Understanding Prime Mover(PM) Productivity in Yard

Li Zhenglong, Lim Kai Chin

Abstract

The study objective is to seek insight from Prime Mover (PM) Operations from port operations to identify common characteristics exhibited by PM with high and low waiting times, through understanding of PM events and operational data. This, in turn, enables us to pinpoint and identify correlated attributes and embark on further study to improve the overall productivity of PM operations and resource utilisation through active targeting of activities contributing to the PM waiting time.

I. Introduction

Maritime trade has been the backbone of international trade since the beginning of time. Today, maritime trade accounts for approximately 92% of world trade (Xu,et al,2020). Singapore's strategic location and political stability has allowed Singapore to command a strategic position as a maritime hub in the global arena.

PSA Singapore As the world's busiest transshipment port, PSA Singapore handles about a fifth of the world's transhipped containers (Saini,2018). As of 2019, the port attracts 130,000 vessel calls on average a year, while the maritime industry accounts for 7% of Singapore's GDP and 170,000 jobs, and this figure is set to grow.(Turner,2019).

Port Operations Upon a vessel's arrival at a berth in the container terminal, containers are discharged from and loaded onto it. A typical discharging operation starts with a quay crane picking up a container from the vessel and placing it onto a prime mover (PM), which then transports the container to a storage yard. At the yard, a yard crane picks up the container from the PM and shifts it to a designated spot. Loading operations involve the transporting of containers in the opposite direction, from the yard to the vessel(Ku, 2018).

PM Productivity is defined as the sum of the total number of containers handled divided by the total

hours. The following are the key terms used by PSA to define PM productivity.

$$Productivity = TotalContainersHandled/TotalTime$$

$$TotalTime = Sum(Est.TravelTime + Est.WaitTime$$
$$+ UnproductiveTime + Non - WorkTime)$$

Total Time is defined as the time difference between the two operation activities.

Estimated Travel Time is the duration between two locations based on distance matrix with fixed speed limit/hr.

Unproductive Time is the time logoff by the same driver.

Non-Work Time is the time taken for a change of driver, meal break, and PM breakdown (if any).

Estimated Wait Time is Total Time minus the summation of Est. Travel Time, Non-Work Time and Unproductive Time.

In 2021, PSA Singapore handled a total 36.9 million twenty-foot equivalent units(Leow, 2021). There are more than ninety thousand PM yard operations daily on a average based the sample Data. With the amount of transactions and connectivity, even a slight improvement in productivity can bring about large savings to the organisation, and the opportunity to increase output, thus providing growth to the organization.

II. Motivation and objectives

Our research and development effort were motivated by the general lack of effective and easy to use web-enabled data visualization tool to conduct data analysis on PM operation data. The project aims to enable operation managers the ability to monitor,drill down and identify key attributes and processes contributing to the PM productivity.

III. Review and critic on past work

Currently, PM operation's performances are tracked using Operation Indicators such as PM productivity and waiting time aggregated by duration, from daily shifts to monthly reports and segregate by individual terminals, equipment, containers and operation types.

Over the years, operational excellence have drive much improvements to various operation functions such as resource, yard and berth planning which has increased the PM productivity. However, daily operation managers are still largely relying on static excel reports, and in more recent years, using Data Visualizations tools such as Qlikview and Power Bi to perform their day to day analytic task on pre-summarized performance figures. Thus, it is difficult to drill down and pinpoint issues, and certainly laborious to monitor outcome of implemented process improvements.

Other previous external studies on PM operation efficiency typically focus on crane productivity by work schedule(Li,Wu,Petering,Goh,Souza, 2006), and resource planning & deployment to find the optimal number of PMs and trucks (haulier) to reduce average PM waiting time.(Eindhoven University of Technology, 2010).

The first study conducted in 2006 only discusses crane productivity and not PM productivity. Additionally, the study is way overdue. The second study by Eindhoven University of Technology in 2010 was conducted in a sandbox environment to find the optimal number of PM per equipment in a single process flow, thus totally ignoring the complexities of live transactions and connectivity of PM operation.

IV. Design Framework

Exploratory data analysis (EDA) with treemaps, bar charts, and histograms are used to give users an overview of PM operations. Treemaps are used for a quick glance at PM operations, so as to quickly visualise any anomalies. This function is envisioned to be used only for high-level visualisation and not for detailed analysis. Bar charts and histograms are used to visualise the distributions and identify any major flags that can then be drilled down further, either with the other facets of the application, or taken offline for more in-depth deep dives into the root causes. Bar charts and histograms are chosen as they are simple to understand without significant statistical knowledge. A key aspect for the EDA plots is interactivity. Considering that it is an exploratory phase, all parameters of the dataset were included as possible to help the user discover findings.

Confirmatory data analysis (CDA) with scatter plots and box-violin plots are used to definitively confirm hypotheses from the user. Scatter plots help to visualise any form of correlation between travel time and waiting time, and identify any segments that require additional root cause analysis. Users can also compare the correlation line across different categories of operations.

Pareto and control charts are used together to help tunnel down into the operation process and to find the root cause of productivity problems. The Pareto chart helps to identify and focus effort on the top portion of the causes to resolve the key contributors of the problem.

By using a variety type of control charts with control limits of 3 standard deviations, we can study the stability of the current process, analyze and make improvements, and monitor the results of the newly implemented process. Furthermore, the data obtained from the process can also be applied to predict future performance of the processes.

The Data visualization in these four components are synchronized through a set of common data filters that allows manipulation of data visual and coordination of interaction between the different views. The linked data presentation enables the user to maintain the filter selection integrity without having to reapply the settings each time when switching to another view. The mouse hover tooltip provides on-demand details without having too much information cluttered on the data visual. These allow users to follow the flow of "overview first, zoom and filter into details," bouncing back and forth, here and there, with ease and without interrupting their train of thought.

Data preparation

The raw data is extracted from PSA's operational database. Data from March 2019 was extracted and annoymised using JMP Pro 15. The data was then imported into R and further data cleaning was carried out using R. Columns not relevant to analysis were removed. The remaining columns were aggregated for waiting time, travel time, unproductive time, and non work time, being grouped by length of container,

move ops status, terminal ID, container type, container empty/full status, shift type, hour, day, and date. These variables were then factorised to ensure that they were considered properly as categorical variables instead of continuous variables.

Shiny Architecture

Development of the application was done on Shiny, an R package used to build interactive web apps. Shiny allows for interactivity as it is able to capture input from users and utilise R code to return output back to the user. Input values can be changed at any time, and the output changes accordingly. Shiny is also able to use the multitude of R packages found on R-cran, which supplements the basic R functions and improves the functionality of the Shiny app.

The design of the Shiny application flows according to business requirements. i) **Home -** A brief overview of the project, key definitions and a basic user guide. This explains the motivations and gives context to the application. A detailed user guide is also appended here. ii) **Exploratory Data Analysis -** This provides a high-level understanding for PM operations and the various parameters involved. iii) **Confirmatory Data Analysis -** This provides statistically-supported charts and graphs for more in-depth hypothesis testing. iv) **Pareto Charts -** This provides detail for the key contributors of waiting time and travel time, enabling users to ask the correct questions. v) **Control Charts -** This allows the user to identify key outliers that can then be targeted offline.

Analysis Techniques

Treemap

A treemap is a type of graphic that aims to show part-whole relationship hierarchically. The size and colour intensity of the boxes determine the scale of the operation. In this case, the number of records are mapped to the size, while the sum of total time spent is mapped to the colour. The grouping variables are shift and day-of-the week for the primary variable, with day-of-the-week and hour of the day being the secondary variable. This helps to visualise the operational scale and volume from a high level perspective.

d3treeR is used to build the treemap in the application, as it allows for interactivity with tooltips and allows the user to drill down into sub-categories by selecting the boxes.

Bar Chart

A bar chart is used to visualise proportions across categories by varying the length of the bar. This chart is used to for comparing different number of records across different categories. The primary category used for the x-axis are the variables mentioned above. The parameters built in the application allow the user to change the structure and make up of the bar charts. Users are able to change the grouping variable, stacking method, and y-axis scale. This allows users to see the same data from different angles.

ggplot2 is used to create the bar chart as it is flexible and allows for faceting by groups. The resultant ggplot2 graph is parsed into **plotly** which allows for interactivity with tooltips. This helps to remove text elements from the bar charts, which tend to clutter the plot as more categories are added into the visualisation from grouping.

Histogram

A histogram is used to display continuous variables. Each bar groups the continuous variables (i.e., waiting time or travel time) into a range. The taller the bar, the more data records are inside the range. A histogram can display the underlying distribution of the dataset. A similar approach to the bar graph is taken, where parameters allow users to modify the structure of the histogram.

ggplot2 is used to create the histogram due to the reasons highlighted above.

Scatter Plot

A scatter plot displays points to represent values for two different numeric variables (i.e., waiting time and travel time). The position of each plot on the x- and y-axes represent the relationship between both variables. This enables the user to see the correlation between the two continuous variables.

Although **ggplot2** is able to create a scatter plot, **ggscatterstats** from the ggstatsplot package was used instead as it includes additional statistical testing using various tests. Only the non-parametric test was used in this case. The rho value represents the Spearman's rank correlation coefficient, which denotes the level of correlation between both values. This plot was not parsed into plotly as the statistical testing output would be lost.

Box-violin Plot

A boxplot is able to display the data using its mean, upper and lower quartiles. A violin plot displays the entire distribution of the data using a kernel density plot. A box-violin plot combines these two components to give the user the full view of the data distribution.

Although **ggplot2** is able to create a box-violin plot, **ggbetweenstats** from the ggstatsplot package was used instead as it includes additional statistical testing using various tests. It has a built-in statistical test to check if the continuous variable varies across different categories. Only significant differences between distributions are shown as the "only significant" argument was parsed when creating the chart.

Pareto Chart

A pareto chart is a chart that displays both bars and a line graph. The values are represented in decreasing order by bars, and the cumulative total is represented by the line.

Control Chart

The control chart enables users visualise the mean, upper limit, and lower limit over time. By comparing current data to control lines, users can visualise whether current data is within expected limits or if there are anomalies that need to be investigated.

V. Demonstration - use case

VI. Discussion

What has the audience learned from your work? What new insights or practices has your system enabled? A full blown user study is not expected, but informal observations of use that help evaluate your system are encouraged.

On average, PM drivers spent more than 50% (8 minutes) of their time waiting, and 28% (4.2 minutes) traveling when transporting containers from point to point in the yard.

Ship Ops(O) events only contributed to approximately 40% of the waiting time $>= 30$ minutes despite only making up close to 70% of the total number of events. We can see priority given to Ship Ops

events which are related to loading and discharge to customer's vessels.

While there are 46% of total PM events deployed to RTG yard cranes, 70% of the long traveling time ($>= 30$min) are contributed by deployment to RTG yard cranes. Further investigation and analysis are required to validate the causes, which could be due to longer planned travel routes or heavy traffic.

Focusing on Operation type, we observed Ship Ops events have tighter upper and lower control limits, better stability, and lower waiting time compared to the other Operation types. Although there are small shifts from the average, the possible causes could be due to different shifts or the way the data are being subgrouped. Inter-terminal (I) event has the widest control limit of 7 to 10 minutes making it less consistent and predictable.

From the I-chart, Operation no 53 is observed to have an unusual large shift from average with 34 min waiting time. Further investigation is required to find out reason of this special cause.

VII. Future Work

It is important to note that whilst the control chart can identify non-random variation, it cannot identify its root cause. The analysis can only show us that PM waiting or traveling time have been reduced or stabilized after an implemented change. However, it cannot be established as the cause of the change.

To implement Check Sheet as part of the data collection plan so as to record down when occurrence of long waiting or traveling time events happens. This will be used to analyze further in detail with Pareto and control chart on the specific targeted category.

The PM waiting time and traveling time are estimated base on distance traveled minus non-work time and unproductive time. The accuracy can be improved if the data set captures the actual traveling time and waiting time using GPS or IoT sensor data. In near future, the Tuas terminals Automated Guided Vehicles (AGV) will replace a majority of the current traditional human-driven Prime Movers to be the new norm. Hence, using a similar approach to what we are currently doing, making use of the **EDA, CDA**, Pareto and Control charts to analyze key attributes and the stability of current and future processes.

Apart from additional analysis being built into the application, more functions can be added. The application currently has a fixed dataset of 1 month.

Future iterations of the application will include a data upload functionality to enable data refresh capability. This ensures that users can operate the application and refresh the dataset without external help, which will be key for continued utilisation of the application. Workshops and brainstorming sessions will be conducted with supervisors and managers of PM operations to identify lacking functions, and to ensure that the managers involved become key stakeholders in developing this application.

VIII. References to be update

Fenner, Martin. 2012. "One-Click Science Marketing." *Nature Materials* 11 (4): 261–63. https://doi.org/10.1038/nmat3283.

Meier, Reto. 2012. *Professional Android 4 Application Development.* John Wiley & Sons, Inc.