

Optimizing Shipping Operations through Real-Time Monitoring and Control: A Decision Support System for Container Stripping Processes

Felix Julio^{1*}, Angelina Vanni¹, Shea Amanda¹, Vitovern Joey¹, Siana Halim¹

¹Industrial Engineering Department, Faculty of Industrial Engineering, Petra Christian University
Jl. Siwalankerto 121-131, Surabaya 60236, Indonesia

Email: felixjulio26@gmail.com, angelinavp221001@gmail.com, ferdiantishea@gmail.com,
vitovern07@gmail.com, halim@petra.ac.id

*Corresponding author

Abstract: The shipping industry plays a vital role in the global economy, with container shipping being one of the critical components. Shipping companies outline the time for customer stripping days in its contracts. The availability of the containers depends on the stripping days. The stripping days' tardiness will hinder the availability of the containers. Therefore, it is fundamental for shipping companies to monitor both the actual condition and the contract condition of stripping days to estimate container availability and prompt customers to expedite the unloading process. However, there has yet to be a tool for monitoring the actual and the contract conditions. In this study, we used the recorded container stripping data to analyze container stripping days, tardiness, and other important parameters that indicate the performance and reliability of stripping containers. These data were post-processed and analyzed using data mining methods, and the resulting information was visualized using a dashboard to facilitate quick and effortless monitoring the dashboard in this study depicts post-processed data on container stripping days and tardiness for each port of discharge, cargo, customer, and other parameters. The dashboard was constructed using Google Data Studio. As a result, the dashboard is expected to help companies monitor, control, and analyze customers with high tardiness, allowing companies to act and ensure that the number of available containers after stripping meets demand at a given time.

Keywords: Shipping logistics, data mining, container stripping days, tardiness, performance dashboard, Google Data Studio.

Introduction

The shipping industry, which includes container shipping as a critical element, plays a vital role in the global economy [1]. According to The International Chamber of Shipping (ICS), the international shipping industry carries approximately 90% of the world's trade, accounting for 11 billion tons of goods annually, equivalent to 1.5 tons per person based on the current global population [2]. Furthermore, with an average annual growth rate of 3.8%, seaborne transport is projected to surpass 16.5 billion tons by 2030 [3].

Container assets, or the containers used to transport cargo, are crucial resources that must be efficiently managed to ensure a timely and cost-effective transport of goods. One factor that can impact container assets' efficiency is container stripping processes or unloading cargo from containers. The longer it takes to strip cargo from a container, the longer it will take for that container to be available for reuse; which can negatively impact the company's overall efficiency and profitability [4]. To address this issue, many shipping companies have established contracts with their customers outlining the number of days allowed for stripping. However, monitoring and controlling the implementation of these contracts in real time has proven to be a challenge for many shipping companies. The current systems and processes often need an integrated, sound decision support system, and structured data sources to quickly identify and address the deviation from the agreed-upon stripping days. To effectively manage its container assets and optimize turnaround time, a shipping company requires a tool or a system that can monitor and control container stripping processes in a practical and real-time manner. Examples can be derived from Martius *et al.* [5] who used machine learning to forecast the worldwide empty container availability, and Gençer and Demir [6] who used mixed-integer linear programming

and scenario-based stochastic programming to optimize the empty containers. In addition, Budipriyanto *et al.* [7] used a simulation approach to solve the empty container problem.

To address this issue, the approach taken in this study involves the utilization of a dashboard – a visual display of critical information integrated on a single screen. [8]. Hapag-Lloyd also utilizes a customer dashboard to provide transparency on vessel port arrivals amidst the chaos caused by Covid-19 congestion on the US West Coast, Port of Yantian, and the obstruction of the Suez Canal [9]. Other researchers have designed dashboards for other wide-ranging aims such as monitoring shipping container repairment [10]. Thus, the dashboard has been applied in various fields.

In this study, the dashboard will be created using data mining methods and visualization techniques and implemented using Google Data Studio [11] and R software [12]. While there is an existing research on using dashboards in shipping operations, much of it focuses on operational metrics rather than strategic ones. In addition, the research is not fully integrated with the company's systems and processes. This study aims to fill this gap by developing a more strategic, integrated, and effective dashboard for monitoring and controlling container stripping processes. The strategic dashboard is aligned with the company strategy and goals, i.e., to track and monitor the real-time stripping process. The designed dashboard presents the key performance indicators (KPIs) to achieve the goal. Additionally, it integrates various datasets, i.e., the customers, marketing, and liner planner datasets. The designed dashboard effectively facilitates data interpretation and comprehension among users, owing to its clear and concise presentation of information.

Furthermore, this study will examine the best practices for implementing and using dashboards in shipping line operations. The Google Data Studio Dashboard can track and monitor the real-time stripping process. The supervisors can predict the availability of the containers and make some policies if the container demands exceed the availability. Overall, developing and implementing a dashboard for monitoring and controlling container stripping days may significantly improve a shipping company's efficiency and profitability by allowing real-time tracking and management for its container assets.

Liner Terminology

Stuffing and Stripping

Stuffing and stripping are activities related to logistics services that involve the unloading and loading of goods from and into shipping containers [13] respectively. Stripping refers to the process of removing goods from shipping containers or containers. Inbound stripping involves removing goods from shipping containers or containers within the warehouse. In contrast, outbound stripping refers to the removal of goods from shipping containers or containers, typically performed outside the warehouse or at the customer's location.

Stuffing, on the other hand, refers to the loading of goods into a container. When goods are loaded or inserted into shipping containers within the warehouse, it is called inbound stuffing. In contrast, when the loading or insertion of goods into shipping containers occurs outside the warehouse or at the customer's location, it is referred to as outbound stuffing. These activities are essential for logistics and supply chain management as they ensure that the goods are adequately loaded, protected, and transported to their destination.



Figure 1. Stuffing process



Figure 2. Stripping process

Port of Loading (POL) and Port of Discharge (POD)

The Port of Loading (POL) is where goods (containers) are loaded onto a ship. It serves as the starting point of the cargo's journey and is typically situated at its place of origin. The cargo is usually loaded onto the ship by a crane or other loading equipment, and the necessary shipping documentation is completed at this point.

The Port of Discharge (POD) is where the ship voluntarily and without any required reason discharges part or all of its cargo [14]. It is the final destination of the cargo's journey and is typically located at the shipment's intended arrival point. The cargo is usually unloaded from the ship by a crane or other unloading equipment, and the necessary shipping documentation is completed at this point. The Port of Discharge is also where customs clearance and other import formalities take place. The Port of Loading and Port of Discharge are essential pieces of information for logistics and supply chain management as they determine the cargo's route and the estimated time for it to reach the destination.

Methods

Tardy

Tardy is an adjective that means late or behind schedule. In a contract, tardiness could refer to lateness or delay in meeting the contract's stipulated terms. The tardiness in fulfilling the terms of a contract can result in various consequences, including financial penalties of even termination of the contract [15].

Data Mining

Data mining is discovering interesting patterns, models, and other types of knowledge in an extensive data collection. Data mining can also be interpreted as a semi-automatic process for extracting and identifying potentially practical and beneficial knowledge stored in databases using statistical, mathematical, artificial intelligence (AI), and machine learning techniques [16]. Therefore, data mining is often regarded as a stage within the broader process of acquiring knowledge. However, the data obtained must be prepared by performing data cleaning, integration, transformation, and selection to reach the data mining stage. Data cleaning aims to remove inconsistent and outlier data. Data integration aims to combine or merge data from various sources. Finally, data transformation is the process of adapting the data's form to fit the requirements of data mining, often through aggregation or summarization [17] [18].

K-Means Clustering

K-Means clustering is the most commonly used unsupervised machine learning method for partitioning a given dataset into groups [19]. It is a popular choice due to its ease of implementation, capability to handle large datasets, and its ability to facilitate understanding and communication of findings. Moreover, it is fast, efficient, and suitable for real-time applications [20]. The analyst can determine the number of groups that emerge beforehand, represented by "k." This classification is performed to ensure that objects within one cluster are as similar as possible, while objects from between clusters are made as different as possible. Each cluster will be represented by the average points given to the cluster, also known as the centroid. The standard k-means algorithm is the Hartigan-Wong algorithm. This algorithm defines the total variation within a cluster as the sum of the Euclidean Distance squares between objects and the corresponding centroid [19].

The Hartigan-Wong algorithm is a classical clustering algorithm, which was developed by Hartigan and Wong in 1979 [21]. The basic concept of k-means is choosing k clusters $\{C\}_1^k$ which minimizes:

$$\sum_{j=1}^k \inf_{y_j \in \chi} \sum_{i \in C_j} \|x_i - y_j\|^2 \quad (1)$$

Where $\chi = R^d - d$ -dimensional real number, $\| \cdot \|$ denotes the Euclidean norm. $\{C\}_1^k$ is the possible cluster from $j = 1, \dots, k$, and k is a given parameter [22].

Dashboard

Visualizing or presenting data in a visually appealing manner will effectively help analyze the data. This analysis serves various purposes, such as conducting an initial investigation, confirming or disputing data models, and explaining it with mathematical or algorithmic concepts [23]. A dashboard is a visual or graphic representation of information. The purpose of creating a dashboard is to assist organizational managers in

planning, designing, implementing, and organizing questions to identify work. Dashboards are currently widely used for real-time monitoring and analyzing business processes [24]. They also help users make effective decisions [25]. However, for this to occur, the design and displayed results must provide transparent information. In many cases, dashboards fail to effectively and efficiently communicate because of poor design implementation, e.g., the dashboard is cluttered with too much information on features, lack of context, inconsistent design such as different fonts, colors, or layout, insufficient information in the visualization, and lack of customization [25].

Yap [26] provides suggestions to improve the design of a dashboard, which include the following: Redundant visual representation, ensuring a clear layout of input and output objects, using colors for indicators, storytelling and creating visual synergy, and defining the scope of the application and range of the data.

Data Preprocessing

This study's data is retrieved from one of the shipping liner company's internal operational data related to container stripping processes, which consists of 26 variables and the customer's contract data. First, the data preprocessing initializes, formatting the raw data into the same format following the annex. Subsequently, it cleans the dataset, joining the stripping and contract data to carry on the data processing process later. The data preprocessing was carried out using R.

As per the request of the shipping liner, the dashboarding was built using the free version of Google Data Studio (GDS). However, the GDS free version has some limitations. For instance, the uploaded file is subject to a 100 MB file size limit per data set [27]. Consequently, the original dataset will be aggregated and pre-calculated using R. The concise datasets will then be uploaded to GDS. These steps will also speed up the GDS performance.

Results and Discussions

Data Processing

The primary dataset comprises 26 variables, including various features such as booking number, customer data, start and end dates of the stripping process, etc. The data was collected over nine months and comprised a total 682,486 rows of data. Additionally, there is a supporting dataset which contains information about customer contracts with the company. This dataset includes details about the stripping deadline or free time provided by the company to the customer for the stripping process. During the preprocessing stage, the supporting dataset was joined with the primary dataset.

The initial step in data processing involves subtracting the start date from the end date of the stripping process, resulting in the stripping time. We compared the stripping time to the deadline to compute the tardiness or delay; then, we aggregated various statistical measures such as mean, standard deviation, and frequency for unique characteristics such as POL, POD, size, cargo, and customer. This aggregation helps us to reveal hidden insights associated with specific unique attributes. Since the frequency of each unique characteristic may vary, the aggregation results are properly weighted to ensure a fair calculation of the mean.

Weighted Mean Tardy

First, the mean and standard deviation aggregation of stripping days, tardiness, and frequency aggregation are calculated for each specific port of discharge (POD), loading (POL), size, cargo, customer, and grade. Next, the frequency aggregation is calculated for customers sharing the same POD and POL. Finally, the weighted stripping day and weighted tardiness are defined as the multiplication of aggregation means and aggregation tardiness times the frequency consecutively.

Clustering

The clustering process aims to group the data based on the average tardiness and stripping volume in TEUs (twenty-foot equivalent units) and explore the characteristics of each group. Since those two features have different measurements, we first standardized them. This standardization prevents any attribute from dominating another attribute due to significantly different values [28]. Then, we deduced the optimal number of clusters based on the within-cluster sum of squares (WCSS) of the standardized dataset. The within-cluster sum of squares (WCSS) measures the total distance between each data point and its corresponding cluster center. The objective of K-means clustering is to minimize this distance, indicating an aim to minimize the WCSS for a given number of clusters [29].

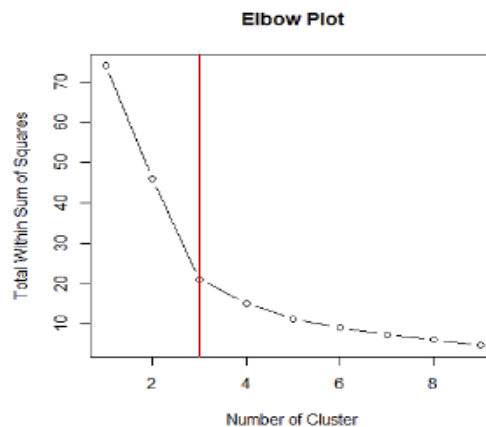


Figure 3. Elbow plot (Container 20 DC grade A)

In this study, we employ K-means clustering as it is essential to consider the characteristics and underlying patterns of the data under analysis [30]. In this research, one approach that can provide more valuable insights is to split the analysis based on the container size (20 DC & 40 HC) and grades (A, B, & C). This approach helps reveal patterns and specific trends to different container sizes and grades. It can also reveal any unseen relationships that may not be apparent when analyzing the entire dataset, enabling a more focused analysis that aligns with the business goals of users [20]. To determine the optimal number of clusters, we can use the elbow method,

which involves plotting the WCSS as a function of the number of clusters and identifying the "elbow" point in the plot where the rate of WCSS reduction begins to level off. This point indicates a good trade-off between the number of clusters and the WCSS, as increasing the number of clusters beyond this point would result in diminishing returns in terms of WCSS reduction [31]. However, it is essential to note that choosing the optimal number of clusters is subjective. The current approach shows that each splitting dataset has three clusters. Figure 3 exhibits an example of an elbow plot for Container 20 DC grade A. Finally, clustering using K-means is performed based on these two features.

Dashboard Design

The dashboard for comprehensive and in-depth analysis of the stripping process comprises three pages, each offering interactive features that allow users to quickly identify key performance indicators such as the average stripping days, stripping frequency, and tardiness. The first page of the dashboard (Figure 4) is a visual that displays this information through a combination of columns, bubble maps, scorecards, doughnut charts, and gauge charts. The columns are divided into three sections, namely POD, Cargo, and Customer, making it simple to pinpoint which ports, cargo, and customers exhibit the highest stripping frequency and tardiness. The bubble map feature is convenient as it allows users to visualize the stripping frequency in terms of bubble size and tardiness in terms of bubble color across all ports. This feature provides valuable insights into regional trends and patterns. The bubbles on the map are color-coded, with greener bubbles representing negative average tardy values, indicating that the stripping was done earlier than the scheduled time. Conversely, bubbles closer to red represent more negative average tardy value, indicating that the stripping was completed later than the scheduled time.

The scorecard feature provides a quick overview of the total stripping frequency and total number of customers, making it easy to identify overall performance trends. In addition, the doughnut chart, which breaks down customers, allows users to see which customers contribute the most to the stripping activity. This information is valuable for identifying opportunities to improve customer relationships. Finally, the gauge chart for the global average stripping days allows users to quickly acknowledge the stripping activity's overall performance. Users can also easily customize the display on this page by utilizing the slicer feature. They can filter the data based on container size, POD, container class, segmentation, and month. This feature lets users focus quickly on specific aspects of the data that are most relevant to their needs, allowing them to make more informed decisions.

The second page of the dashboard (Figure 5) provides stakeholders with a visual representation of the clustering results, enabling them to identify patterns and categories in the data. The bubble chart feature allows users to

quickly identify areas with high stripping frequency and high average tardiness. Similar areas with similar characteristics are grouped into clusters and assigned a uniform color to be easily distinguished from one another. The color of the bubble distinguishes it from other clusters. Additionally, it can help them prioritize resources and make data-driven decisions based on the tendencies of specific combinations of size and container grade. As the stakeholders can focus on specific regions using the POD dimension, they are also allowed to identify patterns and trends in the data that may not be apparent when looking at the data as a whole, allowing them to make more informed decisions. This display can help users to optimize their performance and increase the efficiency of their operation.

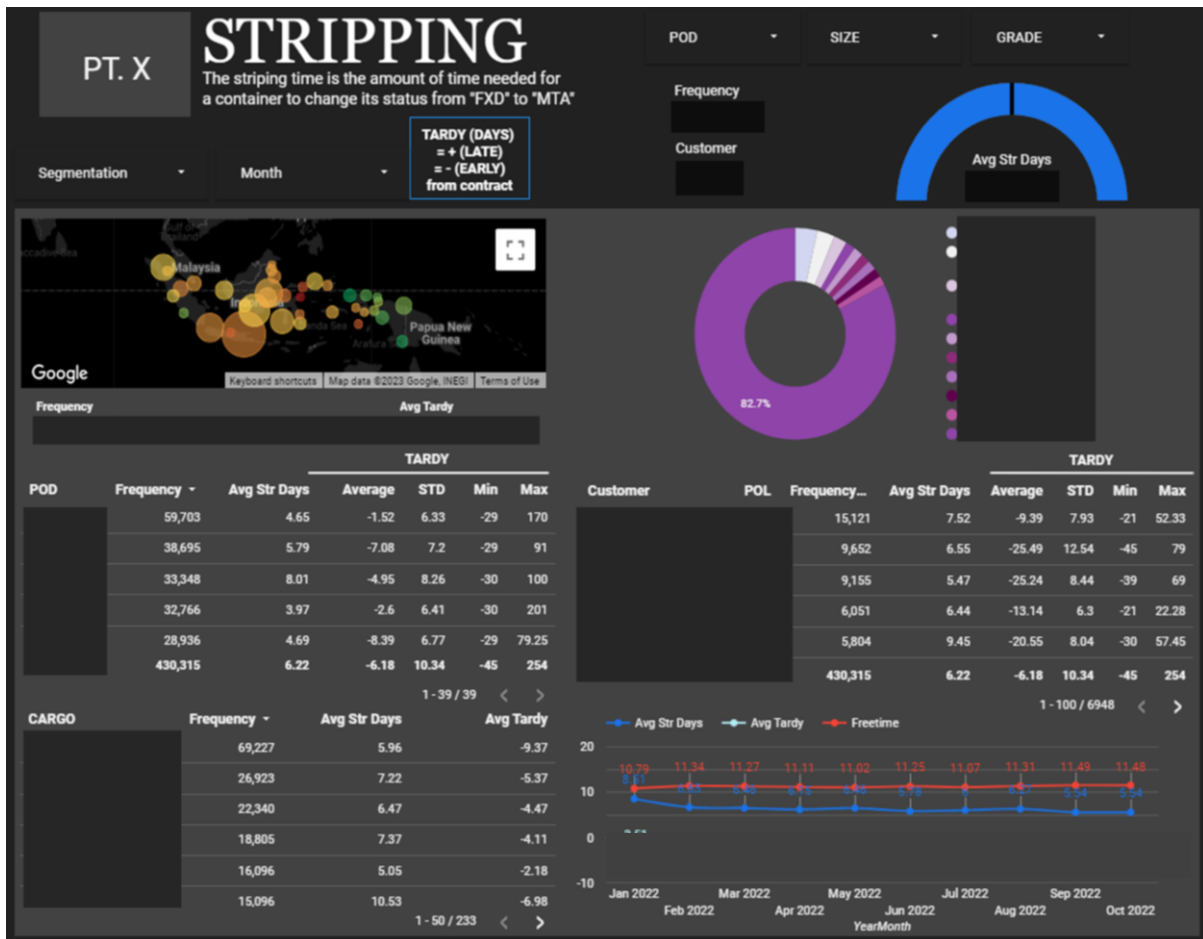


Figure 4. First page of the dashboard

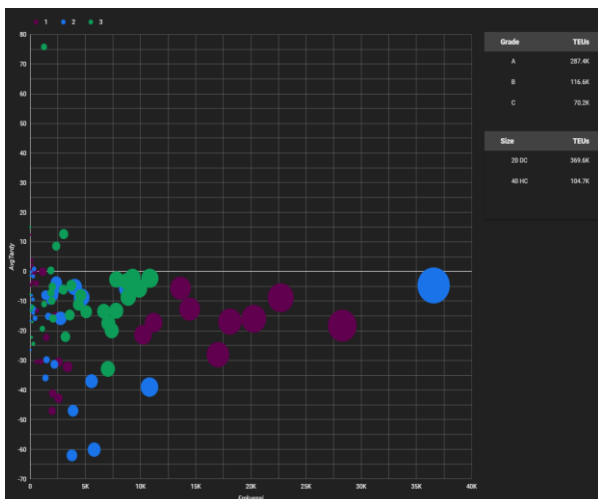


Figure 5. Second page of the dashboard

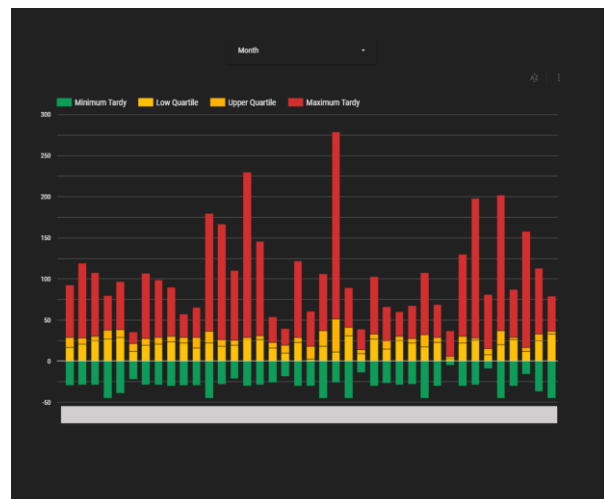


Figure 6. Third page of the dashboard

The last page of the dashboard (Figure 6) employs a stacked combo chart that includes the POD dimension, minimum, maximum, lower quartile, and upper quartile metrics for tardiness. This feature benefits stakeholders by allowing them to review tardiness data more comprehensively. For example, they can analyze the average tardiness and the variance of tardiness within each POD. Additionally, by using the slicer to filter the data by month, users can identify trends and observe tardiness trends over time. Furthermore, a smaller variance/standard deviation indicates a narrower data distribution, which can help inform decision-making and contribute to improve overall performance. The color aspect will help users to identify the range between minimum and maximum tardiness, enabling them to comprehend the tardiness data quickly and highlighting areas that require improvement. Red denotes the maximum tardiness, while green denotes the minimum tardiness. Quartiles are marked with yellow.

Discussion

The cluster summary, obtained after carrying out the clustering process based on stripping volume and average tardy, provides a detailed breakdown of the characteristics of each cluster. This summary is shown in Tables 1-6. These tables typically include information such as the number of observations in each cluster, the mean and standard deviation of the clustering variables (stripping volume in TEUs and average tardy, in this case), and any other relevant characteristics of the observations within each cluster. This information enables a deeper understanding of the patterns and similarities within the data and facilitating further analysis or decision-making.

As seen from Table 1-3, it is clear that the container stripping process for 20 DC containers (grades A, B, and C) tends to complete earlier than the contract's limit time, as indicated by the negative tardy values. However, regarding the 40 HC containers, particularly those in grades B and C, it is evident from Table 5-6 that several clusters display positive average tardy results, indicating a tendency towards delayed stripping processes. This finding can provide valuable insight for the company, suggesting that improving the container stripping process for these 20 DC containers may be a low priority. Therefore, it may be beneficial to further investigate the causes of delay in the 40 HC Grade B and C containers by consulting Tables 7 and 8.

Table 7 illustrates a significant need for improvement in container stripping for 40 HC Grade B containers, particularly in the few PODs identified as ID AAA, ID BBB, ID CCC, and ID DDD, as indicated by the high level of tardiness shown in the table. Similarly, Table 8 suggests that the primary concern for improving the container stripping for 40 HC Grade C containers is in the PODs of ID AAA, and ID BBB. Notably, a particular focus on ID AAA and ID BBB is warranted as this POD has consistently exhibited longer tardiness than other ports. Hence, immediate attention is required to improve overall efficiency and reduce delays in the container stripping process.

Table 1. Clustering container 20 DC grade A

Cluster	TEUs	Avg tardy (days)
1	119,328	-5.66
2	49,787	-3.75
3	30,980	-16.28

Table 2. Clustering container 20 DC grade B

Cluster	TEUs	Avg tardy (days)
1	51,863	-4.06
2	21,454	-1.11
3	12,887	-17.27

Table 3. Clustering container 20 DC grade C

Cluster	TEUs	Avg tardy (days)
1	14,644	-1.74
2	17,937	-3.11
3	2,911	-17.73

Table 4. Clustering container 40 HC grade A

Cluster	TEUs	Avg tardy (days)
1	3,250	-1.74
2	39,670	-3.11
3	6,182	-0.93

Table 5. Clustering container 40 HC grade B

Cluster	TEUs	Avg tardy (days)
1	12,578	-5.28
2	940	2.53
3	2,034	-11.91

Table 6. Clustering container 40 HC grade C

Cluster	TEUs	Avg tardy (days)
1	6,894	-8.39
2	19,196	-4.67
3	228	17.75

Table 7. Cluster 2 container 40 DC grade B

POD	Avg tardy (Days)	Frequency	TEUs
ID AAA	21	14	28
ID BBB	7.68	38	76
ID CCC	6	29	58
ID DDD	3.88	40	80
ID FFF	1.31	139	278
ID GGG	0.54	191	382
ID HHH	-1.13	16	32

Table 8. Cluster 3 container 40 HC grade C

POD	Avg tardy (Days)	Frequency	TEUs
ID AAA	28.84	56	112
ID BBB	6.95	57	114

To ensure that the proposed dashboard is genuinely effective in monitoring, controlling, and analyzing customer data, it is crucial to conduct user testing and evaluations as part of future research. User testing, which involves gathering user feedback, is valuable for understanding satisfaction levels, ease of use, and overall success in achieving desired goals. Additionally, the dashboard's performance can be objectively measured by utilizing various usability metrics such as task completion rates, error rates, and time on task. It is essential to track key performance indicators (KPIs), such as supervisors' success in predicting container availability and making policies in case of exceeding container demands, to gain even deeper insights into the dashboard's effectiveness. However, it is critical to follow standard user experience (UX) research methodologies, such as usability testing, heuristic evaluation, and user feedback surveys, to ensure the validity of the measurement methods.

By combining these techniques, we can gain valuable insights into the effectiveness of the proposed dashboard in achieving its intended goals. Ultimately, this feedback can then be used to refine and improve the dashboard, ensuring its continuity to monitor, control, and analyze customer data.

Conclusion

The designed dashboard helps the company streamline its workflow by allowing them to clean and process data more efficiently. The dashboard also allows companies to conduct more in-depth data analysis. Based on a few samples of data monitored through the designed dashboard, it can be concluded that not all PODs (Ports of Discharge) experience delays as initially suspected. When compared to the agreed-upon contract duration, only a few PODs for certain combinations of size and grade containers experience delays. Out of the six combinations of size and grade containers, only two require the company's primary focus: 40 HC grade B containers and 40 HC grade C containers. These two combinations have a small number of volumes (in TEUs) but a relatively high average delay compared to the other combinations.

Acknowledgement

The authors sincerely appreciate and thank all reviewers for their valuable feedback, which help to improve the paper's quality.

References

- [1]. The Economic Commission for Latin America, "The shipping cycle in the international container market: Which will be the actual shipping cycle in the future of shipping, the traditional or a new one?" [Online]. Available: <https://www.cepal.org/en/notes/shipping-cycle-international-container-market-which-will-be-actual-shipping-cycle-future>., Apr. 05, 2017.
- [2]. International chamber of shipping, "International chamber of shipping: Annual review," [Online]. Available: <https://www.ics-shipping.org/wp-content/uploads/2019/06/ics-annual-review-2019-min.pdf>, 2019.
- [3]. A. Grzelakowski, "Global container shipping market development and Its impact on mega logistics system," *TransNav, the International Journal on Marine Navigation and Safety of Sea Transportation*, vol. 13, no. 3, pp. 529–535, 2019, doi: 10.12716/1001.13.03.06.
- [4]. J. Karmelić, Č. Dundović, and I. Kolanović, "Empty container logistics," *PROMET - Traffic&Transportation*, vol. 24, no. 3, pp. 223–230, Jan. 2012, doi: 10.7307/ptt.v24i3.315.
- [5]. C. Martius, L. Kretschmann, M. Zacharias, C. Jahn, and O. John, "Forecasting worldwide empty container availability with machine learning techniques," *Journal of Shipping and Trade*, vol. 7, no. 1, p. 19, Dec. 2022, doi: 10.1186/s41072-022-00120-x.
- [6]. H. Gençer, and M. H. Demir, "Optimization of empty container repositioning in liner shipping," *Business and Management Horizons*, vol. 8, no. 1, p. 1, Jan. 2020, doi: 10.5296/bmh.v8i1.16327.
- [7]. A. Budipriyanto, M. D. Novianti, and T. Susanto, "Empty container reposition using max-min review system: Simulation approach," *IOP Conference Series: Materials Science and Engineering*, vol. 1072, no. 1, pp. 012048, Feb. 2021, doi: 10.1088/1757-899X/1072/1/012048.
- [8]. M. Chen *et al.*, "Data, information, and knowledge in visualization," *IEEE Computer Graphics and Applications*, vol. 29, no. 1, pp. 12–19, Jan. 2009, doi: 10.1109/MCG.2009.6.
- [9]. Port Technology International, "Hapag-Lloyd to improve vessel transparency following covid congestion and Suez obstruction," [Online]. Available: <https://www.porttechnology.org/news/hapag-lloyd-to-improve-vessel-transparency-following-covid-congestion-and-suez-obstruction/>, Jun. 04, 2021.

- [10]. J. Oktaviani, S. Halim, and Felecia, "Dashboard design of shipping container reparation," in *presented in International Conference on Global Optimization and Its Applications 2022 (ICoGOIA 2022)*, Melaka, Nov. 2022.
- [11]. GDS, "Google Data Studio," [Online]. Available: <https://datastudio.withgoogle.com>, 2023.
- [12]. R-Software, "R-Project," [Online]. Available: r-project.org, 2022.
- [13]. H. Cahyono, and I. N. Pujawan, "Analysis of empty containers supplies in the shipping network: Case study at national shipping company," *IPTEK Journal of Proceedings Series*, vol. 0, no. 5, p. 441, Dec. 2019, doi: 10.12962/j23546026.y2019i5.6390.
- [14]. J. Enström, A. Eriksson, L. Eliasson, A. Larsson, and L. Olsson, "Wood chip supply from forest to port of loading- A simulation study," *Biomass Bioenergy*, vol. 152, pp. 106182, Sep. 2021, doi: 10.1016/j.biombioe.2021.106182.
- [15]. J.-G. Kim, H.-B. Jun, J.-Y. Bang, J.-H. Shin, and S.-H. Choi, "Minimizing tardiness penalty costs in job shop scheduling under maximum allowable tardiness," *Processes*, vol. 8, no. 11, pp. 1398, Nov. 2020, doi: 10.3390/pr8111398.
- [16]. X. Shu and Y. Ye, "Knowledge discovery: Methods from data mining and machine learning," *Social Science Research*, vol. 110, pp. 102817, Feb. 2023, doi: 10.1016/j.ssresearch.2022.102817.
- [17]. C. Fan, M. Chen, X. Wang, J. Wang, and B. Huang, "A review on data preprocessing techniques toward efficient and reliable knowledge discovery from building operational data," *Front Energy Research*, vol. 9, Mar. 2021, doi: 10.3389/fenrg.2021.652801.
- [18]. F. Coenen, "Data mining: past, present and future," *The Knowledge Engineering Review*, vol. 26, no. 1, pp. 25–29, Feb. 2011, doi: 10.1017/S0269888910000378.
- [19]. A. Kassambara, *Practical Guide to Cluster Analysis in R: Unsupervised Machine Learning*, 1st ed. Statistical Tools for High-Throughput Data Analysis, 2017.
- [20]. Y. Lei, D. Yu, Z. Bin, and Y. Yang, "Interactive K-Means clustering method based on user behavior for different analysis target in medicine," *Computational and Mathematical Methods in Medicine*, vol. 2017, pp. 1–9, 2017, doi: 10.1155/2017/4915828.
- [21]. J. A. Hartigan, and M. A. Wong, "Algorithm AS 136: A K-means clustering algorithm" *Applied Statistics*, vol. 28, no. 1, pp. 100, 1979, doi: 10.2307/2346830.
- [22]. M. Telgarsky and A. Vattani, "Hartigan's method: K-means clustering without Voronoi" in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics (AISTATS)*, Chia Laguna Resort, Sardinia Italy, 2010, pp. 820–827.
- [23]. C. Plaisant, "The challenge of information visualization evaluation," in *Proceedings of the working conference on Advanced visual interfaces*, New York, NY, USA: ACM, May 2004, pp. 109–116. doi: 10.1145/989863.989880.
- [24]. R. Toasa, M. Maximiano, C. Reis, and D. Guevara, "Data visualization techniques for real-time information- A custom and dynamic dashboard for analyzing surveys' results," in *2018 13th Iberian Conference on Information Systems and Technologies (CISTI)*, IEEE, Jun. 2018, pp. 1–7. doi: 10.23919/CISTI.2018.8398641.
- [25]. R. Brath, and M. Peters, "Why dashboard design is important," [Online]. Available: http://cs.furman.edu/~pbatchelor/csc105/articles/TUN_DM_ONLINE.pdf, Oct. 04, 2004.
- [26]. A. Y. Yap, "Creating business analytics dashboard designs using visualization methodologies: Case methods for innovative analytics pedagogy," *Information Systems Education Journal*, vol. 18, no. 5, pp. 25–33, 2020.
- [27]. Support Google, "Upload csv files to Looker Studio," [Online] <https://support.google.com/looker-studio/answer/7333350?hl=en#zippy=%2Cin-this-article>, 2022.
- [28]. D. Usman, and F. S. Stores, "On some data pre-processing techniques for K-means clustering algorithm," *Journal of Physics: Conference Series*, vol. 1489, no. 1, pp. 012029, Mar. 2020, doi: 10.1088/1742-6596/1489/1/012029.
- [29]. A. Makles, "Stata tip 110: How to get the optimal K-means cluster solution," *The Stata Journal: Promoting Communications on Statistics and Stata*, vol. 12, no. 2, pp. 347–351, Jun. 2012, doi: 10.1177/1536867X1201200213.
- [30]. M. Ahmed, R. Seraji, and S. M. S. Islam, "The K-means algorithm: A comprehensive survey and performance evaluation," *Electronics (Basel)*, vol. 9, no. 8, pp. 1295, Aug. 2020, doi: 10.3390/electronics9081295.
- [31]. S. Xu, X. Qiao, L. Zhu, Y. Zhang, C. Xue, and L. Li, "Reviews on determining the number of clusters," *Applied Mathematics & Information Sciences*, vol. 10, no. 4, pp. 1493–1512, Jul. 2016, doi: 10.18576/amis/100428.