Data Science and Engineering (DSE) Record, Volume 5, issue 1.

Data Quality Assessment of Economic Indicators in Thailand

Chanwit Chanton [1] Pairach Piboonrungroj [2] and Juggapong Natwichai [3]

[1] Master's Degree Program in Data Science, Chiang Mai University, Chiang Mai, Thailand
[2] Faculty of Economics, Chiang Mai University, Chiang Mai, Thailand
[3] Department of Computer Engineering, Faculty of Engineering, Chiang Mai University, Chiang Mai, Thailand
chanwit_c@cmu.ac.th

Abstract. This study implements data quality assessment framework to rapid economic indicators. Due to the outbreak of the COVID-19 pandemic abruptly halted economic activities worldwide. Assessing its economic im-pact using traditional economic indicators has proven insufficient for the ur-gent analytical and decision-making needs. The advent of Big Data, charac-terized by its diverse sources and frequent reporting for real-time monitoring. However, a critical challenge is the absence of standardized data quality as-sessment frameworks. Neglecting data quality assessment while employing Big Data for decision-making may lead to erroneous decisions. This study evaluates rapid economic indicators, Apple Mobility Index, Global Normalcy Index, and Google search trends. An existing data quality assessment frame-work and data quality dimensions—accuracy, timeliness, and validity—are assessed by Talend Open Studio for Data Quality. Findings reveal the Global Normalcy Index as a promising rapid economic indicator for timeliness and validity. However, accuracy testing yielded inconclusive results due to its fluctuations. This highlights the need for a nuanced approach with consider-ing data characteristics. Future endeavors should diversify data quality di-mensions and refine the assessment framework to enhance data quality as-sessment efficiency.

Keywords: Data Quality, Economic Indicators, Big Data.

1 Introduction

During The COVID-19 pandemic, strict measures led to a sudden economic shutdown, significantly business and household income dropping [1]. To mitigate these impacts, Thai government needs high-quality data for prompt and accurate economic assessments and decision making. The pandemic highlighted needs for the government to move away from traditional, often delayed, and inaccessible statistics. Real GDP growth, reported quarterly and considered a lagging indicator, fails to capture rapid economic changes which were affected by COVID-19 positive rate, making it less effective for timely decision making and trend identification as illustrated in Figure 1.

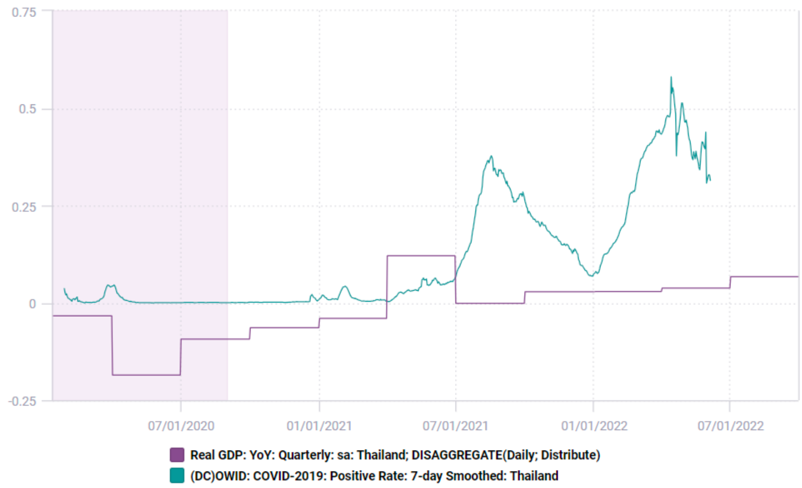

Fig. 1. Comparison of daily COVID-19 positive rate index [26] and quarterly real GDP year-over-year growth (%) in Thailand [27].

The modern approach to tracking economic impacts of the outbreak involves using rapid economic indicators, providing real-time or high-frequency data from sources like credit card transactions. These indicators offer early insights into economic activity, preceding traditional datasets like GDP, which are reported quarterly [3]. In addion, there are a strong correlation between rapid economic indicators and traditional benchmarks during economic crises [4]. Metrics like electricity consumption align well with broader consumption patterns, making them effective indicators of economic activity, as illustrated by their consistent fluctuations with industrial and commercial consumption as depicted in Figure 2.

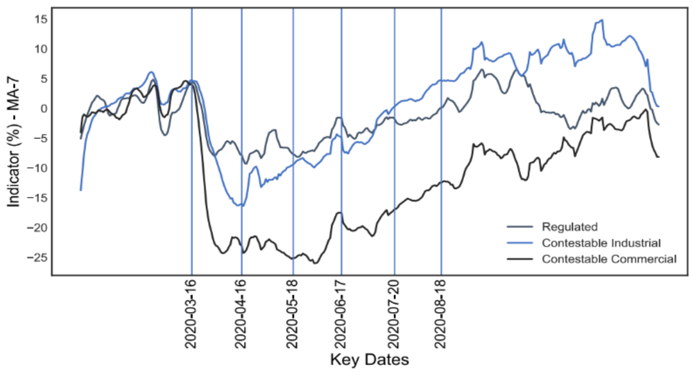

Fig. 2. Electricity consumption indicator for regulated and contestable indicators (%) [4].

However, evaluating the credibility and completeness of data quality is essential to improve the effectiveness of policy implementation. The research on Big Data reveals challenges such as data integration complexity, real-time analysis issues, and lack of quality assessment standards, leading to decision-making problems [5]. According to Gartner's survey, poor data quality costs organizations an average of $12.9 million annually [6]. To eliminate the data quality issues in economic indicator analysis, it's essential to assess data quality. The assessment results should then guide corrective actions to improve data quality, examining various dimensions of data quality issues. This depends on the analysts' data usage needs as illustrated in Figure 3.

Fig. 3. Data quality dimensions (DQD) [15].

Furthermore, evaluating data quality must be conducted within a data quality assessment framework, following the steps of the data quality management process, which involves specifying data usage requirements, conditions, and quality measurement rules to align with data quality dimensions [7]. Therefore, in the context of the COVID-19 pandemic and the data requirements of the Thai government, this study aims to improve comprehension of how to assess data quality in economic indicators for policymakers. Its goal is to provide a framework for assessing data quality and choosing indicators effectively based on their data quality for particular situations.

2 Literature Review

2.1 Review of data quality and development of its framework

Data quality, or "fitness for use," measures how well a dataset meets user requirements and ensures precise decision-making [8]. Issues like manual data entry errors, data integration conflicts, and lack of governance regulations result in poor-quality data, affecting decision-making outcomes as illustrated in Figure 4 [9].

Fig. 4. Pyramid of data quality [10].

To accomplish this, the data quality management process is depicted in Figure 5 involves steps to identify issues, make improvements, and maintain high data quality, aligning with the goal of creating a unified data warehouse for informed decision-making [11].

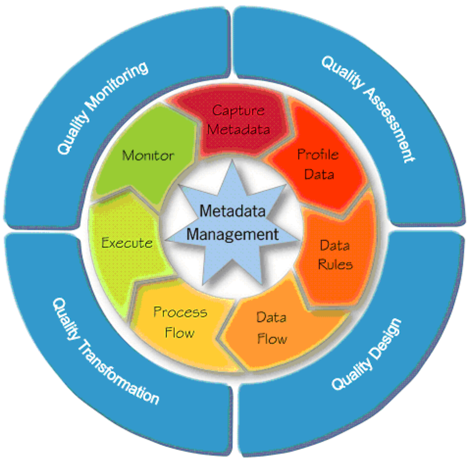

Fig. 5. Pyramid of data quality [12].

The crucial concern is in the initial stage of data quality management, identifying issues involves considering data characteristics and user requirements. A common framework uses testing metrics from data quality dimensions to set improvement goals, benchmarking the current data quality. These benchmarks guide improvement and maintenance efforts, metrics explained through DAMA's six data quality dimensions [13]. Errors in data management can impact multiple data quality dimensions simultaneously as is depicted in Table 1. Data quality rules are essential for assessing whether data points meet specified requirements, directly affecting data quality. Establishing precise criteria improves the efficiency of data quality assessments [16].

Table 1. Comparison of data quality issues and its data quality dimension [17].

	Data Quality Dimensions (DQD)		
Data Quality Issues	Accuracy	Completeness	Consistency
Missing data	X	X	
Incorrect data, Data entry errors	X		
Irrelevant data			X
Outdated data	X		
Misfielded and Contradictory values		X	X
Uniqueness constrains, Functional dependency violation	X		
Wrong data type, poor schema design			X
Lack of integrity constraints	X	X	X

Data quality rules are formulated by data users within each organization, relying on specialized expertise in the dataset. Experts establish these rules based on business operations or objectives to ensure effective analysis and decision-making. In this study, implementing data quality rules for economic indicators is essential for ensuring the accuracy and reliability of metrics. For example, a rule to verify monthly Consumer Price Index (CPI) changes within a ±0.5% range can help maintain data integrity. Any deviation triggers an investigation to correct errors or anomalies, boosting confidence in the data and aiding informed decision-making for policymakers.

To evaluate data quality, data quality profiling is used. This process assesses aspects like accuracy, completeness, consistency, and validity to identify anomalies, aiming to understand the overall data status and uncover potential issues that might impact its effectiveness for analysis, reporting, or decision-making [14]. Data quality profiling facilitates the creation of data quality rules and the choice of suitable assessment tools. Tasks such as examining value distribution and outlier detection are aligned with data quality dimensions, which outline essential features for user verification [18]. This involves examining the results of data profiling techniques to gather statistical data on error counts across various dimensions as illustrated in Figure 6. Evaluation outcomes can then inform strategic discussions aimed at improving data quality [19].

Then, a data quality assessment framework is utilized to evaluate data integrity by setting standards for dimensions like accuracy, completeness, consistency, and timeliness. It includes procedures for data profiling, validation, and cleansing, along with tools and methodologies for analyzing data quality metrics. The goal is to ensure that data meets its intended purpose specifications, providing reliable and actionable insights [20]. Consequently, the development of a data quality assessment framework for economic indicators' data must be adaptable to users' objectives and context. In this case study, the focus is on rapid economic indicators utilized by experts in high-frequency data dissemination for decision-making. Hence, it's crucial to define data quality dimensions that meet the requirements of rapid data usage without compromising data format, range, and accuracy aligned with intended purposes.

Type of profiling	Group of tasks	Task	Area chart	Bar chart	Box plot	Bubble chart	Chord chart	Dashboard	Funnel plot	Geographical map	Heat map	Histogram	Line chart	Matrix plot	Network diagram	Parallel coordinates	Pie chart	Sankey diagram	Scatter plot	Sparklines	Tree map	Violin plot	Volcano plot	TOTAL
Characterize	Cardinalities	Number of distinct values										2					1							3
		Number of rows		2	1			2		3	1	1	2						1					13
		Value lengths										1												1
	Distribution	Frequency measures		6	5			2		4	3	13	3		1		1		1			1		40
		Mean, median, etc.			3			1		1	1													6
		Outliers		2	5			2	1	3		3												16
		Ranges			4						1													5
		Variance, skewness, etc.		1					1			1					1		1					5
	Patterns	Clusters								5	2				1	1			3					12
		Correlation									3		1		1	1			5					11
		Curve fitting											1						1					2
		Primary features								1									1					2
		Trends		1						1			2						1					5
		Value patterns								1														1
Data quality	Completeness	Coverage		1						5		1			1									8
		Missing records		1				1		1														3
		Missing values		2						1	1	1												5
	Correctness	Accuracy													1									1
		Bias										1												1
		Consistency		2									1		1		1		1					6
		Integrity													1									1
		Noise								1														1
		Outlier										1	1											2
		Plausibility								2									2					4
		Use of default values										1												1
		Validity								1									2					3
TOTAL			0	18	18	0	0	8	2	30	12	26	11	0	7	2	4	0	19	0	0	1	0	

Fig. 6. Data quality score as the result of data quality assessment [19].

2.2 Review of economic indicator and emerging of Big Data

Economic indicators are statistical metrics used to assess an economy's performance and dynamics in understanding economic conditions, predicting trends, and devising strategies [21]. They include leading indicators, such as stock market indices, which predict future trends; lagging indicators, like unemployment rates, which confirm past trends; and coincident indicators, such as GDP growth rates, which reflect real-time economic conditions.

Creating an economic indicator involves several stages to ensure its accuracy, reliability, and relevance. This includes clarifying the indicator's aim, defining components, selecting sources, gathering data, processing data, weighting and aggregating components based on significance, normalizing data, documenting methodology. disseminating the indicator to stakeholders, and monitoring and updating the indicator over time [22]. Following these steps ensures the development of a robust economic indicator that aids decision-making.

In contemporary times, there's a surge in economic indicators due to the emergence of Big Data, which serves as an economic state indicator capable of facilitating swift data analysis and detailed exploration for diverse decision-making processes. Traditional data refers to structured information stored in fixed formats within centralized databases, while Big Data deals with large or complex datasets, encompassing various types of data and characterized by the 5Vs: Volume, Velocity, Variety, Veracity, and Value. Big Data aims to extract valuable insights from massive and complex datasets for informed decision-making [23]. The key differences between traditional data and Big Data lie in their quantity, diversity, speed, complexity, and potential significance. Traditional data is typically small-scale, organized, and static, while Big Data is vast, complex, and constantly changing. Consequently, specialized methodologies and tools are required to handle and analyze Big Data efficiently. These distinctions are summarized in Table 2 for easy comparison.

Table 2. Comparison of data quality issues and its data quality dimension [23].

Traditional Data	Big Data
Organization level	Worldwide
Volume ranges from Gigabytes to Terabytes	Volume ranges from Petabytes to Zettabytes or Exabytes
Structured data	Structured, semi-structured, database, and unstructured data
Hour or per day or more	Seconds
Centralized source	Distributed source
Ease in data integration	Difficulties on data integration
Normal functions	Special functions
Static strict schema based	Dynamic flat schema based
Inter relationship	Unknown relationship
ERP and financial data, organizational data, web transaction data etc.	Its data sources include social media, device data, sensor data, video, images, audio etc.

The latest trend in utilizing Big Data for economic indicators involves integrating traditional data with newer sources like social media and satellite images to provide a more comprehensive view of current economic trends. Advancements in artificial intelligence and machine learning are enhancing the accuracy and predictive capability of these analyses, although concerns persist regarding data quality. Overall, leveraging Big Data aids policymakers in understanding and responding to economic shifts, supported by evidence from open available data, particularly from the private sector [24]. This trend demonstrates a consistent rise, as illustrated in Figure 7.

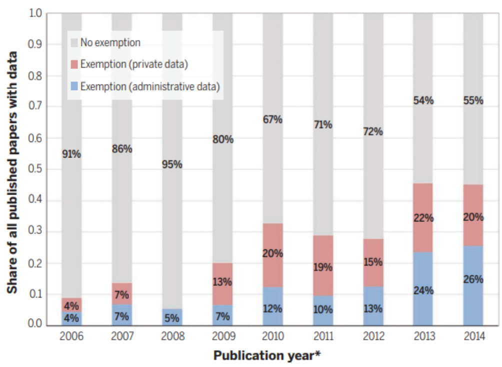

Fig. 7. The increase in applying Big Data in economic studies [24].

The increased demand for Big Data is driven by the COVID-19 pandemic, which has disrupted economic activities and rendered traditional economic indicators incomplete due to disease control measures. Big Data serves as a supplementary resource, enabling continuous economic analysis during the pandemic. For example, the Economic Tracker by the Opportunity Insights team provides real-time tracking of economic activity in the United States using anonymous data from private companies. It offers weekly statistics on consumer spending, business revenues, job postings, and employment rates, analyzed to understand the pandemic's economic impact across different groups [3] as illustrated in Figure 8.

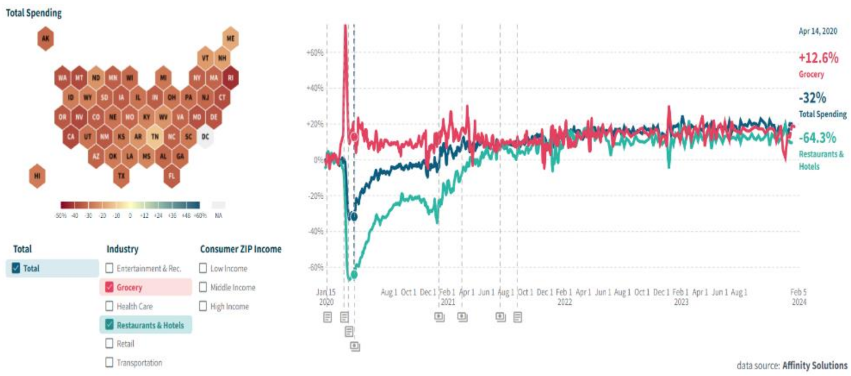

Fig. 8. US total consumer spending by credit card payments [3].

Another one, high-frequency indicators were introduced to monitor economic developments promptly. In Portugal, a new daily economic indicator (DEI) was proposed, combining electric consumption, natural gas, and card-based payments. This DEI, derived from daily datasets within a factor model framework, revealed a significant decline in economic activity coinciding with lockdown measures [2]. Unlike alternative methodologies, this approach detected abrupt fluctuations in economic activity on a daily basis without additional smoothing, providing timely insights into economic trends as depicted in Figure 9.

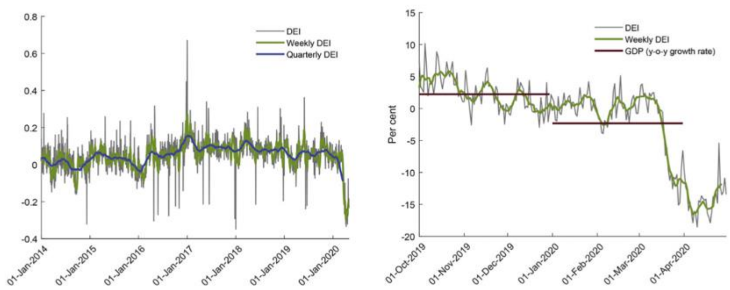

Fig. 9. Daily economic indicators (DEI) [2].

3 Data and Methodology

3.1 Data quality assessment framework

In this study, the data quality assessment framework is implemented on economic indicators. The key steps include firstly, selecting the target economic indicator that needs to be frequently reported and reflect core economic activities. Subsequently, data is collected from accessible sources, and it is essential to verify the data descriptions for clear understanding. Then, the conditions or rules used to assess data quality are considered, and the data quality dimensions are defined comprehensively. Next, the data quality is measured, and the results interpreted. Finally, data quality issues are summarized to be utilized in planning data quality improvements.

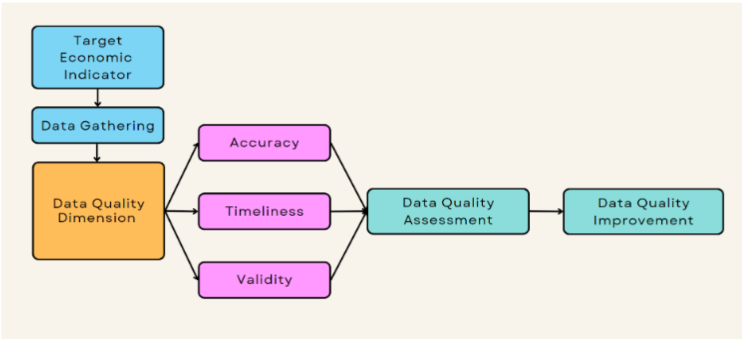

Fig. 10. Data quality assessment framework for rapid economic indicators [7].

3.2 Data collection method

In practical research, utilizes secondary data from the CEIC Database, including high-frequency data of the Apple Mobility Index in Thailand for tracking economic activity based on driving, the Global Normalcy Index in Thailand for tracking economic activity based on time spent outside, and the Google search trends in Thailand for tracking economic activity based on job seeking via LinkedIn. These data are represented for rapid economic indicators and there are procedure for gathering the data as following:

1) Accessing to CEIC Database
2) Filtering the location in Thailand and frequency as daily
3) Searching for target economic indicators including, Apple Mobility Index, Global Normalcy Index, and Google search trend and extract to local storage which define specific range of time duration between January 2020 and September 2022
4) Studying the metadata each dataset and preparing the importable file for loading the datasets to a data quality assessment tool

3.3 Data quality metrics

In this study, establish data quality dimensions in three aspects, including accuracy, timeliness, and validity, to comprehensively cover the characteristics of rapid economic indicators aiming to reflect economic activities, readiness for use, and accuracy of data formats. Calculating various data quality metrics includes the following:

1) **Accuracy (AC)** pertains to the correctness and precision of data in relation to the real-world scenario it represents.

$$AC = \frac{CR}{TOL} \times 100\%$$

Where:

- CR represents the number of accurate or correct data points.
- TOL represents the total number of data points in the dataset.

2) **Timeliness (T)** measures how current or up to date the data is in relation to the time it is needed for decision-making or analysis.

$$T = \left(1 - \frac{TD}{TE}\right) \times 100\%$$

Where:

- TD represents the difference between the time the data is needed and the time it is actually available.
- TE represents the ideal time at which the data should be available.

3) **Validity (V)** refers to the extent to which data conforms to predefined rules, standards, or constraints.

$$V = \frac{AF}{TOL} \times 100\%$$

Where:

- AF denotes the number of accurate data points in terms of data type, range, and formatting.
- TOL represents the total number of data points in the dataset.

3.4 Research instrument

Talend Open Studio for Data Quality

3.5 Data quality evaluation

In this study, several statistics are employed to interpret the results of data quality assessment [14], with the main formats including the following:

1) **Simple ratio**, this evaluates the ratio of achieved outcomes compared to all potential out-comes. Typically, this ratio ranges from 0 to 1, with 1 indicating the most favora-ble outcome. Completeness and consistency can be measured through this ratio.

2) **Min or Max**, this format is created to manage several data quality factors. The minimum value tends to be more cautious, while the maximum value is more lenient. As an example, the minimum value might represent the suitable data level, while the maximum value might indicate timeliness and accessibility.

3) **Weighted average**, This serves as an option instead of using the minimum value and can be ap-plied when organizations understand the significance of each variable in the equation.

4 Result Interpretation

4.1 Accuracy quality assessment

Accuracy verification involves examining precision in real-world scenarios. Data validation, particularly using Benford's Law, is applied in fraud detection. This statistical principle analyzes the frequency of different numbers as the first digit in numerical datasets, where smaller numbers occur more frequently and larger numbers less so [25]. Deviations from this pattern may suggest potential data inaccuracies.

1) **Apple Mobility Index**, reveals that the occurrence of leading digits does not align with the expected distribution curve. Therefore, this dataset has quality issues in terms of accuracy, as shown in Figure 11.

2) **Global Normalcy Index**, reveals that the leading digits do not match the expected distribution, suggesting potential accuracy issues. However, further examination is needed due to the dataset's narrow value range (90-100). The other tools may be required to confirm accuracy-related quality issues, as shown in Figure 12.

3) **Google search trend**, indicates a mismatch in the leading digits' distribution, suggesting potential accuracy issues. However, replacing null values with filled values might align the dataset perfectly with the expected distribution curve, as shown in Figure 13.

Column: metadata.Driving

Benford Law

Leading Digit	Count	%
1	208	20.97%
2	0	0.00%
3	52	5.24%
4	82	8.27%
5	77	7.76%
6	69	6.96%
7	125	12.60%
8	110	11.09%
9	94	9.48%
invalid	175	17.64%

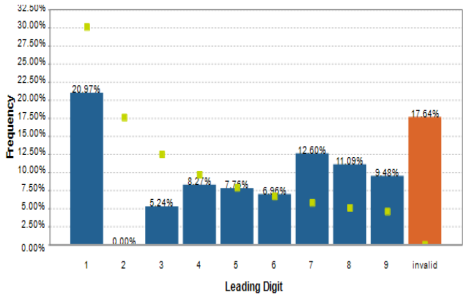

Fig. 11. Accuracy quality result of Apple Mobility Index

Column: Global_Normalcy_Index_TH.Time_Outside

Benford Law

Leading Digit	Count	%
1	111	11.19%
2	0	0.00%
3	0	0.00%
4	0	0.00%
5	0	0.00%
6	0	0.00%
7	0	0.00%
8	181	18.25%
9	621	62.60%
invalid	79	7.96%

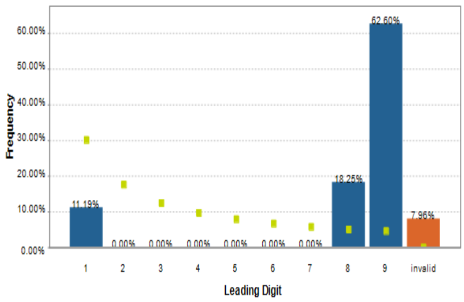

Fig. 12. Accuracy quality result of Global Normalcy Index

Column: metadata.LinkedIn

Benford Law

Leading Digit	Count	%
1	90	9.07%
2	5	0.50%
3	15	1.51%
4	16	1.61%
5	33	3.33%
6	33	3.33%
7	34	3.43%
8	28	2.82%
9	49	4.94%
invalid	688	69.35%

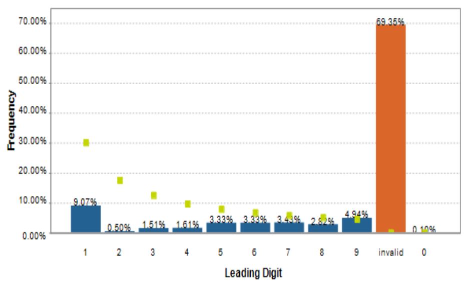

Fig. 13. Accuracy quality result Google search trend

In summary, the accuracy dimension was evaluated to ascertain the distortion of the dataset according to Benford's law, revealing any data inconsistencies. Nevertheless, it is imperative to also consider the dataset's inherent characteristics. From the assessment of all three types of economic indicators, none met the criteria of this test.

4.2 Timeliness quality assessment

Timeliness is assessed to show data availability when needed. The evaluation value is calculated by subtracting the percentage of null values from the total, providing the data quality assessment value for Timeliness using the simple ratio method.

1) **Apple Mobility Index**, there are relatively few null values, indicating readiness for analysis within the specified validation period. The timeliness score is 82%, as shown in Figure 14.

2) **Global Normalcy Index**, there are few null values, indicating the dataset is mostly available for analysis within the specified validation period. The timeliness score is 92%, as shown in Figure 15.

3) **Google search trend**, two out of three values are null, indicating it cannot be used for analysis within the specified validation period. The timeliness score is 31%, as shown in Figure 16.

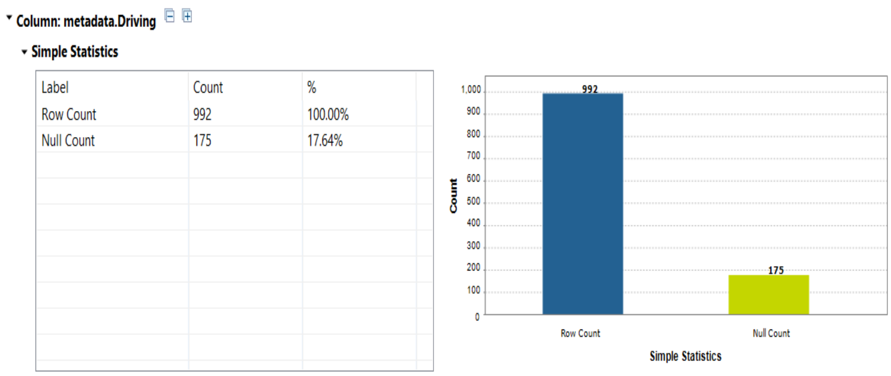

Fig. 14. Timeliness quality score of Apple Mobility Index

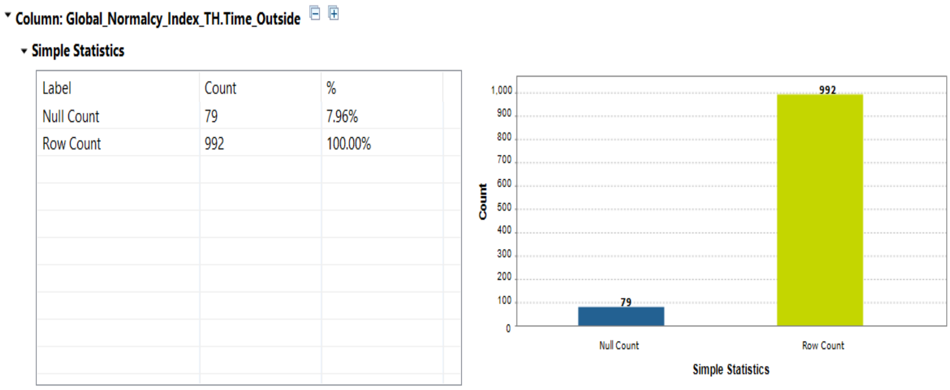

Fig. 15. Timeliness quality score of Global Normalcy Index

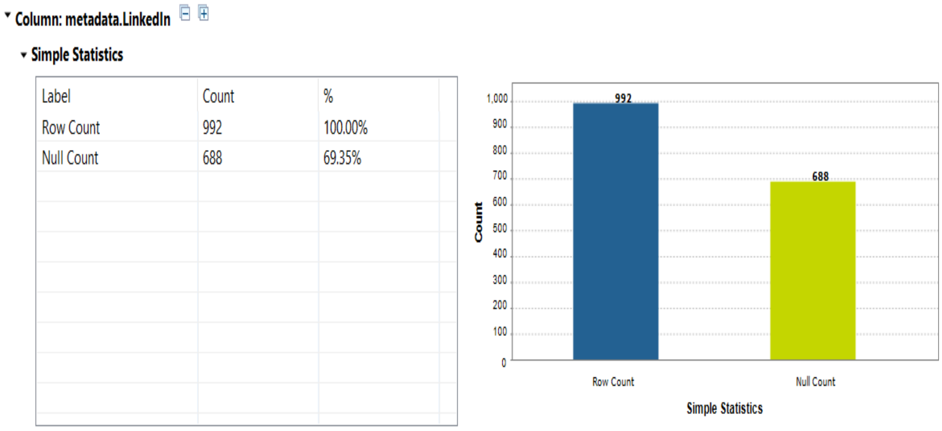

Fig. 16. Timeliness quality score of Google search trend

In summary, the timeliness dimension tested with all three rapid economic indicators, it is found that there are both satisfactory and unsatisfactory results. This is due to missing data, which may cause discrepancies in the data generation timeframe compared to the designated testing period. Nevertheless, the ability of rapid economic indicators to meet urgent data needs remains a favorable option given the circumstances.

4.3 Validity quality assessment

Validity is assessed to verify data conformity to predefined rules, standards, or constraints. The assessment value examines data spread and identifies outliers through anomaly detection, suggesting the dataset might include inconsistent values. This could render the economic indicator unreliable for analysis.

1) **Apple Mobility Index**, the significant differences between the minimum and maximum values, with the maximum exceeding the upper quartile, suggest the presence of outliers. This indicates potential data quality issues regarding validity, as shown in Figure 17.

2) **Global Normalcy Index**, the similarity between minimum and maximum values, with a small range, indicates the dataset likely does not contain outliers. This suggests it successfully passes the validity assessment, as shown in Figure 18.

3) **Google search trend**, a significant discrepancy between the minimum and maximum values, with the maximum exceeding the upper quartile, indicates potential outliers. This suggests possible validity issues within the dataset, as depicted in Figure 19.

Column: metadata.Driving

Summary Statistics

Label	Value
Mean	82.32255813953488
Median	80.24
Inter Quartile Range	41.370000000000005
Lower Quartile	58.8
Upper Quartile	100.17
Range	164.14999999999998
Minimum	35.08
Maximum	199.23

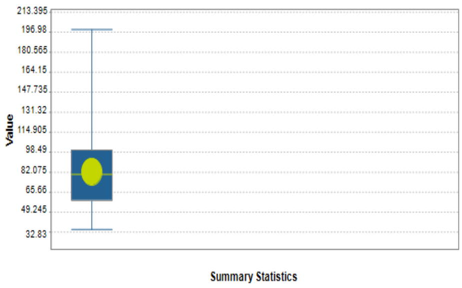

Fig. 17. Validity quality result of Apple Mobility Index

Column: Global_Normalcy_Index_TH.Time_Outside

Summary Statistics

Label	Value
Mean	94.22695968236583
Median	94.5
Inter Quartile Range	7.0
Lower Quartile	90.92857
Upper Quartile	97.92857
Range	21.571430000000007
Minimum	81.28571
Maximum	102.85714

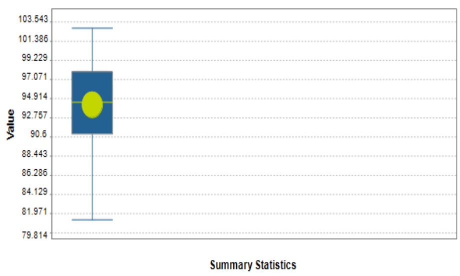

Fig. 18. Validity quality result of Global Normalcy Index

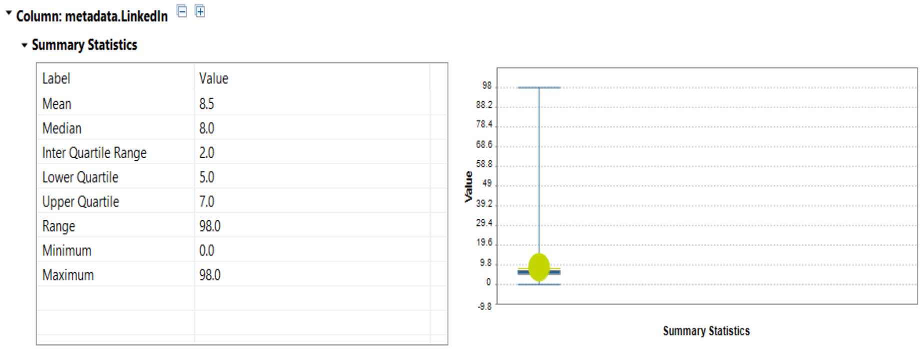

Fig. 19. Validity quality result of Google search trend

In summary, the validity dimension tests whether the data values fall within an appropriate range. It is found that only the Global Normalcy Index meets having no outlier condition. Therefore, utilizing this economic indicator for analysis is likely to be more reliable than others.

5 Conclusion

5.1 Discussion

The evaluation outcomes of rapid economic indicators like the Apple Mobility Index, Global Normalcy Index, and Google search trends highlight accuracy and validity as critical aspects requiring refinement. Addressing outliers can enhance accuracy and align datasets with economic trends, while implementing data quality rules or assessment methodologies can improve accuracy assessment. Comparisons with benchmarking standards of conventional economic indicators are necessary for precision verification. Regarding timeliness, assessing dataset readiness for use is essential, and monitoring time delays in data publication can ensure timely availability.

5.2 Limitation

This study encounters three main limitations. Firstly, accessing rapid economic indicators for the Thai case study is challenging due to limited resources and data disclosure by Thai organizations, resulting in reliance on global enterprise data tailored to measure Thailand's economic situation. Secondly, understanding the rapid economic indicator data nature requires expertise in economic indicator development, suggesting practical application over theoretical study may enhance the quality assessment framework. Lastly, limitations in the tools used for data quality assessment, particularly with free versions, and potential gaps in the assessment framework covering all data quality dimensions highlight the need for continuous development as the use of Big Data for economic indicators evolves.

5.3 Future works

The future research should focus on broadening the dimensions of data quality assessment, particularly through cross-data source checking and demonstrating the substitutability of rapid economic indicators for traditional ones. Involving domain experts, planners, and policymakers is crucial for establishing a robust data quality assessment framework, ensuring effectiveness in maintaining data integrity. Utilizing interview structures or satisfaction surveys can gather evaluations on data quality measurement, with evaluators selected based on their alignment with economic context.

References

1. International Monetary Fund, "Thailand: 2021 Article IV Consultation-Press Release; Staff Report; and Statement by the Executive Director for Thailand," IMF Staff Country Reports, 2021.
2. N. Lourenço and A. Rua, "The Daily Economic Indicator: tracking economic activity daily during the lockdown," Economic Modelling, vol. C, p. 105500, 2021.
3. R. Chetty, J. N. Friedman, N. Hendren and M. Stepner, "The Economic Impacts of COVID-19: Evidence from a New Public Database Built Using Private Sector Data," 2020.
4. F. Menezes, V. Figer, F. Jardim, and P. Medeiros, “A near real-time economic activity tracker for the Brazilian economy during the COVID-19 pandemic,” Economic Modelling, vol. CXII, p. 105851, 2022.
5. L. Cai and Y. Zhu, "The Challenges of Data Quality and Data Quality Assessment in the Big Data Era," Data Science Journal, vol. XIV, p. 2, 2015.
6. M. Sakpal, "How to Improve Your Data Quality," Gartner, Inc., 14 July 2021.
7. P. Kiatkajitmun, C. Chanton, P. Piboonrungroj and J. Natwichai, "Data Quality Assessment Framework and Economic Indicators," in NBiS 2023: Advances in Networked-based Information Systems, Chiang Mai, Thailand, 2023.
8. M. J. Hassenstein and P. Vanella, "Data Quality—Concepts and Problems," Encyclopedia, vol. II, no. 1, pp. 498-510, 2022.
9. J. Patel, "Bridging data silos using big data integration," International Journal of Database Management Systems, vol. XI, no. 3, pp. 01-06, 2019.
10. Y. Jaisingh, "Data Quality," A Medium Corporation, 30 January 2023.
11. S. Vancauwenbergh, "Data Quality Management," in IntechOpen eBooks, 2019.
12. Oracle Corporation, "Oracle® Warehouse Builder User's Guide 10g Release 2 (10.2.0.2)," 2009.
13. DAMA International, DAMA-DMBOK: Data Management Body of Knowledge, 2017.
14. W. Fan and F. Geerts, Foundations of Data quality Management, Springer Nature, 2022.
15. S. Bhandari, N. Ranjan, Y.-C. Kim, J.-D. Park, K. Hwang, W.-H. Kim, Y.-S. Hong and H. Kim, "An Automatic Data Completeness check framework for open government data," Applied sciences, vol. XI, no. 19, p. 9270, 2021.
16. N. Laranjeiro, S. N. Soydemir and J. Bernardino, "A Survey on Data Quality: Classifying Poor Data," in 2015 IEEE 21st Pacific Rim International Symposium on Dependable Computing (PRDC), Zhangjiajie, China, 2015.
17. I. Taleb, H. T. E. Kassabi, M. A. Serhani, R. Dssouli and C. Bouhaddioui, "Big Data Quality: A Quality Dimensions Evaluation," in 2016 Intl IEEE Conferences on Ubiquitous Intelligence & Computing, Advanced and Trusted Computing, Scalable Computing, 2016.

18. Z. Abedjan, L. Golab and F. Naumann, "Profiling relational data: a survey," The VLDB Journal, vol. XXIV, no. 4, pp. 557-581, 2015.
19. R. A. Ruddle, J. Cheshire and S. J. Fernstad, "Tasks and visualizations used for data profiling: a survey and interview study," IEEE Transactions on Visualization and Computer Graphics, pp. 1-12, 2023.
20. C. Cichy and S. Rass, "An overview of data quality frameworks," IEEE Access, vol. VII, pp. 24634-24648, 2019.
21. N. Frumkin, Guide to economic indicators, Routledge, 2015.
22. M. Mazziotta and A. Pareto, "Methods for Constructing Composite Indices: one for all or all for one?," Rivista Italiana di Economia Demografia e Statistica, vol. LXVII, no. 2, pp. 67-80, 2013.
23. GeeksforGeeks, "Difference between Traditional data and Big data," GeeksforGeeks, 2024.
24. L. Einav and J. Levin, "Economics in the age of big data," Science, vol. CCCXLVI, no. 6210, 2014.
25. J. Murtagh, "What Is Benford’s Law? Why This Unexpected Pattern of Numbers Is Everywhere," Scientific American, 8 May 2023.
26. Our World in Data (OWID), "CEIC: (DC)OWID: COVID-2019: Positive Rate: 7-day Smoothed: Thailand," CEIC, 23 June 2022.
27. CEIC, "CEIC: Real GDP: YoY: Thailand: Quarterly, ending "Mar, June, Sep, Dec"," CEIC, 2023.