PAPER • **OPEN ACCESS**

Concepts of Big Data Analysis of Container Terminals in the Digital Era

To cite this article: Tiago Novaes Mathias and Takeshi Shinoda 2022 *IOP Conf. Ser.: Earth Environ. Sci.* **972** 012005

Concepts of Big Data Analysis of Container Terminals in the Digital Era

Tiago Novaes Mathias[1] and Takeshi Shinoda[1]

[1] Department of Marine Systems Engineering, Kyushu University. Kyushu University, 744 Motooka, Nishi-ku, Fukuoka, Japan 819-0395

Abstract. In the last few decades, with the advance of IoT technologies throughout the most different fields have been changing the way some process is handled. However, possessing a great volume of data does not mean the operation will be handled better, in many cases the data lays in the storage and stays untouched. An increasing quantity of freight imposes container terminals to improve their capacities. Therefore, decision-making problems for operation planning, controlling, and evaluation needs to be well defined. Understanding that necessity, managers need to rely on, almost exclusively, the Terminal Operational System (TOS) to identify and solve the bother necks during daily operation. Although the TOS also had some improvements in the past years due to technological advances, it is still not enough to evaluate each operational step for the operation. This paper uses a data mining and data analysis evaluation of the container terminal's daily operation to find out specific operational issues and evaluate the main key performance indicators to keep track of the terminal efficiency.

1. Introduction

A marine terminal or port is planned to facilitate the transshipment of ships' cargo transported to and from inland locations. To provide efficient handling of cargo between ship and shore, the terminal needs to make appropriate usage of the yard equipment. The storage time in the facility should be limited, increasing the possibility for higher throughput in the terminal. Older facilities often have a large yard to storage the cargo due to the lower frequency of the carries. The concept of big facilities to store a higher number of containers is currently inconceivable, the land value, as well as the costs associated to handle the cargo inside the yard, are very high. Also, since cheaper and more frequent land transportation modes became available, this necessity for larger storage areas has decreased, consequently, the need for more efficient cargo handling increased.

Nowadays, containers account for about 60% of deep-sea cargo transportation, with rates of around 400 million TEU (Twenty-Foot Equivalent Units) per annum, increasing at an average of 8% per annum. Even though the actual container box specification has not seemed big changes over the past years, this was not happening with everything else surrounding it. Terminal handling equipment, ship, and business models have evolved as containerization became the predominant method of shipping goods from one county to another, or even domestic transfers [1][2].

Given the size of the investments involved, during the project greenfield, determine the facilities' requirements, and evaluating their economic and financial feasibility are essential for measuring the operational costs. The requirements may be based on queueing theory or computer simulations. After the terminal is built, these technologies can still be applied to evaluate the cargo handling process, but there are limitations regarding the layout or equipment type, for instance. Additionally, with the technological improvement in the past few years, new technologies can also be used in the container terminals, the Internet of Things (IoT), data mining, big data, and data analysis are some examples of new approaches that can be used to evaluate the terminal operations in its core.

A Terminal Operating System (TOS) is a system designed to support the managers to plan, schedule, and control the equipment activities of a container terminal. Therefore, to be responsible for accurate operations within the yard. These activities are the core of terminal operations, the more reliable and agile those operations are, the lower would be the costs associated with the daily base operation within the terminal yard. Since the 70s, the potential of Information Technology (IT) was quickly recognized, it has been acting as a parallel force behind developments in the port and shipping industries [3].

This paper concerns the concept behind the digital transformation of container terminals. We have developed a data mining concept to extract valuable information out of the container terminal yard daily report operations, exposing some possible TOS bother necks and adding new key performance indicators to evaluate waste of resources in the terminal. Moreover, it shows briefly how terminals could evolve and change their operational policies through the digital era.

Two major container terminals were used to verify and validate the data analysis methodology proposed. First, using a Transfer Crane system, Hakata Island City Container Terminal (HICCT). Then, with the Straddle Carrie system, Kashii PPCT. Both terminals are located close to Fukuoka International Airport at Hakata Port and have a direct link to the Fukuoka Interchange of Kyushu Expressway.

Following this introduction, which briefly justifies this paper's overall objective, section 2 reviews the literature on the container terminal's basic operations and the general concepts of data science, including data mining and data analysis. Section 3 shows the methodology applied in this research to conduct the following analysis and results in section 4. Then, section 5 draws conclusions, limitations and makes recommendations for future work in this field.

2. Theoretical Background

Information technologies are being used from many different fields to improve operations efficiency regarding daily management. Container terminals are not different, computing systems are implemented to guarantee better outcomes. Automated terminals are the state of art regarding goods transshipment between transportation modes. But, to introduce a high level of automation, the investment necessary is equally high. Then, many terminals around the globe adopt a semi or no automated system, but there is still a large amount of data generated from the terminal daily operation.

2.1. Container Terminal Operations

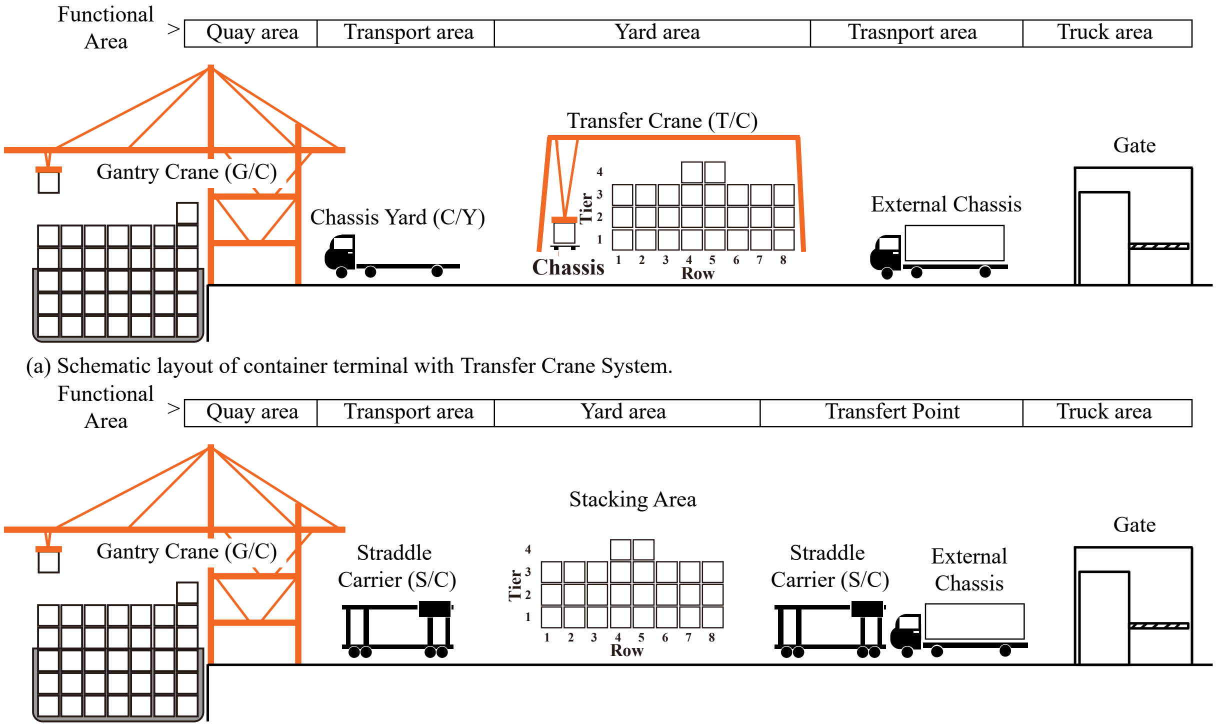
In the past years, the containerized cargo has experienced a fast growth with the hinterland and transshipment expansion occurred in ports. The container traffic growth has caused a great demand for port container terminals, requiring more investment regarding the efficiency in the terminals, intensifying competition between nearby terminals.

It is important to establish a simple definition of Container Terminal, which can be described as a link between two points, furthermore a mode within road, rail, inland waterway, and maritime traffic network. As such, containers are transhipped from and to all different transportation modes.

To ensure higher profit, terminals are searching for different methods to reduce their costs. The operational process to handle containers in the terminal yard is complex to deal with and requires multiple techniques to address the different problems found during the operations. In this sub-section, it is explained how container terminals deal with landside and seaside container operations.

Nowadays, almost every supply chain is integrated with port terminals. Some evaluation methods were developed to verify the integration between seaport container terminals in the supply chain to measure the integration effects on the market competitiveness. Even though there are some benefits in it, there is still a gap between the perspectives of container terminal users and operators [4][5].

The layout of each container terminal consists of a different area with its specific function [6]. There are a few different methods to handle containers within the terminal, in this paper we are focusing on two main ways used by most terminals worldwide, both Transfer Trane (T/C) and Straddle Carrier (S/C) systems were used in this study. Each way has its equipment and operational particularities, from the quayside, Cranes are used to pick up containers from the vessels and then place it in yard trucks or Automated Guided Vehicles (AGVs), therefore used to transport containers



(a) Schematic layout of container terminal with Straddle Carrier System

Figure 1. Schematic layout of both container terminals

through the terminal areas and then Gantry Cranes (GCs) or Transfer Cranes (TCs) to take containers from AVGs (or yard trucks) to the stacks, and to retrieve the container from stack to the vehicles. Another alternative for the transportation process within the terminal is Straddle Carries (SC) or Automated Lifting Vehicles (ALVs), which can perform not only the transport but also the stacking operation in the yard [7].

Figure 1a shows the four operational areas in a container terminal using the T/C system. First, the Quay areas are designed to deal with import and export containers, using the Gantry Crane (C/G) to move the container to and from the moored vessels. Two transport areas are designated to transfer containers, first from quayside to the storage and then from storage to the landside customer, and vice-versa. Note that, different trucks are used in this process, terminal owned trucks, or chassis yards (C/Y) are used to move containers to and from ships, while customer trucks (external chassis) receive and deliver containers directly to the storage area. The yard area is built to store containers into stacks and to be the node between land and sea operations. Finally, a truck area is shown to represent the gate to enter and leave the terminal and could represent the parking lot or wait (queue) to start the operation.

Figure 1b also represents the operational areas in a container terminal but using an S/C system instead. The quay, yard, and truck areas work have the same functions as in the T/C system. The differences are when the external chassis enters the terminal. In the S/C system, there is a specific area for external chassis to deliver and receive the containers, instead of directly going to the stacking area, this place is called Transfer Point (TP) and it is the container link between sea and land. S/Cs are responsible to carry containers between TP and the yards as well as the yard and quayside. Note that in this system, the external chassis does not go until the stacking area, every operation dealing with trucks is done at the transfer point.

The container terminal layout design combined with the equipment type determines how the operations will be conducted while handling containers inward and outward. The storage layout is organized in lanes that are subdivided into bays to delimit the container positions address. Its orientation can be either perpendicular or parallel to the quay. Although the layout orientation does not

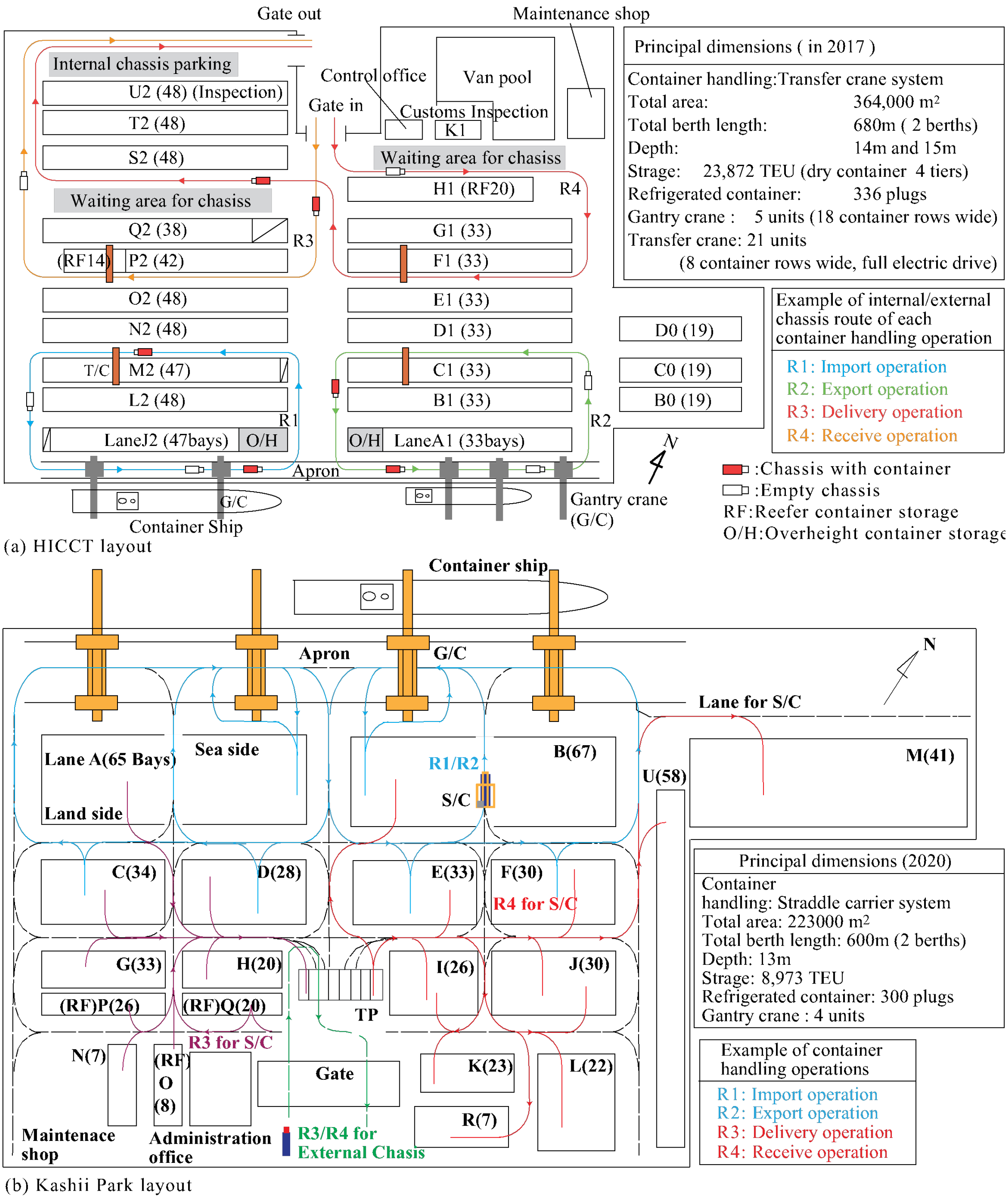


Figure 2. Container terminals layout.

imply significant differences in the methodology, a few adaptations to the data analysis algorithm are necessary. Figure 1 also illustrates the flows of movements to process containers in the terminal. Literature stipulates that the inbound and outbound movements are always from the vessel's point of view. For instance, inbound container refers to movement from the landside to the ship whereas, the outbound container is the other way around. We are focussing on the terminal, therefore those movements Inward and Outward, when the container flow goes from the ship and the land or from land to vessels, respectively. Inward movements are composed of import and export operations, performed from the quayside to stacking lanes. On the other side, outward is done by receive and delivering operations to and from the storage. If the terminal is connected to the rail, outward movements could also be done with C/Y or S/C depending on the terminal equipment. Empty containers may have an exclusive to be placed. Also, refrigerated containers need to be store in a specific area due to the necessity of an electric grid, that will guarantee low temperature for those containers.

To analyse the terminal data, it is necessary to check the basic movements applied to the real terminal layout. As mentioned before, this paper uses two different layout patterns to verify the data mining methodology developed.

Figure 2a and figure 2b show how operators manage the terminal basic operations, Import (R1) and Export (R2) taking the container from and to the storage yard and moving from and to the container ships, using the C/Y in case of HICCT and S/C on Kashii Park. Then, Delivery (R3) and Receive (R4) moving containers inward and outward the terminal. Note that, in the case of the HICCT the link between storage and land is directly executed by the External Chassis trucks, while the Kashii Park needs to use transport the containers to the TP and then transferring from and to the trucks.

2.2. Data Science and Terminal Data Structure

The 21st century has been called a digital era, the revolution that created not only an enormous challenge for companies but also great opportunities. The National Consortium of Data Science defines data science as “the systematic study of digital data using scientific techniques of observation, theory development, systematic analysis, hypothesis testing, and rigorous validation.” In other words, the main goal of data science is to describe, explain, and predict events, produce knowledge, and develop new methods to analyse data, and then improve the process. An efficient combination between information technology, statistics, distributed systems, and mathematical approaches is also part of data science definition. These methods nowadays are vital for any company that wants to remain competitive, take the available data, and then transform it into valuable information for individuals, society, or organizations might be the biggest advantage for modern companies [8][9].

Studies about how to track features related to customer consumption, products and production is becoming a trend in the past years. New ways to collect data, for example, social media, blogs, product reviews, or product rating, creates important insights on how products and services affect consumer behaviour while buying and post-purchase behavior. Aligned with that, the Internet of Things (IoT) is a powerful system of gathering data about how processes are executed. Using the internet, wireless communication, micro-electromechanical systems (MEMS), embedded systems, automation, global position system (GPS), and others can collect enormous data of positioning and status of every product in the most different business [10][11][12][13].

Firstly, to describe Big Data, it is necessary to check a few common characteristics, called the three V’s model, standing for Volume, Velocity, and Variety. Later, this model was revised and reported that “Big Data technologies describe a new generation of technologies and architectures, designed to economically extract value from a very large volume of a wide variety of data, by enabling the high-velocity capture, discovery, and/or analysis”. Thereafter, the definition can be summarized as the 4 Vs: Volume, Velocity, Variety, and Value. A few authors also add Veracity to the previous definition [14][15][18].

The term ‘data mining’ derives from machine learning, artificial intelligence, and statistics, and it is becoming popular in recent decades. Although data mining came from the early ’70s, after the internet became more popular and accessible, the technologies to extract information from data started to stand and gain relevance. Data mining is a tool to determine the relationship, the patterns between datasets, and extract hidden information. However, it does not exclude the necessity of knowing analytical methods and understanding the business environment. Recently, it is cheaper for companies to invest in computing processing power and to acquire some level of automation, which makes it easier data collection and storage. Nonetheless, with the huge amount of data, the main challenge remains on how to understand and extract useful information from it, to make a prediction and better decisions in science and markets [16][17][18].

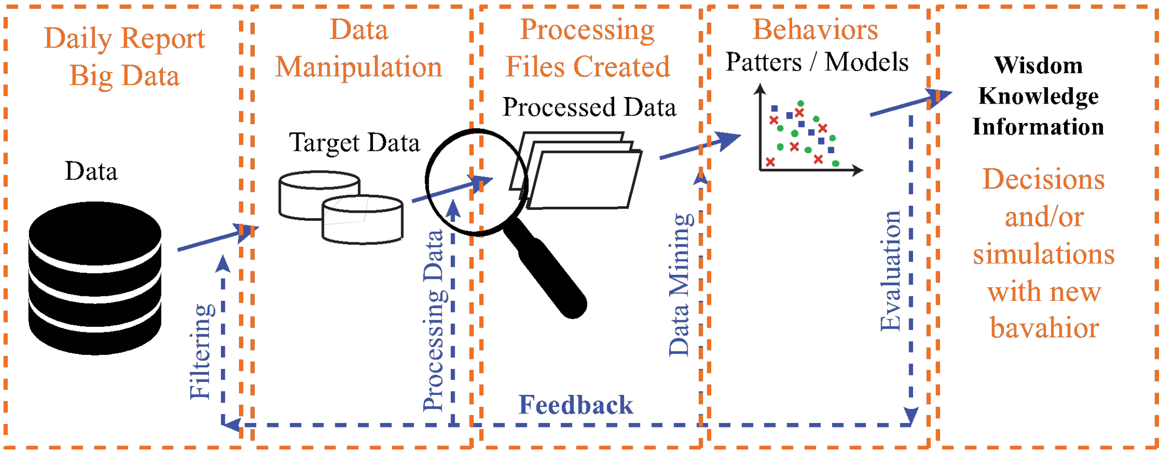


Figure 3. Steps of Knowledge Discovery Process.

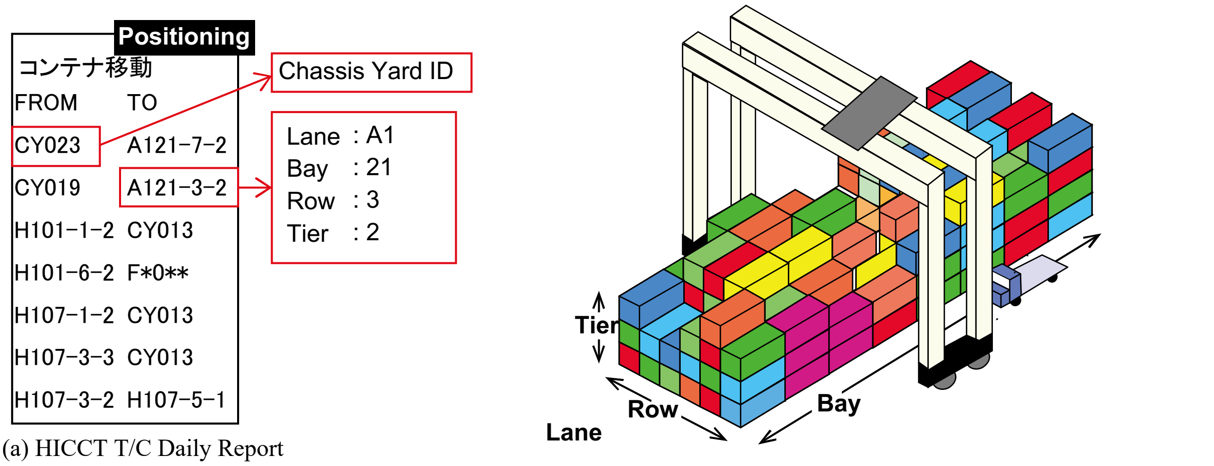
Different terms have been given to the concept of finding useful meaning from data sets: data mining, knowledge extraction, information discovery, information harvesting, data archaeology, and data pattern processing. Nevertheless, the term Knowledge Discovery in Database (KDD) has been widely used to emphasize that extracting knowledge out of the data is the main goal of data-driven analysis. Then, data mining can be considered a step in the KDD process to apply different methods and algorithms to create new models over the data. Figure 3 shows the KDD process scheme where each step is linked by a process to help the analysts to create knowledge from the database. From the data (commonly Big Data), target data is selected to be processed. This stage is typically a pre-processing phase, in this process, for many cases, the data needs to be converted to a friendly format for the next phase of generating a processed data. After that, data mining algorithms play their role in finding different patterns, behavior, or new methods to be categorized, interpreted, and evaluated by a specialist. Every process is composed of a loop between themselves, and even though that is a basic flow toward knowledge creation, it is necessary to make multiple interactions, feedbacks, and loops between them [17][20].

2.3. Terminal Data Structure

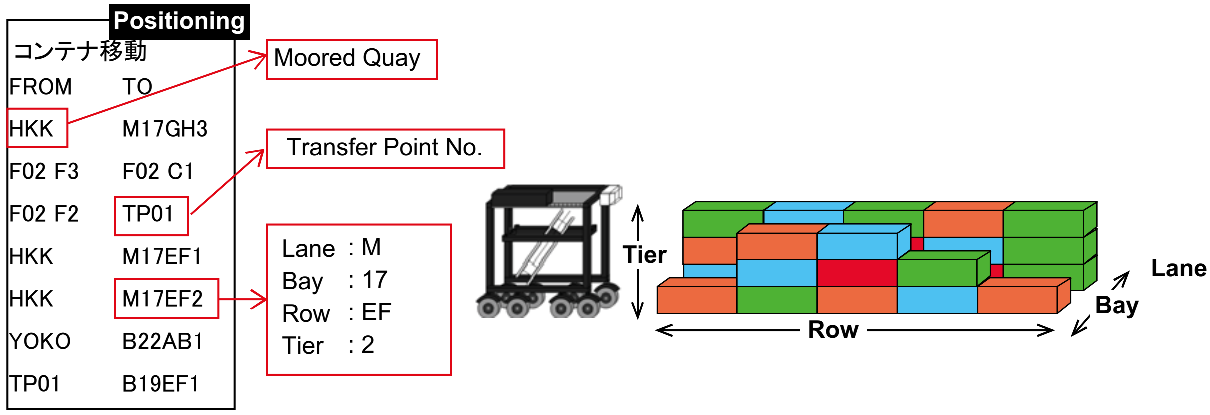
To extract useful information from any database, first, it is necessary to understand the data. Figure 4 (a) and (b) show an example of the T/C daily report of HICCT before the pre-processing procedure. The raw data in Microsoft Excel is generated from the TOS, in which each sheet represents one operational day and about 20 columns are representing each operation done by the equipment when moving containers within the terminal. It describes from the moment the order to move the container is received until the operator finishes the requested operation. Since each row in the database represents a single movement, tracking a container in the terminal is not a straightforward process, heavy processing is required to retrieve such information. It is important to highlight that the time the operator receives the order might not be the same as the begging of the operation. The container characteristics, such as type, size, height, and IMO (container ID number) also can be found in the data. Moreover, the T/C or S/C number, GC, C/Y, external truck chassis, TP, companies, and operators responsible to execute that order are inserted in the data.

Even though both terminals database has the same macro structures, some important differences must be taken into consideration when analyzing and evaluating the terminal operational behavior. Figure 4 also shows those differences in the equipment and the positioning records are the main differences between both terminal's data records.

First, regarding the equipment naming, figure 4 (a) represents the T/C system terminal therefore, the transfer crane number needs to be recorded in the data. While figure 4 (b) is using the S/C number to show which equipment was used for each operation. Also, the positioning system needs to be distinct for each transportation system. Then, due to the nature of both systems, the method to address containers in the stack is also changing. Although both systems have the same macro structure naming the Lanes, Bays, Rows, and Tiers where each container is placed, figure 4 (a) shows how the T/C



(a) HICCT T/C Daily Report



(b) Kashii Park S/C Daily Report

Figure 4. Main differences between T/C and S/C systems positioning.

reports the positions of the containers in the yard. The lane name is given a capital letter followed by a number representing which quay this lane is located, then a two digits number for each bay, as shown before, lanes have different capacity, therefore different numbers of bays in it. It is important to underline that 20 feet containers are placed on odd bays while 40 feet containers are on even bays. After that, the row for each bay is placed and the height (tier) where the container is addressed to be placed or retrieved. Figure 4 (b) on the other hand, has just a capital letter to represent the lane, and then also a two digits number for each bay. To record the row, it is used one or two capital letters, for a 20ft or 40ft container, respectively. Finally, the tier works similarly to the T/C system, indicating which height the container will be placed.

The operations are done by the terminal need to be understood and properly evacuated according to each terminal peculiarity. Some operations might have different naming and application depending on the terminal, but the core of each of them is the same. The data indicates all types of container handling work, a total of 12 types of operations were found while evaluating the database. Table 1 shows not only each type of operation found on both terminals but also explains, respectively, the data type as shown in the data, which terminal the type may be applied, the data name abbreviation we are using here forth, the name of operation translated from Japanese, the direction of the container movement, whether inward, outward or within the terminal lanes. Then, finally, the detail explaining the operation itself.

The main difference between both terminals when dealing with the types of movement is that on Kashii Park instead of Shift-In and Shift-Out operations, the terminal uses only Shift operation followed by Export. Then, need to evaluate a few more parameters to know if it is the equivalent of HICCT Shift-In and Shift-Out movements.

Table 1. Types of container cargo handling operation.

Data Type	Terminal	Abbr. in paper	Name of operation	Direction of Container Movement	Detail
揚	HICCT/ Kashii	Import	Import	Inward	Unloading container from ship to importing container from ship to storage lane through internal chassis.
出	HICCT/ Kashii	Deliv	Delivery	Outward	Delivering container from storage lane to external chassis.
リ	HICCT/ Kashii	RH.Dl	Rehandle for delivery	Inside of storage lane	Re-handling container for delivery. It is instructed automatically by the system, closest place, and lowest tier.
積	HICCT/ Kashii	Export	Export	Outward	Loading container to the ship to export a container from storage lane to ship through internal chassis.
入	HICCT/ Kashii	Recv	Receive	Inward	Receiving container from external chassis to storage lane.
SO	HICCT	ShOut	Shift out	Outward from storage lane	Shift out container to transfer to another storage lane. It is paired operation with SI.
動	HICCT/ Kashii	RHSf	Re-handling for shift	Inside	Re-handling container with SO, convenience for operations, or protection for high wind. It is instructed by the yard operator, YO, closest place, not above today's container, and lowest tier.
SI	HICCT	ShIn	Shift in	Inward to storage lane	Shifting in the container from another storage lane. It is the arrangement for ship loading faster. Paired operation with SO.
無	HICCT/ Kashii	RH.Dr	Rehandle by the driver	Inside of storage lane	Re-handling by Equipment driver when they see dangerous stacking.
仮	HICCT/ Kashii	Tp	Temporary	Inward, from vessels	Tentative storage containers from ship to rehandle containers inside ship's cargo hold. It is paired operation with a return operation.
戻	HICCT/ Kashii	Rt	Return	Outward, to vessels	It is paired operation with a temporary operation.
カ	HICCT/ Kashii	Es	Escape	Inward/Outward	Escaping empty container from VP (vanpool) to storage lane (or reverse).

To evaluate the container timeline in the terminal, it is necessary to process each container IMO separately. First filtering by ID, then sorting by the date, after that checking each movement with its pair. For example, if a container arrived at the terminal from a ship, it is an Import type container, then if a Delivery is required, this container will be taken by the external chassis to be transported to its owner. Additionally, it is calculated how long the container spent in the terminal, also how much rehandling was necessary until this box leave the yard. After repeating this process for every container, and all possible types of movements combinations it is possible to measure the container average time in the terminal.

It is important to emphasize that the data was not built to this purpose, then, some types of operations, that do not reflect reality may happen and it is necessary to use exceptions to deal with those kinds of combinations. For instance, it might appear a container arriving in the terminal by truck,

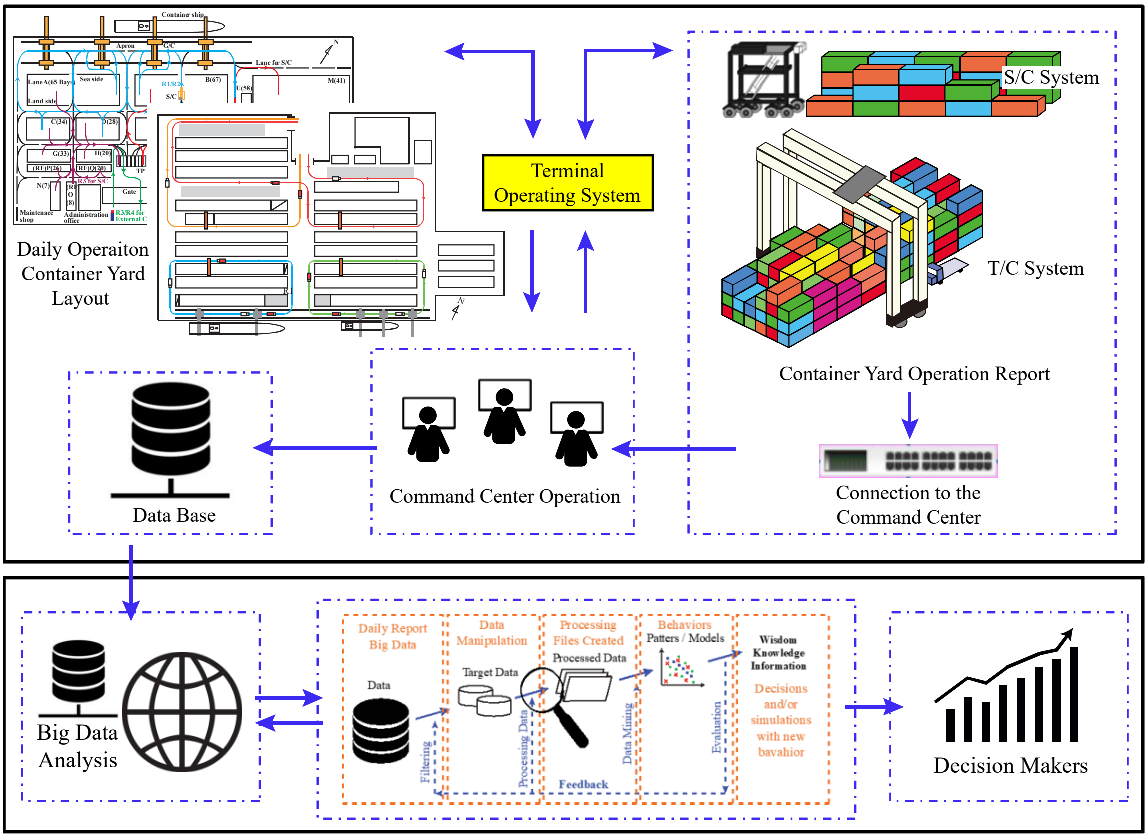


Figure 5. Big data analysis and datamining methodology.

receive operation, then leaving also by truck, as delivery operation. That may happen due to an operator mistake or need to be investigated that container to check its real path.

3. Methodology

In this paper, the concepts of the data value chain, data mining, big data analysis, and statistics were applied to develop a procedure to understand and evaluate the container terminal daily report data to support decision-makers to take the best actions possible when planning the daily yard operations.
Figure 5 shows the flow chart of the methodology to extract useful information out of the terminal daily operational data. It was used both T/C and S/C systems to express the methodology effectiveness, the concept of Big Data value chain while collecting and analysing the data is the core of this procedure [21][22][23].

We divided this methodology into two parts, as shown in figure 5. First, data collection plays its role in recording each container's operational data. This data is taken from the moment the order arrives until the end of the movement itself when the operator press a button to record the exact time that process was completed. The TOS through the Command Centre Operation is responsible for the address system, manage the yard capacity and operation, and sending each job information to the equipment operator that will complete the task as requested, note that the operator may need to change the address according to the terminal condition, weather, or his experience. If any change is made, the data is manually changed to the newest location, but the timing process remains the same. The data from HICCT and Kashii Park are the record of each T/C or S/C operation completed in the yard, but it also addresses other equipment or quay needed during the operation. The external chassis number (C/Y), G/C, S/C, TP, Quay, company, and driver responsible for that operation are also taken in the same record. This means that by using this proposed methodology, not only are the key performance indicators (KPIs) from the T/C or S/C can be taken, but it also becomes possible to estimate all resource indicators.

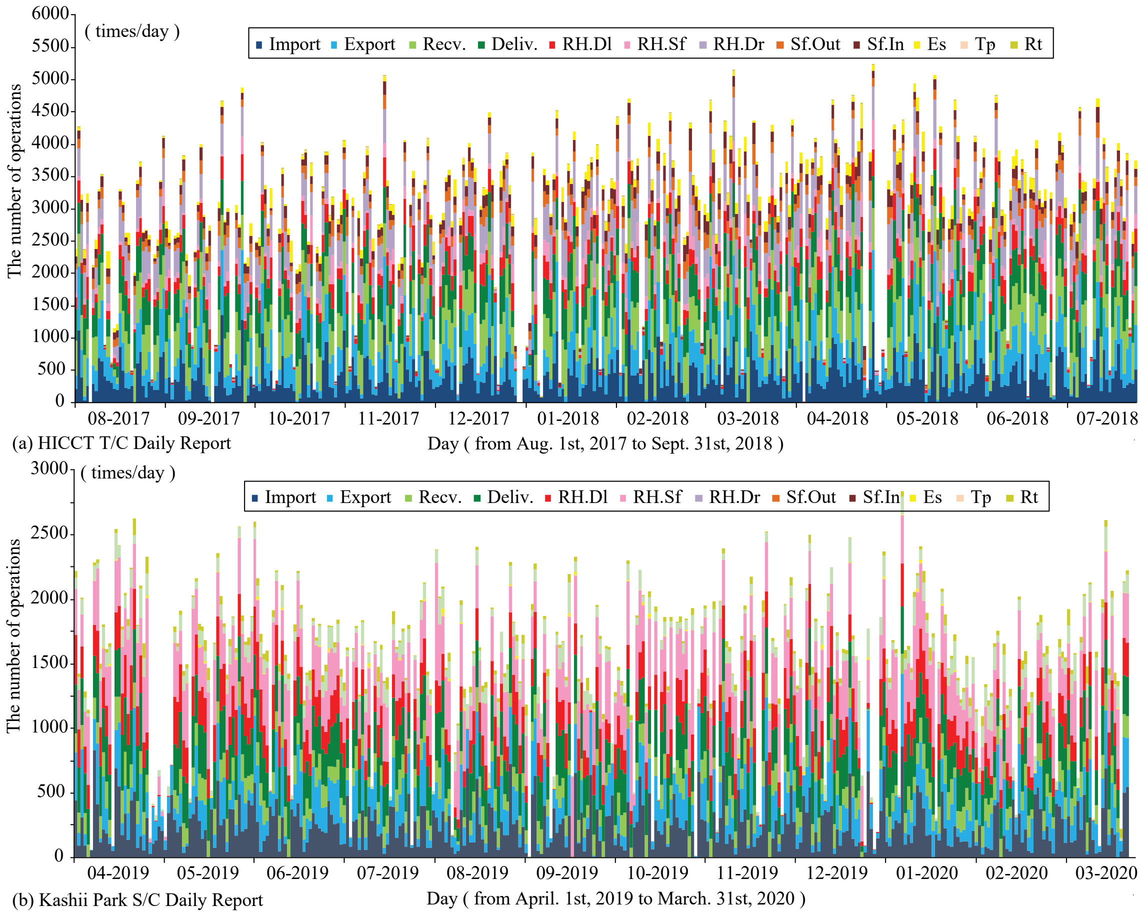


Figure 6. Container terminals types of movements throughout one year operation

The following block has its start with the database of the daily operation; using appropriate approaches to extract information and analyse this data, some key performance indicators can already be calculated and compared. It is in this step that the core of this methodology is applied, mining the data to find out unknown patterns and behavior to propose operational changes.

All information collected from the T/C or S/C daily work report is gathered; the analysis is based on the turnover container's movements, which is the composition of all movements each container had in the terminal, also how many times those containers visited the yard. Not only the containers turnover information can be taken, but also the terminal throughput, yard and/or TP utilization, G/Cs productivity, C/Y and external chassis, average stacking tear, container status, etc [24][25][26].

4. Results and Data Analysis

A comprehensive description of operational behavior is mandatory to be examined. In this study, the HICCT data of approximately 1.4 million container handling operations from August 1st, 2017 to July 31st, 2018. And Kashii Park data with around 600,000 rows representing the operation, working from April 1st, 2019 to March 31st, 2020. After combining both datasets, removing the outliers and missing values, still, about 1.8 million rows remained in the database. Note that the tools most used in the market, such as Microsoft Excel, to analyse datasets has a row limitation of around 1.1 million rows. Then, to be able to proceed with this research, it was developed Python modules to extract useful information from data. Several relations between equipment operations, storage positioning, etc were seen from the data analysis, particularly when the equipment operator change the address based on his experience. To comprehend the system, initially, it was analysed the daily number of operations divided by its handling type, as mentioned before. Figure 6 shows the breakdown of each cargo classification for both terminals. According to this figure, there is a large fluctuation in the amount of cargo handling per day throughout the year. Also, when comparing FIgure 6 (a) with Figure 6 (b) it is

IOP Conf. Series: Earth and Environmental Science **972** (2022) 012005 doi:10.1088/1755-1315/972/1/012005

possible to visualize that the number of rehandling by shift on Kashii Park (b) is bigger than HICCT. That difference is due to the nature of Kashii Park data, as previously explained, there is no shift in or shift out operations, being everything recorded under rehandling by shift.

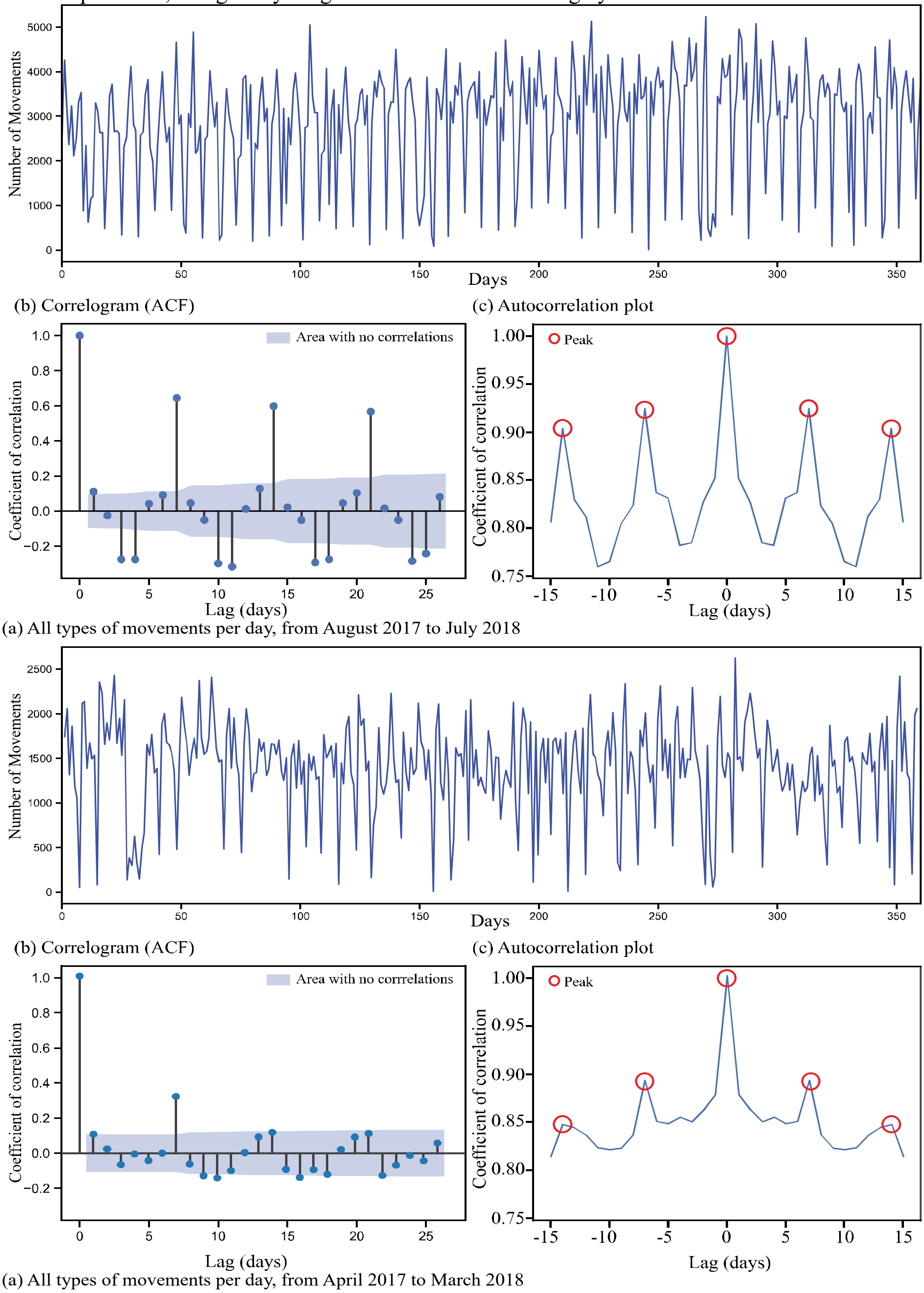


Figure 7. Container data cyclic analysis

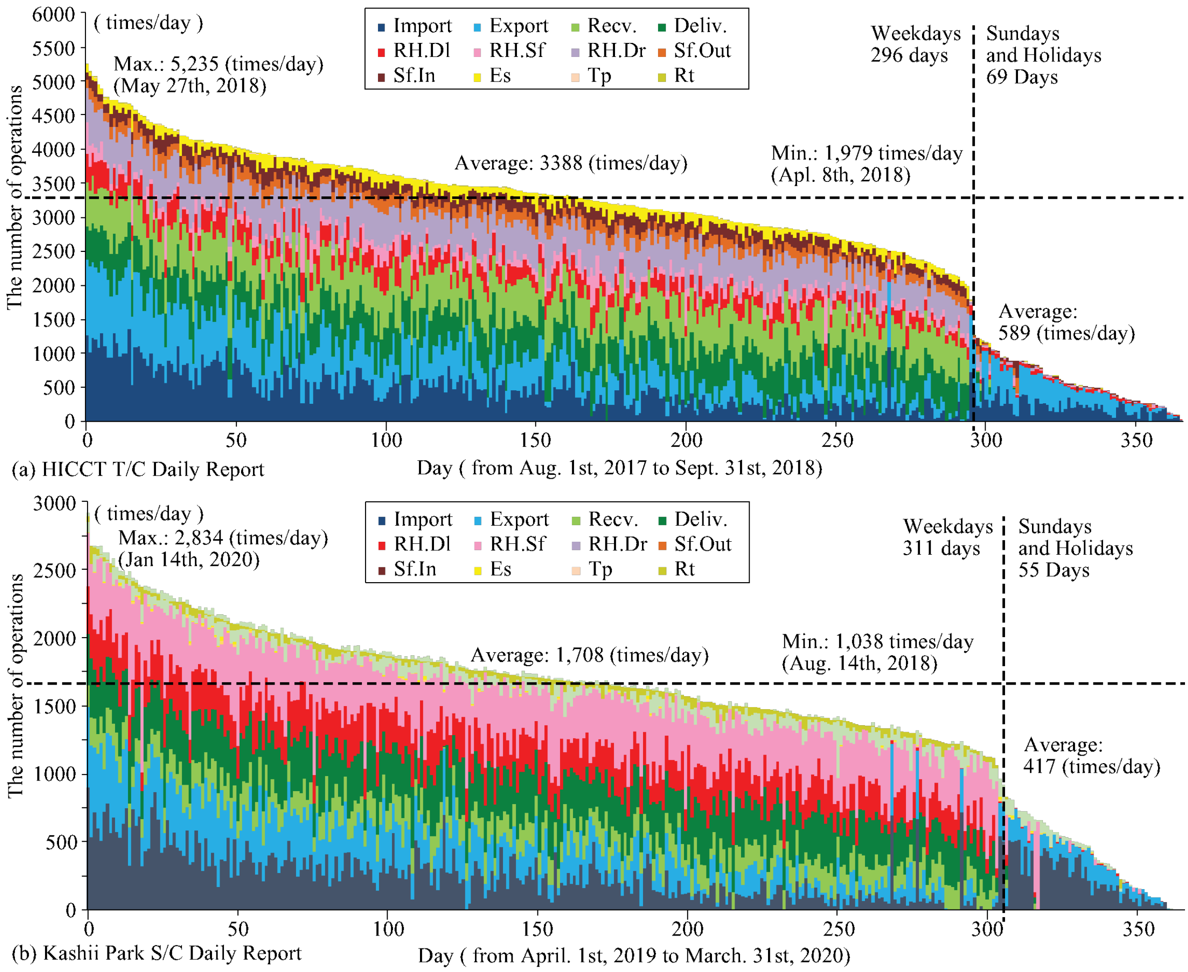


Figure 8. The daily number of operations is sorted by the highest to the lowest number.

In addition, Figure 6 suggests a hidden cyclic pattern, to investigate that, Figure 7 represents the self-correlation analysis of the total number of movements, and a weekly cycle of operational behavior regarding the combination between incoming vessels and external trucks with holidays as the boundary appears.

To understand the properties of the total of operations per day, on both container yards, it was considered not only the operational status and equipment utilization rate considered, but also the time-series relationship. Figure 7 (a) shows the total movements per day for one year, starting day 1 and finishing day 365 for HICCT and 366 for Kashii Park because of the leap year. Figure 7 (b) uses the Auto Correlation Function (ACF) to check randomness in the data set. This randomness is determined by computing autocorrelations for data values at varying time lags. Which is the calculated correlation with the observed time-series with its previous step. If random, such autocorrelations should be near zero for all time-lag separations. Otherwise, then one or more of the autocorrelations will be significantly non-zero. Also known as correlogram, the autocorrelation of this time series is done using a daily lag. The confidence interval is drawn in the light-blue shade area with a 95% interval. Furthermore, when the variable fits the Gaussian distribution, Pearson's correlation coefficient can be used to calculate the correlation between those values. The autocorrelation at lag zero is always equal to 1 because this represents the autocorrelation between the lag and itself. Price and price with lag zero are the same variables. Each spike that rises above or falls below zero can be considered statistically significant. Additionally, Figure 7 (c) computes the autocorrelation of the movements within 15 days positive and negative, a weekly periodicity of the data is shown for both terminals. It means that every seven days the same behavior tends to repeat, but with a different proportion [23][24][29].

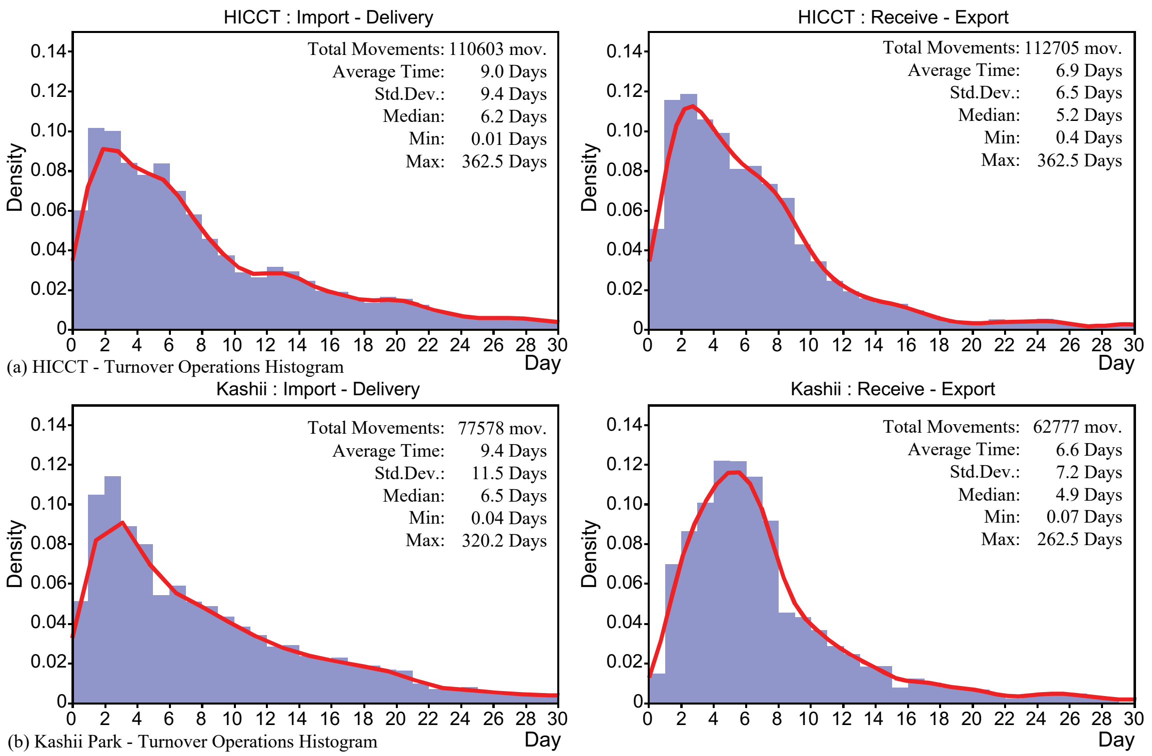
Figure 8 shows the data in Figure 7 sorted by decrescent order of handling cargo volume. Explaining HICCT, Figure 8 (a) shows a large gap between working days (296 days) and holidays (69 days). Evaluating the no working days first, Sundays and Holidays, it is possible to see that there is no trucks activity on those days, external chassis is not coming to the terminal, but vessels still operating normally. An average of 589 movements is calculated, about 5.7 times smaller than normal days. Then, when the terminal is fully operational, the peak happened on Wednesday, with 5,235 movements/day, and the minimum handling on Saturday, 1,979 movements/day. Which is around 2.7 times different, and the average is 3,338 movements.

Figure 8 (b) characterizes Kashii Park operations, and the same approach was applied to verify the terminal operational behavior. Since the datasets do not share the same dates, the working days are 311 against 55 holidays. Similarly, the Kashii Park also does not allow external chassis into the yard during Sundays and Holidays, while the ship's operations keep running. An average of 417 containers moved during no working days, 4 times lower than weekdays. A peak and valley operational days also have about 2.7 times differences between them, being 2,834 times/day and 1,038 times/day, respectively, and an average of 1,708 times/day.

The movements analysis is important to have an overview about how the terminal resources (equipment or personal) are utilized during the operations, considering the terminal was sized to operate in its full capacity, more than 90% of the time there will be idle equipment, or non-essential operations being conducted, which could represent hidden damage to the operation efficiency. Another feature that can be detected is related to seasonality; the first trimester had a higher movements rate than others but following a cyclic behavior returning to the peak on the following year.

A different aspect worth evaluating is how long the container stays in the terminal, also computing how many times that container makes a cycle. Here forth, this timing will be referred to as turn-over movement. TOS was built to manage the system operations and to offer some parameters to post-evaluate some parameters most terminals are pursuing, for instance, throughput. Then, to take the turn-over movement time for each container, data processing is required, as explained in the last section. From those 12 basic movements found in the database and combining every container type, also empty and full containers, it was analysed each cycle and built a timeline for every time the container reaches the terminal, whether land or seaside. A few combinations that do not make sense were considered outliers, therefore it was not computed for this specific analysis. When dealing with an import container, the natural flow until the same container is prepared to be exported, would be imported, and delivered, then after being process by its owner, it would take the reverse flow (to be exported) and the operations would follow the receive and export order. Figure 9 illustrates the turn-over histogram for import and delivery, and receive and export operations, also for both container terminals studied in this paper. Figure 9 (a) shows that import containers using HICCT have an average of 9.0 days before being delivered to external chassis, but the density shows that most containers are delivered within 6.0 days. On the other side, when exporting cargo, the average time spent in the terminal is about 6.9 days before being loaded into the container ship, and density showing that export containers tend to spend less time stored in the yard. Containers using Kashii Park, Figure 9 (b), take an average of 9.4 days to be delivered, and 6.6 days to be exported. Note that, the Kashii Park density is slightly different from HICCT, import containers are delivered faster and export containers stay longer warehoused in the terminal. There is no indicator in the data that can explain this feature, it could be due to some commercial agreements, or any other reason that is still not possible to visualize on this dataset. Container owners might be using the terminal as an extension of their storage, damaging the terminal operation. Then, if the storage period of the container is restricted, it may lead to the improvement of the container throughput of the terminal and the reduction of the storage area, possibly dropping the waste of resources. Even though some containers might stay in the terminal for almost a year, it is shown just the 1st 30 days behavior.

IOP Conf. Series: Earth and Environmental Science **972** (2022) 012005 doi:10.1088/1755-1315/972/1/012005



(a) HICCT - Turnover Operations Histogram

(b) Kashii Park - Turnover Operations Histogram

Figure 9. Analysis of residence time of container in container terminal

5. Conclusions, limitations, and future research

This paper uses a datamining framework to extract information and evaluate a container terminal database daily report. It was used datasets from two terminals with different systems, T/C and S/C transportation systems from HICCT and Kashii Park, respectively. It was shown examples of how powerful this methodology could be to take useful information out of the daily operations records.

Daily movements were analysed and presented in different formats to check the cyclic behavior during one-year operation, and how the equipment is being utilized. The analysis presented suggests that the terminal equipment was oversized, terminal can handle peak days, but since days with hight number of movements do not happen frequently, many important resources is underutilized. Also, it was demonstrated how long containers are left in the terminal. Some container owner might be using the terminal yard as an extension of their storage, but this attitude increases the terminal handling work. If the container leaves the terminal as soon as possible, there would be less chances for this container to damage the overall operation. Therefore, reducing the number of rehandling and the average tier in the terminal yard bays.

From another perspective, the terminal has still space for improvement and it is not saturated, in other words, it might be possible to achieve better efficiency adjusting operational metrics and operational behavior that is dragging down the terminal key performance indicators, and consequently the throughput.

It was mentioned before that the TOS was designed to control the terminal operations; therefore, it is not scope of the system to provide multiple parameters to evaluate many metrics. Even though it is possible to estimate the KPIs through calculations, that is an important limitation to be considered. And, for further studies it is suggested to use a bigger database, multiple years would confirm the behavior trend and might show different hidden patterns.

6. References

[1] Angeloudis P and Bell M G H 2009 *An uncertainty-aware AGV assignment algorithm for automated container terminals* Port Operations Research and Technology Centre, Department of Civil and Environmental Engineering, Imperial College London, United Kingdom.

[2] Steenken D, Voss S and Stahlbock R 2004. *Container terminal operation and operations research – a classification and literature review* OR Spektrum 26 (1), 3–49.
[3] Boer C A and Saanen Y A 2012 *Improving container terminal efficiency through emulation* Journal of Simulation 6, 267–278. doi:10.1057/jos.2012.10
[4] Panayides P and Song D 2008 *Evaluating the integration of seaport container terminals in supply chains* International Journal of Physical Distribution & Logistics Management 38 (7) 562-584.
[5] Tongzon J Chang Y and Lee S 2009 *How supply chain oriented is the port sector?* International Journal of Production Economics 122 (1) 21-34.
[6] Meisel F 2009 *Seaside Operation Planning in Container Terminals* Contribution to Management Science, Physica-Verlah, HD.
[7] Brinkmann B 2011 *General Considerations on Container Terminal Planning* Handbook of Terminal Planning - Operations Research Interfaces Series 49 25-39 DOI: 10.1007/978-1-4419-8408-1_2.
[8] van der Aalst W 2016 *Data Science in Action. In: Process Mining* Springer Berlin Heidelberg.
[9] Novaes Mathias T Shinoda T Hangga P Inutsuka H 2019 *Big Data Approach to Identify the Waste Management of Container Terminal Resources* Asian Transport Studies 5 (4) 653–678.
[10] Tirunillai S Tellis G J 2012 *Does chatter matter? Dynamics of user-generated content and stock performance* Marketing Science 31(2) 198–215.
[11] Leeflang P S H Verhof P C Dahlström P Freundt T 2014 Challenges and solutions for marketing in a digital era European Management Journal 32(1) 1-12.
[12] van Rijimenam M 2013 A Short History of Big Data Available at: https://datafloq.com/read/big-data-history/239.
[13] Foote K D 2017 A Brief History of Big Data Available at: https://www.dataversity.net/brief-history-big-data/#
[14] Xia F Yang L T Wang L Vinel A 2012 Internet of Things International Journal of Communication Systems 25 1101–1102.
[15] Erl T Khattak W Buhler P 2015 Big data fundamentals: concepts, drivers & techniques. Prentice-Hall
[16] Bhatia P 2019 Data Mining and Data Warehousing Principles and Practical Techniques Cambridge University Press.
[17] Fayyad U Piatetsky-Shapiro G Smyth P 1996 From Data Mining to Knowledge Discovery in Databases American Association for Artificial Intelligence.
[18] ang L Fu X 2005 Data Mining with Computational Intelligence Springer-Verlag Berlin Heidelberg.
[19] Laney D 2001 *3-d data management: controlling data volume, velocity, and variety* META Group Research Note.
[20] Piatetsky-Shapiro G 1991 *Knowledge Discovery in Real Databases: A Report on the IJCAI-89* Workshop AI Magazine 11(5) 68–70.
[21] Bruce P Bruce A and Gedex Peter (2020) *Practical Statistics for Data Scientists* Second Edition O'Reilly.
[22] Wiley J (2015) *Data Science & Big Data Analytics: Discovering, Visualizing and Presenting Data* EMC Education Services.
[23] McKinney W (2018) *Python for Data Analysis* O'Reilly.
[24] Tamhane A C and Dunlop D (2000) *Statistics and Data Analysis from Elementary to Intermediate* Prentice-Hall.
[25] Wickham H and Grolemud G *R for Data Science* O-Reilly.
[26] Hatcher L *Advanced Statistics in Research: Reading, Understanding, and Writing Up Data Analysis Results* Shadow Finch Media.
[27] Cowpertwait P S P and Metcalfe A V (2009) *Introductory Time Series with R* Springer.