If there is enough accumulation in the conveyor prior to shipment label application, a worker can be staffed at the manual station, alleviating the bottleneck. Bags are then conveyed to the sorting area and automatically diverted to the section corresponding to
their delivery window and node (3). As bags arrive, sorting workers take them to their assigned route rack.

### 3.3 Simulation Results

After simulating the base case and proposed state operations for one week and comparing the performance of both models during that time frame, we noticed that the proposed state led to approximately 34% decrease in area required for outbound operations, driven by the reduction of WIP inventory. Figure 25 shows the WIP build-up for both the base case and proposed state after simulating operations for seven days. The proposed state benefits from a higher number of batches, increased to eight from four in the base case. This leads to a reduction in the maximum number of bags that operations need to hold (as seen in Figure 25). In addition, the flow in the warehouse is smoothened, as the higher number of cycles (or peaks) leads to a flatter WIP inventory profile.

If the released space is used to store additional products in the warehouse, the capacity of an entire building can increase by approximately 14%, reducing total fixed costs per order as the utilization of assets is improved (and assuming that demand scales with capacity, as stated in previous chapters). With this increase in capacity, e-grocers can potentially postpone investments in new buildings, preserving their cash flow and improving profitability.

![WIP Build-Up Comparison](image)

*Figure 25 – WIP Build-Up Comparison*

The proposed state resulted in ~47% reduction in variable costs per order, enabled by shorter distances walked by sorters and loaders, improvements on packing and shipment label application stations, automation and elimination of buffer induct and route labelling processes.
Table 4 shows the relative variable costs reduction for each outbound process. The 21% reduction in Packing and Shipment Label Application is enabled by the introduction of the feeding conveyors before packing and by the automated shipment label application stations. Costs related to the buffer induct and route label application processes are completely eliminated, since bags go directly to the sorting area in the proposed state.

Finally, the 31% cost reduction in the sorting and loading processes are enabled by the significantly shorter distances that workers need to walk between conveyors and route racks during the sorting process and between the route racks and delivery trucks during the loading process. Even though a higher number of batches requires workers to perform the loading task more frequently, the benefits from the reduced distances offset this penalty.

<table>
<thead>
<tr>
<th>Total Variable Costs per Order</th>
<th>Relative Decrease in Variable Costs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Packing &amp; Shipment Label Application</td>
<td>21%</td>
</tr>
<tr>
<td>Buffer Induct</td>
<td>100%</td>
</tr>
<tr>
<td>Route Label Application</td>
<td>100%</td>
</tr>
<tr>
<td>Sortation and Loading</td>
<td>31%</td>
</tr>
</tbody>
</table>

*Table 4 – Simulated Improvements in Variable Costs*
SECTION IV: CONCLUSIONS & RECOMMENDATIONS

4.1 Summary of Findings

The primary goal of this project was to study and improve e-grocery’s asset utilization, which was identified in the literature as one of the key success factors for the second wave e-grocers that chose to own and operate their own warehouses. The project specifically focused on identifying areas for improvement within outbound processes, which represent the mechanical throughput bottleneck for current operations. The improvements identified went beyond utilization increase and also enabled a reduction in variable costs by changing operations layout and automating manual processes. The project also studied the optimization algorithm used by e-grocers with the goal of assessing the impact of changes in outbound process on delivery costs. Finally, discrete-event simulation was used to assess variable costs benefits of changing the process, test different parameters and correctly size material handling equipment (e.g., conveyors), without needing to disrupt current operations.

The number of batches that e-grocer’s warehouses choose to operate was identified was one of the key drivers for the high area requirement of outbound operations. The fixed costs per order was shown to decrease as a power function of number of batches, a direct consequence of utilization improvement. On the delivery cost side, the number of deliveries per route (DPR) metric outputted by a routing heuristic algorithm was studied under different parameters, including delivery window width, vehicle capacity and order volume. It was shown that DPR saturates as vehicle capacity increases and that there is an optimal capacity for a given delivery window width and vice versa. It was also shown that, contrary to intuition, an increase in order volume has little impact on DPR after a certain volume threshold is exceeded. By combining the findings from the fixed and delivery costs per order analysis, we were able to determine an optimal number of batches for warehouses with different capacities.

After studying the batching policies, we analyzed how modifications in the layout (e.g., elimination of buffer area, reduction of walking distances in sorting and loading processes) and the use of automation can improve variable costs in outbound operations. The discrete—time simulation model was used to simulate the proposed state and indicated a potential reduction of 47% in variable costs when compared to the simulated base case.
4.2 Recommendations and Next Steps

An e-grocer should adopt a more flexible approach when designing outbound operations for their warehouses. When a warehouse starts operating, it most likely won’t have enough volume to support the optimal number of batches corresponding to its full capacity and will have to consolidate more batches in order to lower its delivery costs. However, when demand ramps-up, the recommended number of batches must increase until it reaches an optimal number, as highlighted by this document. The warehouse outbound design should allow for this flexibility. As an example, phasing the design in three different stages (launch, ramp-up and full capacity) could be a good way for e-grocers to anticipate these changes and be ready for the time when the volume thresholds are attained.

In addition, before investing additional capital to expand warehouses facing space constraints or building new buildings to satisfy additional demand, e-grocers should consider increasing number of batches in their current network in order to improve flow and alleviate bottlenecks. By adopting lean methodology concepts in their operations and tackling wastes such as excess inventory, over-processing, waiting and motion, e-grocers have the opportunity to increase capacity without additional investments in infrastructure.

As different process changes and parameters are considered, discrete-time simulation software can be utilized to quantify improvement and test feasibility without disrupting current operations. It can be a powerful tool to filter out and select configurations to be piloted in a real warehouse, significantly reducing implementation costs.

While they were not the focus of this project, retail levers could also be adopted to reduce intra-day/week demand variation and further increase asset utilization. By incentivizing customers to place orders with unattended delivery, in less popular days and delivery windows, e-grocers can further level load at warehouses, increase asset utilization and reduce delivery costs.
REFERENCES


