IE 5801 Capstone Project

Fall 2020

"Final Report"

Sponsor: Daikin Applied

Industry Mentors: ***Nathan Erdman, Thomas Keller, Nadia Silva***

Faculty Advisor: ***Prof. Darin England***

Team: ***Aparajita Kar, Seungyoon Lee, Weiqi Wang***

Industrial and Systems Engineering (Analytics Track)

University of Minnesota, Twin Cities

December 15th, 2020

ACKNOWLEDGEMENT

We would like to express our deepest gratitude and appreciation to Professor Darin England for his continuous support and able guidance throughout the course of this project.

We would also like to extend our gratitude to all mentors Nathan Erdman, Nadia Silva and Thomas Keller for their effort and time besides their busy schedule and also for providing us with all the information required for the successful completion of this project.

In addition, we are extremely grateful to Professor William L Cooper for giving us this opportunity to work with Daikin Applied and providing valuable feedback throughout this project.

Aparajita Kar, Seungyoon Lee, Weiqi Wang

Executive Summary

Daikin Applied Inc. is a subsidiary of Daikin Industries, a global leader in the market of commercial and industrial use air conditioning systems. Daikin Applied is focused on providing HVAC (heating, ventilation, and air conditioning) solutions for applied commercial and industrial customers in North America among various other business areas of Daikin. A manufacturing plant of Daikin Applied is in San Luis Potosi (SLP), Mexico. The shipping orders produced by the plant are transported from SLP to Laredo (LRD), Texas, and are then shipped to customers across the USA & Canada.

The objective of this project is to design a cost-effective consolidation and distribution method using Less than Truckload and Truckload mode to transport shipments from Laredo, Texas to various destinations across the U.S.A and Canada. The strategy is to minimize the total costs effectively by loading items into one or more trucks and use a multi-stop planning method to deliver the shipments.

INDEX

Chapter 1: Project Description

1.1 Background

Current State: Daikin used full-truckload (FTL) mode to ship shipping orders from SLP to LRD. These shipments were then distributed to customers via Less-than-truckload mode of shipments across the U.S and Canada. Less-than-truckload sometimes referred to as “Less than Load (LTL)” denotes transportation of products that do not require the use of a full trailer. The shipper pays the amount for the portion of the trailer that their shipment occupies. This also means that the shipments are delivered along with shipments from a different shipper. In contrast, Full Truckload referred to as “Truckload (TL)” denotes a mode of transportation wherein a single truck is used to transport a dedicated shipment, where each shipment could be a consolidation of multiple shipping orders. For LTL mode of transportation, Daikin selects the best quotes provided by different carriers. Different carriers have different ways of determining the LTL cost based on their base rates, discounts provided, and the weight of the shipments. Unlike LTL, costs of TL shipments depend upon the origin and destination and are not a function of shipment weight.

Challenges: For shipments originating from Laredo, there is no consolidation of larger shipments shipped via LTL and smaller TL shipments going to the same geographical region. Moreover, transporting shipments via LTL mode increases transit time and poses the risk of damaged freight.

Future State: Daikin is looking forward to saving costs for shipments by consolidating shipping orders going to the same geographical location. These would also mean lowering damages and hence the cost of claims incurred.

1.2 Objective

Starting October 2019, Daikin shipped a small percentage of its shipments from Laredo, via single stop (a single order) TL mode. These shipments resulted in fewer to no damages and eventually reduced the cost of claims. Hence, the logistics department at Daikin is planning to move towards TL mode of shipments for most of its shipping orders originating from Laredo in the future. However, TL shipments have significantly higher shipping rates than LTL. Hence, a strategic analytical study is required before Daikin plans to improvise TL shipments for all of its shipping orders. In addition to that, shipping orders can be shipped through the TL mode of transportation via a single stop or a multi-stop model. Currently, shipments made via TL only have a single stop. According to a case study by Daikin, shipping a consolidation shipment of three shipping orders via TL mode has a 40% lower cost than shipping via three different LTL shipments. Therefore, shipping costs can be reduced by utilizing the multi-stop (shipments for two or three customers having different destinations combined together) TL model by consolidating two or more orders going to the same geographical location.

The objective of this study is to:

1. Identify delivery zones for multi-stop TL shipments
2. Provide shipping orders consolidation recommendations for multi-stop TL shipments

3. Compare the costs of consolidated TL shipments to the traditional LTL mode of transportation.

The aim is to understand if Daikin can reduce the costs by shipping through the multi-stop TL transportation model. Daikin can then decide which consolidation options of TL shipments they would use given the recommendations to achieve the goal of minimizing costs.

1.3 Opportunity Value

Data Analysis & Cost Regression Model: Exploratory Data Analysis is conducted on the historical shipment dataset to understand the process of the supply chain, costs incurred from shipping via traditional LTL mode, and the claims and damages involved due to mishandling of shipments. Several regression models are formulated by the use of predictive analytics to estimate the cost of shipments via TL and LTL modes. The best performing regression model is used in the heuristics algorithm to find the optimal solution and estimate the total cost which is required to be minimized.

Shipment consolidation and Route optimization: For the purpose of route optimization, the optimal delivery zones are found wherein all shipment consolidations are done. An optimization model is formulated which minimizes the total cost and gives the best possible combination of shipments bound by capacity constraints of the truck. Later, a heuristic approach is presented for finding the feasible solutions.

1.4 Proposed Solution

The solution as discussed with the mentors at Daikin includes the following:

a. An exploratory data analysis is required to understand the current model.

b. Fit regression model to estimate the costs of shipments through TL and LTL modes.

c. An unsupervised clustering model to identify the delivery zones.

d. An optimization model which gives the optimal shipping orders consolidation option and minimizes the total shipping cost.

e. Present a heuristic algorithm that finds solutions to the optimization model.

f. Develop a Python program which outputs an Excel file that contains all possible consolidations of orders. By using pivot tables of Excel, the consolidations can be sorted by the cost-saving in descending order given a delivery zone.

g. Develop an interface showing optimal delivery zones and shipping order consolidation.

1.5 Deliverables

The product deliverables include:

a. Data-driven cost analysis and estimation summary for both TL and LTL modes

b. A Python project which outputs an Excel file that contains all possible consolidations of orders of all delivery zones defined by a clustering method.

c. A dashboard that shows the options of shipping order consolidation given a clustering-defined delivery zone.

The project deliverables include:

a. Frequent communications with mentors, faculty advisors

b. Demonstration of work through reports and presentations throughout the semester

c. Final presentation of our model and cost analysis along with suggestions for the problem

1.6 Major Milestones

- Plan
  - Understand the current process of shipments via LTL and TL modes.
  - Conduct an Exploratory Data Analysis
- Cost Estimation
  - Preprocess using the Shipments data.
  - Fit a regression model to estimate TL and LTL transportation costs required for the purpose of shipment consolidation.
- Shipment Consolidation and Multi-Stop Load Planning
  - Identify the optimized delivery zones for multi-stop TL shipments by using HDBSCAN method of clustering for geospatial data.
  - Formulate an optimization model to select optimal routes for shipments through TL by minimizing the total shipping costs.
  - Present a heuristic algorithm to solve the consolidation of shipments and the multi-load planning problem.

Chapter 2: Technical Approach

2.1 Exploratory Data Analysis

Two datasets were used for initial analysis of the shipments, a production dataset and a shipment dataset. The production dataset includes information about orders ready to be shipped between September 2020 and October 2020 whereas shipment dataset includes information about shipments that were already shipped between January 2020 and August 2020. The shipment dataset consists of information regarding the shipments from the source city Laredo, Texas to destination cities, across the United States and Canada. It also provides important information related to the type of order, mode, time, and distance travelled.

In addition, a ZipCode dataset and Distance Data are used in later analysis. ZipCode dataset is an open-source dataset provided by ServiceObjects, a validation company that provides validated datasets and real-time APIs. It contains the latitude and longitude of each US and Canada Zip code which are required to identify the delivery zones. Since production data does not contain driving distance of the shipping orders, the open-source Google Distance Matrix API is used to get the driving distance from origin and destinations. This is required for cost computation in the heuristic algorithm.

Initial preprocessing on the shipment dataset was performed such as removing irrelevant columns, deleting duplicates, removing missing data, and combining the claims column. The final dataset had the following information:

- Shipment ID, Order ID
- Mode: LTL and TL
- Shipping Cost, Weight, Volume, Distance, Cost of Claims
- Shipping Date, Delivery Date
- Destination information: City, State, Zip code

A histogram function from the *matplotlib* package is used to visualize the distribution of the shipping cost across the two modes of shipment.

The following graph shows that TL shipment evenly distributed in the higher cost range (2000, 6000) and most of the LTL shipments have a lower cost. This emphasizes the fact that a single TL shipment outweighs the cost of a single LTL shipment.

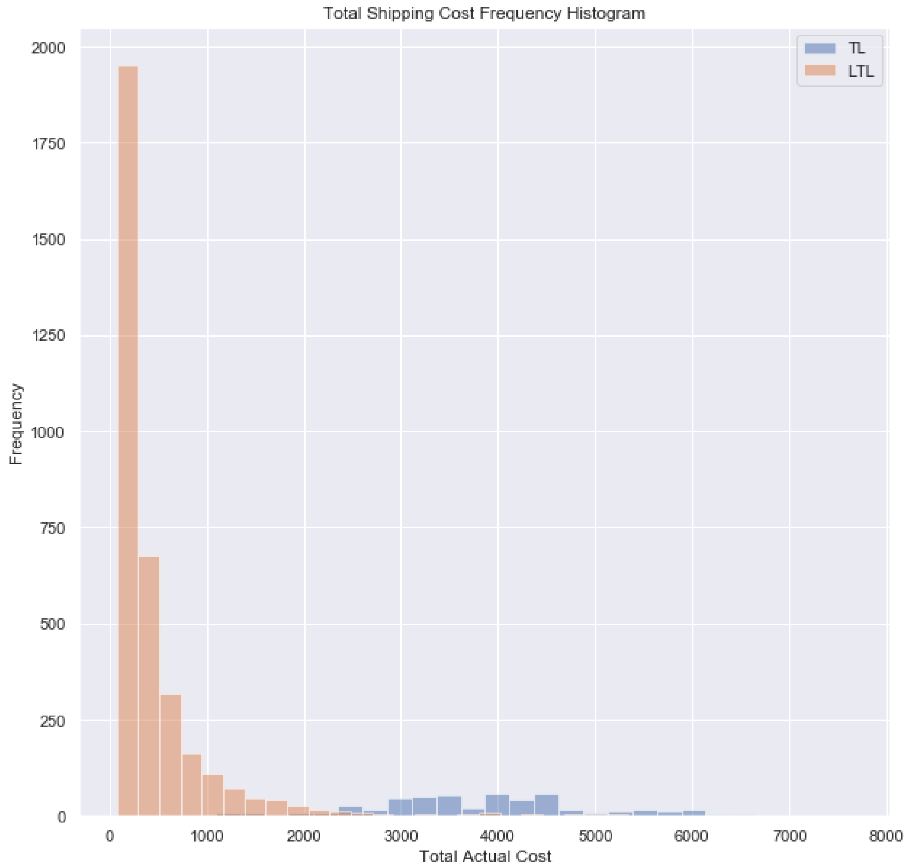

Figure 1: Histograms of Total Shipping Frequency

The following time series graph provides information about the shipping cost via TL mode per unit weight of shipments. It can be said that when large shipments are using TL mode, it incurs a lower cost than shipping through separate LTL mode. However, a spike in the month of September may occur due to the transportation of smaller individual shipments through TL.

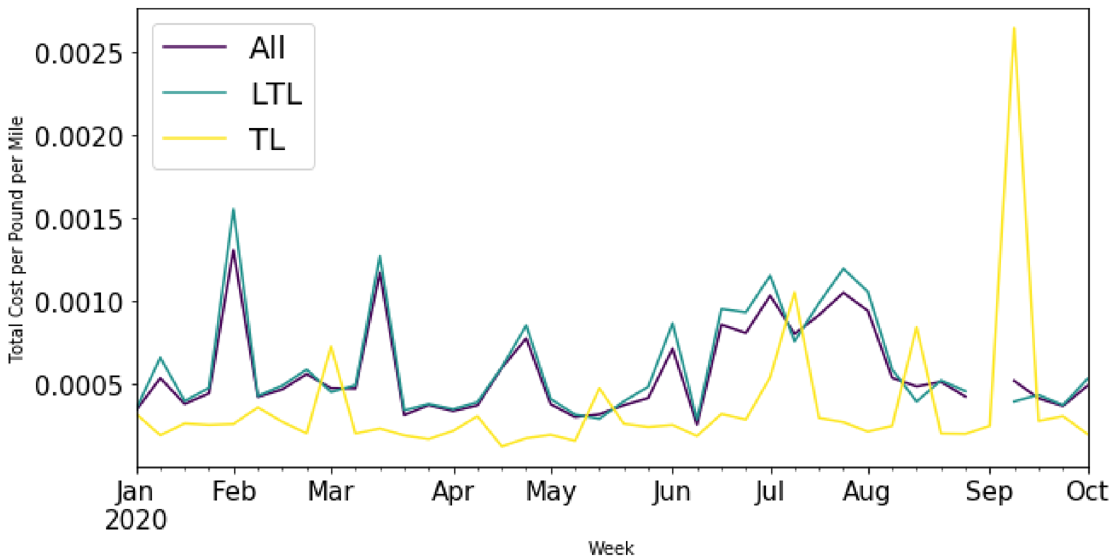

Figure 2: Time Series of Total Cost per pound per mile

The following map gives a detailed breakdown of the major geographical regions having a high frequency of shipments. This gives an idea of the optimal zones wherein TL shipments be concentrated. It can be concluded that the majority of Daikin customers are from the Midwest and east coast of the United States. However, Daikin also has a fair number of customers from the extreme west coast of California.

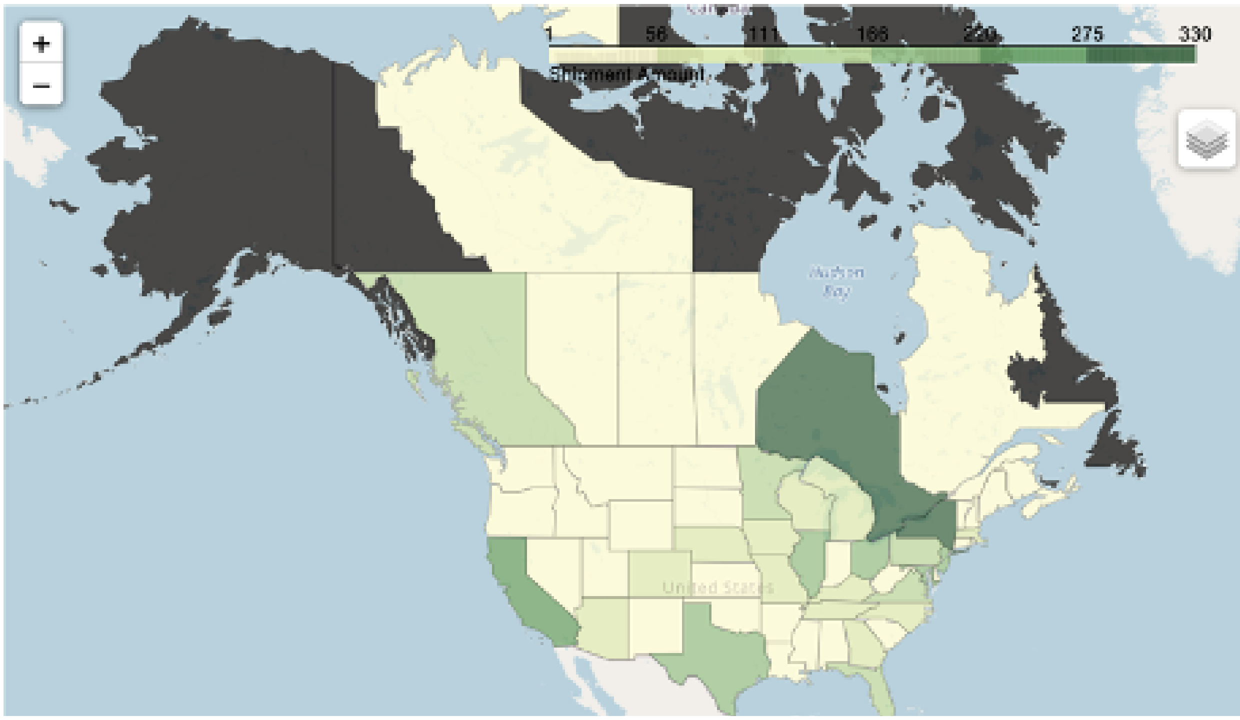

Figure 3: Choropleth Maps of Shipment distribution by states

1.4 Regression Modeling

To find the shipping cost of future shipments, a cost model is required to estimate the total shipping cost for both modes of shipments. It then can be used to compare cost savings if LTL is selected or TL mode is selected for consolidated order shipments. For the purpose of this study, the cost structure is obtained from a regression model implemented on the historical shipment dataset. Various regression techniques such as Linear, Lasso, Ridge, Random Forest, and Elastic Net were performed to fit the data. R-squared is used to estimate the accuracy of the model. The higher the R squared value, the better the fit to the data. A summary of the R-squared value for all the regression models are attached below:

Table 1: R-squared value for different regression models

Model	R-squared (%)
Linear Regression	70
Lasso Regression	77
Ridge Regression	72
Random Forest	
Elastic Net	73

It is evident from the R-squared value that the Lasso Model gives the best fit to the data. However, all other R-squared values are quite comparable. When these models are implemented to estimate the LTL cost, Elastic Net provides the minimum test error rate on the production dataset. Hence, for further evaluation, the Elastic Net model is used to estimate the LTL/TL cost. This is due to the fact that the elastic net model is a regularized regression method that linearly combines the L_1 and L_2 penalties of the lasso and ridge methods. Hence, it takes into account all the limitations of the other two methods.

A summary of the coefficients of the predictor variables from Elastic Net model are attached below:

Table 2: Coefficient of predictor variables

Parameters of Shipments	Coefficients
Intercept	-613.12
Mode (1 if using TL mode, 0 otherwise)	977.09
Distance	0.2839
Weight	0.0614
Volume	0.0003

Hence, the final model used for the purpose of estimating the shipping costs is defined by:

$$Cost =- 613.12 + 997.09\times(Mode = TL) + 0.2839\times Distance + 0.0614\times Weight + 0.0003\times Volun$$

2.3 Optimal Delivery Zones

The optimal delivery zones are specific geographical regions wherein a multi-stop truckload mode of shipments is concentrated. It is important to find optimal delivery zones because of the fact that a specific truck cannot travel destinations from one part to other parts of the country. Hence to ensure feasibility, optimal zones are found. Prior to delivery, Daikin would have the knowledge of the production schedule which is ready to be delivered. Hence, the clustering of regions is performed on the destinations provided on the production dataset. Moreover, a 5-day planning horizon is considered for the purpose of this study. This is due to the fact that each of the shipments' shipping dates are scheduled prior depending upon the customer's requirement. Hence, Daikin would want to keep a buffer of 5 days which means any scheduled delivery should be shipped no later than 5 days from the original promise date. A 5-day shipment schedule from September 21st to September 25th is selected from the production dataset.

For the purpose of finding the optimal delivery zones, various clustering algorithms were tested such as K-means clustering, DBSCAN and HDBSCAN algorithms. However, due to the geospatial nature of the data, HDBSCAN algorithm was able to classify most accurately.

HDBSCAN stands for *Hierarchical Density-based Spatial Clustering of Applications with Noise*, an unsupervised machine learning algorithm. It is a non-parametric method that looks for a cluster hierarchy which are shaped by multivariate modes of the underlying distribution. Instead of looking for clusters with a particular shape, it looks for regions of the data that are denser than the surrounding space. Hence, this algorithm is better suited for geospatial data. The parameter "eps" is set at 300 miles (by default) which denotes the radius of the neighborhood. The parameter *MinPts* is the minimum number of neighbors within "eps" radius. It is set at 3 (which comes from the fact that at most three multi-stops are to be included in a truckload mode of transportation).

Latitudes and longitudes of the destinations were computed using a ZipCode dataset provided by ServiceObjects, a validation company that provides validated datasets and real-time APIs. This was the input to the HDBSCAN algorithm. The HDBSCAN algorithm was implemented separately for shipments to be delivered across the United States and Canada.

For the purpose of visualizing the clusters on a real time map, the dataset containing the clustered zones was uploaded on a google maps platform. Each colored zone represents an optimal delivery zone where all truckload shipments for a multi-stop method are concentrated.

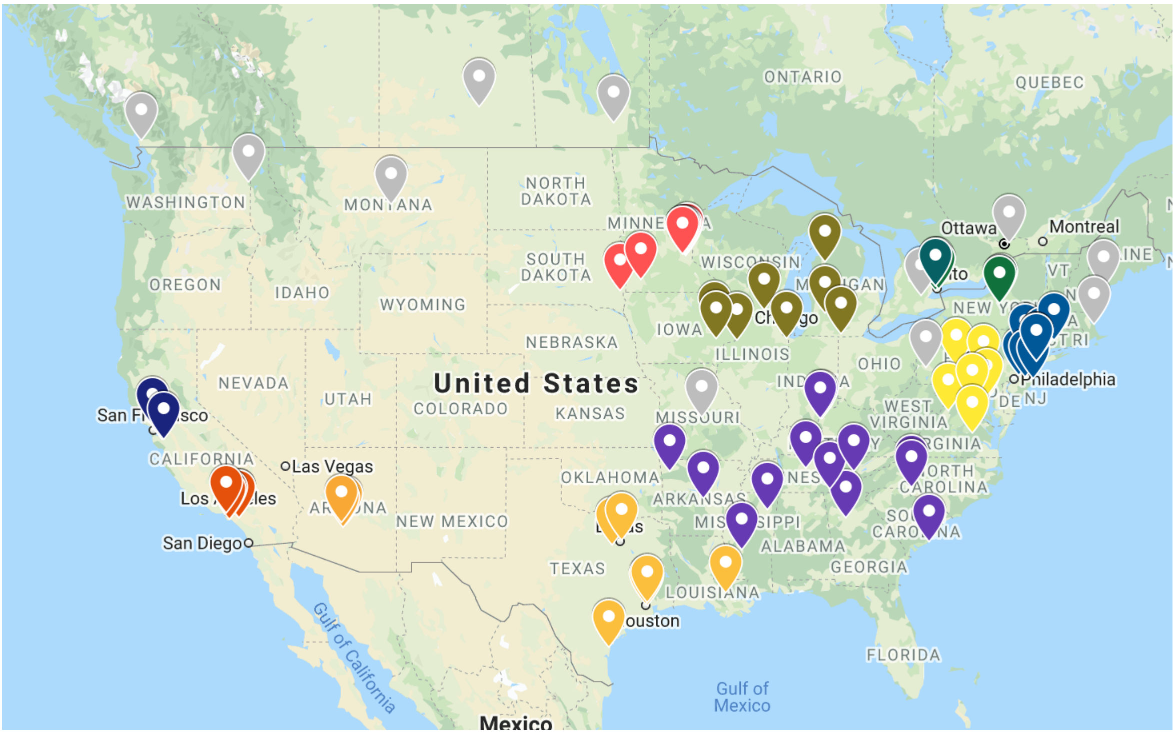

Figure 4: Optimal Delivery Zones across the United States and Canada for shipments from 21st September to 25th September

Here, the grey colored locations do not belong to any zone and hence the shipments belonging to these destinations are not a part of the multi stop TL delivery mode of shipments.

2.4 Optimization Model

In this section, a mathematical formulation is presented. The objective is to minimize the total costs such as LTL and TL cost. In addition, the volume constraints of the truck are considered. The weight parameter is not considered since it is assumed that the weight of the shipments in one full truck never exceeds the truck weight capacity. Definitions of variables and a mathematical model are presented in the next subsections.

2.4.1 Variables and Parameters Definitions

In order to develop a mathematical model, the notations of variables and parameters are given as follows:

Parameters:

$N = total\ number\ of\ shipments$

$V_i = volume\ of\ order\ i$

$TL = capacity\ of\ the\ truck$

$C_i^{LTL} = Cost\ to\ send\ order\ i\ via\ LTL$

$C_{ij}^{TL} = Cost\ to\ send\ order\ i, j\ via\ TL$

$C_{ijk}^{TL} = Cost\ to\ send\ order\ i, j, k\ via\ TL$

Decision Variables:

$x_i =$ $a\ binary\ type\ having\ a\ value\ equal\ to\ 1\ if\ shipment\ i\ goes\ via\ LTL\ and\ 0\ otherwise,$

$i \in N$

$y_{ij} =$ $a\ binary\ type\ having\ a\ value\ equal\ to\ 1\ if\ shipment\ i, j\ goes\ via\ TL\ and\ 0\ otherwise,$

$i, j \in N\ i \neq j$

$z_{ijk} =$ $a\ binary\ type\ having\ a\ value\ equal\ to\ 1\ if\ shipment\ i, j, k\ goes\ via\ TL\ and\ 0\ otherwise,$

$i, j, k \in N\ i \neq j, j \neq k$

2.4.2 Model Formulation

The integrated shipment consolidation with volume constraints can be formulated as:

Minimize

$$\sum_{i=1}^{N} C_i^{LTL} + \sum_{i=1}^{N} \sum_{j=1,\, j\neq i}^{N} (C_{ij}^{TL} + 150)\times y_{i,j} + \sum_{i=1}^{N} \sum_{j=1,\, j\neq i}^{N} \sum_{k=1,\, k\neq i, k\neq j}^{N} (C_{ijk}^{TL} + 300)\times z_{i,j,k}$$

(2.1)

Subject to:

$$x_i + \sum_{j=1, j\neq i}^{N} y_{i,j} + \sum_{j=1, j\neq i}^{N} \sum_{k=1, k\neq i, k\neq j}^{N} z_{i,j,k} = 1 \qquad \forall i \qquad (2.2)$$

$$y_{i,j}(V_i + V_j) < TL \qquad \forall i, j,\ i \neq j$$

(2.3)

$$z_{i,j,k}(V_i + V_j + V_k) < TL \qquad \forall i, j, k,\ i \neq j \neq k$$

(2.4)

In this formulation, the objective function (2.1) minimizes the total costs of shipping. The first term in (2.1) is the cost of shipping if a shipment is transported via LTL mode of shipment. The second term in (2.1) is the shipping cost involved if two shipments are transported through a single truck via multi-stop TL delivery including the cost of stops equals to 150. The third term in (2.1) is the shipping cost involved if three shipments are transported via multi stop TL delivery including the cost of stops equals to 300. For this study, only three shipments with two intermediate stops are considered for the multi stop TL mode of shipment.

Constraints (2.2) ensures that each shipment is assigned to either a direct TL mode or with one or two different shipments via TL mode or with LTL mode. Constraints (2.3) and (2.4) ensures that the consolidated shipments do not exceed the capacity limit of the truck. For the sake of simplicity, only volume constraints are considered. The constants added are due to the charges incurred at stops. Due to the rectangular nature of the shipments, it is assumed that the geometric orientation of the shipments is preserved when volume is taken into account. This means, the total volume considered will accommodate the desired length, width and height of the shipments consolidated.

Due to the complex nature of the optimization problem, a heuristic approach is provided in the next section to get good solutions. The heuristic approach is derived from this optimization model.

2.5 Heuristic Algorithm

To solve the shipping order consolidation problem, a heuristic algorithm is developed based on the optimization model. The algorithm enumerates all possible shipping orders consolidation of two and three orders. The consolidations that do not exceed the TL truck capacity are considered as feasible consolidations. Daikin uses standard 53 ft long trailers for TL deliveries, whose volume capacity is 3,489 cu ft.

Here, Algorithm 1 shows how the algorithm traverses the shipping orders and considers every possible solution of consolidation.

A Python script is created to implement this algorithm. Finally, the program sorts the order consolidation options for each zone by cost-saving in descending order, which works as a recommendation list of shipping order consolidation.

2.6 Dashboard

The dashboard is used to effectively visualize the possible solution of consolidations. The purpose of data analytics is not only to solve the problem but also to present data more natural for the human mind to identify trends easily. The order consolidation options are displayed through the Tableau, which is one of the widely used visualization tools. As observed, the graph of the Orders with Cost Savings, the amount of cost savings can be recognized in terms of the set of orders. Furthermore, the Recommend Zone graph ranks the average cost saving depending on the zones, so users can intuitively understand which zone has an edge on TL shipping.

Figure 5: Tableau Dashboard

Chapter 3: Results and Conclusion

3.1 Conclusion

The recommendation list of shipping orders could be identified by developing the heuristic model. The top results are the shipments consolidations which have the highest cost savings when transported through one truck. The results show that LTL cost is relatively low near the Texas area, meanwhile there are a lot of cost savings near the New York region. It is observed that shipments having fewer orders to an isolated zone should be transported through LTL. The truckload shipping cost increases if trucks are not filled to its maximum capacity. Hence, consolidation of multiple shipments is necessary. Furthermore, when the set of orders have volume close to the truck capacity, consolidation works best as it has more chance to bring larger cost savings. Costs incurred due to claims and damages were not included in the scope of the study. This is due to the fact that sufficient data was not available to come to a conclusion. However, it can be said that shipping through TL mode lowers cost of claims and damages to shipments. Hence, some shipments which have comparable cost differences when shipped via LTL, can be shipped through TL due to low possibility of damages to shipments.

ZONE	ORDER 1	ORDER 2	ORDER 3	LTL COST	TL COST	COST DIFF
18	6023446.0	6023692.0	6023843.0	8290.666406049150	7475.131429237210	815.5349768119400
18	6023446.0	6023692.0	6023846.0	8220.359512807720	7512.997340059990	707.3621727477220
18	6023446.0	6023843.0	6023846.0	2464.4080918319900	1757.045919084270	707.3621727477200
18	6023446.0	6023692.0	6023822.0	8026.024201240150	7506.103335035220	519.9208662049240
18	6023446.0	6023822.0	6023843.0	2270.0727802644200	1750.151914059500	519.9208662049220
18	6023446.0	6023822.0	6023846.0	2199.765887022990	1788.0178248822800	411.7480621407040
18	6023446.0	6023692.0		7611.185114423680	7258.056372184100	353.12874223958500
18	6023446.0	6023843.0		1855.2336934479600	1502.1049512083700	353.12874223958500
18	6023446.0	6023846.0		1784.9268002065200	1539.9708620311500	244.95593817536700
18	6023446.0	6023822.0		1590.591488638950	1533.0768570063800	57.51463163256840

Figure 7: Final order consolidation with their highest cost savings

The shipments with order ID 6023446, 6023692, 6023843 are shipped in one truck via two separate intermediate stops TL delivery. The total cost in this case is 7475 $ approximately. Whereas, if these three orders are shipped in multiple LTL deliveries, it will incur a total cost of 8290 $. Hence, TL deliveries amount to around 10.9 % cost savings. The same shipment amounts to 4.6% cost savings when transported in a one stop intermediate TL delivery. Hence, this type of consolidation achieves significant transportation costs savings for Daikin.

3.2 Assessment

The project performance was evaluated based on the original objectives.

a) **Data-driven cost analysis:** Successful. The project team identified the Elastic-net regression model to calculate the LTL and TL shipping costs instead of using Czarlite.

Estimation of shipping cost with Czarlite should be calculated individually, but Elastic-net regression allows for users to get the multiple shipping costs at one time.

b) **Optimal Delivery Zone and Consolidation:** Extremely successful. At the end of several clustering methods trials, HDBSCAN algorithm showed the best optimized clustering. By applying the heuristic algorithm based on the each zone, the team was able to get the recommended consolidation order.

c) **Dashboard:** Successful. The options of shipping order consolidation given a clustering-defined delivery zone could be visualized as graphs in the Tableau.

d) **Communications**: Extremely successful. Even though zoom was the only method to communicate, the project team could hold the meeting following the meeting plan.

e) **Team Assessment:** Successful. Thanks to each member's unique talent, the project could be completed using the various analytic skills.

f) **Sponsor Assessment**: Successful. The industry mentor gave the positive feedback from the results.

3.3 Impact of the Project

This project was a great learning opportunity for us. It gave us an insight to the world of logistics and to apply the data analytics skill into a real-world problem. This project brings primary impact to the Daikin's Logistics Team. After acknowledging the demand quantities from the customers, the team implemented it to a 5-day delivery planning based on the manufactured products. This is because 5 day are the appropriate lengths for the Logistics team to make a weekly delivery schedule. By bringing the optimal delivery zone and shipping orders recommendation for multi-stop TL shipments, the logistics team will be better able to plan the schedule considering the saving of the total costs. The average of saving costs in terms of zones can give insight to the schedule planner which region should be considered to save cost at first, and the trends of opportunity costs can be utilized to forecast the future shipping strategy against the claimed cost.

Chapter 4: Lessons Learned

In this project, the project team has exploited every opportunity to enrich experience. The project team has gained knowledge and experience in supply chain management, shipment and load planning, mathematical modeling, and data analysis. Every member of the team utilized what they learned in Analytics' courses and applied the techniques and knowledge to solve a real-world shipment planning problem.

For future capstone project teams, here are some suggestions:

1. Keep communicating and present status updates with faculty advisors and industry mentors.
2. Always be prepared to take risks and be ready to change the plan.
3. Utilize experience and knowledge you learned in the past.
4. Discuss with team members, provide feedback and share your ideas.
5. Be ready to learn new skills and implement them in the project.

REFERENCES

1. Manser, R. (2020). Download Free ZIP and Postal Code Database with GeoCoordinates for US and Canada. https://www.serviceobjects.com/blog/free-zip-code-and-postal-code-database-with-geocoordinates/
2. Pooley, J., & Stenger, A. J. (1992). Modeling and evaluating shipment consolidation in a logistics system. Journal of Business Logistics, 13(2), 153.
3. Attanasio, A., Fuduli, A., Ghiani, G., & Triki, C. (2007). Integrated shipment dispatching and packing problems: a case study. Journal of Mathematical Modelling and Algorithms, 6(1), 77-85.
