NOVA SCHOOL OF SCIENCE & TECHNOLOGY

DEPARTMENT OF MECHANICAL AND INDUSTRIAL ENGINEERING OF NOVA School of Science and Technology | FCT NOVA
AND DEPARTMENT OF HUMANITIES AND MANAGEMENT OF ESCOLA NAVAL

WANJING WANG

BSc in Economics

DASHBOARD IN PORT AND ORGANIZATION DECISION-MAKING

MASTER IN MARITIME LOGISTICS

Universidade NOVA de Lisboa / Escola Naval

September, 2024

FACULDADE DE CIÊNCIAS E TECNOLOGIA

DEPARTMENT OF MECHANICAL AND INDUSTRIAL ENGINEERING OF NOVA School of Science and Technology | FCT NOVA
AND DEPARTMENT OF HUMANITIES AND MANAGEMENT OF ESCOLA NAVAL

DASHBOARD IN PORT AND ORGANIZATION DECISION-MAKING

WANJING WANG

Bachelor in Economics

Adviser: Doctor Pedro Manuel Geada Borda de Água
Associate Professor, Escola Naval (Portuguese naval Academy)

Examination Committee:

Chair: Doctor Helena Carvalho,
Assistant Professor, Coordinator of the Master in Maritime Logistics
UNIDEMI - Department of Mechanical and Industrial Engineering
Faculdade de Ciências e Tecnologia da Universidade NOVA de Lisboa (NOVA FCT)

Rapporteurs: Doctor Armindo Dias da Silva Frias,
CFR NA / Military Lecturer in the Department of Humanities and Management of Escola Naval

Adviser: Doctor Pedro Manuel Geada Borda de Água,
Associate Professor, Escola Naval (Portuguese naval Academy)

MASTER IN MARITIME LOGISTICS

Universidade NOVA de Lisboa / Escola Naval
September, 2024

Dashboard In Port and Organization Decision-Making

This document was created with Microsoft Word text processor and the NOVA thesis Word template [1].

DEDICATION

This report is dedicated to Politecnico de Setubal, especially the Department of Systems and IT (DSI), for providing me with the fellowship to work on Nexus project.

Special thanks to my professor Pedro Manuel Geada Borda de Água from Escola Naval, my team leader Rui Madeira from Nexus project, whose expertise, patience, and encouragement that made this research possible.

ABSTRACT

This report explores the development and implementation of a data visualization dashboard, designed as part of the Nexus open data platform under the Nexus project.

The theoretical framework incorporates the ECI model, emphasizing its primary principles, models, and concepts. The significance of data visualization is analyzed, particularly in its ability to empower organizations with actionable insights. However, the technical challenges encountered during dashboard development – such as handling duplicated or missing data, quantifying complex operational metrics, and ensuring user-friendly data accessibility – highlight the need for targeted technical solutions. Proposed solutions include advanced methods for data cleansing, approaches to quantify previously hard-to-measure operations, and strategies to optimize data delivery to front users.

Methodologically, critical thinking is employed as a qualitative research tool, demonstrating its value in designing dashboards that address user needs effectively. A detailed comparison of digital tools, including Power BI, underscores the technological choices made during development.

The result showcases the dashboard's purpose, utility, and target audience, emphasizing its contribution to improved decision-making. The report concludes the implications of the findings, highlighting how the developed solutions advance the state of data visualization and serve as a foundation for future projects within the Nexus open data platform.

By addressing technical challenges and offering practical, scalable solutions, this report underscores the transformative potential of data visualization in organizational contexts.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

API	Application Programming Interface
AI	Artificial Intelligence
BSC	Balanced Scorecard
DTLF	Digital Transport and Logistics Forum
DIT	Domain-independent tokenization
DIDD	Domain-independent duplicate detection
DAX	Data Analysis Expressions
ECI	Electronic Circuits Inc.
EDI	Electronic Data Interchange
GPS	Global Positioning System
GROW	Goal, Reality, Options, Will
HR	Human Resource
IBM	International Business Machines
IT	Informatics
KPI	Key Performance Indicator

POR	Place Of Receipt
POWER BI	Power Business Intelligence
R&D&I	Research, Development, and Innovation
SQL	Structured Query Language
SWOT	Strengths, Weaknesses, Opportunities, Threats
VP	Visualization Platform
WP	User Interface

1

INTRODUCTION

This report is part of the mandatory deliverables for the attainment of the Master in Maritime Logistics Degree offered jointly by NOVA University of Lisbon and the Portuguese Navy Academy *(Escola Naval)*.

1.1 The motivation, job role and scope

As to the job role, I have been working as a research fellow (fellowship) for Nexus project, and the project is led by Porto de Sines (the Portuguese major maritime port). I am based in the *DSI - Dep. de Sistemas e Informática* (Department of Systems and IT) of the *Politecnico de Setubal*. The main research area is **open data platforms/dashboard design** and **prototyping**. Including me, the team consists of 4 people. I work with the teammates, who are skillful personnel with their specialization in Computer Science, to achieve the digital transformation goal in the context of ports and terminals (Port of Sines).

As Nexus is a large project at national level, and proposed to take years of time, involving diversified stakeholders, so regarding the general job scope, it can be concluded as:

Table 1 - Team Task Scope

Team objective	Draft out a proposal
Action plan in details	[1] Examining how data can correlate with each other.
	[2] Formulize prototyping product for end-users
	[3] Analyze/Test the collected/existing raw datasets from port by putting in prototype

Regarding the motivation to join the Nexus project, it can be facilitated into several considerations. First, in this end-to-end era, to visualize operational performance of multimodal

transportation has been increasingly critical for port and terminal. However, it remains still a blank space at especially inland transport, to be specific, trucking service performance visualization. As the trucking service can always act as the first leg or the last mile leg of the whole supply chain, it's with particular importance to be visualized. Hence, this research will take the opportunity to explore the development to resolve this professional pending issue. Second, before joining Nexus, I developed a SWOT (Strengths, Weaknesses, Opportunities, Threats) analysis of what's the edge that can contribute to the project, as well as what will benefit personal development for my future career. It's a mutual selection basis. The preferred sub-project spans Nexus' WPs (Work Package) - WP1, WP2, and WP3.

Considering the below self-asking questions, WP1 was the final choice. The self-questions are based on GROW model:

- Goal: A clearly defined goal to be achieved.
- Reality: Analysis of the current reality, assessing the existing resources, obstacles, and progress.
- Options: Exploration of possible options and strategies to find the best path to achieve the goal.
- Will: Decision on the specific actions to be taken and commitment to take those actions to achieve the goal.

So, several questions were elaborated for helping the consideration.

- Which goal do we want to achieve?
- What's the challenge or issue we want to overcome?
- What does success look like?
- Why this goal is important?
- How will we know when we achieve our mission?

In the following paragraph, each of the question will be facilitated by linking to Nexus project goals, objectives and proposed deliverables that consist of and formalize the final choice of which team to join. Results of the personal SWOT analysis can be concluded as following aspects for answering self-asking questions:

- The goal to be achieved is always in accordance with the team – Draft out a proposal.
- Driven by curiosity, I want to acquire Computer Science techniques and acknowledge new information by working with Computer Science specialists.
- Success here looks like the team will be capable of delivering a satisfied result, from cost-saving perspective as well as profit-driving.
- By participating a national-level project that is related to the most cutting-edge and forward-looking technology fields today, it offers the most up-to-date knowledge about technology innovation.
- By delivering as much as possible as the WP1 deliverable requires. The deliverables so will be available to be verified its expediency more comprehensively and to be improved.

1.2 WP1 – Nexus Open Data Platform

As previously noted, I've been working on WP1.2 – Visualization Component, which is one of the sub-sectors of WP1 main package - Nexus Open Data Platform. In this section, an introduction of WP1 main package will be outlined, together with WP1.2 objectives, main tasks, and proposed deliverables.

The objective of the WP1 main package is to initiate 'architected' open data models that are available, being 'embedded' into a platform which provides ultimate data sharing between the logistics parties. These platforms are normally known as 'open data platforms'.

WP1 Nexus vision is to gather all data (static and real-time) concerning multimodal transportation into one single platform, which is partly known as 'Logistics Single Window', at national level. The multimodal transportation includes but are not limited to rail and road transport (such as last mile), intra-domestic feeder and barge transport, airlines, etc.

The 'Logistics Single Window' refers to a digital platform, where all information that will be used for subject open data platform will be digitized. The stakeholders, generally include, but are not limited to, shipping companies, customs and port authorities, terminal operators, and other regulatory agencies, from internal or external. The open data platform will allow each party member to submit and access necessary information through a 'single entry point'. Figure 1 provides a graphical explanation of an example of data flow in 'open data platform'.

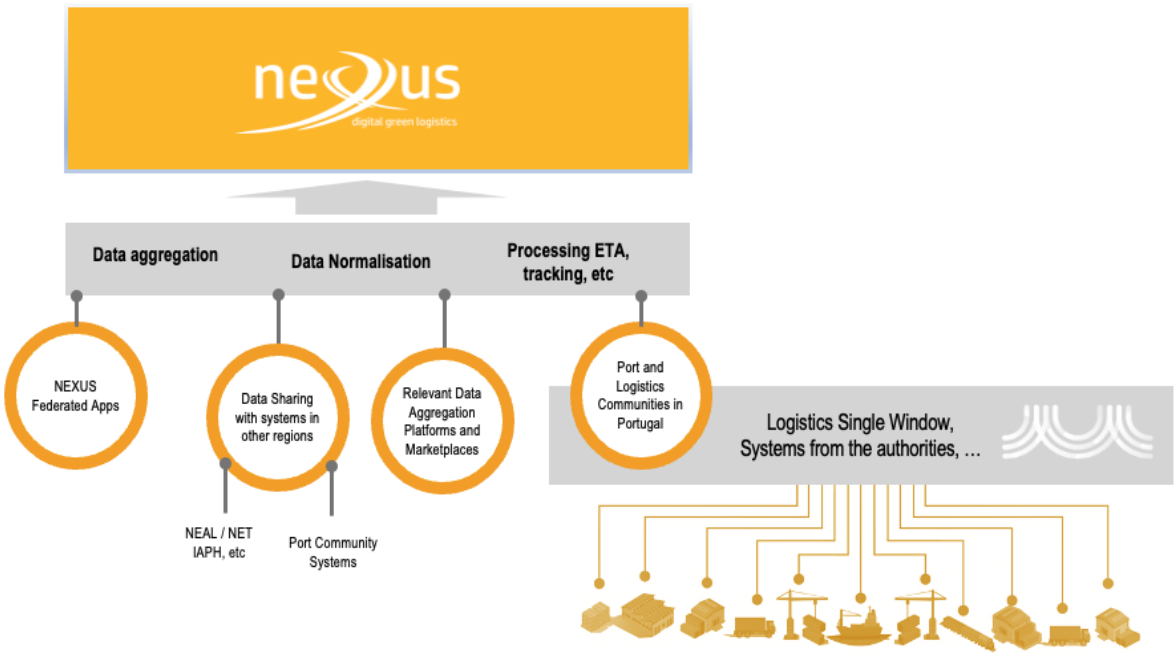

Figure 1 - Diagram of an example of data flow [Adapted from [1]]

By constructing an open data platform, below listed are the key objectives of WP1 main package shall achieve:

- Public data to re-use for new products and services released by companies within Nexus ecosystem, for commercial or other purposes (such as international exploitation by local companies), namely, to foster digital innovation by data sharing.
- Development of new technologies based on AI (Artificial Intelligence).
- CO2 emission decrease.

It is worth mentioning that all new products and services from WP1 will also be 'combined utilized', namely, to operate together, in certain cases, under a federated system approach. Therefore, the 'working together' will adopt the principles proposed by the European Commission's DTLF (Digital Transport and Logistics Forum), which are:

- Plug and Play: provide each stakeholder with concepts and processes to share data in accordance with contentment.
- Technology Independent Services: to produce a connective platform that allows independent technology to support common procedures, business interactivity, as well as to comply with legislations.
- Interoperability of platforms: to create a federated platform that will allow collaboration of different platforms, even if each have multifunctional, technical realization and varied business patterns.
- Trusted, Safe and Secure: to build up a neutral governance structure for data sharing, including peer-to-peer solutions.

In the following section, the tasks distributed in WP1 main package is introduced, as well as the one I was assigned for, which is WP1.2 that is part of WP1 main package. Tasks for WP1 main package consist of 3 key sub-task sections in general, which are:

- Nexus open data app marketplace design and development
- Visualization solutions, digital mobility, process scanning
- Connectivity services

Tasks for WP1.2 mainly focuses on 3 aspects, which are the visualization solution, digital mobility, and process scanning. In general, includes:

- Monetization functionality establishment in visualization platforms.
- Workflow visualization by connecting more than one data products. For instance, putting data product A without location information into data product B that contains location information, and produce a map.
- Well-defined and filed API (Application Programming Interface) that nodes other IT component (domestic or international) with the open data platform, by calling API endpoints with a 'visualization definition' embedded in the data products produced in visualization platform, to interface the results (embedded visualization) in its own systems.
- Customizable digital product with automatic assistance and friendly user-interface productions with the most common functionality for the end-users.
- International standardized geographical visualization capability build-up, to enable location viewing and analysis of every movement in supply chain. For example, the spatial location data sharing.

In details:

Table 2 - WP1.2 tasks categories and respective initiatives

Categories	Initiatives
Identify and evaluate datasets and processes.	- Geospatial capability verification - Documentation review - Meeting with stakeholders
Technical solution specifications/configuration	- Definition of location intelligence - Integration and interoperability between location technology and open data platform - Eligible business patterns and procedures identification for digitization.
Nexus location intelligence implementation and tests	- On stand-alone solution platform - On location-based capabilities (GeoApps, GeoAPIs and GeoServices)
Follow-up and compliance assessment	- Report generation to identify deficiency.

In line with the WP1.2 objectives, key deliverables are as following:

Table 3 - WP1.2 deliverables review

Key Deliverables	Description
Report design and generation	- Preparation of datasets to integrate in VP (visualization platform). - Processes that require location data. - Review on test and implementation according to the configuration set.
Nexus location intelligence platform specification	- Smart assistance - Monetary mechanism

1.3 Nexus Project

In this section, a quick and simple guide of the Nexus project will be facilitated, including the project structure and key stakeholders involved. In total 10 WPs are included in the Nexus project, while WP1 – open data platform, is the main package among them. The open data platform provides a 'common framework' which is critical for the Nexus agenda in gathering enormous volumes of high-quality data that will be utilized not only to develop big data and AI solutions but will foster technology innovation for new services and federated applications.

Nexus agenda / project structure in review (Table 4):

Table 4 - R&D&I Nexus ecosystem of products and services for digital and green transition

WP1	Nexus Open Data Platform
WP2	Synchro modal Logistics Network (AI powered applications)
WP3	Smart Gates and Smart Terminals
WP4	Smart Train Operations
WP5	Smart Logistics
WP6	5G Port of the Future
WP7	Cybersecurity and Cyber-Resilience
WP8	Green and Energy Transition in Ports and Logistics
WP9	Exploitation, Dissemination, and Internationalization
WP10	Advanced Training and HR (Human Reource) Qualification

The stakeholders involved in Nexus project can be mainly divided into four streams, which are academia and universities (such as research centers), industry and business entitles (such as technology companies), government and public sectors (such as port authorities), civil society (such as end-users). Figure 2 displays the main project partners (consortium), who also act as involved stakeholders in review:

:he

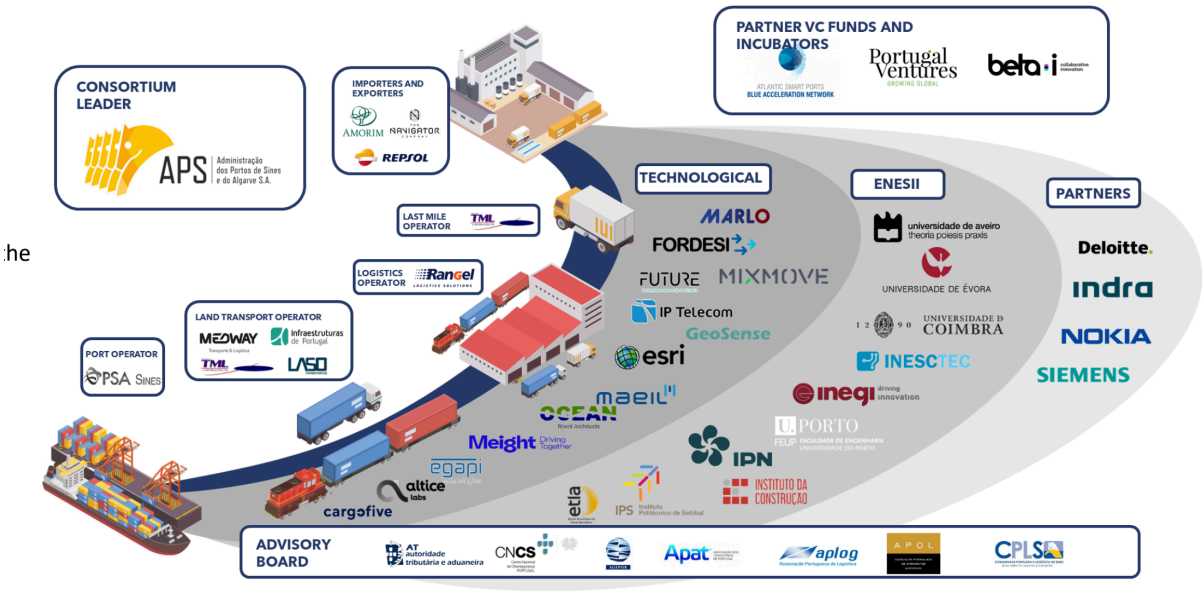

Figure 2 - main project partners (consortium) of Nexus [Adapted from [2]]

1.4 Report Structure

In this section, an introduction of the 'skeleton' of the project report is made. The report can be divided into 6 sections, which are listed in the table of contents. The 'introduction' part focus on facilitating the overview of the Nexus project, including the project agenda and key stakeholders involved. The motivation, job role and scope, together with objectives, tasks, and key deliverables of the sub-working package I've been participating in will also be introduced in this section. The 'Theoretical Framework' will be focusing on the theoretical foundation of dashboards development from scorecards.

The 'Methodology Used and Main Activities Taken' section will be focusing on the conducted activities during the fellowship, including the used digital tools. In addition, the utilized methodology for making all activities happen, which was critical thinking. The 'Results and Work Done' part will be focusing on displaying the final achievement by the activities taken in the fellowship, including the description of the completed dashboard, regarding its purpose, utility, and audience. Technical difficulties (limitations) and respective proposed solutions during the work will also be facilitated in this section. In the end, it will come to the 'Conclusion and Discussion' section, which will wrap up with closing statements which reiterates the contribution of the activities taken and results achieved.

2

THEORETICAL FRAMEWORK

The Balanced Scorecard lays a solid foundation in the development of dashboards. It can be said that dashboards are based on the conceptual framework of scorecard and has achieved a higher level of functional expansion in terms of complexity of applicable scenarios and the evolution of user needs. Therefore, the conducted professional activity is totally originated from thinking about this question 'what makes scorecards/dashboards special for port/organization decision making'? The contemplation revolves around 3 aspects referring to this question:

- What to change?
- What to change to?
- How to cause the change?

Regarding contemporary dashboard definition, it can be described as an integrated real-time view system; a control panel set within the central console of diversified transportation, such as a spacecraft, vessel, vans, or aircraft [3]. Regarding open data platforms definition, it can be described as an infrastructure that provides data access through technical and legal frameworks, aiming to promote data reuse and value creation. This platform, combined with a dashboard, can help end-users quickly gain insights and support decision-making by summarizing and visualizing key business datasets [4].

In the upcoming sections, answers to those questions will be provided. Hereby, it will be introduced an overview of the ECI (Electronic Circuits Inc.) model to become as a strategic management system that represents an evolutionary stage between the traditional BSC (Balanced Scorecard) and contemporary dashboards that deals with the integration of information systems and decision-making processes. While BSC primarily offers periodic review and is often static in nature, ECI model emphasizes the importance of continuous data flow and real-time monitoring. While dashboards offer the technological capabilities for real-time monitoring, they are rooted in the strategic alignment principles established by the BSC and further developed by the ECI model.

2.1 Main View, Models, Concepts, and Working Principle of the ECI model

The acronym ECI stands for Electronic Circuits Inc., which is an integrated framework associated with the information and business intelligence system in organizations, to describe how information flows within such organizations. Therefore, it emphasizes the importance of an integrated approach to decision-making, where data, technology, and human factors converge to provide a clear performance target to enhance organizational effectiveness.

The ECI model features for 4 measurable perspectives, which are: (1) financial performance, (2) customer satisfaction, (3) internal processes, and (4) the organization's innovation and improvement activities. These 4 dimensions are typically designed for ensuring all critical aspects of an organization are considered.

Key concepts of ECI model, such as system integration and user-centered interface design, are critical when implementing dashboards in practice. On behalf of the traditional financial performance, the other 3 are considered complementary measurements to the financial performance – operational measures that are the drivers of future financial performance. Figure 3 displays the 4 perspectives in details.

- The *Financial perspective* evaluates how well the organization is generating value for its shareholders and sustaining profitability.
- The *Customer perspective* examines how well the organization is serving its target audience by compound indicators setting.
- The *Internal processes perspective* assesses the efficiency and effectiveness of key ongoing internal procedures.
- The *Innovation and learning perspective* focus on the capability to innovate, improve and grow of an organization, at several aspects such as employee's professional development or company culture for long-term success.

Dashboards should be adopting working principles such as continuously refinement based on user feedback and be adaptable to constantly changing business circumstances. With the guidance of the ECI model's 4 perspectives, dashboard design will be more reasonable. The respective working principles of the ECI modal can be summarized as:

- *Financial perspective*: not only tracking financial performance in a conventional way by revenue, profit, or return-on-investment review, but understand how those figures align with broader strategic commercial goals.
- *Customer perspective*: customer-centricity and engagement. An organization is supposed to deliver value that meets or exceeds customer expectations and market responsiveness.
- *Internal process perspective*: focus on optimizing internal operations and continuous improvement to support strategic objectives and deliver value effectively.
- *Innovation and learning perspective*: investing in employee training, fostering a culture of continuous innovation to solid adaptability to changing business world for both human and technology.

ECI's Balanced Business Scorecard

Financial Perspective

GOALS	MEASURES
Survive	Cash flow
Succeed	Quarterly sales growth and operating income by division
Prosper	Increased market share and ROE

Customer Perspective

GOALS	MEASURES
New products	Percent of sales from new products
	Percent of sales from proprietary products
Responsive supply	On-time delivery (defined by customer)
Preferred supplier	Share of key accounts' purchases
	Ranking by key accounts
Customer partnership	Number of cooperative engineering efforts

Internal Business Perspective

GOALS	MEASURES
Technology capability	Manufacturing geometry vs. competition
Manufacturing excellence	Cycle time Unit cost Yield
Design productivity	Silicon efficiency Engineering efficiency
New product introduction	Actual introduction schedule vs. plan

Innovation and Learning Perspective

GOALS	MEASURES
Technology leadership	Time to develop next generation
Manufacturing learning	Process time to maturity
Product focus	Percent of products that equal 80% sales
Time to market	New product introduction vs. competition

Figure 3 - ECI model in 4 perspectives [5]

2.2 Significance of Data Visualization in port/organization decision-making

Echoing the second question: What to change to? - that was proposed in the theoretical framework part – data visualization can be regarded as a 'revolution' by dashboards in a new era that helps ports/organizations for reliable decision making.

Therefore, what is data visualization? According to Michael Friendly, cited by [11], data visualization converts complex data information into graphical representations, making the acquisition and understanding of information more intuitive and efficient. The application of this technology is particularly critical in organizational decision-making because it can help decision makers quickly extract useful information from large amounts of data and make scientific decisions. By integrating a variety of visualization tools and technologies, organizations can comprehensively improve their data analysis and processing capabilities, thereby achieving more accurate decisions.

Nguyen and Jung [12] explore in their article on data visualization as an emerging trend in business analytics and its importance for organizational decision making, as for instance:

- Data visualization significantly enhances operational efficiency in ports by offering real-time insights that help in managing and optimizing traffic flow, and efficiently allocating resources.
- Data visualization plays a vital role in strategic planning by enabling scenario analysis and forecasting.
- Data visualization bridges the gap by presenting complex data in a more understandable format, ensuring that all stakeholders, from investors to operational teams, are on the same page when it comes to decision-making processes.

Heer and Schneiderman [13] work offer a comprehensive and valuable insights into how data visualization tools and techniques enhances decision-making processes for ports and organizations in general. Key pillars to bear in mind:

- The interactivity of data visualization allows engagement of users from cross-functional and will assist especially ports or organizational leaders to explore diversified scenarios to identify key influencing factors from case to case, and to manipulate and update data views more effectively.
- Data visualization techniques simplify the complexity inherent in multidimensional datasets in ports or terminal operations. By transforming intricate data into understandable and comprehensive visual formats, those techniques enable decision-maker to detect some signals such as outliers, or trends and patterns in a more reasonable way.

Tools and techniques application in review:

- Brushing and linking to analyze relationships between different operational aspects, such as the correlation between cargo type and handling times across various terminals.

- Dynamic query filters help to rapidly filter through large volumes of operational data. The dynamical filter enables organizational managers to filter on a specific timespan basis, to quickly evaluate performance metrics and identify bottlenecks through each step of tasks in daily port operations.
- History and Undo/Redo. This tool is particularly useful in testing various scenario and compare outcomes. For example, for the comparison on the impact between 'before' and 'after' scenarios of specific laws or regulations change on terminal operations. The ability to adopt this tool in port and organization supports iterative exploration and decision-making process refinement.
- Faceted Browsing techniques can be applied to organizational contexts where decision-makers need to explore complicated datasets and then to drill down the most relevant ones to support specific decisions. For example, it will enable a leader from customer service function to facet customer satisfaction commentary data across different service lines in port. This technique helps decision-makers to quickly address specific operational challenges and reallocate resources to achieve a balanced daily work in ports and organizations.

2.3 Technical Difficulties (limitations) during work

During the groupwork to initiate the dashboard, three key technical difficulties were identified as the most concerning challenges for the accuracy and optimization of the results.

First, there are duplicated and empty (void) data. Because the raw data was entered manually, there is a chance that the data source's accuracy was compromised by incorrect or out-of-date data. Secondly, Certain operational procedures are difficult to quantify since each internal stakeholders have distinct guidelines. Thirdly, some data is not acquirable since the frontend users are not eligible to attain such data.

In the following section, it will be discussed the solutions/resolutions corresponding to respective obstacle aforementioned to gain some insights for future improvement of our dashboard.

2.4 Technical solution/resolution to improvement of data quality in dashboard

This section describes several technical solutions and resolutions to address the main barriers encountered during the project work.

2.4.1 Handling duplicated/empty data

In the experimental work of Medeiros and Silva [14], they introduced effective methods to detect potential data with similarity. DIT (domain-independent tokenization) and DIDD (domain-independent duplicate detection).

The methods explain how to use similarity calculation and pattern recognition techniques to identify duplicated data generated during manual input. Both algorithms aim to be unrestricted by specific fields but to apply multi-functional areas. Key features of these two methods are:

1, Both adopts measure of tokenization at the first stage to process data ready for experimentation. Regarding the tokenization, the differences in between are the DIT uses delimiters, such as spaces or punctuations, to 'token' each record in the datasets, while DIDD uses standardization of characters to 'token' each record.

2, Upon the completion of the tokenization of each record, it goes to the comparison of the 'remarked' datasets, where the aim is to divide the sample datasets into groups for analysis.

3, Regarding the 'grouping' of datasets, which can be called 'clustering', the differences between the subject two techniques are: DIT adopts a calculative measure called 'edit distance', with which the potential duplicated records are calculated and will be grouped as one. DIDD adopts a threshold of record similarity, if it hints that the similarity of a certain record pairs is above the set threshold, then the pairs can be seen as potentially duplicated. With the rules of specific techniques, such as transitive closure, the DIDD technique is also able to identify indirect repetition of records amongst samples, and then remove it, to achieve the final accuracy of the experimental result.

In conclusion, DIT and DIDD algorithms are proven to be measurable and actionable technical solutions in data cleaning regarding some aspects. For example, it improves the accuracy of datasets clustering, which allows the grouping to be representative. By eliminating the void datasets, it also enhances the trustfulness of data views, which makes it convincing for trustworthy decision-making for end-users in dashboard building.

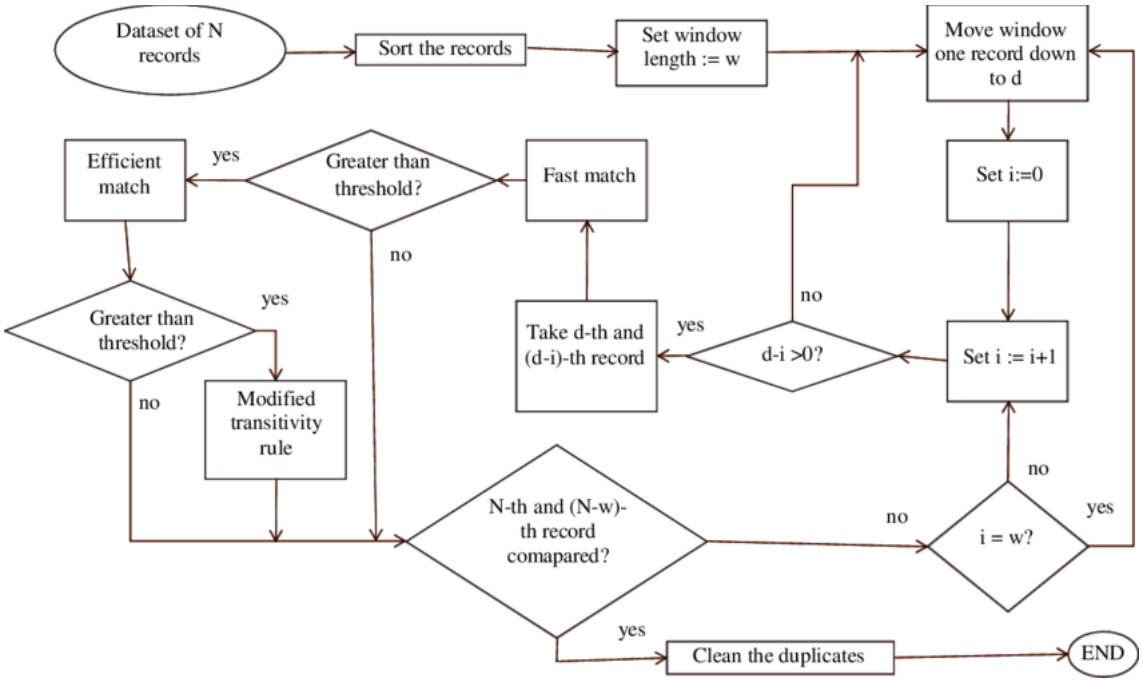

Figure 4 - Flowchart of the proposed DIDD algorithm [14]

Another worth-to-mention measure to avoid duplicated and empty data input by data governance is introduced by IBM (International Business Machines) as a data cleansing tool called IBM InfoSphere. It has the capability to maintain high data quality over time. Key features of InfoSphere can be summarized as [15]:

- Real-time automatic detection of data input to identify whether the inputted data is correct or not.
- Support regular data cleansing and deduplication processes scheduling, which helps the implementation of continuous monitoring of data to track its quality metrics.

2.4.2 Quantifying Hard-to-Measure Operations

In the CloverDX [16] article, the author emphasized the importance of 'data unification' for a real-time dashboard. Catering to different guidelines from respective internal stakeholders, the data metrics needs to be standardized. By standardizing data metrics, the consistency of data will be improved, which will eventually benefit each stakeholder for a standardized view of the panel to make quick and reliable commercial decision. The author also mentioned four key factors to achieve the goal (standardization):

- Clear definitions and rules.
- A process in place to keep systems up to date.
- Necessary automations are in place.
- Good data error handling measures

In Atlan [17], it emphasizes the importance of establishing clear guidelines of data governance and standardization of processes in data metrics for hard-to-measure operations to

reduce the variability and improve data consistency. 12 actionable steps were outlined, and can be summarized by the following aspects:

- Data governance policies set up for cross-department common understanding.
- Build up data quality assessment guideline and establish monitoring mechanism.
- Data validation, standardization, and integration.
- Continuous feedback collection from employees based on data analysis result and encourage an 'speak-out' environment for data continuous improvement.
- Collaboration with internal and external data providers.

The viewpoint from *SunTec Data* [18] developed the data cleansing strategies. Tactics introduced in articles can help with quantifying operations under different guidelines and regulations. Key effective ways are concluded as:

- Cross-referencing to detect incohesive datasets.
- Anomalies and outliers' detection.
- Imputation to eliminate redundancies.

Monte Carlo [19] mentioned in his article '8 data quality issues: challenges and solutions' on how data observability powers data quality. The document also offers insights into solutions for data integrity from several guidelines. Key aspects can be summarized as:

- Detecting broken pipelines due to upstream schema changes.
- Build up a data volume test method to prevent too little or too much data that will negatively impact the visualization results.

2.4.3 Making Data Accessible to Frontend Users

Atlan [20] proposed some best practices for data management, ensuring data accessibility across the organizations for front-users securely and effectively access the data they need. Some key practices are:

- Development of a centralized data catalog for searching documented mega data.
- Classification of each document before saving it as: public, internal, or confidential, for employees easily tracking the confidentiality before using it.

3

METHODOLOGY USED AND MAIN ACTIVITIES UNDERTAKEN

This section briefly describes the used approach or followed methodology during the assigned behind this report.

3.1 Critical Thinking as a Qualitative Research Method

Figure 4 presents the 'research onion diagram', which was proposed by Mark Saunders, Adrian Thornhill, and Philip Lewis in their co-authored book 'Research Methods for Business Students'. Critical thinking acts as a qualitative research method. On behalf of ''critical thinking, in the layer 'strategies', other methods are also demonstrated in the diagram.

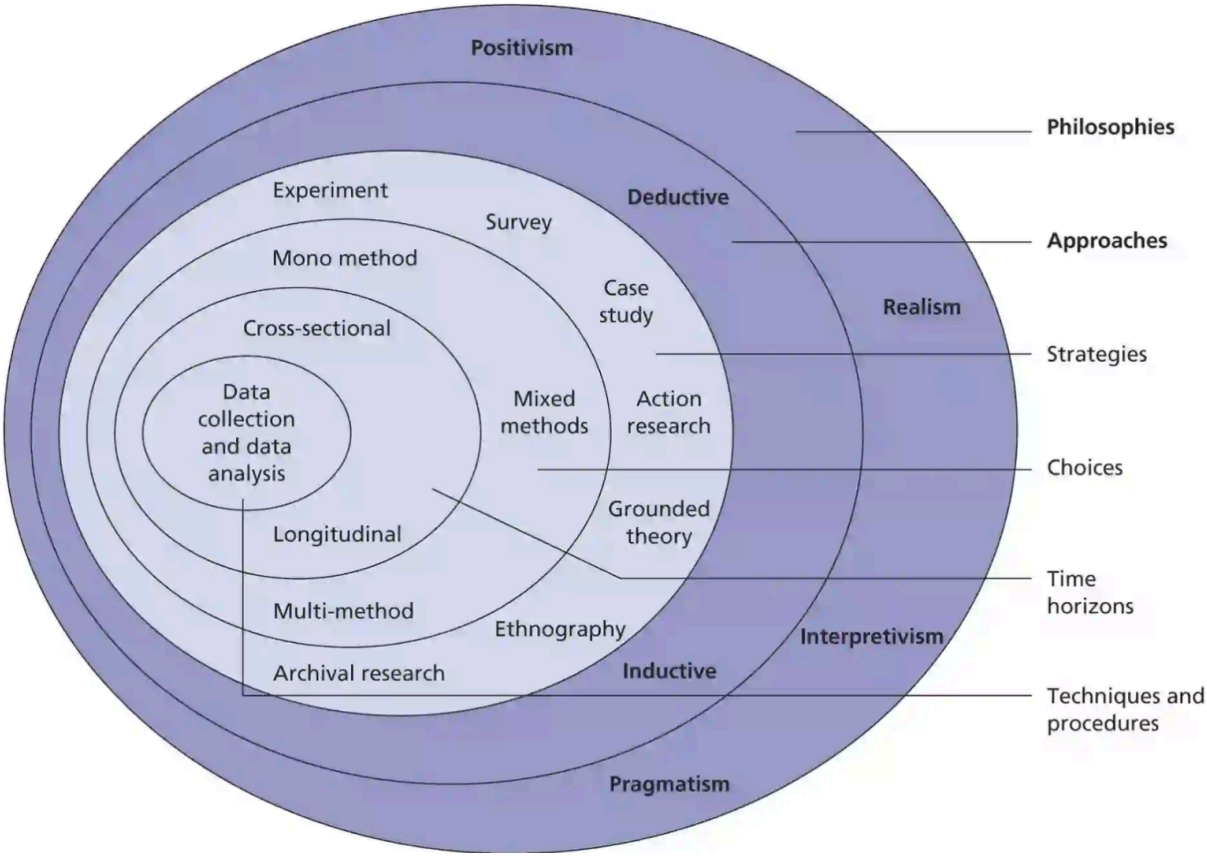

Figure 5 - research onion diagram [Adapted from [6]]

Typically, the methodology consists of a qualitative and/or a quantitative approach and is concerned with the philosophy and overall research design. Each methodology may include varied methods, such as critical thinking. Methods refer to specific tools, techniques, or procedures adopted to collect and analyze data. Different research methodologies use different research methods.

For the qualitative approach, frequently used methods are:

- Case study
- Critical thinking
- Grounded theory
- Ethnography
- Observation
- Interview/diaries (for instance, in-depth, focus group or semi-structured)

For the quantitative, frequently used methods are:

- Experiment/project
- Statistical analysis
- Several mathematical methods

Some approaches can be regarded as 'mixed' ones of qualitative and quantitative, for example:

- Questionnaire/survey (depends on the type of questions it includes – more direct or more open)
- Secondary data (for instance, categorical or numerical)

Regarding the definition of critical thinking, Schwandt defines critical thinking as a core tool in constructing meaningfulness from data [7]. Miles and Huberman highlight critical thinking as a key method for uncovering underlying patterns and relationships in complex datasets [8]. Critical thinking as an analysis tool, is the process of applying unbiased, rational evaluation to observations, facts, and evidence to form a judgement.

3.2 Rationality and Advantages of Critical Thinking in Dashboard Building

As for the dashboard building, the used methodology, which was mentioned is critical thinking, plays a critical role for leading to the proof-of-concept stage. To be specific, the proof-of-concept stage can be seen as the fortress of the whole dashboard building 'infrastructure'. Rationally identifying always leads to actionable insights. Key aspects can be concluded as:

Firstly, the *proof-of-concept* stage stands for the pre-requisites for gaining deeper understanding of the user needs and expectations (analysis of client's demand) – helps the initiator to figure out 'the basis behind' - what is in truly demand from the stakeholders or end-users when they talk about commercial decision making?

Secondly, when it comes to the question - *what sort of information to be displayed on a final dashboard?* Then, critically evaluating the quality of data, as well as its relevance seems inevitable. Because, with a clear panel view of higher quality of data, the trustworthiness of decision-making will be realized.

The critical thinking helps the initiator by analyzing the data source, the timeliness of raw data input, together with its accuracy, to determine the most-relevant data that is appropriate for eventual selection. Referring to Laudon and Laudon (2015), the critical thinking helps in the thorough process of assessing these aspects, which assured the only pertinent data that are being displayed on a dashboard.

Thirdly, in a sense of the way to display data – what kind of panels should be utilized to effectively help the end-users to understand what meaning the data are conveying. If the end-users find it holds up for them to position the information they need, then the deferral for making quick decisions can result.

As mentioned by Laudon and Laudon, the selection of data visualization tools and techniques is substantially influencing how information is perceived by who receives it [9].

Turan, Fidan, and Yıldıran [10] highlights that critical thinking serves as a qualified decision-making tool by enhancing the evaluation and interpretation of information, which is critical for determining what to build in a dashboard. The paper also proposed several

methods and strategies of applying critical thinking in dashboard building. Critical thinking helps in avoiding common pitfalls, such as information overload or the inclusion of irrelevant data, by focusing on what truly matters for decision-making. Key methods and strategies can be concluded as:

- *Questioning Assumption*. Urge users to challenge the initial assumptions about what should be included in the dashboard. This could involve rethinking which metrics are most important or considering alternative data sources.
- *Scenario Analysis*: Use scenario analysis to explore how the dashboard will perform under different conditions or with varying data inputs, which can help identify potential weaknesses or areas for continuous improvement.
- *Iterative Design and Feedback*: Implement an iterative design process where the dashboard is regularly reviewed and updated based on user feedback and critical reflection.
- *Stakeholders Involvement*: Engage with various stakeholders to gather diverse perspectives, which can help in critically assessing the dashboard's design and functionality.

Data visualization is fundamental for dashboard design and implementation. By using critical thinking, the initiator has the chances to press a 'pause' bottom to think about what the most proper methods are to display the processed data to end-users, so that all the information will be facilitate in an understandable way and, to achieve a better end-users experience.

In conclusion, on top of the comprehensive acknowledgement of the full picture mentioned above, some key aspects from critical thinking before any executive operations, the initiator will certainly be more capable of 'sketching' the tailor-made demo dashboard that is not only better in meeting up with end-users' needs and expectations to align with their business objectives with high-quality data, but also better managed to provide the end-users with a dashboard with enhanced clarity and understandability.

3.3 Power BI and Other Digital Tools Description and Comparison

The dashboard created in this project work is initiated by PBI (Power BI), a trending digital tool with quick development in recently years. Echoing the third question that was proposed in the theoretical framework part - Power BI acts as an effective digital tool which realize 'ideas' to become 'reality' which will cause the change. Hereby a brief description is provided about the adopted tool for building the dashboard, as well as other candidate digital tools that were considered in the working process to achieve data visualization.

Power BI is an abbreviation of an interactive data visualization software product that was developed by Microsoft. BI stands for as business intelligence, and it is a powerful and efficient tool. To be specific, it can combine various sources of data and turn all of it into static data visualization in one page and make all the visualizations interactive. The sources of data can include but are not limited to software services, apps, and connectors.

While utilizing it during the performed work as part of the Nexus WP 1.2 project, the following benefits of Power BI outstand. Firstly, it's simple enough to use, even for less-skilled end-users. For example, it provides functionality as an 'automatic robot', allowing users to do one click or drag-and -drop to upload data that is needed for analysis. Secondly, just like other 'human-computer interactions', the Power BI also has the capability to interact with other Microsoft products, as it is part of the Microsoft Power Platform, which provides an accessibility for its product, such as SQL (Structured Query Language), Access, and spreadsheet to integrate. To sum up, the way POWER BI realizes the 'user-friendly' interface and the agility of integrating data undoubtedly make it distinguished among other digital tools.

Tableau is also a power digital tool for data visualization, which enables a broad range of data input to be made and make it connected and shareable. Particularly, it provides highly customizable panel view especially for detailed and complex data analysis. In addition, Tableau has a solid 'supportive community' background, as Tableau was born earlier than Power BI, which is indeed one of the reasons why its community support is strong. Tableau was first released in 2003, while Power BI was launched in 2015. Moreover, it has gone through multiple versions of iterations. Therefore, Tableau has a longer time in the market which helped in building its user base and continues the improvement as a product.

Looker Studio is a digital tool for data visualization that was rebranded from Google Data Studio, for the purpose of merger and acquisition. The old-version one is perfect for creating reports or dashboards for small-sized enterprise or individuals, as it is online based and free. The modified one has the capability to do data modelling. As it is Google Suite based, both products are also capable of integrating or synchronizing with other Google products.

Zoho Analytics is a kind of robust digital tool even though it is not that prevalently known. It is Cloud-based and user-friendly, and it offers end-users an AI-powered little robot to give suggestions during report or dashboards creation. Though it is not free, the price remains competitive, so it is regarded ideal for individuals, small or medium-sized enterprises, for their business development purpose in markets.

In this conclusive paragraph, the comparison of the advantages and disadvantages of all the candidate tools mentioned above will be discussed in one glance.

Table 5 - Comparison of candidate digital tools

Name of Digital Tool	Advantages	Disadvantages
Tableau	- Versatility of data types of input and its real-time integration - Capability of processing complex and datasets - Solid user base worldwide	- Expensive, particularly for enterprise-level deployment and its cost-saving strategy - Turn time is long for operating mass datasets. - Time-consuming learning curve for

		newcomer/green-hand users.
Google Data Studio/Looker	- [Google Data] Cloud-based, free, perfect for small-sized enterprise and individual development purpose - [Looker] allow for data modelling	- Limited mass data input and processing - Only basic features are available with free version, lacking customization capability
Zoho Analytics	- Competitive price among similar products - AI powered consultancy -	- Fewer advanced features such as modelling and customization. - Incompetency of massive data process.
Power BI	- Drag-and-drop or one-click cost fewer turn time for data modelling. - Seamless interactivity with other Microsoft products for data integration purpose. - Iterated data input friendly, fit for enterprise-level users.	- Cost considerations for full functionality.

To conclude all above in comparison to other available alternatives, Power BI stands out as the final option for creating the dashboards for the work assigned, mainly for two tiers of reasons:

Tier one, concerning the Power BI efficiency applied to build the dashboard. With its high-efficiency realization from its deep-technical features, one aspect that reflects in the way is that it allows for quick mass data import and simple constructing (modelling). In this way it enhances the speed-up to data integration and transformation. The other aspect is its conception of user interface, which is clean and broad. End-users have access to a wide range of options (panel views) referring to their respective business demands for visualization. The panel views offer include but are not limited to bar charts, line charts, pie charts etc.

Tier two, what is the standard of evaluating for the Power BI efficiency? Key benchmarks can be summarized as:

- *Turn time*, represents as the total time consumed to establish and deploy a final dashboard, including the learning curve for the initiators of the chosen digital tool.
- *Variety of functionality*. For example, for the availability or accessibility to data processing and data modelling at a time, with a certain level of complexity of datasets (for example, varied types of data input).
- *Capability to integrate and synchronize iterated datasets*.

4

RESULTS

This section will demonstrate the results of the project work.

4.1 Description of purpose, utility, and audience of designed dashboard

Linking to the Nexus project goals and objectives, this section, provides further about the description of purpose, audience, and the utility of the designed dashboard.

In ports or terminals, inland transportation products are key elements. The inland transportation products are typically relating to intermodal transportation (rail, truck, feeder, barge, etc.), warehousing and distribution, etc. To build up an adequate SLA (service level agreement); visibility based on the products mentioned above always plays a pivotal role in ensuring the seamless physical movements of goods and containers from ports or terminals to final destinations, because the inland transportation serves as the first stage of a supply chain upon vessel's arrival at destination. Hence, making the operation smoothly running is undoubtedly important.

The purpose of the developed visualization panel is designed to analyze and real-time monitor the performance of external vendor trucking, and of their in-service trucking on-time ratio. By calculating, the patterns of delays of a certain container or sort of goods can be identified. The panel enables port and terminal operators to make data-driven decisions and provide insightful hints to improve the efficiency at operational level. The audiences suitable for using the subject panel can also be the trucking vendors themselves (by modifying restricted data or erase the confidential data, referring to the requirements for certain end-users), or other external users. Visibility for sharing with both the internal and external stakeholders will help both sides in simultaneously optimizing their SOP (standard operation processes).

Regarding the utility, which specifically is revolving round track-and-trace (especially for the cases that one truck carrying multiple containers) the real-time track and monitor are indispensable. To realize it, the implementation of GPS (Global Positioning System) trackers is a pre-requisite, which allows for precise remote handling of exceptional cases on the road, as well as the 'time nodes' recording for every event that happened.

The definition of the subject panel is related to the ratio of accuracy of trucking time of arrival or using the abbreviation as 'Trucking On-Time Ratio'. The purpose of design is to

visualize the performance of cooperated or external vendors trucking, down to each truck or van that belongs to the vendor service.

How does this panel work? By setting up a benchmark of on-time target, for example, a truck or van can be seen 'on time' if its time of arrival at factory or loading spot minus the time of departure (the time it departed from port or terminal with carrying empty box ready for loading cargo), is within 60 minutes (1 hour). Any span of time of any truck or van that exceeds 1 hour will be determined as 'not on time'. In the end, to sum up all the trucks or vans in service allows the calculation of the final ratio of on time trucks or vans to get a result of the vendor performance.

On the working principal basis, the proposed calculation formula of 'Trucking On-Time Ratio' panel is:

If on-time or not = (actual time of arrival at the factory - estimated time of arrival =<one hour |)

Any variation = not on-time

Count number of shipment has variation= X

Total export shipment = Y

Ratio of Truck arrival not on-time = X/Y x 100%

Ratio of Truck arrival on-time = 1- (X/Y x 100%)

Denominator = Truck arrival target = 1

Numerator = Ratio of Truck arrival on-time

Based on the definition and proposed calculation formula, in the following section, an explanation is provided on what indicators are chosen for raw data input and the raw data liaise with the designed algorithm. The proposed algorithm and its utility for applying to the final visualization in dashboard panel (outcomes) is also provided.

target CTNR#	Estimated stuffing tim	Time of arriving factor	shipment#	ContructualCustomerCod	ContructualCustomerName	PORArea	ContructualCustomerName+PORArea
HASU4343621	2022-05-14 09:00:00	2022-05-14 08:55:21	292741858	***151296	SHENZHEN AUKEY E-COMMERCE LIMITED	PRN	SHENZHEN AUKEY E-COMMERCE LIMITEDPRN
TCLU5413685	2022-08-05 09:00:00	2022-08-05 07:13:39	220550571	***151296	SHENZHEN AUKEY E-COMMERCE LIMITED	PRS	SHENZHEN AUKEY E-COMMERCE LIMITEDPRS
MRKU5636452	2022-06-05 08:00:00	2022-06-05 03:20:21	291872931	40600494281	NINGBO HANDIAN ELECTRIC APPLIANCE CO	PRE	NINGBO HANDIAN ELECTRIC APPLIANCE COPRE
HASU4529576	2022-06-05 08:00:00	2022-06-05 03:20:21	291872931	40600494281	NINGBO HANDIAN ELECTRIC APPLIANCE CO	PRE	NINGBO HANDIAN ELECTRIC APPLIANCE COPRE
TCKU6529712	2022-06-22 14:00:00	2022-06-22 08:00:22	219421146	***09010	KUMHO & CO INC	PRN	KUMHO & CO INCPRN
SUDU6993104	2022-06-21 09:00:00	2022-06-21 10:13:51	291873223	40600494281	NINGBO HANDIAN ELECTRIC APPLIANCE CO	PRE	NINGBO HANDIAN ELECTRIC APPLIANCE COPRE
MRSU5151259	2022-06-20 15:00:00	2022-06-20 10:05:52	291873223	40600494281	NINGBO HANDIAN ELECTRIC APPLIANCE CO	PRE	NINGBO HANDIAN ELECTRIC APPLIANCE COPRE
MRSU5598313	2022-06-20 15:00:00	2022-06-20 09:21:26	291873223	40600494281	NINGBO HANDIAN ELECTRIC APPLIANCE CO	PRE	NINGBO HANDIAN ELECTRIC APPLIANCE COPRE
MSKU0740569	2022-06-22 14:00:00	2022-06-22 08:46:01	219421146	***09010	KUMHO & CO INC	PRN	KUMHO & CO INCPRN
MRSU3528092	2022-08-05 09:00:00	2022-08-05 08:05:30	220550609	***151296	SHENZHEN AUKEY E-COMMERCE LIMITED	PRS	SHENZHEN AUKEY E-COMMERCE LIMITEDPRS
MRKU6447900	2022-08-08 09:00:00	2022-08-08 07:01:02	220550483	***151296	SHENZHEN AUKEY E-COMMERCE LIMITED	PRS	SHENZHEN AUKEY E-COMMERCE LIMITEDPRS
PONU7931654	2022-08-09 08:00:00	2022-08-09 09:05:47	220707413	***151296	SHENZHEN AUKEY E-COMMERCE LIMITED	PRE	SHENZHEN AUKEY E-COMMERCE LIMITEDPRE
MSKU6948170	2022-07-11 10:00:00	2022-07-11 09:15:17	219212614	102-8621	CCM HOCKEY AB	PRE	CCM HOCKEY ABPRE
MSKU6464067	2022-07-11 10:00:00	2022-07-11 08:19:43	219212614	102-8621	CCM HOCKEY AB	PRE	CCM HOCKEY ABPRE
MRKU0954949	2022-07-12 10:00:00	2022-07-12 09:04:49	219212614	102-8621	CCM HOCKEY AB	PRE	CCM HOCKEY ABPRE
FSCU7216140	2022-07-11 08:00:00	2022-07-11 07:40:33	219500679	***144968	CROCS	PRS	CROCSPRS
TCNU6155392	2022-06-28 09:00:00	2022-06-28 08:25:44	291872968	***151296	SHENZHEN AUKEY E-COMMERCE LIMITED	PRN	SHENZHEN AUKEY E-COMMERCE LIMITEDPRN
TCLU2315849	2022-06-15 10:00:00	2022-06-15 11:20:37	291873214	***29775	MICROSOFT CORPORATION		MICROSOFT CORPORATION
MRSU6033625	2022-06-15 10:00:00	2022-06-15 08:55:26	291873213	***29775	MICROSOFT CORPORATION		MICROSOFT CORPORATION
PONU1350755	2022-07-11 13:30:00	2022-07-11 09:10:26	219500691	***144968	CROCS	PRS	CROCSPRS
MVIU0025431	2022-07-11 15:00:00	2022-07-11 09:10:26	219500677	***144968	CROCS	PRS	CROCSPRS
MRKU4258717	2022-07-27 14:00:00	2022-07-27 08:02:00	219783364	***144968	CROCS	PRS	CROCSPRS
CIPU5211209	2022-07-04 09:00:00	2022-07-04 11:04:18	219495192	***29775	MICROSOFT CORPORATION	PRE	MICROSOFT CORPORATIONPRE
TTNU1111429	2022-07-04 09:00:00	2022-07-04 11:35:48	219493658	***29775	MICROSOFT CORPORATION	PRE	MICROSOFT CORPORATIONPRE

Figure 6 - indicators for raw data input [Author's elaboration]

In Figure 5, it can see that the indicators are listed on top for manual raw data input regarding the trucking on-time ratio. Those indicators are all mandatory for the application algorithm regarding the next steps of data processing. Despite the indicators, the rest of content in Figure 5 are sample datasets selected randomly for dashboard running testing. As highlighted in yellow, the chosen indicators are as follows:

- Target Container Number
- Estimated Stuffing Time
- Time Of Arriving Factory
- Shipment Number
- Contractual Customer Code
- Contractual Customer Name
- Port/Place of Receipt Area
- Combined Contractual Customer Name + POR Area

When it comes to the reasons why the above-mentioned indicators are selected and the meaningfulness of each indicator and how it all work together, a brief description of it is elaborated in the following section – algorithm.

The Contractual Customer Code and Name (as well as the combined one) are set up for modelling with the 'main' customer list in the Power BI modelling section. *The number of customers* can be one or more, depending on the requirements for end-users and how many KPIs (Key Performance Indicator) for the customers (or in the case, vendor trucking) are required to be analyzed. *The POR(Place of Receipt) Area, (Port/Place of Receipt Area),* together with *Shipment Number* column, are set up for identifying a certain shipment. *The Target Container Number* is set up for identifying the status of a certain container. *The Estimated Stuffing Time* is the accurate timepoint the cargo is loaded onto truck or van or say the departure time of loaded container from factory to port or terminal. *The Time of Arriving Factory* is the timepoint the

truck or van arrives at factory carrying an empty container from a port or terminal, ready for the cargo to load.

One thing that needs attention is that among all the indicators, the *Estimated Stuffing Time* and *Time of Arriving Factory* are only available to be recorded on the condition that the containers are equipped with GPS tracker, for data transmission purposes. An interface to transmit all the real-time recorded data into the internal system is also inevitable. For example, API node settings and EDI (Electronic Data Interchange) transmission channel opening to connect different layers of data transmitting and synchronization. Once the data are transmitted into the internal system, a one-to-one correspondence, for example, to link a certain timepoint pairs with its shipment number or container number should also be processed. A work flowchart to further explain was sketch (Table 6):

Table 6 - Work flowchart of data transmission

Layer/Step	Step 1	Step 2	Step 3	Step 4
Layer 1	GPS Module	Data Transmission Module	Central Server	Data Processing Module
Layer 2	API Nodes/Interface	Internal System's Database		
Layer 3	EDI Module	EDI Transmission Module	Internal System's Database	

In Table 6, the central server processes the data received from GPS recordings, as layer 1 shows. Then it goes to the next step - the central server transmits the data (unrestricted of data type) to the internal system's database through the API nodes/interface, as layer 2 shows. When the electronic data needs to be shared with external partners, then as layer 3 illustrates, the EDI module will be set up to standardize the data received into EDI format and transmitted via EDI network.

The proposed algorithm for calculating the on-time ratio is as follows:

1}

Date Diff in Days =

IF (

ISBLANK ('sample'[Estimate Stuffing Time]),

1,

```
DATEDIFF (

    'sample'[Estimate Stuffing Time],

    'sample'[Time of Arriving Factory],

    DAY

)

)
Date Diff in Minute =
IF (
ISBLANK ('sample'[Estimate Stuffing Time]),
1,

DATEDIFF (

    'sample'[Estimate Stuffing Time],

    'sample'[Time of Arriving Factory],
```

MINUTE

)

)

{2}

Y/N = IF ([Date Diff in Minute]>60,"Y","N")

Measure_Total_Number_Of_Container_L/ = COUNT ('sample'[Y/N])*

On_Time_The_Number_Of_N_L/ = IF([Y/N] ="N","1")*

*Measure_On_Time_The_Number_Of_N_ll/** = COUNTAX (FILTER ('sample', 'sample'[Y/N] ="N"),'sample'[Y/N])*

Measure_SortUp_On_Time_The_Number_Of_N_ll/ = SUM('sample'[On_Time_The_Number_Of_N_L/*])*

{3}

Monthly_On_Time_The_Number_Of_N_L/ = CALCULATE([SortUp_On_Time_The_Number_Of_N_ll/*],'sample'[Estimate Stuffing Time]) --- The formula means to filter by month.*

{4}

Measure_On_Time_Number_Of_Container_L/ = CALCULATE([SortUp_On_Time_The_Number_Of_N_ll/*],'sample'[Estimate Stuffing Time])*

Measure_SortUp_Monthly_On_Time_The_Number_Of_N_ll/ = SUM([Monthly_On_Time_The_Number_Of_N_L/*])*

5}

Measure_On_Time_Ratio_L/ = [SortUp_Monthly_On_Time_The_Number_Of_N_ll/]/[Total_Number_Of_Container_L/] --- The ratio is the number of guests who have not exceeded the time limit / the total number of guests. This ratio is then added up.*

Not_On_Time_Ratio_New = 1-[On_Time_Ratio]

The above algorithm can be divided into several stages, as already marked by curly parentheses and numbered. The purpose of the used syntax (DAX (Data Analysis Expressions)), together with its definition of each stage, and how it consists for achieving the final dashboard visualization.

In conclusion, with the combination of staged calculation in DAX, a final panel view of 'Trucking On-Time Ratio' was successfully achieved. The proposed result is illustrated in Figure 6:

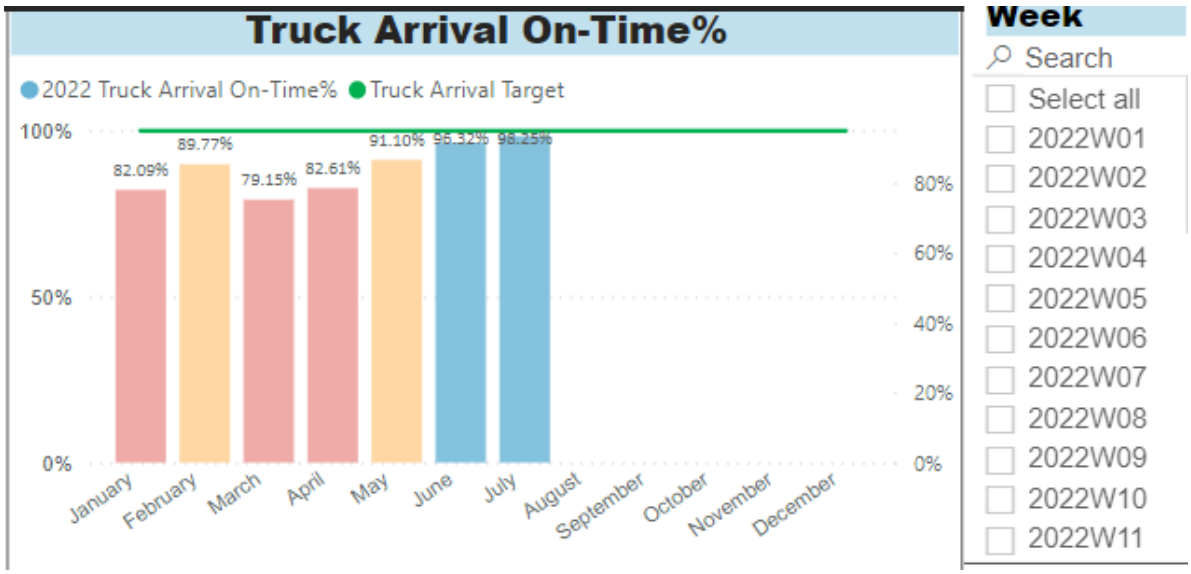

Figure 7 - proposed panel view of trucking on-time ratio [Author's elaboration]

In Figure 6, it can be seen an overview of the trucking vendor's truck arrival on-time ratio of subject client/stakeholders. The top line represents the target, for example 100%. The column stands for the certain on-time ratio of all the truck/vans in service. Statistics are distributed by month and highlighted by the respective color referring to the stage on-time ratio. For example, in June and July, the ratio is over 95% and filled in blue, which means 'good performance'. Instead, in January, March and April, the ratio is under 85% and filled with red, which can be seen as an alert, which means these results are not that satisfying from the stakeholders' viewpoint, and there is space to improve the efficiency of operating truck/vans.

On the right side of the figure, it also set up as a separated navigation bar for 'time dimension', where the end-users can see the vendor trucking performance results on a weekly basis by selecting the specific week in the navigation bar.

Regarding the cadence of uploading new raw data, which means the quantity of raw data input will be time-to-time reiterated. The total amount of raw data input will increase from time to time. However, the frequency of updating raw data inputs in the Power BI panel is subject to certain limitations. As the volume of reiterated raw data grows, processing these updates via Power Query can significantly slow down the performance of the system. This latency occurs because Power Query must reprocess and transform large datasets each time new data is uploaded, leading to longer refresh times and potential bottlenecks in data availability.

Therefore, normally in the practices, several data processing techniques to mitigate these negative impacts are adopted. For instance, to archive older data to spare some space for the system to run quicker; or batch upload and update newly input datasets. For an alternative, using other applications such as SQL to take over the data clean work first the upload the processed datasets into Power BI for modelling and visualization.

5

CONCLUSION AND DISCUSSION

5.1 Conclusions

In conclusion, this research has demonstrated the importance of dashboards in port and organizational decision-making, particularly in data visualization and real-time monitoring. By using tools like Power BI, a user-friendly and powerful dashboard that enables end users to make more informed decision in a complex data environment has been successfully created.

The importance of dashboards can be observed in: Firstly, the dashboard simplifies the way data is displayed, allowing decision-makers to quickly comprehend and process information. Secondly, the real-time updating make data analysis more flexible and efficient, improving the timeliness of decision-making. Thirdly, data visualization enables end-users' ability to gain insights to identify key trends and potential issues.

However, there are areas where dashboards can be further improved. Firstly, dashboards can be equipped with intelligent tools to be able to support trend forecasting. Secondly, automatic modelling can be developed to reduce the manual intervention on recognition of related datasets from different spreadsheets. Future research and development can further optimize the management of data sources, improving the data quality and the accuracy of analysis results, as well as to integrate more advanced features into dashboard functions, such as automation and prediction. For example, machine learning algorithms can offer automated trend forecasting and decision recommendations, helping decision-makers make more forward-looking decisions.

REFERENCES

[1] NEXUS Project, Work Package description: WP1-NEXUS Open Data Platform, presented in Oct 2022.

[2] NEXUS Project, Introduction to NEXUS Project, presented in May 2022.

[3] S. Few, "Dashboard Confusion: What Is a Dashboard?" Perceptual Edge, 2004.

[4] M. Janssen, Y. Charalabidis, and A. Zuiderwijk, "Benefits, Adoption Barriers and Theths of Open Data and Open Government," *Inf. Syst. Manag. *, vol. 29, no. 4, pp. 258-268, 2012, doi: 10.1080/10580530.2012.716740.

[5] R. S. Kaplan and D. P. Norton, "The balanced scorecard: Measures that drive performance," Harvard Business Review, vol. 70, no. 1, pp. 71-79, Jan.-Feb. 1992.

[6] M. Saunders, A. Thornhill, and P. Lewis, Research Methods for Business Students, 7th ed. Harlow, UK: Pearson, 2016.

[7] T. A. Schwandt, "Dictionary of Qualitative Inquiry," 2nd ed., SAGE Publications, 2001.

[8] M. B. Miles and A. M. Huberman, "Qualitative Data Analysis: An Expanded Sourcebook," 2nd ed., SAGE Publications, 1994.

[9] K. C. Laudon and J. P. Laudon, Management Information Systems: Managing the Digital Firm, 14th ed. IEEE Computer Society, 2015.

[10] U. Turan, Y. Fidan, and C. Yıldıran, "Critical Thinking as a Qualified Decision-Making Tool," Journal of History Culture and Art Research, vol. 8, no. 4, pp. 1-18, 2019, doi: 10.7596/taksad.v8i4.2291.

[11] What is Data Visualization?, https://www.mastersindatascience.org/learning/what-is-data-visualization/, accessed Aug. 5, 2024.

[12] Nguyen, T. T., & Jung, C. (2018). Applying data visualization techniques to enhance the operational efficiency of ports. IEEE Access, vol. 6, pp. 37955-37963, 2018, doi: 10.1109/ACCESS.2018.2853837.

[13] Heer, J., & Shneiderman, B, “Interactive dynamics for visual analysis,” Communications of the ACM, vol. 55, no. 4, pp. 45-54, 2012, doi: 10.1145/2133806.2133821.

[14] K. S. N. Ripon, A. Rahman, and G. M. A. Rahaman, “A Domain-Independent Data Cleaning Algorithm for Detecting Similar-Duplicates,” J. Comput., vol. 5, no. 12, pp. 1800–1809, Dec. 2010, doi: 10.4304/jcp.5.12.1800-1809.

[15] E. Jones, "6 Pillars of Data Quality and How to Improve Your Data," IBM Blog, May 30, 2023. [Online]. Available: https://www.ibm.com/blog/6-pillars-of-data-quality-and-how-to-improve-your-data/. [Accessed: Aug. 5, 2024].

[16] CloverDX, "Why You Should Build a Data Quality Dashboard: Benefits and Tips," CloverDX Blog. [Online]. Available: https://www.cloverdx.com/blog/why-you-should-build-a-data-quality-dashboard-benefits-and-tips. [Accessed: 05-Aug-2024].

[17] Atlan, "How to Improve Data Quality: A Comprehensive Guide," Atlan. [Online]. Available: https://atlan.com/improve-data-quality/. [Accessed: 05-Aug-2024].

[18] SunTec Data, "7 Effective Data Cleansing Strategies to Improve Data Quality," SunTec Data Blog. [Online]. Available: https://www.suntecdata.com/blog/7-effective-data-cleansing-strategies-to-improve-data-quality/. [Accessed: 05-Aug-2024].

[19] Monte Carlo, "8 Data Quality Issues: Challenges and Solutions," Monte Carlo Data Blog. [Online]. Available: https://www.montecarlodata.com/blog-8-data-quality-issues. [Accessed: 05-Aug-2024].

[20] Atlan, "Data Quality Measures: How to Improve Your Data Quality," Atlan. [Online]. Available: https://atlan.com/data-quality-measures/. [Accessed: 05-Aug-2024].

APPENDIX

Further explanation of the proposed algorithm in 'Results' chapter can be detailed as following:

In stage 1, the purpose of codes checks if the column 'sample'[Estimate Stuffing Time] is blank (for example, with empty value). If it is blank, the formula returns 1. If it is not blank, the formula calculates the difference in days and minutes between the 'sample'[*Estimate Stuffing Time*] and 'sample'[*Time of Arriving Factory*].

The ISBLANK function in DAX is used to check whether a given expression or value is blank (for example, with empty value). It returns TRUE if the expression is blank, and FALSE otherwise. The DATEDIFF calculates the difference between two dates, returning the difference in units specified by the third argument (for example, days, months, or years).

In stage 2, the purpose of codes checks whether certain events occurred within a specific time frame, in this case, to check whether the truck or van arrived at the destination (factory) within 1 hour (60 minutes) from its departure, for example, from a port, with empty container carried for loading cargo. Detailing:

- Y/N = IF ([Date Diff in Minute]>60,"Y","N"), this formula creates a calculated column that checks if the value of [Date Diff in Minute] is greater than 60 minutes. If it is, it returns "Y"; otherwise, it returns "N".
- Measure_Total_Number_Of_Container_L/* = COUNT ('sample'[Y/N]), this measure calculates the total count of the values in the 'sample'[Y/N] column.
- On_Time_The_Number_Of_N_L/* = IF([Y/N] ="N","1"), this formula creates a calculated column that checks if [Y/N] equals "N". If it does, it returns "1".
- Measure_On_Time_The_Number_Of_N_ll/** = COUNTAX (FILTER ('sample', 'sample'[Y/N] ="N"),'sample'[Y/N]), this measure uses the COUNTAX function to count the number of rows in the 'sample' table where 'sample'[Y/N] equals "N".
- Measure_SortUp_On_Time_The_Number_Of_N_ll/* = SUM('sample'[On_Time_The_Number_Of_N_L/*]), this measure sums up the values in the 'sample'[On_Time_The_Number_Of_N_L/*] column.

In stage 3, a measure called 'Monthly_On_Time_The_Number_Of_N_L/*' was created. It uses the CALCULATE function to evaluate the measure.

- [SortUp_On_Time_The_Number_Of_N_ll/*] within the context of the 'sample'[Estimate Stuffing Time] column, to consider the value of [SortUp_On_Time_The_Number_Of_N_ll/*] in relation to the [Estimate Stuffing Time] field, then to aggregate it monthly.

In stage 4, The code defines two DAX measures used for analyzing data related to on-time performance. Detailing:

- Measure_On_Time_Number_Of_Container_L/* = CALCULATE([SortUp_On_Time_The_Number_Of_N_ll/*],'sample'[Estimate Stuffing

Time]), This measure calculates the value of [SortUp_On_Time_The_Number_Of_N_ll/*] within the context of the 'sample'[Estimate Stuffing Time] column.

- Measure_SortUp_Monthly_On_Time_The_Number_Of_N_ll/* = SUM([Monthly_On_Time_The_Number_Of_N_L/*]), this measure sums up the values of [Monthly_On_Time_The_Number_Of_N_L/*]. The SUM function aggregates all the values returned by [Monthly_On_Time_The_Number_Of_N_L/*], which represents the total number of on-time containers for a given month.

In stage 5, The code defines two DAX measures that calculate the on-time ratio and the not-on-time ratio. Detailing:

- Measure_On_Time_Ratio_L/* = [SortUp_Monthly_On_Time_The_Number_Of_N_ll/]/[Total_Number_Of_Container_L/], this measure calculates the ratio of on-time containers to the total number of containers. It does this by dividing [SortUp_Monthly_On_Time_The_Number_Of_N_ll/*] (which represents the number of on-time containers in a month) by [Total_Number_Of_Container_L/] (which represents the total number of containers).
- Not_On_Time_Ratio_New = 1-[On_Time_Ratio], this measure calculates the not-on-time ratio by subtracting the On_Time_Ratio from 1. This gives the proportion of containers that were not on time.