LANCASTER UNIVERSITY

Lakeland: Optimisation of E-Commerce Packaging Options

Tejas Khandale

36141687

MSc Business Analytics

Supervisor

Prof. Jamie Fairbrother

08 September 2023

This dissertation is submitted to the Department of Management Science and the Board of Examiners of Lancaster University in partial fulfilment of the requirements for the degree of Master of Science in Business Analytics. I hereby attest that this dissertation is entirely my own work and that all sources of information have been fully referenced.

Abstract

The optimisation of e-commerce packaging requires careful consideration and integration of factors such as operational efficiency, financial expenses, and environmental sustainability. The objective of this dissertation is to enhance the process of selecting shipping boxes for online orders at Lakeland through optimisation techniques. An optimisation model is formulated and trained using the order data from Lakeland in the year 2022. The model proposes an optimised range of box dimensions with the objective of increasing packing density, minimising box volume or minimising cardboard usage, and aligning with the objectives of the sustainable economy.

The optimisation uses a sampling methodology to train on subsets of orders, afterwards evaluating the ability to generalise the model on the complete dataset. The underlying process comprises a list of potential boxes, the determination of dimensionally appropriate options for each order, and the optimisation of box selection to either enhance packing efficiency or minimise total volume. The model is assessed in various scenarios with a range of 12 to 19 boxes.

The key findings indicate that there is a notable enhancement in packing efficiency, ranging from 8% to 13% when compared to the volume-based heuristic employed by Lakeland. The modified layouts of the boxes resulted in a decrease in average box volumes ranging from 7% to 8%. The analysis conducted indicated that there were diminishing marginal returns observed when the number of uniquely sized boxes exceeded 19. The most optimal approach involves utilising Lakeland's existing boxes in conjunction with two newly designed additions, resulting in a notable enhancement of 19.5% in packing efficiency from current packing efficiency.

In summary, the model offers a data-centric methodology for identifying additional boxes that enhance packing efficiency when integrated with existing inventory. The enhanced efficiency and decreased utilisation of resources are in accordance with the objectives of sustainability. The limitations of the current approach include the dependence on rigid product orientations and disregard for projected demand patterns. Future research should focus on these areas and additionally consider the inclusion of procurement and transportation cost considerations.

Acknowledgements

Engaging in the process of conducting this dissertation has proven to be a very beneficial and enlightening experience for me. The gaining of new skills and information during this process was made possible due to the crucial assistance provided by the faculty of Lancaster University Management School.

I would want to extend my utmost appreciation to my supervisor, Professor Jamie Fairbrother. He has played a crucial role in shaping this experiment with his extensive knowledge and expertise. The continuous encouragement and assistance provided by him have played a significant role in assisting me in defining the extent of my experiment. The invaluable contribution of Professor Jamie's ongoing input has played a crucial role in enhancing the quality of this dissertation. Without his expert assistance and consistent recommendations, the completion of this dissertation would not have been feasible.

I would like to express my gratitude to my family, friends, and colleagues for their constant moral support and invaluable guidance.

Contents

List of Figures

List of Tables

Dedicated this to my parents.

Chapter 1

Introduction

1.1 Overview

The world of e-commerce has recently seen many changes. The rapid changes in conditions can be attributed to a variety of COVID-related lockdown measures, increasing living expenses, and changes in consumer behaviour. However, the adaptability of packaging solutions in the context of e-commerce has been limited. This presents a challenge for companies to modify their packaging to correspond with current consumer preferences, adopt more environmentally friendly practices, and comply with forthcoming legislation.

At Lakeland, there has been a significant transformation in online orders over the course of the last four years. This is likely associated with the aforementioned issues. Given the increasing prevalence of online purchasing and the dynamic nature of the business, it is evident that an overhaul of our e-commerce packaging strategies and procedures is needed. It is a cause for concern that a significant number of box sizes currently employed remain unchanged over the past decade, despite substantial transformations in the business landscape.

There are many important reasons for demanding an overhaul of packaging. In recent years, there has been an observable shift in consumer behaviour, with an increasing emphasis placed on sustainability and the preference for eco-friendly options. Opting for smaller boxes not only yields immediate cost savings but also contributes to the ongoing pursuit of Environmental, Social, and Governance objectives in the long run. Given the imminent adoption of Extended Producer Responsibility laws and the next packaging tender, it is imperative to prioritise the optimisation of our box sizes in order to comply with regulations and adhere to prevailing industry standards.

In the near future, it is likely that the sector will go through significant transformation as a result of the implementation of Extended Producer Responsibility (EPR) rules along with modifications in packaging procurement processes. In order to effectively handle the modifications, it is crucial to optimise the assortment of containers that are accessible for e-commerce shipping operations. The goal is to optimise operational processes in a sustainable manner to accommodate a diverse variety of orders while aligning with our overarching environmental and social goals.

In order to facilitate the execution of this task, Lakeland provided us with previous order data, courier restrictions, and current box information. The datasets will be utilised to construct an optimisation model that proposes an optimal set of boxes. The recommended boxes will ensure an optimal fit for our diverse range of orders, maximising packing efficiency and minimising waste.

The primary aim of this project was to deliver Lakeland with a consolidated selection of 12 box sizes that would facilitate the efficient and economically viable delivery of customer orders. Lakeland aimed to decrease expenses related to cardboard utilisation and transportation while simultaneously ensuring the secure packaging and shipping of their products.

1.2 Brief Structure of the Dissertation

The dissertation is structured over six chapters as follows:

1. **Introduction:** This section provides a comprehensive overview of the challenge at hand and highlights its significance, emphasising the need for its resolution.

2. **Literature Review:** This section provides an explanation of the fundamental concepts and terminology that provide the basis for understanding the subject at hand. Additionally, it looks into the current state-of-the-art research that serves as the foundation for the present study.

3. **Problem Description:** This section presents the current strategy used by Lakeland for packing orders. Alongside this, exploratory data analysis is presented, which is useful for Lakeland for inventory management.

4. **Methodology:** This section provides a comprehensive description of the experimental setup necessary for conducting this experiment. The subsequent sections within this chapter provide a comprehensive approach to the steps performed to achieve the desired results.

5. **Numerical Test:** This section presents the outcomes derived from the models described in the methodology and examines each scenario, deep-diving into the model's capacity to generate an optimal set of boxes.

6. **Conclusion:** This section presents the recommended model and how it addresses the issues raised in the introduction section. It further discusses the future works or scope for improvement in the current model.

1.3 Collaboration

Together with my classmate Abinauv Selvaraj(36366243) from the MSc Business Analytics Programme, I worked on this project. In regard to his contribution to certain aspects of the project, we mutually agreed to work together and jointly write the "Chapter 5: Results" section of this report. The collaboration with Abinauv proved to be very helpful as his valuable abilities and ideas mutually complemented my own efforts. Through collaboration, we successfully produced more robust outcomes that incorporate the combined efforts of both parties involved in the project. Collaborating with a partner such as Abinauv elevated the project's level of involvement and served as a wonderful opportunity for learning. I consider myself fortunate to have had the opportunity to engage with a highly skilled colleague on this important part of the report.

Chapter 2

Literature Review

A critical consideration in optimising packaging selection is determining the feasible box options for fulfilling a particular order. According to (Wäscher, Haußner, and Schumann, 2007), this problem is classified as a 3D container loading problem (CLP). In CLPs, the large packing items are termed containers, while the smaller items to be loaded are designated as boxes. Based on heterogeneity among the boxes, two primary categories - "identically or weakly heterogeneous" CLPs involve many boxes of relatively few types, while "strongly heterogeneous" CLPs feature few boxes across a wide diversity of dimensions or types. For the box packing use case, each unique order represents a distinct CLP, with the objective being the identification of the optimal container to efficiently package the boxes. The heterogeneity in box types and product sizes adds complexity in evaluating viable options.

2.1 Rules and Arrangement of Products

According to (Gehring et al., 1997), to achieve a stable arrangement when packing boxes into a container, these guidelines should be followed:

- Each box should rest on the container floor or be stacked on top of another box such that its center of gravity is supported.

- All boxes should be positioned parallel to the container walls, with no angular packing.
- The entire box should lie completely within the interior dimensions of the container.

These guidelines ensure boxes are packed in a way that minimises shifting and provides maximum stability during shipping and handling. As the arrangement of products in the box is not in the scope of this project, we will not be focusing on the center of gravity.

Also, according to (Gehring et al., 1997), in order to produce feasible solutions for the box packing problem, the arrangement of boxes must adhere to five critical constraints: First, an orientation constraint stipulates that certain subsets of boxes cannot be oriented vertically in one or two dimensions. Instead, these boxes must be placed in predefined restricted positions and orientations. This accounts for cases where box dimensions limit vertical stacking options. Second, a top placement constraint specifies particular boxes that are prohibited from bearing any weight or having other boxes stacked directly atop them. This constraint designates boxes that cannot be positioned on the bottom supporting other items. Third, a weight limit constraint caps the total aggregated weight of all packed cargo. The combined weight of all boxes cannot surpass this maximum threshold. This restricts the overall loading capacity. Fourth, a stability constraint enforces a minimum stability ratio for each individual box. The stability ratio is calculated as the bottom surface area in contact with supporting boxes divided by the total bottom area. Requiring a minimum ratio ensures boxes have adequate support below them and do not tip over or fall. Finally, a balance constraint requires the center of gravity of the cargo load along both x and y dimensions to be within the defined tolerances of the container's midpoint. Keeping the overall center of gravity near the midpoint promotes balanced weight distribution and prevents tipping.

Together these five constraints address the practical considerations of box orientations, weight limits, equilibrium, and stability necessary for generating feasible packing solutions given the physics of the loading environment. A viable box arrangement must satisfy all constraints simultaneously.

2.2 Types of objectives

The 3D container loading problem exhibits two primary objective function formulations. In an input minimisation model, the goal is to pack a given set of rectangular boxes into the minimum number of larger rectangular containers. This aims to fully utilise the volume of a limited container fleet. In contrast, an output maximisation model assumes a fixed set of containers and seeks to pack the highest total value or volume of boxes from a larger box set. The output maximisation formulation applies when there is flexibility in box selection, and the aim is to maximise utilisation of the available containers.(Zhao, Bennell, and Song, 2016)

Based on our project objective of input minimisation, there are six unique problem types:

Single stock-size cutting stock problem (SSSCSP): Containers are identical, and boxes are weakly heterogeneous. Single bin-size bin packing problem (SBSBPP): Containers are identical, and boxes are strongly heterogeneous. Multiple stock-size cutting stock problem (MSSCSP): Containers and boxes are weakly heterogeneous. Multiple bin-size bin packing problem (MBSBPP): Containers are weakly heterogeneous, and boxes are strongly heterogeneous. Residual cutting stock problem (RCSP): Containers are strongly heterogeneous, and boxes are weakly heterogeneous. Residual bin packing problem (RBPP): Both containers and boxes are strongly heterogeneous.

Relying on the context of the work presenting the dissertation, box and product are used in place of container and box, respectively, for ease of communication and understanding with Lakeland, as these are the terminologies used by them.

Depending on the contents in an order, the type of problem would vary for each order. As most of the orders contained 4 or fewer products and all of the box types are strongly heterogeneous, these orders fit into either the MBSBPP or RBPP category.

2.3 Sampling

According to (Sharma, 2017), sampling represents a critical technique when conducting research on large populations. Since studying an entire population is often unfeasible, sampling provides an easier way to examine a subset that accurately reflects the population. Proper sampling leads to the benefits of reduced costs, faster implementation, and convenience. There are several approaches for sampling which are mentioned below:

- Probability Sampling - In this approach, the probability of choosing an element from the population is random.
- Non-Probability Sampling - This approach is based on judgment.

Simple Random Sampling is a technique in probability sampling in which each element has an equal chance of being selected from the population. The randomness introduces no systematic biases, allowing simple random sampling to generate representative data and support robust statistical inferences.(Sharma, 2017) However, the approach is still susceptible to sampling error if the sample size is too small, as the sample may not fully reflect the population characteristics. The smaller the sample, the higher the risk that key traits could diverge from the broader population. While simple random sampling provides an unbiased snapshot, thorough testing is required to ensure the sample accurately represents the population on relevant parameters.

2.4 Stochastic Optimisation

Stochastic programming refers to optimisation problems where some parameters are uncertain and described by probability distributions, in contrast to deterministic optimisation, where all parameters are known.(Kaut and Stein W. Wallace, 2007)

Scenario generation refers to creating a discrete representation of the uncertainty called a scenario tree, where each path through the tree represents a potential realisation of the stochastic parameters (Høyland and Stein W Wallace, 2001). The quality of the randomness impacts the solution, so assessing scenario-generation methods is important.(Kaut and Stein W. Wallace, 2007)

There are several approaches to scenario generation. Sampling methods simply draw random samples to build the scenario tree. As the sample size increases, convergence to the true distribution is guaranteed. (Høyland and Stein W Wallace, 2001) Key criteria for evaluating scenario methods are stability, where solutions are consistent across different scenario trees, and optimality gap, measuring the difference between the true stochastic solution and approximate solution.(Kaut and Stein W. Wallace, 2007)

2.5 Stability

According to (Kaut and Stein W. Wallace, 2007) stability refers to whether a scenario generation method produces consistent solutions across different scenario trees generated for the same stochastic programming problem. There are two main notions of stability:

In-sample stability:

- This tests whether the optimal objective function value reported by the stochastic programming model is similar across different scenario trees.
- It is checked by solving the optimisation model with multiple sampled scenario trees.
- If the optimal objective values vary significantly, it indicates instability.

Out-of-sample stability:

- This tests whether the true objective value is similar across solutions obtained from different scenario trees.
- The true objective value is evaluated by simulating the stochastic solution on a large reference scenario tree.

- Differences in the true objective indicate the solutions themselves are unstable.

Ideally, a scenario generation method should demonstrate both in-sample and out-of-sample stability. In-sample stability ensures the model solution itself is consistent. Out-of-sample stability ensures the objective value in the real stochastic environment is consistent.

Instability can arise from having too few scenarios or a scenario generation method that does not adequately capture the underlying uncertainty. Methods like sampling will tend to improve stability as the number of scenarios increases.

Chapter 3

Problem Description

The purpose of this section is to provide an in-depth analysis of the current process followed by Lakeland as this will form the basis of our optimisation model. The gathering of this knowledge will serve as the basis for the creation of a packaging model that is defined by improved efficiency. This approach will enable us to improve the allocation of boxes while simultaneously improving the important key metrics.

3.1 Current Strategy

Lakeland currently uses a "Volume Fill" heuristic for box assignment. This simple approach selects the smallest box that can fit the order. Specifically, it compares the total order volume to the volume of available boxes. The box with the lowest volume still exceeding the order volume is chosen. While straightforward, this greedy volume-based heuristic has clear limitations. It ignores box dimensions and geometric packing constraints. The optimization aims to develop a more sophisticated packing logic that efficiently utilizes box space and dimensions. This will minimize wasted volume and improve packing density. The optimized model should substantially outperform the current volume fill baseline.

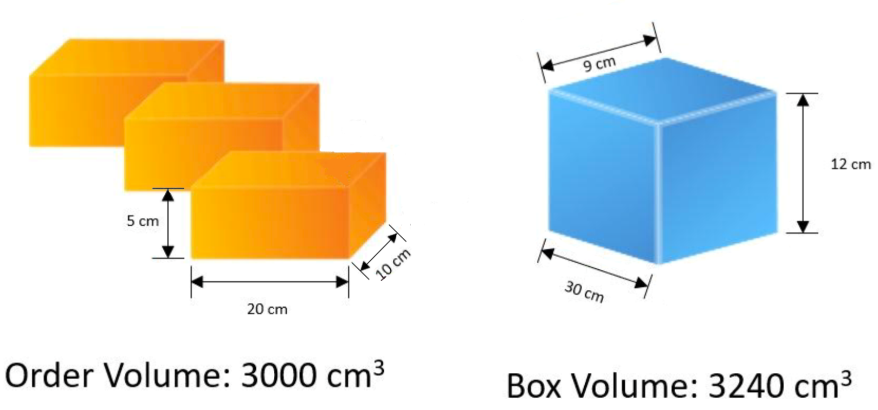

FIGURE 3.1: Failure of Volume Fill Case

Figure 3.1 highlights the shortcomings of the volume fill heuristic; the box volume of 3240 cm^3 exceeds the order's 3000 cm^3. Yet the order cannot actually fit in the box due to dimensional mismatches.

Specifically, the box's length and width are insufficient to accommodate the order's product. So volume alone fails to capture fit. This illustrates the need to consider dimensional constraints during packing optimization. The model must go beyond volume fill to develop logic that pairs orders and boxes based on 3D geometric compatibility.

3.2 Data Description

Lakeland provided two datasets, mentioned below.

3.2.1 Order Data (OD)

The Order Data (OD) dataset encompassed all orders received by Lakeland in 2022 - over 3.5 million rows representing over 1.2 million orders. Each order row contained rich attributes - order date, ID, product ID, dimensions, weight and v_box. This v_box field warrants explanation. It specifies the box recommended by Lakeland's current packing system for that order. Packagers are instructed to use this suggested v_box for fulfilling the order. So the v_box captures the existing heuristic logic for box assignment. The optimisation aims to improve on this baseline. In summary, the OD dataset provides complete order information over a full year. The volume of 3.5 million rows offers ample data to train and evaluate the optimisation model. The attributes enable the development of optimised box recommendation logic, demonstrably improving on the current v_box

3.2.2 Box Data (BD)

The Box Data (BD) dataset covers all packaging used by Lakeland in 2022. It focuses on the 12 standard cardboard boxes compliant with courier restrictions. Lakeland supplements these with some oversized JUMBO boxes and Jiffy bag envelopes as needed.

Importantly, Lakeland employs 2 machines to alter 2 of the 12 standard boxes. By crimping and folding, these "auto-boxes" can match product height and reduce wasted volume. This dynamic sizing increases packing efficiency.

In summary, the BD dataset provides the dimensions of Lakeland's packaging options. The 12 standard boxes are central, with the special adjustable auto-boxes being key for optimisation. Understanding these mechanical capabilities will allow the model to exploit them for tight, efficient packing. Table 3.1 shows the boxes used by Lakeland currently.

3.3 Box Allocation

In this section, we will be performing aggregations and filtering the dataset to get an understanding of the data to identify potential opportunities for improvement and optimisation.

Figure 3.2 illustrates the relative utilization of the suggested box types for all orders in 2022. The data excludes 9544 series of boxes which consist of jiffy bags (envelopes), as these mailers are not included in the list of suggested boxes.

Box Name	Depth	Width	Height
9493	50	42	20
9494	107.5	31.5	13
9495	72.5	49.5	18.2
9496	50.5	41	12.5
9497	60.4	41.4	35
9500	70	32.5	19
9515	27.5	20	10
9594	39	24.5	22
9544B	20	11.5	1
9544C	20	14.5	1
9544D	24	17.5	0.8
9544E	23	20.5	1
9544F	33.5	21	1.6
9546	46	34	3
9547	66	43.5	3
9566	48.2	42.5	35.5
9567	40	25	22
9567H	40	25	11
9567I	40	25	15
9567J	40	25	19
9569	50.8	40	29
9569H	50.8	40	15
9569I	50.8	40	20
9569J	50.8	40	25
JUMBO01	170	140	48
JUMBO02	140	85	48
JUMBO03	85	70	48

Table 3.1: Current Box Table

Products that are fragile or glass are typically not shipped in jiffy bags, with the selection of these boxes left to the discretion of the packers. Excluding the jiffy bag shipments provides a clear picture of how frequently each of the recommended box types is used for typical package shipments.

The box sizes named ‘JUMBO01’, ‘JUMBO02’, ‘JUMBO03’, and ‘TOO HEAVY’ were excluded from further analysis, as they represent orders requiring more than one box. Since orders with multiple boxes are outside the scope of this project, these sizes were removed based on a discussion with Lakeland.

Boxes 9567 and 9569 are referred to as ”auto-boxes” because Lakeland has the

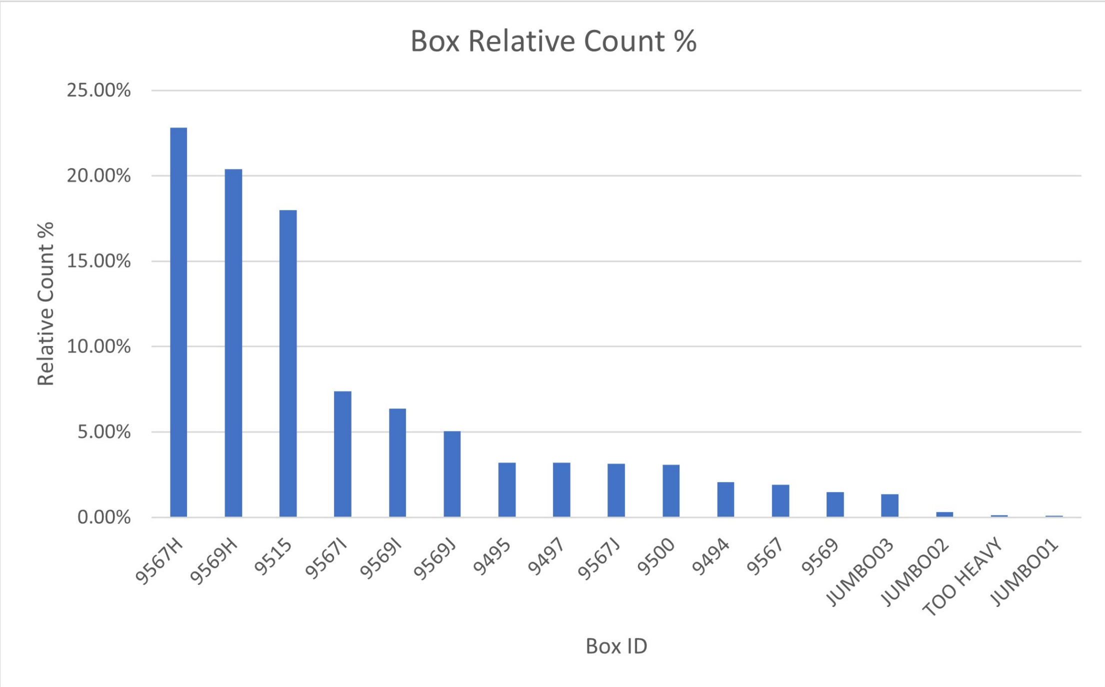

FIGURE 3.2: Box Relative %

ability to adjust their height based on the products being shipped. Using bespoke machines, these boxes can be crimped and folded to match the height of the products inside them. This auto-box capability allows for the reduction of empty space inside the box and an overall decrease in package volume. The automated adjustability makes these unique among the box options.

While the adjustable auto-boxes help minimize empty space and reduce package volume, they do not actually decrease cardboard usage and associated EPR fees. The crimping process simply folds any excess cardboard within the adjusted box, meaning the overall amount of cardboard per package remains unchanged. As such, the auto-box capability provides no savings in cardboard materials or EPR costs, despite the reduced package sizes. The folded cardboard still accompanies each order during shipping. Further analysis is required to quantify potential EPR savings from right-sizing box dimensions prior to adjusting through crimping.

Box 9567 contains 9567H, 9567I and 9567J as a subset of this particular box, with the height being 50%, 68% and 86%, respectively, compared to the 9567 box keeping the other 2 dimensions same. Box 9569 contains 9569H, 9569I and 9569J as a subset of this particular box, with the height being 50%, 68% and 86%, respectively, compared to the 9569 box keeping the other 2 dimensions same. Updated updated relative count percentage is shown in **Figure3.3**.

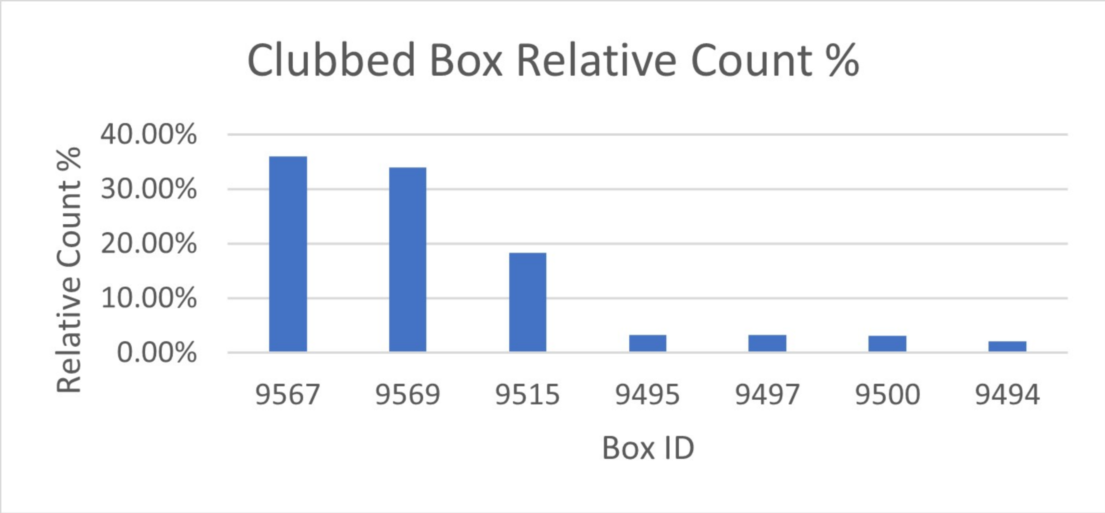

Figure 3.3: Clubbed Box Relative %

3.4 Monthly Order Analysis

Figure3.4 represents the number of orders received per month over a one-year period. Initial observation reveals a clear seasonal pattern in order volume. The highest number of orders occurred in November (181,986) and December (163,240), the probable reason being the winter holidays and people ordering more during winter break. In contrast, the lowest number of orders occurred in June (63,320). The probable reason could be because of the summer cool-down. Order volume peaked again in September at 106,679, as sales picked up heading into the fall. These patterns have important implications for retail businesses as Lakeland has to forecast for ordering the boxes from the manufacturer. This analysis leads to better inventory management of boxes along the workforce allocation.

Figure3.5 provides deeper insight into the usage of various boxes across months. For example, category 9567 had the highest order count in all months, with 128,955 orders placed in January. Categories 9494, 9495, and 9497 also exhibited high order volumes in certain months. Understanding these seasonal patterns by month allows Lakeland to manage inventory accordingly.

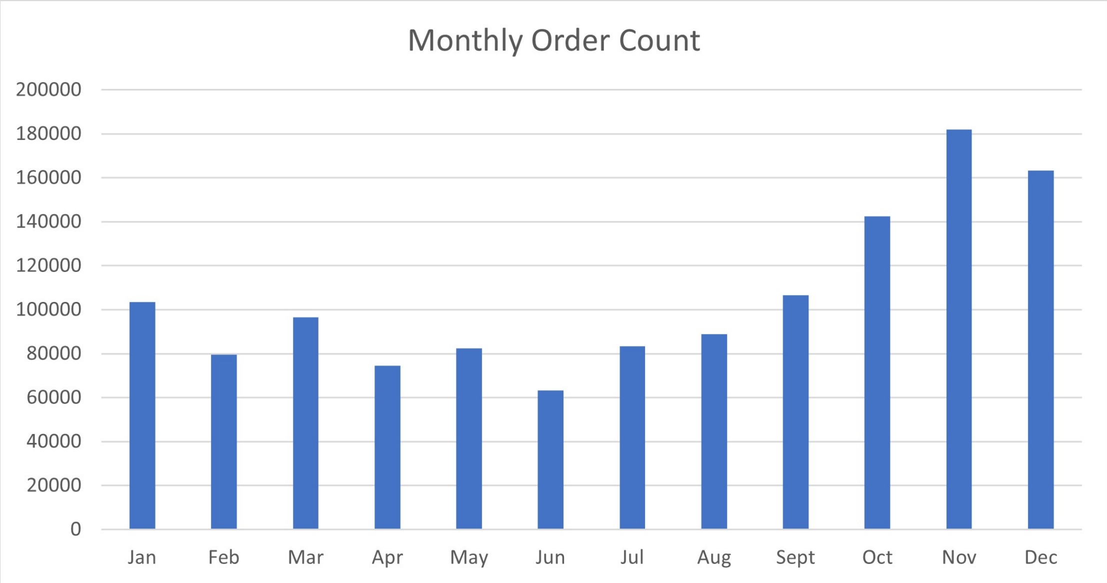

FIGURE 3.4: Monthly Order Count

Figure3.5 provides a detailed depiction of the monthly usage patterns of different box types within Lakeland. The data gathered from this figure would be significant in forming decisions related to inventory management and the allocation of resources. Operations management and sustainability planning could both benefit from this information. One of the key findings that can be derived from this data relates to the seasonal patterns in box utilisation. The patterns seen over the course of several months give important insights into inventory management and resource allocation. As an example, the utilisation of box type 9494 is at its highest level in the month of November, but box type 9495 reaches its peak in December. These data are consistent with the rise in buying during holiday seasons, demonstrating the influence of consumer behaviour on packaging demands. Additionally, the data indicates that several box types, namely 9515, 9567, and 9569, exhibit more consistent demand over the course of the year. The understanding of these variations in box usage might inform decision-making processes related to box procurement and storage.

Furthermore, the diagram illustrates the correlation among different types of boxes. In the month of October, there was a significant increase in the use of box type 9497 and, at the same time, a decline in the use of box types 9494 and 9495.

Furthermore, a certain amount of detail in the data allows for an assessment of the environmental impacts related to packing. Through the process of quantifying the use of different types of boxes, Lakeland has the ability to generate an estimation of the carbon footprint which is associated with this type of usage. This enables the company to take proactive measures in implementing packaging practices that are environmentally sustainable. For instance, if box type 9569 is found to have a relatively high usage compared to others, we could look into making minor changes to this particular box, which could help in increasing packing efficiency or minimising volume while being able to accommodate a similar number of orders.

The same figure could also be used in the context of predictive modelling. By using time series analysis, Lakeland has the capability to predict upcoming packaging requirements by examining past patterns. This holds significant value in efficiently handling periods of high demand, particularly during peak seasons like the holiday period. Creating accurate projections could help in preventing stockouts and minimising bottlenecks within the whole supply chain. In addition, the utilisation of predictive modelling has the potential to enhance just-in-time inventory management, resulting in decreased storage expenses and the mitigation of wastage.

3.5 Packing Efficiency

Figure 3.6 shows the average packing efficiency for each box along with the overall packing efficiency for the complete order dataset. On analysing the graph, we can see that boxes 9497, 9495, 9567 and 9569 have better packing efficiency than the overall average. While boxes 9494, 9500 and 9515 have lower average packing efficiency. We will use this information to optimise the dimensions of the latter boxes to increase the overall packing efficiency.

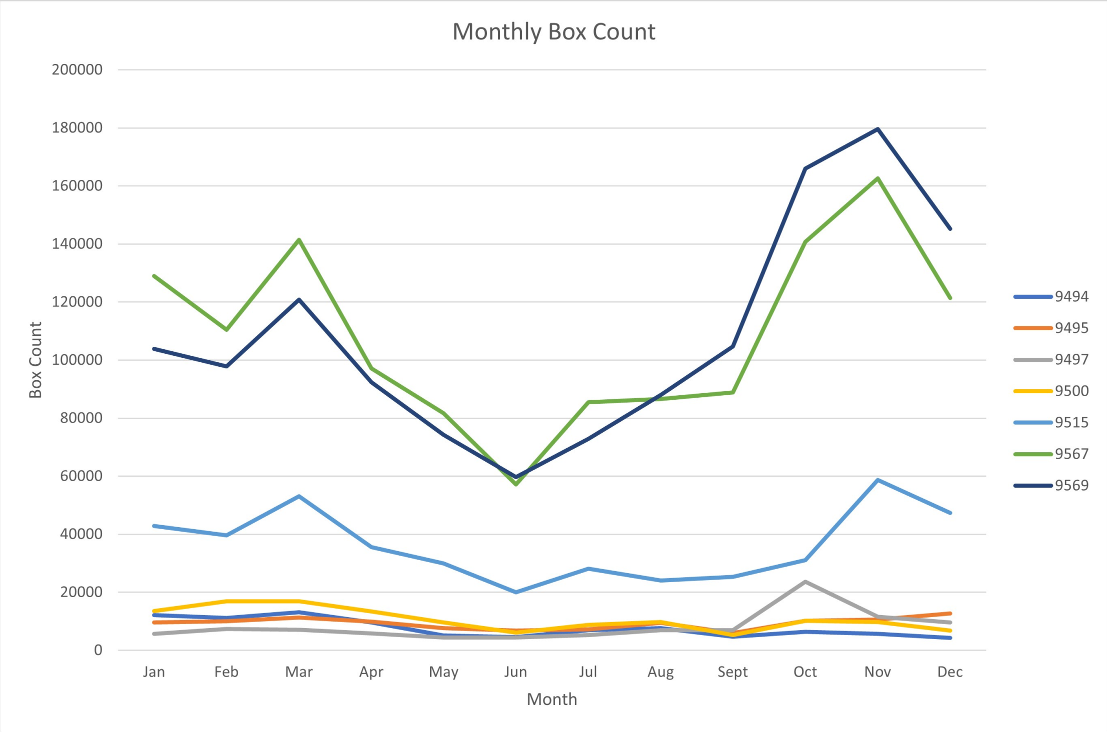

FIGURE 3.5: Monthly Box Count

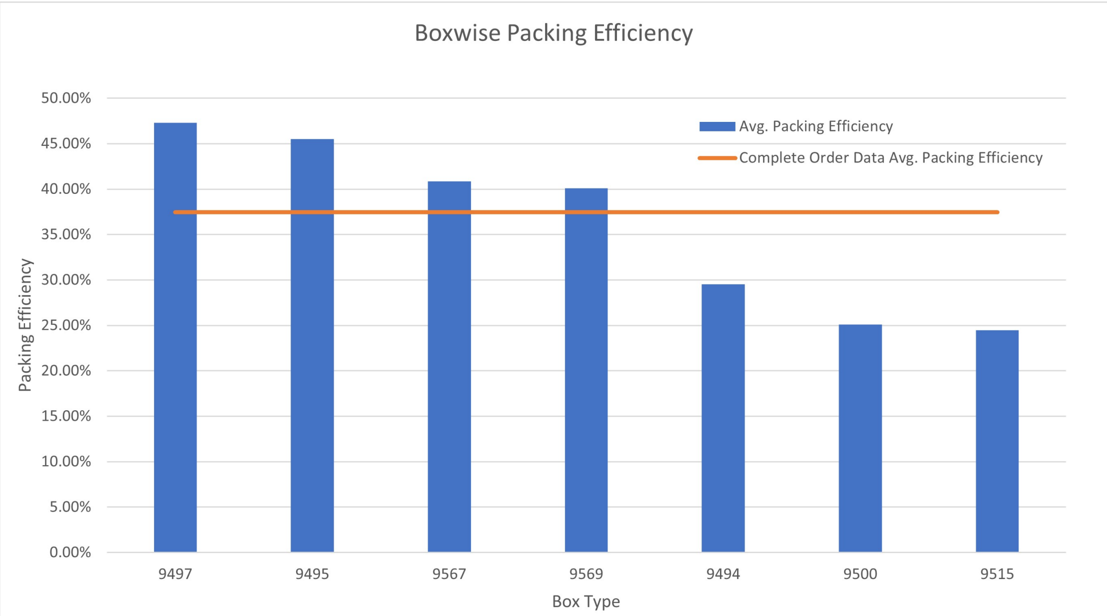

FIGURE 3.6: Monthly Box Count

Chapter 4

Methodology

4.1 Experimental Setup

The sampling approach is used for optimisation tasks, and simulations are computationally and memory-expensive. To perform optimisation on the complete data, we would require state-of-the-art systems that are highly expensive and not feasible. Hence, personal computers and university-provided High-End Computers(HEC) were used. The configuration of a personal laptop is mentioned in **Table 4.1**.

The model was implemented in Python 3.10 using a virtual environment. **Table 4.2** contains the list of primary libraries used for this project. Along with this list of libraries, we developed the code using python-poetry which gives us the benefit of packaging the code and could easily be deployed by Lakeland in their environment.

Sr. No.	Hardware Device	Memory Size
1	NVIDIA GeForce GTX1050Ti GPU	4GB
2	DDR4 RAM	16GB
3	Intel i7-8750H CPU	6 Cores 2.2 GHz Clock Speed

Table 4.1: Hardware Requirements

Sr. No.	Package Name	Version
1	pandas	2.0.2
2	pulp	2.7.0
3	highspy	1.5.3
4	gurobipy	10.0.2
5	matplotlib	3.7

TABLE 4.2: Software Requirements

The visualisations for reporting these experiments were also done using Python by utilising the suitable packages.

4.2 Data Description

To perform optimisation, we were provided with two datasets by Lakeland. The details for those datasets are mentioned below:

- **Order Data (OD)**- This dataset contained all the orders received by Lakeland from 1st Jan 2022 till 31st Dec 2022. It consisted of over 3.5 million rows for over 1.2 million orders.
  Each row consisted of Order day, Order month, Order ID, Product ID, Product dimensions, Product weight and v_box. The v_box mentioned is the recommended box from the current box packing system that Lakeland suggests the packers pack the order.

- **Box Data(BD)** - This dataset contained all the boxes along with dimensions used by Lakeland during the dates mentioned above.
  Currently, Lakeland has 12 cardboard boxes in its inventory which adheres to courier restrictions. Alongside, there are some JUMBO boxes and Jiffy Bags (Envelopes) as well.
  Lakeland also has 2 specialised machines which are able to alter the height of 2 of the 12 boxes mentioned. This allows for crimping and folding the box to the height of products leading to reducing the volume of the order. These boxes would be referenced as auto-boxes further in the text.

4.3 Data Processing and Assumptions

The data provided by Lakeland contained a few issues which needed to be solved before performing optimisation. The detailed steps on these are mentioned below:

1. Removal of JUMBO orders - These are orders that don't fit into boxes adhering to the courier rules. This could be due to too many products that wouldn't be able to fit in a single box and need to be split into multiple boxes. Also, some products are too big to fit in the existing 12 boxes and need special boxes. These orders were filtered out from the dataset as it was not in the scope of the project.

2. Removal of Jiffy bags - These are a kind of envelope used for packing. It is used to pack orders which contain small products. Currently, the system that Lakeland uses doesn't recommend these bags in the OD, as several things have to be considered before using Jiffy bags. One of them being glass or fragile products can't be sent using these. The products need to the flexible enough to send through these bags. Hence it is left at the discretion of packers to use a Jiffy bag whenever the situation seems fit for it. As these bags are not being recommended, the current Lakeland's logic allows the smallest box available for these orders, which are bigger than Jiffy bag. Due to this, the packing efficiency decreases for these orders.

3. Removal of Zero Volume Products - These were a few products present in the database, such as stickers, promo codes or cashback offers. These things are recorded as products in the database, which has been removed from OD.

4. Addition of columns - There have been a few attributes added to the dataset, such as product volume in OD and box volume in BD to aid our analysis.

5. Arrangement of Dimensions: The dimensions for all products in the OD table are ordered such that depth is always listed as the longest dimension, followed by width and then height as the shortest dimension. This ordering assumes that rotating the product does not affect its packing, so the absolute orientation of the dimensions does not matter when placing the product in any type of box.

4.4 Sampling Strategy

As the optimisation process is resource-intensive, we have sampled 100,000 orders using **simple random sampling** method from the OD provided by Lakeland, referred to as the evaluation set. This subset was used to conduct simulations with the optimisation model. To evaluate the performance of the optimisation model, the evaluation set was further sampled to sizes of 500, 1000, 2000, 5000 and 10000, referred to as in-sample datasets.

To evaluate the performance of the optimisation model, we sample subsets of size 500, 1000, 2000, 5000 and 10000 (in-sample) from the evaluation sample. The optimal solutions derived from the in-sample data sets were subsequently validated through testing on the out-of-sample evaluation set.

To account for inherent variability and uncertainties, the optimisation simulation is executed for 30 replications, with each replication representing a unique run of the model using the given data and parameters. Running multiple replications tests the stability of the model's output across different sample sizes, which is necessary for producing reliable solutions. Executing 30 replications provides an indication of output consistency and guards against drawing conclusions from spurious results that may occur in any single model run. By evaluating performance over repeated optimisations, the stability and reliability of the overall approach can be verified despite unpredictable fluctuations that can occur across iterations.(Kaut and Stein W. Wallace, 2007)

During each iteration of the simulation, the optimisation model processes the provided data, such as the order and box information, and attempts to find the most efficient way to allocate products into the available boxes. However, since real-world scenarios can be unpredictable, running the simulation only once might not provide a comprehensive understanding of the model's performance. This is where multiple iterations come into play.

By running the simulation 30 times, the model is exposed to a range of potential outcomes. These outcomes might differ due to the random nature of certain processes or fluctuations in the data. For example, if there's variation in the sizes and shapes of orders or the dimensions of available boxes, each simulation iteration might yield slightly different results.

Once all 30 iterations are completed, the next step is to analyse the outcomes. Instead of focusing on any single run, the median value of the results from all 10 iterations is used as the key metric to evaluate the efficiency of the model. The median is a statistical measure that sits in the middle of a dataset when it's sorted in ascending or descending order. It provides a more robust representation of the data's central tendency compared to using the mean (average), as it is less sensitive to extreme values or outliers.

Using the median of all the outputs is a way to address the variability inherent in the simulation process. It helps to mitigate the impact of outliers or unusual results that might occur in individual iterations due to chance. By considering the median, the evaluation becomes more stable and less influenced by specific simulation runs that might not be representative of the general performance of the model.

In summary, the approach of running the simulation for 30 iterations and then using the median of all the outputs to determine the efficiency of the model allows for a more comprehensive and reliable assessment of how well the optimisation process is working in various scenarios, considering the inherent variability present in the system.

4.5 Generation of Probable Boxes

The study aims to get a set of boxes that would create the highest packing efficiency and lowest total surface area of the cardboard used for manufacturing these boxes. The boxes used in Lakeland have to follow a set of rules which are laid by the courier company, which are as follows:

- Box Weight < 15 kgs
- Box Length < 1.2 meters

Probable Box Name	Depth	Width	Height
PB_20_10_5	20	10	5
PB_20_10_10	20	10	10
PB_20_15_5	20	15	5
PB_20_15_10	20	15	10
PB_20_15_15	20	15	15
PB_20_20_5	20	20	5
.	.	.	.
.	.	.	.
.	.	.	.
PB_115_35_15	115	35	15
PB_115_35_20	115	35	20
PB_115_40_5	115	40	5
PB_115_40_10	115	40	10
PB_115_40_15	115	40	15
PB_115_45_5	115	45	5
PB_115_45_10	115	45	10
PB_115_50_5	115	50	5

Table 4.3: Probable Box Table

- Box Length + 2 $\times$ Box Width + 2 $\times$ Box Height $<$ 2.25 meters

Following the dimension guidelines, a series of 788 probable box sizes were generated as potential packing solutions. The starting dimensions were a depth of 20 cm, a width of 10 cm, and a height of 5 cm. Each dimension was incrementally increased by 5 cm to create the full list of box sizes. This generated set of boxes was then used to find the optimal packing solution. It may be possible to obtain better solutions by adjusting the starting box dimensions or changing the size increments to create a larger pool of potential boxes to select from. **Table 4.3** shows a snippet of the probable boxes generated for the experiment execution.

4.6 Calculation of Eligible Boxes

In pursuit of an initial solution, a pool of 100,000 sampled orders underwent box allocation from a pool of 788 potential boxes generated. This allocation process was executed using two distinct heuristics, as outlined below:

- Volume Fill - This method entailed ensuring that the cumulative volume of the order was less than or equal to the volume capacity of the box.
- Dimensional Feasibility - This method involved verifying that the dimensions of all products within the order were each smaller than or equal to the corresponding dimensions of the chosen box.

The two heuristics would generate a list of candidate boxes that potentially meet the requirements of each order. However, since these are approximate heuristics, not all of the proposed boxes may actually be feasible for fulfilling every order. To verify the true feasibility of the boxes, an exact optimisation model for the packing problem needs to be solved (Padberg, 2000). This packing model takes the candidate boxes as inputs and checks if each order can be packed into one of its proposed boxes. By solving the optimisation, we can filter out any infeasible box suggestions from the heuristics and identify the set of boxes that definitively work for each order. The final output is verified feasible boxes for each order that align with the dimensions of the product. A detailed explanation of this method is explained in Abinauv Selvaraj's dissertation.

4.7 Key Metrics

Let O_i represent the volume of an individual order, B_i represent the volume of the box having dimensions d_i, w_i, h_i used to pack O_i and n represent the total number of orders.

Two key metrics were used to analyse model performance:

- **Total Volume** - This measures the total cubic volume occupied by all packed boxes.

$$\text{Total Volume} = \sum_{i=1}^{n} B_i$$

- **Total Surface Area** - This measures the total surface area of all packed boxes.

$$\text{Total Packing Efficiency} = \sum_{i=1}^{n} 2 \times (d_i \times w_i + w_i \times h_i + d_i \times h_i)$$

- **Total Packing Efficiency** - This metric calculates the percentage of box volume used by the packed items. A higher efficiency % indicates less unused space and better utilisation of box capacity.

$$\text{Total Packing Efficiency} = \sum_{i=1}^{n} \frac{O_i}{B_i}$$

To obtain the best set of boxes, the total surface area is used as the ranking parameter. As European Packaging Regulations (EPR) guidelines stipulate that companies are charged fees based on the total cardboard surface area utilised for packaging. Therefore, minimising total box surface area directly reduces the packaging charges incurred by Lakeland. This aligns with broader sustainability initiatives aiming to limit packaging waste.

By ranking the optimal box sets from each replication based on total surface area, the set with the lowest total area can be selected as the best solution. For replications where multiple box sets achieve similar minimum surface areas, secondary metrics like total box volume could be used as the tiebreaker, as this will reduce the courier charges.

4.8 Optimisation Model

In this section, we will present a mathematical framework aimed at the task of optimising the allocation of boxes for orders within given containers. The purpose of this model is to optimise the allocation of boxes by minimising either the overall volume of boxes utilised, minimising the total surface area of the boxes, or maximising the packing efficiency. The objective is to identify an optimal solution that effectively distributes boxes across orders, taking into account required constraints.

4.8.1 Sets and Parameters

$$
\begin{aligned}
\mathcal{B} &= \text{Set of all available boxes} \\
\mathcal{O} &= \text{Set of all orders} \\
\mathcal{B}_o \subseteq \mathcal{B} &= \text{feasible boxes for order } o \in \mathcal{O} \\
V_b &= \text{Volume of box } b \in \mathcal{B}
\end{aligned}
$$

4.8.2 Decision Variables

Let x_b be a variable defined as:

$$
x_b = \begin{cases} 1 & \text{if we use box 'b'} \\ 0 & \text{otherwise} \end{cases}
$$

Let y_b^o be a variable defined as:

$$
y_b^o = \begin{cases} 1 & \text{if we use box 'b' for order 'o'} \\ 0 & \text{otherwise} \end{cases}
$$

4.8.3 Optimisation Formulation

$$
\text{minimize} \sum_{o \in O} \sum_{b \in B_o} V_b y_b^o \tag{4.1}
$$

subject to

$$
y_b^o \leq x_b \tag{4.2}
$$

$$
\sum_{b \in B_o} y_b^o = 1 \qquad \text{for all } o \in \mathcal{O} \tag{4.3}
$$

$$
\sum x_b = 12 \tag{4.4}
$$

The objective is (4.1) minimises the total volume of boxes assigned to all orders. The constraint (4.2) models the fact that each order in $\mathcal{O}$ get a box present in x_o.

The constraint (4.3) models that every order in $\mathcal{O}$ gets exactly 1 box present in x_o. The constraint models the total number of boxes, in this case, 12.

Alternatively, we could replace the minimise volume objective with maximise packing efficiency having a new variable parameter P_b^o, where

$$P_b^o = \text{Packing Efficiency of box b for order o } b \in \mathcal{B}o \in \mathcal{O}$$

Using the new parameter, the objective function would also change to packing efficiency for all orders, making the new function as

$$\text{maximize} \sum_{o \in O} \sum_{b \in B_o} P_b^o y_b^o \tag{4.5}$$

while maintaining the same constraints mentioned in (4.2), (4.3) and (4.4).

4.8.3.1 Issues with current model

On analysing the optimisation model's output, an issue was observed. Some orders in the out-sample dataset were left with no allocated box. On deeper analysis, it was observed that the in-sample datasets lacked orders needing large boxes. Thus the model's optimal box selection excluded bigger boxes. When applied to the out-sample orders, few orders were not allotted any box.

This is caused due an inherent limitation of the staged sampling approach. If the in-sample lacks edge cases, the model overlooks those needs. The optimized boxes are biased by what the model sees, blinding it to unseen needs. This sample bias manifests in suboptimal generalisation.

To overcome this shortcoming in the model, 2 methods are devised:

1. Minimal Subset
2. Rejection of Orders

4.8.3.2 Minimal Subset

The general idea of this method is to find orders in the evaluation set whose eligible boxes are minimal. So, satisfying optimised box requirements for these orders will ensure feasibility for all orders. To achieve this, we use the following setup.

The proposed approach aims to identify a subset of orders from the evaluation set that have the most constrained eligible box requirements. By ensuring feasibility for these orders, feasibility will therefore be guaranteed for the entire order evaluation set.

Specifically, the method involves first analysing the eligible box across all orders in the evaluation set. Orders are then selected such that allocation of at least 1 box to these orders will ensure allocation for all boxes in the evaluation set. This method covers the shortcoming in the above approach where few orders were not allotted any box.

$$\overline{\mathcal{O}} = \text{Evaluation Order Subset}$$
$$\mathcal{O} = \text{Order in-sample subset}$$

We can ensure feasibility for all o in $\mathcal{O}$ by using set covering inequalities:

$$\sum_{b \in \mathcal{B}^o} x_b \geq 1 \qquad \text{for all } o \in \mathcal{O}$$

Now let $\mathcal{B}^{o_1}$ and $\mathcal{B}^{o_2}$ contain the set of eligible boxes for orders o_1 and o_2 respectively.

Note that if

$$\mathcal{B}^{o_1} \subseteq \mathcal{B}^{o_2}$$

Then

$$\sum_{b \in \mathcal{B}^{o_1}} x_b \geq 1 \implies \sum_{b \in \mathcal{B}^{o_2}} x_b \geq 1 \qquad \text{for all } o_1, o_2 \in \mathcal{O}$$

For all

$$o \in \overline{\mathcal{O}} \setminus \mathcal{O}^* \quad \text{there exists} \quad o^{'} \in \mathcal{O}^* \text{ such that} \quad \mathcal{B}^{o^{'}} \subseteq \mathcal{B}^{o}$$
$$\text{and for all } o_1, o_2 \in \mathcal{O}^* \text{ we have } \mathcal{B}^{o_1} \not\subseteq \mathcal{B}^{o_2}.$$

For such a set $\mathcal{O}^*$, we need add only the following set covering constraints:

$$\sum_{b \in \mathcal{B}} x_b \geq 1 \quad \forall o \in \mathcal{O}^*$$

A shortcoming of this method is when catering to all the orders, the optimal cost for the optimisation model gets impacted, which ends up creating a sub-optimal solution.

By focusing strictly on fulfilling the minimal subset order's box requirements, the overall optimisation model is limited in its ability to find cost savings. Due to these specific order, the model has to select bigger boxes leading to creating a sub-optimal solution in terms of cost. So in catering the solution exclusively to the needs of these orders, the packing efficiency for the overall order set is reduced.

In summary, the weakness of this approach is that while it guarantees feasibility, it comes at the expense of achieving the mathematically optimal cost for packing the full set of orders. Further development into balancing feasibility with overall packing cost-effectiveness is needed.

4.8.3.3 Rejection of Orders

To address the sub-optimal cost performance of the Minimal Subset method, this new approach involves selectively rejecting a subset of orders to reduce total volume or improve packing efficiency.

Specifically, the algorithm analyzes the set of orders and identifies the ones with larger, more outlier box dimension requirements. These edge case orders represent outliers that contribute disproportionately to unused wasted space when packed for the majority of the smaller orders.

By rejecting these edge case orders from the evaluation set, the total volume needing to be packed is reduced. This allows the remaining orders to be packed more efficiently, with less unused space. Even though some orders are rejected, improved packing density is achieved for the overall order set. While order rejection reduces revenue opportunities, the improvement in packing efficiency and cost reduction for the remaining orders may offset the lost revenue.

In summary, selectively rejecting oversized edge case orders represents a straightforward method to overcome the cost sub-optimality of the Minimal Subset approach. Removing exception cases improves packing density and cost performance for the remaining orders. Implementation would require careful analysis to balance order rejection with packing efficiency gains.

Chapter 5

Numerical Test

5.1 Experimental Set-up

In this section, we will be discussing the various experimental set-ups, including the sample sizes and types of solvers used to generate results, results of the Padberg check on current box recommendations, followed by the generated eligible box lists, created the optimised new boxes and analysed for multiple scenarios.

5.1.1 Solvers

A very important part of solving a large optimisation model is to Selection of a solver. In our experiment, we have used the PuLP package, which contains a default solver called CBC. However, the number of constraints in the model grows exponentially as the sample size increases. Refer **Figure 5.1**. CBC slows down the execution of models with a higher number of constraints.

To overcome this issue, we used different solvers that could solve problems faster. In our experiment, we have used solvers named 'HiGHS' and 'GUROBI'. All the 3 solvers use the branch and bound method to reach the optimal solution for the model. 'HiGHS' is a free-to-use solver, whereas 'GUROBI' is a paid version for commercial use but is free for academic use.
Table 5.1 shows the median time required for the 'CBC', 'HiGHS', and 'GUROBI'

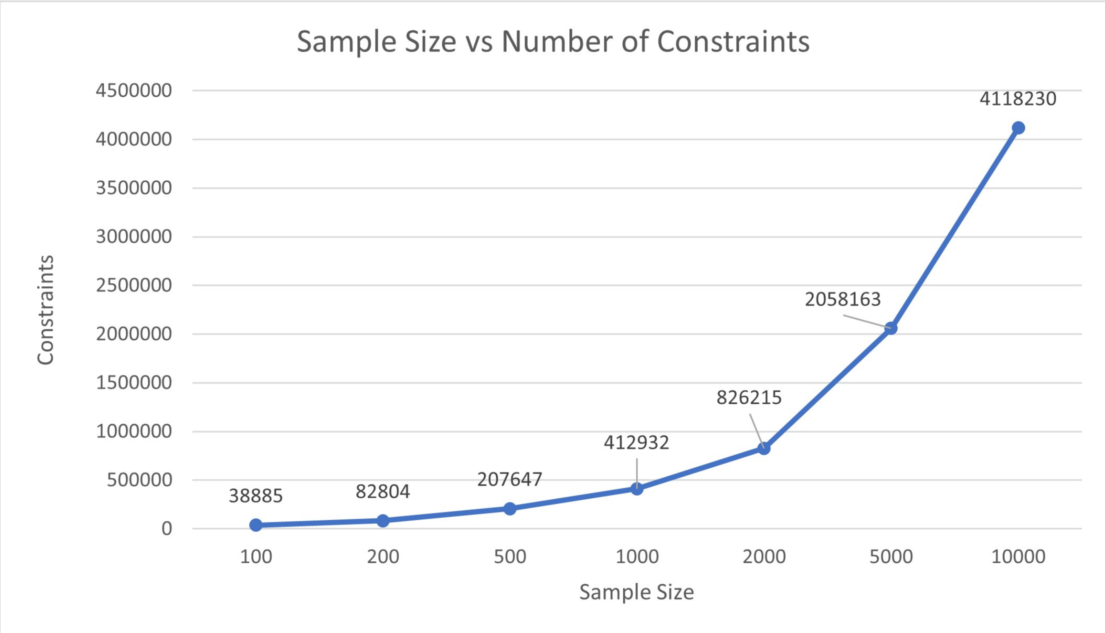

FIGURE 5.1: Box Relative %

Sample Size	CBC	HiGHS	GUROBI
100	39.66	5.04	0.64
200	56.47	7.13	0.9
500	56.36	10.1	1.81
1000	188.16	26.88	3.84
2000	565.57	66.46	7.81
5000	735.82	96.06	12.54
10000	2318.26	244.8	25.85

TABLE 5.1: Current Box Table

solvers across different sample sizes. Due to computational resource constraints, 'CBC' was tested with 5 replications, 'HiGHS' with 20 replications, and 'GUROBI' with 30 replications. The values reflect solving the model for each sample size using 12 different box set configurations.

From **Table 5.1**, we can analyse that GUROBI is required when the sample size is high, whereas the other solvers are sufficient if we have moderate-sized problems.

5.2 Feasibility Check with current recommendations with Padberg

We wanted to test the Padberg model on the order data we had received from Lakeland, which had the suggested boxes. Hence, we sampled 100,000 orders with the suggested boxes. We had coded this test such that the Padberg model would take in the order data and the suggested box data as input while returning a True or False as a result, based on whether the order could be packed inside the box or not. We then executed this code.

The result was an output CSV file with columns that included order details, the suggested box and the Padberg result. We found that 96.7% of the time, the suggested box fitted the order. This was close to the value that Lakeland had given us, which was 95%, based on the warehouse packers' close observations. The matching values on a considerably significant sample meant that the Padberg model closely emulated the actual packing scenarios.

The result was crucial as it gave us a confirmation of the reliability of our packing feasibility algorithm and, by extension, the Padberg model. We wanted to do an extensive analysis of this result. We were particularly interested in seeing the variation in the correct suggestions with an increase in the number of products per order. It's also important to note that the orders with more products are relatively fewer.

The variation of the correct suggestions with the number of products in the order is shown in **Figure 5.2**. It was evident that increasing the number of products per order reduces the accuracy of box suggestions. This was expected because, with an increase in the number of products, the combinations of alignments of the products inside boxes get complicated, and it gets difficult to pack all the products of the order inside the box.

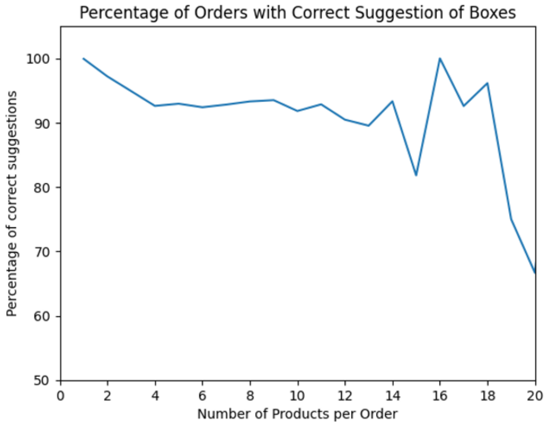

FIGURE 5.2: Percentage of Orders with Correct Suggestion of Boxes

5.3 Finding Eligible Boxes for Probable Boxes

With the reliability of our Packing feasibility check algorithm proven, we moved on to find the eligible box list for each order. We used the method described in **Section 4.3.3** from Abinauv's dissertation to obtain this result.

The output file with the eligible box list for each order was obtained in the form of a CSV file. The generated CSV file had 9 columns. Of which 2 were input columns which included order ID and order volume. At the same time, the other columns were generated by the packing feasibility check algorithm. To explain the output in a better way, we have taken one particular order and shown the various parameters of the output in **Table 5.2**.

Parameter	**Value**
orderId	1
orderVolume	15855.84
eligibleBoxesId	['PB_55_30_15', 'PB_55_30_20', ., .,'PB_115_40_15']
eligibleBoxesPackingEfficiency	[0.64, 0.48, ., ., 0.23]
mostEligibleBox	PB_55_30_15
mostEligibleBoxVolume	24750
mostEligibleBoxPackingEfficiency	0.64
PadbergResult	['PB_55_30_15', 'PB_60_30_15', ., ., 'PB_75_40_35']
PadbergPackingEfficiency	[0.6406, 0.5873, ., ., 0.1510]

TABLE 5.2: Eligible Box List Sample Output

From the first step of the packing feasibility check algorithm (the volume fill and dimensionality check), the 'eligibleBoxesId' column was created. It was a list of boxes that were eligible for the order to be packed based on the dimensions. These boxes were sorted in ascending order of volume, which can be clearly seen in **Table 5.2**. This column was used primarily in the further steps.

Based on the 'eligibleBoxesId' column, the volumes of these boxes were calculated and their respective packing efficiencies for the order were added to the 'eligibleBoxesPackingEfficiency' column. We could observe the packing efficiency decrease as we moved towards the right - affirming the fact that the boxes were indeed sorted in ascending order of volume.

The best box from the 'eligibleBoxesId' column (based on packing efficiency) was shown in the 'mostEligibleBox' column. The volume and packing efficiency of this box were shown in the 'mostEligibleBoxVolume' and 'mostEligibleBoxPackingEfficiency' columns, respectively.

From the second step of the packing feasibility check algorithm (the Padberg check), the 'PadbergResult' column was created. It's important to note that this list was not sorted by increasing volume. Based on this column, the 'PadbergPackingEfficiency' column was added.

We wanted to do an extensive analysis of the most eligible box results. The average packing efficiency of the most eligible boxes was found to be 66.29%. The distribution of the same is shown in **Figure 5.3**.

The packing efficiency of the most eligible box gives us a good idea of the upper bound on the packing efficiency. In other words, it tells us the absolute maximum packing efficiency we can achieve, given the courier restrictions on box sizes. **Figure 5.3** can be interpreted as the ideal case where all the 788 probable boxes were available for packing. We can see a left-skewed distribution, with the peak occurring in the 60% to 80% range. This was the exact opposite of the packing efficiency distribution we saw in **Figure 3.1** from Abinauv's Dissertation., where we had a right-skewed distribution. With this being the ideal case in hand, we had set this result as the benchmark for our optimisation results.

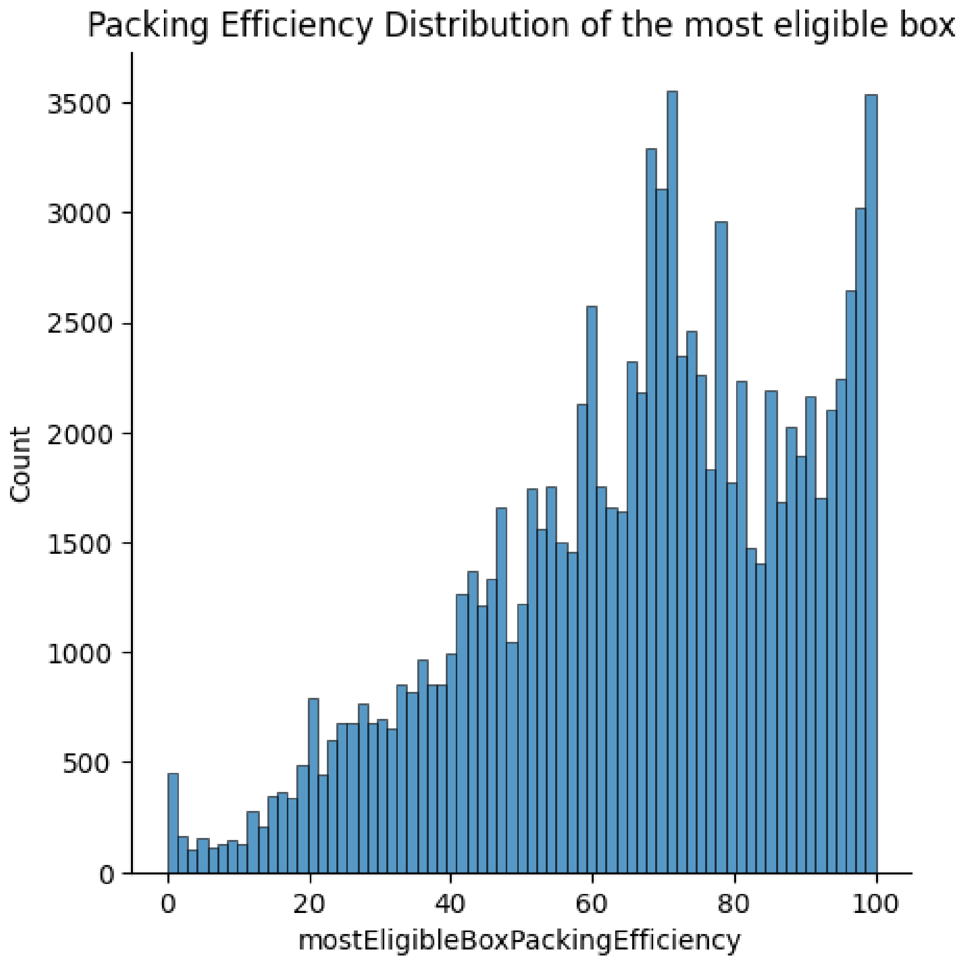

FIGURE 5.3: Packing Efficiency Distribution of the most eligible box

Apart from this, we were interested in the frequency of the boxes in the most eligible box column. To study this, we made **Table 5.3**, which contains the top 10 most eligible boxes in terms of relative frequency. Since the sample size of the orders was 100,000, the relative frequency is represented as a percentage of that.

All the top 25 boxes in this list had a frequency of over 1000 (which is above 1% of the orders in this sample). However, we had to show a limited number of rows here to simplify the explanation. We could see some unique boxes like PB_55_30_30, being the most eligible box for a significant number of times.

It was evident from **Table 5.3** that multiple boxes could potentially be grouped together into a single box, and Lakeland's height adjustability feature could be employed to increase the packing efficiency. This was, however, not quite right

mostEligibleBox	orders
PB_35_25_10	4.88%
PB_35_30_10	2.94%
PB_30_20_10	2.23%
PB_25_20_5	2.19%
PB_25_20_10	2.15%
PB_20_15_10	2.13%
PB_20_15_5	1.91%
PB_20_10_5	1.75%
PB_55_30_30	1.69%
PB_35_20_10	1.53%

TABLE 5.3: Frequency of top 10 most eligible boxes

as these were not the optimised dimensions. They were simply the best boxes for the particular order, and that doesn't essentially mean that they have optimised dimensions.

It was interesting to note that the box PB_35_25_10 was the best box for 4.88% of the orders. The dimensions of this box, however, closely resemble the dimensions of the 9567 box in **Table 3.2** from Abinauv's dissertation. The dimensions $40cm \times 25cm \times 22cm$ not only resemble the dimensions of this box, but since 9567 is one of the 2 height adjustable boxes in the current Lakeland inventory, the height could be altered to give a better fit as well. This takes the dimensions even closer. This also affirms Lakeland's argument of the 9567 box type being one of the most used ones in their inventory.

The approach of grouping the boxes together to make the best out of the height adjustability feature was used on the boxes with optimised dimensions. This is discussed in detail alongside optimising the box dimensions in the following section.

has context menu

5.4 Optimisation of New Boxes

Now, we use the model from **Section 4.8.3** to suggest new subsets of boxes. We start the analysis by optimising for maximum packing efficiency.

The analysis is divided into 4 scenarios:

- **Scenario 1:** Number of Boxes = Infinity
- **Scenario 2:** Number of Boxes = 12
- **Scenario 3:** Number of Boxes = 19
- **Scenario 4:** Number of Boxes = 19 with existing boxes

5.4.1 Scenario 1

Firstly, to obtain the optimal packing solution, the constraint limiting the number of box types will be relaxed. This allows each order to be packed in the best box across all potentially available boxes mentioned in **section 4.5**. By relaxing this constraint, the model assigns each order the optimal box that maximises packing efficiency and minimises box volume. Solving this unconstrained way reveals the maximum packing density that could theoretically be achieved if any box size could be available in inventory.

Figure 5.4 shows the packing efficiency histogram for unlimited boxes. The average packing efficiency in this scenario is 58.7%, and the average box volume is 15,663 cm^3.

5.4.2 Scenario 2

This scenario is implemented considering the total number of box types remains the same in inventory for Lakeland and would not need changes in operational management.

After the optimal packing efficiency has been achieved by using an unlimited number of boxes, the goal will then move towards achieving practical outcomes that could be used in real-world scenarios. As explained in **section 3.5**, our objective would be to improve on the mentioned figures.
Figure 5.5 and **Figure 5.6** shows the list of 12 optimal boxes and the histogram

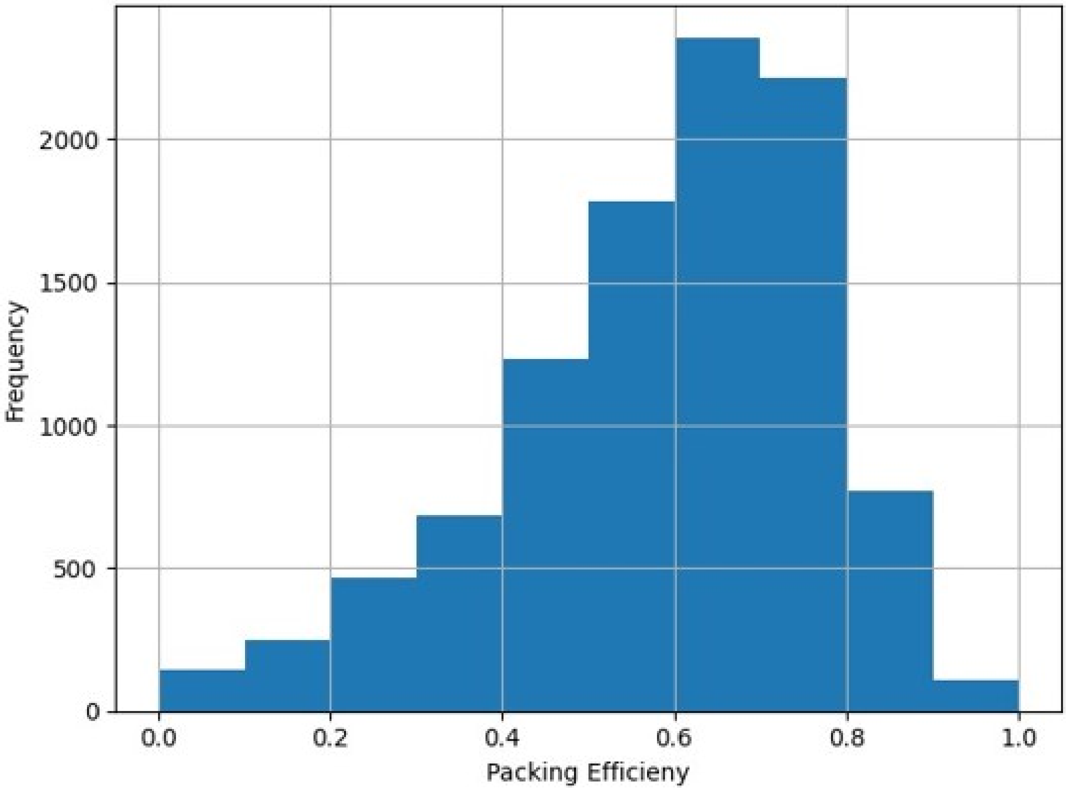

FIGURE 5.4: Packing Efficiency for Scenario 1

of packing efficiency for the same set of boxes. In the current scenario, the packing efficiency obtained is 40.47%, while the average volume of the boxes is measured to be 24,312 cm^3. Further analysis of the optimal set of boxes reveals that a few boxes can be combined, and the largest box can be used as an auto box. In this particular scenario, it is feasible to use the base box labelled as 'PB_35_25_20' and, after that, obtain a modified box denoted as 'PB_35_25_10' by making height adjustments in a way similar to the existing technique used for the 9567 and 9569 series of boxes. This strategy can be further extended by including the utilisation of 'PB_35_20_5', considering the fact that the dimensions of the box are smaller compared to the previous one. However, the difference in size will not cause problems during the process of packaging the orders within the larger box. A possible disadvantage associated with this approach is a decrease in packing efficiency, which will require some extra calculations to determine the feasibility of implementing this technique at Lakeland.

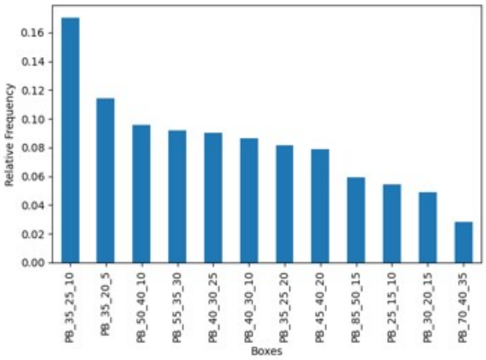

FIGURE 5.5: Optimal Boxes for Scenario 2

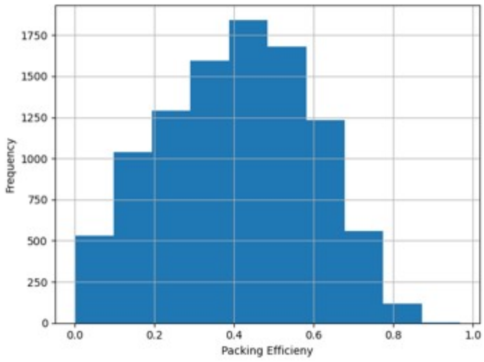

FIGURE 5.6: Packing Efficiency for Scenario 2

5.4.3 Scenario 3

After achieving metrics better than the current metric, we moved towards increasing the count of box types. One of those scenarios is explained here, where we have optimised for 19 box types.

Figure 5.7 and **Figure 5.8** shows the list of 19 optimal boxes and the histogram of packing efficiency for the same set of boxes. In the current scenario, the packing efficiency obtained is 42.10%, while the average volume of the boxes is measured to be 24,472. We can see the metrics have improved by increasing inventory, as each order would be getting a better box than the previous scenario.

Similar to the previous scenario, we can club a few boxes to get auto boxes. In this particular scenario, it is feasible to use the base box labelled as 'PB_40_30_15' and, after that, obtain a modified box denoted as 'PB_40_30_10' by making height adjustments in a way similar to the existing technique. Alongside this, we have another possibility where we can use 'PB_45_40_20' as the base box and obtain boxes 'PB_45_40_15' and 'PB_45_40_10'.

This strategy can be further extended by including the utilisation of 'PB_40_35_30' as the base box to obtain 'PB_40_30_15' and 'PB_40_30_10'. The drawbacks that were present in the prior situation are still relevant, as the same limitations that prevented achieving optimality continue to exist.

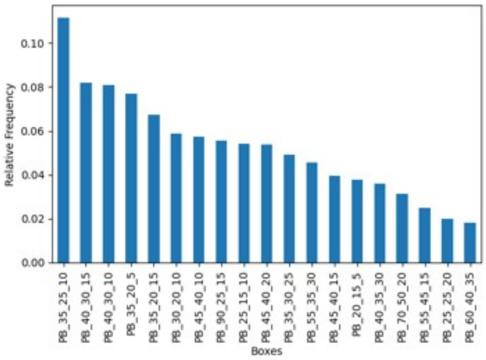

FIGURE 5.7: Optimal Boxes for Scenario 3

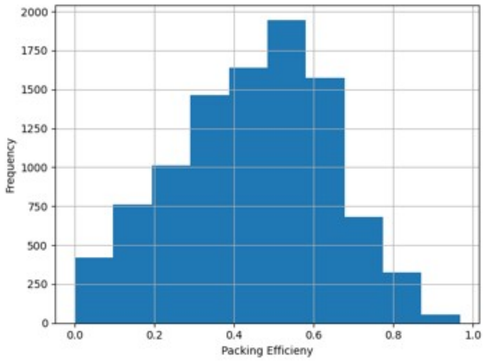

FIGURE 5.8: Packing Efficiency for Scenario 3

5.4.4 Scenario 4

For this scenario, we have incorporated the existing boxes utilised by Lakeland into the collection of potential boxes considered within the optimisation model. This test was done in order to determine if any of the existing boxes are included in the ideal subset of boxes, as this would result in Lakeland avoiding making modifications to the current inventory. In addition, during the analysis phase, we aggregated the boxes labelled 9567H, 9567I, and 9567J into a single category denoted as 9567. Similarly, we combined the boxes labelled 9569H, 9569I, and 9569J into a consolidated category referred to as 9569.

Figure 5.9 and **Figure 5.10** shows the list of 13 optimal boxes and the histogram of packing efficiency for the same set of boxes after clubbing 9567 and 9569 box types. In the current scenario, the packing efficiency obtained is 44.70%, while the average volume of the boxes is measured to be 21,229 cm^3. We can see the metrics have improved by adding existing boxes to the model.

Similar to the previous scenario, we can club a few boxes to get auto boxes. In this particular scenario, it is feasible to use the base box labelled as 'PB_30_20_15' and, after that, obtain a modified box denoted as 'PB_30_20_10' by making height adjustments in a way similar to the existing technique.

This strategy can be further extended by including the utilisation of 'PB_35_20_10' as the base box to obtain 'PB_30_20_15' and 'PB_30_20_10'. The drawbacks that

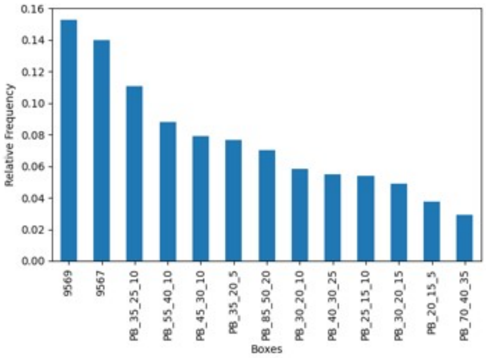

FIGURE 5.9: Optimal Boxes for Scenario 4

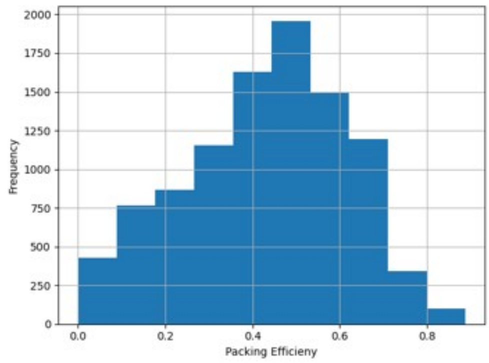

FIGURE 5.10: Packing Efficiency for Scenario 4

were present in the prior situation are still relevant, as the same limitations that prevented achieving optimality continue to exist.

This particular scenario has produced the most favourable outcome thus far in terms of practicality, as it has achieved the highest packing efficiency observed so far. Furthermore, there has been a significant gain of 19.5% in packing efficiency compared to the current level, achieved by the utilisation of 11 distinct types of boxes.

5.4.5 Packing Efficiency vs Number of Boxes

Figure 5.11 shows a positive correlation between packing efficiency and the increase in the number of boxes. If we were to extend this trajectory hypothetically to an infinitely large box set, it can be extrapolated that the efficiency of packing would approach a theoretical maximum of 58.7%, as shown above. However, the existence of practical limitations prevents the possibility of endless variations in boxes. It would be advisable to conduct a more comprehensive examination using a cost-benefit approach in order to determine the most effective threshold for unique box configurations. This analysis should consider the trade-offs between achieving higher packing density and the logistical challenges associated with increased product variation, such as higher inventory costs and increased complexity in warehouse management. While the goal of maximising packing performance is

Boxes	Maximising Packing Efficiency		Minimising Volume	
	Packing Efficiency	Volume(cm^3)	Packing Efficiency	Volume(cm^3)
Infinity	58.70%	15,663	56.15%	19,856
12	40.47%	24,312	39.20%	22,903
19	42.10%	24,472	42.61%	21,439
19 with Existing	44.70%	21,229	42.80%	20,381

Table 5.4: Performance Summary for Different Objectives

still desired, it is important to consider real-world aspects that determine a satisfactory threshold. Beyond this threshold, the additional expenses incurred may no longer be justified by marginal advantages.

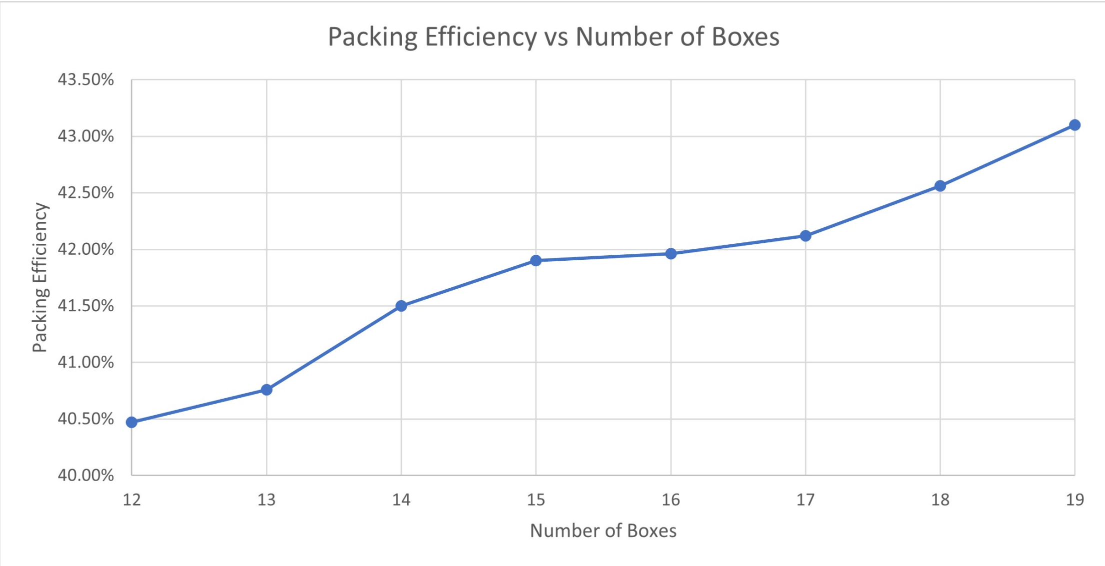

Figure 5.11: Packing Efficiency vs Number of Boxes

5.4.6 Summary for different metrics

All the analyses mentioned above used the maximum packing efficiency. Along with this, we also have other objective functions. **Table 5.4** shows the comparison for the same scenarios using the minimum volume objective. From the same table, it can be inferred that while Scenario 4 is the best, depending on the selection of objective function, the performance of key metrics change.

Chapter 6

Conclusion

6.1 Summary of Work

This project aimed at developing an optimisation model to improve the allocation of boxes for e-commerce shipping orders at Lakeland. The model was trained and tested on order data from 2022 to evaluate its ability to improve packing efficiency and reduce cardboard wastage in comparison to the current heuristic used by Lakeland.

The experiments revealed several key findings. Firstly, if we allowed unlimited boxes, it could achieve a maximum packing efficiency of 58.7% on average. This established an upper bound on theoretically achievable efficiency given complete box flexibility.

Practical scenarios with only 12-19 boxes yielded lower packing efficiencies of 40-45% but still significantly exceeded the 37.44% efficiency from Lakeland's volume-based heuristic. The optimised configurations consistently reduced box volumes by 7-8% as well. These results quantify the potential material savings and efficiency improvements from upgrading to an optimisation-based approach.

Analysing the correlation between packing efficiency and the number of boxes showed diminishing returns past around 19 boxes. Balancing marginal packing gains against rising inventory costs indicates that 19 uniquely sized boxes provide a good efficiency-cost tradeoff.

The scenario leveraging Lakeland's existing boxes plus new optimised additions achieved the best results at 44.7% efficiency using only 13 box types. This demonstrates current boxes, when combined in an optimised mix, can provide substantial efficiency gains without operational overhaul. The model provides data-driven guidance for selectively supplementing Lakeland's box portfolio.

For real-world applications, Lakeland could adopt the optimized 13-box configuration from the final scenario. This would require introducing just 2 new box sizes to complement current offerings. The consolidated auto-box approach could streamline box flexibility while controlling box variations. Expected benefits include 19.5% packing efficiency gains from current packing efficiency, 8% box volume reduction, and material waste savings - aligning with sustainability objectives.

While promising, the model has limitations. It uses a staging sampling approach which can neglect edge cases. It focuses solely on packing efficiency, whereas factors like shipping and procurement also affect costs. Product dimensions are fixed when rotating may enable better fit. Predictive modelling of order patterns is needed to optimise preparations for demand fluctuations.

6.2 Future Work

After successfully implementing the optimisation and getting improved results. There are some areas whose improvement could improve the model significantly. These are listed below:

- **Height adjustability feature:** Currently, we have to analyse the optimal set of boxes and deduce which boxes could be clubbed in 1 box to derive auto boxes. This feature could be helpful such that those boxes get clubbed in the model, which would make way for more boxes to be added in the optimal solution.

- **Jiffy bags in optimisation model:** During processing, we removed these bags as all the products couldn't be packed in those. The intelligence in the products database, which could flag, if the product is eligible or not to go into Jiffy bags, would be helpful in improving the model's performance.

- **Adding more boxes:** As the probable boxes are currently being generated in increments of 5, on lowering this increment, we would get a bigger set of eligible boxes, leading to a better box for each order. This change could make significant improvements in the key metrics.

Bibliography

Gehring, Hermann et al. (1997). "A genetic algorithm for solving the container loading problem". In: *International transactions in operational research* 4.5-6, pp. 401–418.

Høyland, Kjetil and Stein W Wallace (2001). "Generating scenario trees for multistage decision problems". In: *Management science* 47.2, pp. 295–307.

Kaut, Michal and Stein W. Wallace (2007). "Evaluation of Scenario-Generation Methods for Stochastic Programming". In: *Pacific Journal of Optimization* 3.2, pp. 257–271.

Padberg, Manfred (2000). "Packing small boxes into a big box". In: *Mathematical Methods of Operations Research* 52, pp. 1–21.

Sharma, Gaganpreet (2017). "Pros and cons of different sampling techniques". In: *International journal of applied research* 3.7, pp. 749–752.

Wäscher, Gerhard, Heike Haußner, and Holger Schumann (2007). "An improved typology of cutting and packing problems". In: *European journal of operational research* 183.3, pp. 1109–1130.

Zhao, Xiangtong, Julia A Bennell, and Xinyu Song (2016). "A Review of Solution Methodologies for 3D Container Loading Problems". In: *International Transactions in Operational Research* 23.3, pp. 287–320. DOI: 10.1111/itor.12094.

Declaration of Generative AI and AI-assisted technologies used in the research and writing process

1. **Generative AI:** The AI tool ChatGPT and Claude was used in the initial, idea generation phase. No part of the main text or content has been AI-generated.

2. **Translation Software:** No part of the main text or content has seen the usage of any Translation Software.

3. **Grammar Checkers:** Grammarly was used for the purpose of grammar checks and correcting sentences across the document.