Bol. Soc. Paran. Mat.
©SPM – E-ISSN-2175-1188
SPM: www.spm.uem.br/bspm

(3s.) **v. 2026 (44) 5** : 1–12.
ISSN-0037-8712
doi:10.5269/bspm.79920

Transforming Unstructured Customs Data into Strategic Insights with Robotic Process Automation

Kerem Yalcinkaya, Onder Sahinaslan and Ender Sahinaslan

ABSTRACT: Strategic decision-making in logistics and foreign trade relies on fast, accurate, and analytically actionable data. However, many commercial data platforms do not provide data through an Application Programming Interface (API) and must be manually obtained. Manual processes are error-prone, labor-intensive, and costly for companies. In this study, a sample data analytics framework based on Robotic Process Automation (RPA) has been developed to automatically extract, clean, and transform unstructured customs declaration data into analytical insights. Data was obtained from Datamyne.com using Python and Playwright, cleaned and standardized with Pandas and NumPy, stored in SQL Server, and visualized with Microsoft Power BI dashboards. Empirical findings show that the proposed system reduces daily data collection time by 93%, increases data accuracy from 95% to 98%, and reduces annual operational costs by approximately 94%. Furthermore, the integration of automated data pipelines with dynamic BI dashboards significantly increases analytical agility by enabling real-time monitoring, detailed analysis, and threshold-based KPI alerts. These results demonstrate that RPA-supported automation not only eliminates repetitive manual processes but also provides a measurable and sustainable competitive advantage for logistics and foreign trade operations by strengthening data-driven decision-making.

Keywords: Customs data, business Intelligence, data transformation, RPA.

Contents

1. Introduction

The globalizing trade ecosystem confronts businesses with increasingly complex supply chains, rapidly changing market dynamics, and multi-source data environments. Technological advancements and increasing customer and market expectations have led to the need for radical digital transformation and advanced data analytics in the logistics sector [1, 2]. In this environment, timely access to accurate and up-to-date data has become a critical element for businesses to gain competitive advantage in strategic decision-making processes [3]. Much of the data used in the logistics and foreign trade sectors exists scattered across the internet in semi-structured or unstructured format. Manually collecting, verifying, and reporting such data leads to significant time and cost losses [4]

Customs data are a strategic data source that contains rich information on countries' import and export transactions. This data encompasses many critical variables, such as consignment, shipper, non-cocoa, container volume, item quantity, and transportation method. However, these valuable datasets are often stored on commercial platforms that lack API access and access is limited to manual scanning or paid services. This creates both technical and operational challenges in data-driven decision-making processes [5]

In recent years, Robotic Process Automation (RPA) technologies have become widely used to increase operational efficiency and reduce human error in businesses. By automating repetitive rule-based tasks using software robots, RPA offers significant savings in both time and cost [6, 7]. Furthermore, RPA is

2020 *Mathematics Subject Classification:* 68-11, 68T09, 68P205.

Submitted November 05, 2025. Published February 14, 2026

increasingly being used not only in operational processes, but also in data analytics processes such as data collection, cleansing, and integration [8].

This study aims to evaluate RPA technology from a data analysis perspective. The focus is on systematically extracting, cleansing and restructuring customs data without API access, using RPA, and visualizing them with business intelligence (BI) tools. This creates a decision support infrastructure that can generate strategic insights from unstructured raw data.

The primary research question of this study is: "Can an RPA-based approach transform unstructured customs data obtained from trading platforms without API access into usable analytical insights for strategic decision-making?" To address this problem, Python's Playwright libraries were first used. RPA automatically extracted unstructured customs data from web platforms without API access. The data were then processed using Pandas, NumPy, and SQL Server scripts. Missing or inaccurate records were corrected and the data was made ready for analysis. Finally, dynamic reports and dashboards were created with Microsoft Power BI, which supports managers in data-driven decision-making. The RPA-based data flow diagram used is shown in Figure 1).

Figure 1: RPA-Based data flow diagram

Thanks to this integrated structure, reports generated from unstructured customs data allowed managers to quickly and accurately perform analyzes based on product, country, supplier or time. Furthermore, the system's SQL Server-based architecture allowed for data to be stored centrally, both securely and up to date. These centralized databases enabled secure data recording, processing, and transmission [9].

This study aimed to fill two important gaps in literature. First, there is a limited number of empirical studies on the integration of RPA in the context of data analytics, particularly with data sources lacking API access. Second, there are limited holistic and methodological approaches to the process of transforming unstructured web data into business intelligence systems in the fields of logistics and foreign trade.

The subsequent sections of the study are structured as follows: The second section examines the relevant literature and summarizes studies on the analytical use of RPA, business intelligence, and customs data. The third section details the methodology and system architecture used in the study, and the fourth section presents the findings and discussions. In the last section, conclusions based on the study findings and future research directions are discussed.

2. Literature

Robotic process automation and enterprise applications

Robotic Process Automation (RPA) is defined as software-based agents capable of executing repetitive, rule-based, and high-volume tasks without human intervention [5]. Initially used in operational processes such as accounting and finance, RPA has recently been widely adopted in areas such as data integration, process mining, and analytical reporting [3]. This technology increases process speed by minimizing human-related errors, advancing automation maturity in businesses [4].

Van der Aalst and others emphasized that Robotic Process Automation (RPA) is not only an operational tool but also a cognitive transformation mechanism that provides important input into data-driven decision-making processes. Similarly, Hofmann and colleagues demonstrated that the integration of RPA with artificial intelligence (AI) and machine learning (ML) technologies plays a critical role in automating data flows and increasing knowledge extraction capacity [8]. Kitsantas et al. examined the integration of RPA and AI in terms of analytics, implementation, and business process automation and discussed methodological limitations in this context [10]. In their research, they evaluated 89 academic studies on RPA and AI applications using qualitative analysis methods, comprehensively presenting current trends.

The use of RPA in data collection processes is considered an effective solution, especially on web platforms with limited API access [11]. Process suitability, data quality, user participation, and the level of system integration are prominent factors determining the success of RPA. However, studies in the literature that examine RPA in integration with data mining, data cleansing, and reporting layers are limited. This creates a significant research gap, especially in the context of transforming unstructured web data into strategic decision support systems.

Data analytics and business intelligence systems

Data analytics and business intelligence (BI) systems are fundamental technologies that enable businesses to improve decision support processes by extracting meaning from historical and current data. BI tools play a critical role in organizational structures by visualizing data, monitoring performance indicators, and supporting strategic planning decisions [12]. While these systems are recognized as a strategic tool for economic growth and corporate competitiveness, the factors determining BI-based innovative development have not yet been fully elucidated [13]. However, unlike traditional reporting systems, modern BI ecosystems offer real-time data integration, automatic updating, and interactive analysis capabilities. Tools such as Power BI, Tableau, and Qlik Sense enable the analysis of data from different data sources through dynamic dashboards, giving managers access to continuously updated insights into operational performance, supply chain efficiency, and market trends [14].

Businesses continue to seek methods and tools that can leverage big data analytics and business intelligence technologies to strengthen strategic business processes and increase operational efficiency [15, 16]. In this context, big data, data analytics, and business intelligence solutions are at the core of information-based management models and directly contribute to decision-making processes [17]. BI is considered a type of decision support system that enables organizations to achieve their goals, create value, and improve performance [18]. Business intelligence platforms visualize data sets and present them to users through reports, indicators, and interactive dashboards, facilitating decision-makers' rapid, intuitive, and holistic access to information [19, 20]. Thanks to dynamic reporting features, users can instantly access the information they need through flexible reports and filterable dashboards updated with real-time data [21]. This allows for monitoring not only historical performance analyses but also current trends and operational anomalies [14]. Thus, decision support mechanisms are evolving into a more agile, up-to-date, and data-driven structure.

Data quality is a decisive factor in the success of BI applications. Unstructured, poorly formatted, incomplete, or noisy data undermines the reliability of analytical output; therefore, data cleansing, transformation, and integration steps are considered critical stages of the analytical value chain. In this context, libraries such as Pandas and NumPy are frequently used in data preprocessing processes within the Python ecosystem, and direct integration with BI tools is possible via SQL Server or cloud-based databases.

In recent years, research on the integration of RPA technology with BI systems has revealed significant gains, such as reduced data preparation costs, lower error rates, and increased analytical agility. This integration not only provides automation efficiency but also strengthens data-based learning processes through human-robot collaboration models. This allows businesses to adapt to changing market conditions with greater flexibility, responsiveness, and insight.

In this context, common BI tools such as Google Data Studio, Qlik Sense [22], Tableau [23] and Power BI [24] were evaluated in comparisons [25]. In practice, Power BI was chosen for its compatibility with existing IT infrastructure, rapid data integration, powerful reporting features, and user-friendly interface. Furthermore, Power BI's metadata management, automatic correlation, and complex data modeling capabilities accelerate analytical processes and provide users with easy access to advanced data analysis functions. These features increase efficiency in logistics and supply chain operations, transforming strategic decision-making processes into a data-driven and agile structure.

Analytical use of customs and trade data

Customs data is strategically important for understanding countries' trade performance and supply chain movements. This data includes multidimensional information such as product codes (HS codes), countries of origin, import-export quantities, freight rates, and transportation methods [26]. However, because much of this data is presented in semi-structured or text-based PDF and HTML formats, it is not directly suitable for analytics. However, using advanced analytics, decision-makers can derive insights from unstructured data [27].

There are three main obstacles to effectively analyzing customs data. First, because many platforms do not offer API access, data collection processes are often manual, reducing efficiency. Second, the diversity of declaration formats applied in different countries leads to data heterogeneity, complicating integration processes. Finally, the presence of incomplete, inconsistent, or duplicate records complicates the data cleansing process and significantly reduces analytical accuracy. At this point, RPA-supported data collection systems offer a significant alternative. High-volume data can be extracted from sources without API access using web scraping methods [28]. This data is then cleaned with libraries such as Python and Pandas, transferred to SQL-based storage, and converted into strategic indicators using BI tools. Furthermore, the use of customs data is increasing not only in economic analysis but also in areas such as risk management, supplier performance, and sustainable trade policies [29]. However, examples in the literature where RPA is effectively used in such data preparation and transformation processes are quite limited.

Literature Gap

The literature review reveals three key research gaps on which this study is based. First, the position of RPA in the context of data analytics has not been sufficiently explored. Most existing studies consider RPA solely as a process automation tool and ignore the dimensions of data cleansing, integration, and integration with business intelligence layers [3, 4]. Second, the unstructured nature of customs data has not been adequately addressed in the literature. While technical approaches exist for converting customs declarations to PDF, HTML, or CSV formats, a comprehensive methodology for automatic data extraction from platforms without API access and transforming this data into analytical models is lacking [30, 31]. Third, empirical studies on RPA–BI integration are quite limited. Existing research generally remains at the conceptual level; Practical examples demonstrating the integration of RPA with tools like Power BI or Tableau in real-world data scenarios have been limited.

3. Method

Data Analytics Architecture

The data analytics solution proposed in this study is built on an integrated architecture consisting of three primary layers. The basic process steps of the developed solution are presented in Figure 2.

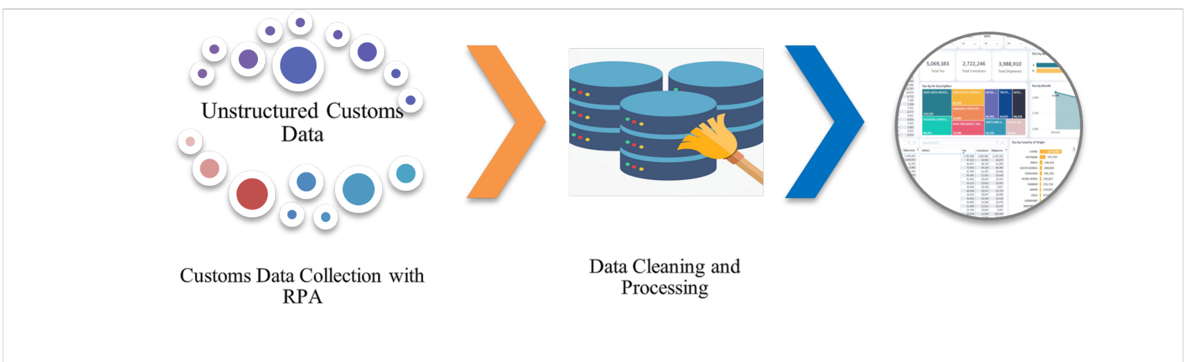

Figure 2: Integrated data analytics architecture

In this framework, the Datamyne.com platform, which lacked API access, was used as the data source. In this layer, RPA-based bots automatically performed webpage crawling, HTML parsing, and data extraction. The collected raw data was transferred to the data cleansing and transformation layer, where datasets were normalized using Pandas and NumPy libraries, and missing and inaccurate records were corrected. Finally, the cleaned data was loaded into an SQL Server database and integrated with Microsoft Power BI to create dynamic dashboards for managers.

Data Collection Process

The Python-based Playwright library was used to extract data from commercial data platforms lacking API access. This method offers an RPA-based data collection approach that mimics human user behavior by automatically checking web browsers. During the process, target URLs were identified, and access permissions were checked, web pages were navigated to select the necessary data tables, and raw data was extracted in HTML table or list format and temporarily saved in CSV or JSON format. These steps covered basic variables of customs declaration data, such as product group, country of origin, quantity, and value. bots automatically performed data collection tasks on a daily or weekly basis without user intervention.

Data Cleaning and Processing Findings

Raw data often contains incomplete, duplicate, or inaccurate records, which can negatively impact the accuracy and reliability of analytical processes. Therefore, the data cleaning process is a critical step in ensuring analytical accuracy and data integrity. During this process, missing values were filled using appropriate statistical methods (e.g., median, mean, or estimation-based approaches), duplicate records were identified and removed from the dataset, date and numeric formats were standardized, and text-based categorical data was normalized.

All these processes are conducted within an automated data processing pipeline built using the Pandas and NumPy libraries. Data quality metrics are recorded at every stage of the process, and accuracy checks are performed before data is transferred to the SQL database, ensuring data integrity

Database Integration

The cleaned data sets were transferred to a central SQL Server database. This layer is designed to maintain data integrity, provide secure storage, and enable multi-user access. The database structure was structured to ensure the organized and relational management of core data components related to customs transactions. To achieve flexibility and increased performance in reporting processes, a snowflake schema was chosen in the data warehouse architecture. To this end, a fact table and multiple dimension tables connected to this table were designed.

a) Fact_Customs Table: The central fact table contains the atomic data representing each customs shipment or bill of lading line. This table contains numerical measurements and foreign keys that establish relationships with other dimension tables. The table's basic components are: i) Fact Values: Kilograms, metric tons, container count, TEU value ii) Dimension Keys: Contains main product group, subcategory, and description information. It consists of the fields MasterShipperId, MasterConsigneeId, ConsolidatedConsigneeId, NvoccId, VoccId, ShipperDeclaredId, ConsigneeUnifiedId, ShipperUnifiedId, PortOfArrivalId, PortOfDepartureId, ShipmentTypeId, BillTypeId, and CarrierId.

b) Dimension Tables: i) DimMasterShipper: Contains a standardized master list of shippers, cleaned with fuzzy logic. (MasterShipperID, MasterShipperName) ii) DimMasterConsignee: Contains a standardized list of master receivers. (MasterConsigneeID, MasterConsigneeName) iii) DimNVOCC / DimVOCC: Contains unique lists of carriers NVOCC and VOCC companies. (NVOCC_ID, NVOCC_Name) iV) DimPortOfArrival: Contains a list of destination ports and their respective regions (US Region). (DestinationPortID, PortName, RegionName) v) DimPortOfDeparture: Contains a hierarchical list of departure ports, their respective countries, and their respective world regions. (DestinationPortID, PortName, Country, WorldRegion) vi) DimShipmentType: Contains a list of shipment types (e.g., FCL, LCL). (ShipperTypeID, ShipperTypeName) vii) DimBillType: Contains a list of bills of lading types (e.g., Master, House). (ConsignmentTypeID, ConsignmentTypeName) viii) DimHsDescription: Lists the Hs Codes and their descriptions for multiple items included in the declaration.

This structural arrangement not only ensures that data queries are executed quickly and consistently but also allows for the creation of an optimized data model for BI tools.

Visualization and Presentation

The Power BI platform facilitates the transformation of data, which has been meticulously cleaned and loaded into a SQL Server database, into dynamic and interactive dashboards that integrate both visualization and advanced analytical capabilities. These dashboards are designed to support data-driven decision-making processes by enabling comprehensive insights across multiple dimensions. Within the scope of this study, the developed dashboards encompass several critical analytical areas, including the identification of potential customers, competitor analysis and strategic customer acquisition, enhancement of existing customer relationships, and monitoring of market trends to detect new opportunity areas.

In a study conducted by Barsan for the Chicago Region in the United States, spot customers with a monthly volume of 20 TEU or less, originating from Asian countries where the company is strong (China, Vietnam, Malaysia, Thailand, and India), and located in Illinois, Michigan, and Ohio, were targeted. After filtering the report, these potential customers were contacted, and approximately 1% received positive feedback. This report also accesses the consignee and country information of NVOCCs conducting high-volume business in the US. It allows them to see the volume of business they are doing with which companies, as well as the customers they have lost or gained, thus creating new opportunities. The Chicago regional government actively uses this report when making decisions about which sectors or countries to focus on. Filtering and drill-down features allow users to quickly access necessary insights at the operational and strategic level. The proposed methodology has an integrated architecture that includes RPA-based data scraping, data cleansing, SQL-based storage, and Power BI integration. Automation and data validation checks implemented at every stage of the process significantly reduce data processing time and minimize error rates. Furthermore, the system's modular design increases the solution's scalability by allowing for the easy integration of different data sources or additional analytical layers.

4. Findings and Discussion

Data Collection Findings

Customs data extracted from the Datamyne.com platform, which has limited API access, using RPA-supported Python-Playwright-based bots, covers United States maritime import data for 2025. The raw data set contains approximately 15,000,000 records and covers fields such as consignment, shipper, port of origin, country of origin, carrier, container quantity, weight, number of declarations, and number of items. A sample data set of raw customs data is shown in Table 1.

Table 1: Sample Records from the Raw Customs Datasets

Date	Shipper	Receiver	Bill of Lading No	HS Code Definition	Port of Departure	Country of Origin	Container	Kg	Teu
9/5/2025	MAXXX	ADXXX	MEDXXX	610110 – MEN'S OR BOYS'	VUNG TAU, VIETNAM	CAMBODIA	1	9060	2
9/5/2025	ROXXX	ADXXX	MEDXXX	640110 – WATERPROOF F	VUNG TAU, VIETNAM	CAMBODIA	1	6981	2
9/5/2025	CAXXX	ADXXX	MEDXXX	640110 – WATERPROOF F	VUNG TAU, VIETNAM	CAMBODIA	1	6418	2
9/5/2025	CAXXX	ADXXX	MEDXXX	640110 – WATERPROOF F	VUNG TAU, VIETNAM	CAMBODIA	1	6584	2
9/5/2025	XOXXX	ADXXX	MEDXXX	640110 – WATERPROOF F	VUNG TAU, VIETNAM	CAMBODIA	1	6253	2

Raw data was automatically extracted using RPA, requiring no manual intervention. The process increased data collection speed by 93% compared to traditional manual methods. Furthermore, data accuracy was increased by eliminating human errors.

Data Cleaning and Processing Findings

The unstructured raw customs dataset inherently contained incomplete, duplicate, and inaccurate records. As a result of the cleaning and standardization process performed on this dataset using Pandas and NumPy libraries and SQL Server scripts in the database layer, duplicate records were reduced by approximately 3%, missing information was removed, misspelled categorical values were corrected using a fuzzy logic algorithm, missing truth values were filled using predictive methods, and date and numeric formats were normalized for consistency. A sample of the cleaned dataset is presented in Table 2.

Cleansed data was imported into SQL Server, optimizing data integrity, access control, and query performance. This structure enabled business intelligence (BI) tools to dynamically extract data, ensuring

Table 2: Sample records from the cleaned dataset

Date	ID	HS Code	Bill of Lading No	Sender ID	Receiver ID	Arrival Port ID	Departure Port ID	Container Quantity	Kg	Origin Country ID	Teu
9/5/2025	1911205	640110	MEDXXX	311286	14581	75	40	1	8878	87	2
9/5/2025	1911193	640110	MEDXXX	311286	14581	75	40	1	6981	87	2
9/5/2025	1911196	640110	MEDXXX	49163	14581	75	40	1	6418	87	2
9/5/2025	1911198	640110	MEDXXX	49163	14581	75	40	1	6584	87	2
9/5/2025	1911201	640110	MEDXXX	418147	14581	75	40	1	6253	87	2

its timeliness and security. After the cleansing process, data quality measurement accuracy reached 98%, demonstrating that RPA-powered automation provides high reliability not only in data extraction but also in data preparation processes.

Mathematically, the completeness and consistency improvements in the dataset can be formalized through several data quality indicators. The missing value rate for an attribute a_j was computed as $r_{\text{miss}}(a_j) = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}\{x_i(a_j) = \varnothing\}$, while the duplicate rate was defined as $r_{\text{dup}} = \frac{N - |\text{unique}(\mathcal{D})|}{N}$. After the cleaning pipeline was applied, the improvement in duplicate reduction was calculated as $\Delta r_{\text{dup}} = r_{\text{dup}}^{\text{before}} - r_{\text{dup}}^{\text{after}} \approx 0.03$. Furthermore, overall validation accuracy—based on logical consistency, referential integrity, and format correctness—was expressed as $\text{Acc} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}\{\text{“validate”}(x_i) = \text{True}\} \approx 0.98$. These metrics empirically confirm that the automated cleaning framework ensures systematic enhancement of data quality while increasing analytical reliability.

Power BI Dashboards and Strategic Insights

Dashboards developed using Power BI enable managers to perform comprehensive analyses across products, countries, and suppliers, providing both strategic and operational insights. Example dashboards include identifying potential customers, analyzing competitors, assessing existing customer development, and analyzing logistics trends over time. A sample screenshot of the BI dashboard created for this purpose is presented in Figure 3.

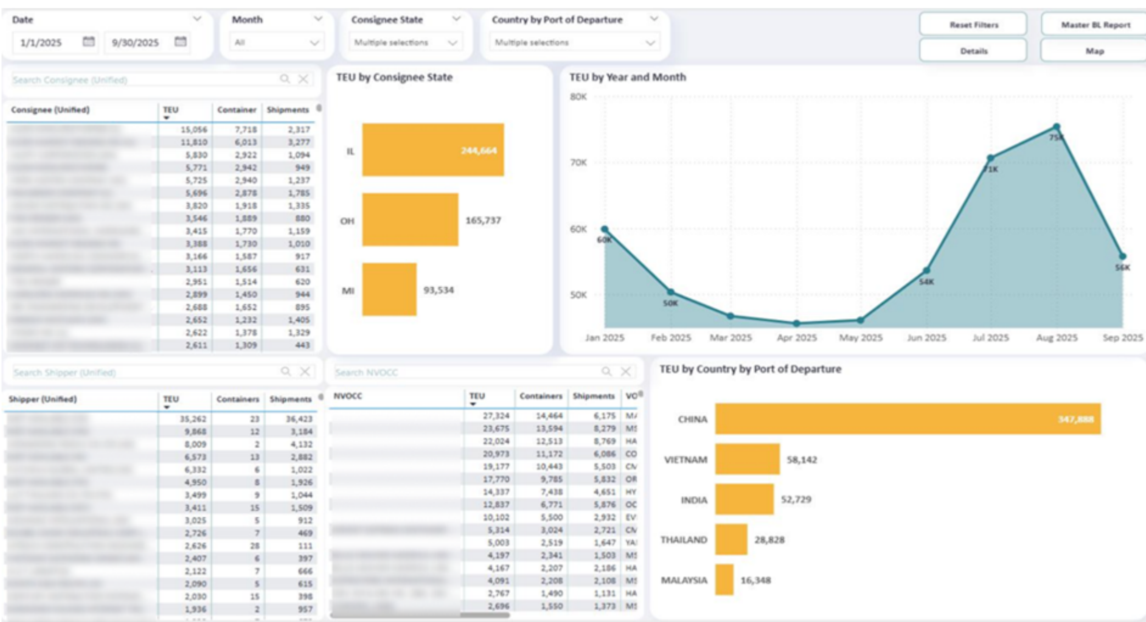

Figure 3: Integrated data analytics architecture

The Power BI dashboard in Figure 3 is designed as a dynamic decision support interface that enables multidimensional data analysis related to import processes. IL, OH, and MI were selected for the State filter, while China, Vietnam, India, Thailand, and Malaysia were selected for the Origin Country filter. Consignee values based on these criteria are presented in the top left table, state-based TEU distributions are presented in the middle bar chart, time-based trend analysis in this line is presented in the top right, and TEU distribution by origin country is presented in a bar chart in the bottom right. The business intelligence dashboard makes operational performance indicators interactive, allowing users to instantly filter and visualize data by sender, receiver, VOCC/NVOCC, shipment type, ports, and period. The drill-down feature allows managers to gain detailed insights into a specific product group or country, while KPI indicators generate automatic alerts when set thresholds are exceeded, providing instant notifications to decision makers. This transforms the RPA-automated data stream into real-time analytical output in the BI environment, resulting in measurable improvements in data freshness, accuracy, and agility in decision support systems. This integrated framework is designed to help increase the analytical maturity and competitiveness of a business by facilitating managers' data-driven, rapid, and strategic decisions. It provides strategic insights into decision-support processes.

From a mathematical perspective, the analytical behavior of the dashboard can be formalized through aggregation and threshold functions. For any selected period t, the total imported TEU volume is defined as:

$$V_t = \sum_{i=1}^{N_t} val_i \tag{4.1}$$

where val_i denotes the TEU value of shipment i. The proportional contribution of a specific origin country c to total import volume is expressed as:

$$\text{Share}(c) = \frac{\sum_{i:\,\text{country}_i = c} val_i}{\sum_{i=1}^{N} val_i} \tag{4.2}$$

Equations (1) and (2) are calculated dynamically within Power BI as the user applies filters or detailed actions. To operationalize administrative decision rules, an alert function based on a threshold value is defined as follows:

$$\text{Alert} = \begin{cases} 1, & \text{if Share}(c) > \tau_c \text{ or } V_t < \tau_v \\ 0, & \text{otherwise} \end{cases} \tag{4.3}$$

where τ_c and τ_v represent business-defined thresholds associated with country-concentration risk and temporal volume decline. Similar KPI-driven formulations have been utilized in BI-supported logistics analytics research [31]. These expressions formalize how the RPA-automated data stream is transformed into real-time, anomaly-sensitive indicators, thereby strengthening the analytical maturity of the decision-support framework.

Performance Analysis Results In this study, significant performance gains were achieved because of the implementation of RPA and BI integration. Daily data collection time was reduced by approximately 93% compared to manual methods, resulting in significant time savings. Error rates in terms of data accuracy decreased from 5% to 2% increasing analysis reliability. Furthermore, thanks to the automatic updating of BI dashboards, managers gained access to real-time insights and significantly improved analytical agility. These gains strengthened the data-driven decision-making culture within the organization, creating a sustainable competitive advantage.

The findings comparing the process performance values of the developed method with the results of the manual method are presented in Table 3.

Table 3: Findings Regarding Process Performance Comparison Between Methods

Measure	**Manual Process**	**RPA + BI Process**	**Difference (%)**
Data Collection Time	3 hours per day	0.2 hours per day	-93%
Data Accuracy Rate	95%	98%	+3%
Reporting Update	weekly	daily	-
Operational Cost	70,404 $ / Year	4,272 $ / Year	-93%

When the reporting process is done manually, the data collection and cleaning phase is the most significant difference. Downloading 5,000-line Excel files from the website, merging them, and manually cleaning them takes an employee approximately three hours each day. Because this process is automated end-to-end with RPA, it doesn't require a daily workforce. However, when there's a change to the website interface or an improvement, there's an immediate workforce requirement. Since the report went live on March 1, 2024, less than an hour per week has been spent on such maintenance and development. This demonstrates that the 15 hours of weekly manual effort (3 hours/day x 5 days) has been reduced to less than one hour, achieving a time efficiency of over 90%.

Evaluating this time's savings from a financial perspective more clearly demonstrates the tangible benefits of the project. When comparing labor costs based on the gross minimum wage (20,002.50 $) set for 2024, a modest comparison is made between the three hours of manual effort spent per day and the equivalent of approximately 66 working hours per month. The annual cost of this time based on the gross minimum wage exceeds 70,000 $. In contrast, the annual cost of maintaining the system after RPA automation, which amounts to approximately four hours per month, is around 4,300 $. Comparing these two scenarios, it appears that RPA automation provides direct annual labor savings of approximately 66,000 $, even based on the most basic cost estimate. However, the real financial benefit of the project arises beyond these savings, as it redirects the time of qualified personnel from repetitive tasks like data collection to high-value-added activities like generating strategic insights from the obtained data.

The performance improvements obtained with RPA–BI integration can be mathematically formalized through time-efficiency and cost-efficiency metrics. The time-saving ratio is defined as:

$$G_{\text{time}} = 1 - \frac{T_{\text{RPA}}}{T_{\text{manual}}} \tag{4.4}$$

where T_{manual} represents the daily duration of manual data collection and T_{RPA} denotes the automated processing time. Substituting the observed values (3 hours per day manually, 0.2 hours per day with RPA), Equation (4) yields $G_{\text{time}} \approx 0.93$, indicating a 93% reduction in time spent.

Similarly, labor-based cost savings are expressed as:

$$G_{\text{cost}} = 1 - \frac{C_{\text{RPA+BI}}}{C_{\text{manual}}} \tag{4.5}$$

where C_{manual} denotes the annual personnel cost required for manual execution and $C_{\text{RPA+BI}}$ denotes the annual maintenance cost of the automated system. Using the reported values (70,404 $ vs. 4,272 $), Equation (4.5) results in $G_{\text{cost}} \approx 0.94$, demonstrating significant financial gains.

In addition, data-quality improvement can be expressed by the accuracy function:

$$\text{Acc} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{I}\{\text{validate}(x_i) = \text{True}\} \tag{4.6}$$

where Acc increased from 0.95 to 0.98 after process automation, improving analytical reliability. Empirical validation of Equations (4) and (4.6) confirms that the automated pipeline not only reduces operational workload but also increases analytical precision, which aligns with evidence reported in recent RPA-driven data analytics studies [31].

Discussion

The findings reveal that RPA-based data scraping and BI integration provide significant contributions to logistics and customs data analytics. Compared to manual or semi-automated data collection methods reported in the literature, the developed approach demonstrated superior performance in terms of both operational efficiency and analytical accuracy [3]. While the heterogeneous and unstructured nature of customs data poses a significant challenge in traditional analytical processes, RPA's automated data cleansing and transformation capabilities have largely eliminated this obstacle. Thus, the rapid and accurate integration of data into BI platforms has significantly accelerated decision-making processes. RPA is a technology that automates repetitive business processes, but its application in customs management is limited due to the unique nature of the field [32]. The modular architecture of the developed system allows for the easy addition of different data sources or new customs reports, increasing the scalability of the solution. In this respect, the study offers both a methodological and practical contribution to the topic of RPA–BI integration, which has been addressed only limitedly in the literature. This proposed approach, which transforms data sources without API access into analytical values, provides a guiding framework, particularly for the development of data-driven decision support systems [8]. RPA-supported data scraping processes increased process efficiency by saving approximately 93% of the time compared to manual methods. Thanks to data cleansing and BI integration, error rates decreased from 5% to 2%. The findings demonstrate that RPA–BI integration provides significant advantages at the operational, analytical, and strategic levels in logistics and customs data analytics.

5. Conclusion and Future Study

Conclusion

This study examined the transformation of unstructured customs data without API access into strategic insights using Robotic Process Automation (RPA) and Business Intelligence (BI) tools. As part of the research, automated data scraping processes were conducted using Python and Playwright libraries from the Datamyne.com platform. The resulting raw data was cleaned and restructured using Pandas and NumPy. This data was then imported into a SQL Server database and integrated with Microsoft Power BI to create dynamic, interactive dashboards. Thus, RPA-based data automation and BI-based visualization processes were brought together under an integrated decision support architecture.

Research findings demonstrate that RPA–BI integration offers significant advantages at the operational, analytical, and strategic levels in logistics and customs data analytics. RPA-supported data scraping processes have increased process efficiency by saving approximately 93% of time compared to manual methods, and error rates have decreased from 5% to 2% thanks to data cleansing and BI integration. Dynamic dashboards have strengthened a data-driven decision-making culture by providing managers with instant analytical insights based on products, countries, and suppliers. This integrated structure has not only provided businesses with operational agility but also achieved sustainable competitive advantage and effective supply chain management.

The results provide both theoretical and practical contributions to the topic of RPA–BI integration, which has limited application examples in the literature. The study demonstrates that unstructured customs data can be used not only in economic analysis but also in supply chain management, risk analysis, and sustainable trade strategies. From an academic perspective, the integration of RPA-supported data extraction and BI-based visualization fills a methodological gap in the literature, proposing a reusable framework for data processing and analytical transformation. From a practical perspective, the developed model provides a scalable solution for the effective use of data sources without API access by providing time savings in data collection and reporting processes, reduction in error rates and agility in decision support mechanisms.

Future Study and Recommendations

While this study has successfully demonstrated the applicability of RPA and BI integration in the field of customs data analysis, it also presents several areas for future development. Initially, the data management process was implemented on SQL Server; however, future integration with big cloud-based data infrastructures (e.g., Microsoft Azure, AWS, or Google Cloud) could offer significant advantages in terms of scalability and the ability to respond to increasing data volumes. Furthermore, applying machine learning algorithms to cleansed datasets can strengthen the system's decision support capabilities in advanced analytical scenarios such as demand forecasting, price trend analysis, and risk prediction. Furthermore, current RPA processes focus on collecting data at specific time intervals. Integrating web sockets or stream-based real-time data streams can ensure BI dashboards are constantly updated, increasing agility in decision-making processes. The study's limitation to Datamyne.com data constitutes a limitation in terms of scope; therefore, integrating different customs and trade data sources will increase the generalizability and robustness of the proposed model. Finally, improvements focused on user experience (UX) and interaction design in Power BI dashboards can support organizational adoption of the system by increasing the speed and accuracy of managers' analysis processes. Overall, this study demonstrated that RPA and BI integration is an effective and innovative approach for transforming unstructured customs data into strategic insights. RPA eliminates manual errors by automating the data extraction process, while Pandas and NumPy-based data cleansing steps ensure analytical accuracy and data consistency. The dynamic reporting infrastructure provided by the integration of the SQL Server database and Power BI provides managers with real-time, interactive, and decision-oriented analysis. This integrated approach maximizes the analytical value of customs data, supporting data-driven, rapid, and reliable decision-making processes in logistics and supply chain management. Therefore, the study not only contributes methodologically to academic literature but also serves as a concrete example of digital transformation-focused data analytics solutions in practical applications.

Acknowledgments

The authors acknowledge that some of the findings presented in this paper were previously shared at the 9th International Conference on Mathematical Sciences (ICMS 2025) held at Maltepe University, Istanbul, Türkiye. The authors would also like to express their sincere gratitude to the management and expert teams of Barsan Global Logistics Inc. for their valuable support, resources, and contributions throughout the completion of this study.

References

[1] J. Fellenstein and A. Umaganthan, *Digital transformation: How enterprises build dynamic capabilities for business model innovation: A multiple-case study within the logistics and transportation industry.*, 2019.

[2] S. Roden, A. Nucciarelli, F. Li, and G. Graham., *Big data and the transformation of operations models: A framework and a new research agenda. Production Planning and Control*, 28(11–12), 929–944. (2017). doi: 10.1080/09537287.2017.1336792

[3] R. Syed, S. Suriadi, M. Adams, W. Bandara, S.J.J. Leemans, C. Ouyang and Arthur H.M., *Robotic process automation: Contemporary themes and challenges. Computers in Industry* (2020).doi: 10.1016/j.compind.2019.103162

[4] M. Lacity, and L.P. Willcocks, *Robotic process automation and risk mitigation: The definitive guide. SB Publishing..* (2017).

[5] S. Aguirre, and A. Rodriguez., *Automation of a business process using robotic process automation (RPA): A case study. In Workshop on engineering applications*, (pp. 65-71) (2017).

[6] V. Bhardwaj, *A systematic review of robotic process automation in business operations: Contemporary trends and insights. IJ. Intell Syst. Control*, 2(3), 153-169 (2023). doi: 10.56578/jisc020304

[7] W.M.P. Van der Aalst, M. Bichler and A. Heinzl., *Robotic process automation Business and Information Systems Engineering*, 60, pp.269–272 (2018). doi: 10.1007/s12599-018-0542-4

[8] P. Hofmann, C. Samp, and N. Urbach., *Robotic process automation.*, *Electron Markets*30, pp. 99–106 (2020). doi: 10.1007/s12525-019-00365-8

[9] E. Şahinaslan, E. and Ö. Şahinaslan., *Microsoft SQL sunucusunda veritabanı kurtarma teknikleri.*, *International Journal of Innovative Engineering Applications*, Vol. 6(1), pp. 158-169 (2022) doi: 10.46460/ijiea.1070325

[10] T. Kitsantas, P., Georgoulas, & E. Chytis, *Integrating robotic process automation with artificial intelligence for business process automation: Analysis, applications, and limitations*, *Journal of System and Management Sciences*, 14(7), pp 217–242 (2024). doi: 10.33168/JSMS.2024.0712

[11] V. Sharma, A. Khang, P. Hiwarkar, & B. Jadhav, *Robotic process automation applications in data management. In AI-centric modeling and analytics*, pp. 238-259 (2023).

[12] H. Chen, R.H.L. Chiang, & V.C. Storey., *Business intelligence and analytics: From big data to big impact. MIS Quarterly*, 36(4), 1165–1188 (2012). doi: 10.2307/41703503

[13] C.M. Olszak, *Business intelligence systems for innovative development of organizations. Procedia Computer Science*, Vol. 207, 1754-1762, ISSN 1877-0509 (2022). doi: 10.1016/j.procs.2022.09.233.

[14] R. Akyüz R., Y. Güzelaydın, E. Şahinaslan, Ö. Şahinaslan., *Information-based management in logistics: The strategic role of dynamic reporting Tools. International Graduate Students Interdisciplinary Symposium-2025*, Maltepe University, Istanbul, Turkey.

[15] V. Grover, R.H. Chiang, T.P. Liang, & D. Zhang, *Creating strategic business value from big data analytics: A research framework. Journal of Management Information Systems*, 35(2), 388–423. (2018). doi: 10.1080/07421222.2018.1451951

[16] Y. Niu, L. Ying, J. Yang, M. Bao, & C.B. Sivaparthipan, *Organizational business intelligence and decision making using big data analytics. Information Processing & Management*, 58(6), 102725. (2021). doi: 10.1016/j.ipm.2021.102725

[17] A. Al-Okaily, A.P. Teoh, & M. Al-Okaily, *Evaluation of data analytics-oriented business intelligence technology effectiveness: an enterprise-level analysis. Business Process Management Journal*, 29(3), 777-800. (2023). doi: 10.1108/BPMJ-10-2022-0546

[18] M. Alnoukari, *A framework for big data integration within the strategic management process based on a balanced scorecard methodology. Journal of intelligence studies in business*, 11(1), pp 33-47. (2021).

[19] Ö. Şahinaslan, C. Karakaş, and E. Şahinaslan., *An innovative approach to order picking and positioning in logistics warehouses. Verimlilik Dergisi*, 57(3), 491–512. (2023). doi: 10.51551/verimlilik.1188635

[20] M. Günerkan, E. Şahinaslan, and Ö. Şahinaslan., *Examining the efficiency of learning-based algorithms in the process of declaring customs. Acta Infologica*, 6(2), 175–188. (2022). doi: 10.26650/acin.1057060

[21] E. Şahinaslan, *İş zekâsı ve veri analizi.* In G. Telli & S. Aydın (Eds.)*Bilgi Çağı'nda Müşteri İlişkileri Yönetimi*, pp. 197–211. İstanbul: Kriter Yayınevi. ISBN: 978-625-7130-19-6 (2022).

[22] Qlik. (2025, July 24). Qlik Sense. https://www.qlik.com/us/products/qlik-sense

[23] Tableau. (2025, July 25). Tableau. https://www.tableau.com/

[24] Microsoft. (2025, July 24). *Power BI. https://www.microsoft.com/en-us/power-platform/products/power-b*

[25] S. Gaenger, *E10 business intelligence tools to visualize and analyze your data. Atlassian. https://www.atlassian.com/data/business-intelligence/10-data-visualization-tools*

[26] World Customs Organization, (2025, June 11). *What is HS Code? HS Code Explained https://www.tradefinanceglobal.com/customs/what-is-an-hs-code/*

[27] [27] UNCTAD (2025, September 25). 2025 Technology and Innovation Report, https://unctad.org/system/files/official-document/tir2025_en.pdf, United NationsConference on Trade and Development, United Nations Publications.

[28] V. Pichiyan, S. Muthulingam, S. G, S. Nalajala, A. Ch, M.N. Das, *Web scraping using natural language processing: Exploiting unstructured text for data extraction andaAnalysis. Procedia Computer Science*, V 230, pp 193-202. (2023). doi: 10.1016/j.procs.2023.12.074

[29] OECD. (2024). OECD. (2024). Risks and Resilience in Global Trade. OECD Publishing. https://doi.org/10.1787/1c66c439-en.

[30] J. Wewerka, & M. Reichert, *Robotic process automation - a systematic mapping study and classification framework. Enterprise Information Systems*, 17(2) (2023). doi: 10.1080/17517575.2021.1986862

[31] L. Patrício, L. Varela, & Z. Silveira, *Integration of artificial intelligence and robotic process automation: Literature review and proposal for a sustainable model. Applied Science*, 14(21), 9648. (2024). doi: 10.3390/app14219648

[32] Q. Gao, & Z. Kuang, *Can Robotic Process Automation Technology Enable Risk Data Analysis for Customs' Post-Clearance Audit: A China Customs Case Study. World Customs Journal*, 17(2), 93–104. (2023). doi: 10.55596/001c.88821

Kerem Yalcinkaya,
Barsan Global Logistik
Information Technology
New York, USA.
E-mail address: kerem.yalcinkaya@barsan.com

and

Onder Sahinaslan
Maltepe University
Department of Informatics
Istanbul, Turkey.
E-mail address: ondersahinaslan@maltepe.edu.tr

and

Ender Sahinaslan
Mudanya University
Department of Management Information Systems
Bursa, Turkey.
E-mail address: ender.sahinaslan@mudanya.edu.tr