



Data Collection Improvement: Daily Self-Enumeration Accommodation Survey

I A Setyadi

Directorate of Statistical Information System, BPS - Statistics Indonesia - Jakarta,
Indonesia

*Corresponding author's e-mail: adityasetyadi@bps.go.id

Abstract. Until now, BPS - Statistics Indonesia has conducted monthly accommodation surveys both for the star and non-star accommodation categories to provide information on commercial accommodation activities at the national and regional levels. Both star and non-star accommodation categories are done by complete enumeration in each region. Statistics include guest night and room capacity to obtain the occupancy rate of a hotel room. The data contains daily accommodation information that is collected every month, so then it will be entered completely in each region following the observation month. Due to the timeliness requirements for monthly press releases, BPS has implemented online data entry since 2017. It may seem obvious, regions that have more interest will have an impact on a bigger number of accommodations, which also affects the number of enumerators and may lead to such problems especially in response burden. Unfortunately, the same problem is also not easily avoided by regions with less accommodation, mostly due to the distance issues to the accommodation area and its spread in the region. Therefore, a new data collection strategy is required to provide respondents with convenience in order to increase response rates, as well as to reduce the workload of enumerators which also leads to lower cost. The outbreak of COVID-19 has posed unprecedented problems for National Statistical Offices (NSOs) around the world, including BPS – Statistics Indonesia. This crisis has led us to think in new ways and make decisions that will change our statistical operations in order to meet ongoing data needs even throughout the epidemic. The purpose of this paper is to discuss the evolution of accommodation surveys, which are designed to not only solve problems but also achieve objectives. Currently, there are nearly 180 active users of this self-enumeration accommodation survey for about 142 distinct accommodations across Indonesia. Moreover, this addition has proven to have succeeded in increasing the response rate average from 57.17% in 2020 to 68,35% in 2021.

1. Introduction

BPS - Statistics Indonesia has conducted monthly accommodation surveys in order to provide information on commercial accommodation activities at the national and regional levels, and it is targeted to establishments that provide short-term accommodation services. The survey collects daily accommodation information such as guest night and room capacity mainly to determine the occupancy of a hotel room.

The survey is conducted typically on a census basis in all accommodation categories, among both star and non-star accommodations. There are numerous classifications for star accommodation categories, which are gradually based on their license star hotel, namely from one-star, two-star, and up to five-star categories. Non-star accommodation categories, on the other hand, are defined as



accommodations that do not have a license to operate as a star hotel and are also known as budget hotels. The accommodation classifications are based on the Regulation of the Ministry of Tourism and Creative Economy of the Republic of Indonesia No. PM. 53/HM.001/MPEK/2013 concerning Hotel Business Standards [1]. The classification also corresponds to the Regulation of the Chief of the Statistics Indonesia No. 95 of 2015 on Indonesian Standard Business Classification (Klasifikasi Baku Lapangan Usaha Indonesia - KBLI) [2]. Furthermore, the assessment of hotel business standards is carried out by the Business Certification Agency (Lembaga Sertifikasi Usaha - LSU) in the Tourism Sector.

The survey produces numerous accommodation statistics indicators published monthly for tourism official statistics, including average length of stay; guest per room; room occupancy rate; bed usage rate; night guests stay; and room nights used (see Table 1).

Table 1. Accommodation statistics indicators

Indicator	Description
<i>average length of stay</i>	The average length of stay of guests which is divided into foreign, domestic, and foreign and domestic guests.
<i>guest per room</i>	The average number of guests in one rented room which is divided into foreign, domestic, and foreign and domestic guests.
<i>room occupancy rate</i>	Comparison between the number of room nights used and the number of available room nights.
<i>bed usage rate</i>	Comparison between the number of beds used and the number of available beds.
<i>night guests stay</i>	The number of nights guests stay.
<i>room nights used</i>	The number of rooms used or rented every night.

Since 2017, data processing has shifted from regional processing with distributed desktop applications, to a centralized solution based on web technology. This was driven by the need to comply with the latest technology at that time, which was followed by the development of the ability to ensure data security during processing. Apart from that, there is a basic need to provide timely data to assist the process of publishing tourism official statistics in the monthly period, both at the national and regional levels.

During the data processing phase, it was noted that technology was shown to be able to respond as well as accelerating the process of providing data collection results up until to be specific indicators for official statistical purposes. Nonetheless, the true issue was discovered during the data collection phase, where it was extremely difficult to acquire respondents' confidence in providing information, including a commitment to be ready to be asked for the following months, and despite the fact that the difficulty in reaching accommodation locations was a major obstacle during the survey.

As is known, many accommodations exist and are usually placed to support tourism areas, although not a few are located to assist urban areas. Whereas, most of those accommodations in tourism areas are located far from residential areas and thus are difficult to reach, which could be a real challenge, particularly for monthly data collection. Although if accommodations are easily accessible, they may still pose issues, particularly due to the burden factor. On the other hand, the increasingly diverse number and types of surveys now put a burden on surveyors at the regional level which also has a negative impact on this survey activity. As a result, it is believed that there is a need for innovation and the creation of new data collection techniques that are designed to not only solve the challenges listed above but also to meet and even enhanced the needs of the main objective.

Adetia et al. examine the elements that influence the response rate of BPS e-surveys and make recommendations for improving it [3]. The results suggest that device preference and trust are major factors that positively influence the intention of the company to participate in the BPS e-survey. The device preferences factor identifies the respondents' preferences for accessing the survey through various methods and platforms, while trust factor highlighted BPS's confidentiality as a legitimate and credible entity for statistical data collection, as well as its ability to ensure the confidentiality of



respondents' identities and responses. These findings [3] can be used as a starting point for a technological improvement solution approach, as well as for supporting the rejuvenation process and enriching the data collection flow, which is expected to meet the needs of non-response solutions that are still being faced.

2. Data collection during the COVID-19 pandemic

For years, statisticians around the world have discussed new data collection methods to replace or complement direct, face-to-face measurements, as well as to increase the efficiency and timeliness of statistical processes and products. The outbreak of COVID-19 has made unpredictable difficulties for National Statistical Offices (NSOs) all across the world, including BPS. As the pandemic broke out at the beginning of 2020, we needed to suspend face-to-face interviews, and even the employees were asked to work from home. Simultaneously, policymakers desperately need information on the developing pandemic and its effect, not exemption on the tourism statistics. This crisis has taught us to think differently and make decisions that will alter our statistical operations in order to satisfy continuous data needs even during the pandemic.

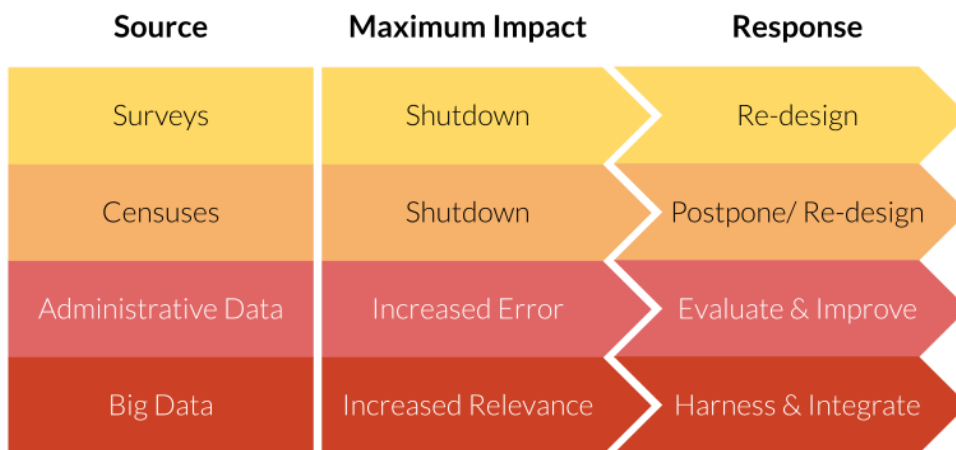


Figure 1. Response recommendations on the pandemic maximum impact on statistical data collection by UNESCAP

In particular, the crisis scenario necessitates adaptive techniques that have the least possible impact on quality. Due to the situation, the Statistics Division of the United Nations Economic and Social Commission for Asia and the Pacific (UNESCAP) has proposed ideas for adopting rapid changes in survey design and taking necessary efforts to mitigate possible quality implications [4]. The goal is to adopt an approach that may be inferior to the original survey but fills the data gap with reasonable accuracy (see Figure 1).

However, as the phrase goes, “the show must go on” — BPS - Statistics Indonesia is expected to continue to issue results in the form of a press release on every first working day of the month. As known, BPS - Statistics Indonesia has made a commitment to the public to have continual statistics production, as the government requires timely data to make judgments about how to mitigate COVID-19's impact in numerous sectors.

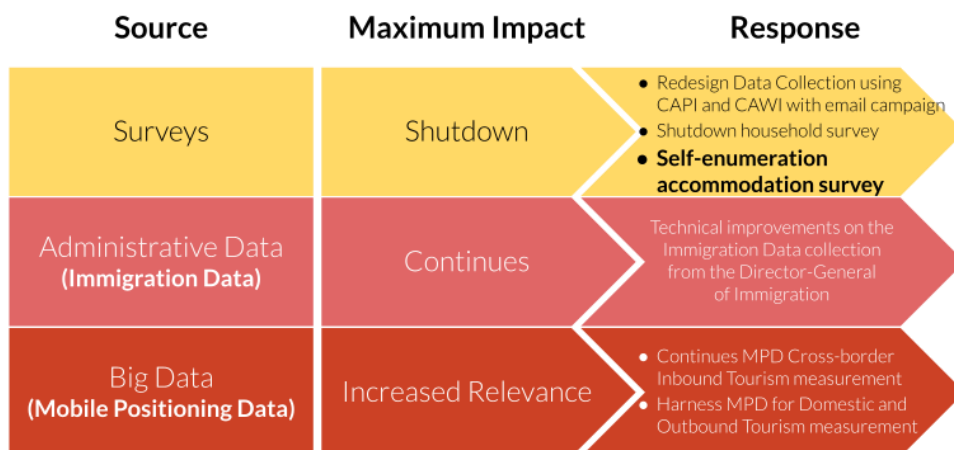


Figure 2. Pandemic maximum impact responses on tourism statistics data collection in BPS - Statistics Indonesia

Specifically, regarding tourism statistics, data collection during the pandemic was altered in numerous ways, including the delay in the implementation of several surveys, but the utilization of other data sources including administrative data and big data were extensively exploited (see Figure 2). Concerning to accommodation survey, as mentioned in Figure 2, the response to the pandemic is the creation of a self-enumeration data collection platform for accommodation, which serves as a complement to the existing data processing system. The platform is intended to be available as an alternative for respondents to choose an independent response through the platform or to be prepared to receive surveyors' arrivals every month in order to be asked for information related to accommodation.

3. Self-enumeration platform

Since 2020, Indonesia, in this case, BPS - Statistics Indonesia, has officially begun integrating data collection modes with self-enumeration methodologies. Although still in the early stages of production, BPS - Statistics Indonesia has been working on this mode for several years before it was officially released to the public.

When compared to face-to-face interviews, the self-enumeration method offers economic advantages because it can save on expenses. This strategy also can eliminate bias caused by subjectivity or interviewer questioning in addition to saving expenses. The time for filling out the questionnaire becomes more flexible from the respondent's standpoint since they can pick the appropriate time for completing out the survey. Furthermore, respondent privacy has been protected, making self-enumeration appropriate for sensitive instances. The difficulty in describing the concepts and definitions of each item in the questionnaire is, however, one of the self-enumeration method's shortcomings. Respondents will understand questions more quickly if they are provided with proper pronunciation and intonation rather than simply reading them. The low level of understanding of respondents can cause respondents to be wrong in answering questions, the implication is that the resulting data becomes less accurate [5].

In the Industrial Revolution 4.0 era, the use of the internet is growing stronger every year. The use of the internet as a source of information and a market for transactions continues to increase, with significance for the wider economy. Nowadays, just about every accommodation has a website, and at once as the most important lead generator, booking channels, marketing tools, and revenue manager. Following the success of the existing data processing systems based on web technology, as well as being regarded as familiar for accommodation actors, the development of a self-enumeration platform was also built on web technology.

This platform is not necessarily built to take the place of an existing data processing system. The two are clearly differentiated in terms of security, as the self-enumeration platform is open to the public, whilst the existing data processing system is created according to the institution's security standards and



can only be accessed through a limited Virtual Private Network (VPN). Despite being open to the public, the self-enumeration platform has been constructed to the security standards which are permitted to be published for production and has also met the due diligence conducted by the division responsible for this security issue. As per the user actors, the existing data processing system is designed for internal employees and hired trustworthy partners, whereas the self-enumeration platform will be available only to accommodation users.

3.1. User registration

Data integrity is critical for any successful platform in web technology, and a user registration/login system is the gateway to most web applications. User registration systems are screens, forms, or profile pages that request information from a user to create a web-based account or profile. A user registration system generally asks a user to create a username and password, and possibly answer other security questions as well.

Since this accommodation survey is conducted on a monthly basis, this self-enumeration platform is designed to be used not only for one-time enumeration but also for filling up the accommodation information on a continuous basis. As a result, the creation of this self-enumeration platform also aids in the creation of user management for any accommodation that accesses the platform. In order to match the existing accommodation directory, and ensure that the registered user is really the accommodation concerned, the registration system in the self-enumeration platform must go through a series of integration processes with existing data processing systems.

The user registration scheme must include a unique token generated by internal users in the existing data processing system after obtaining approval or interest from the accommodation party to want to join the self-enumeration platform. The user will start the registration process by submitting the given token, which will be followed by providing their email credentials like a username and password, as appropriate. This registration process also serves as a sign of approval for respondents, indicating that they do not need to be contacted again each month to provide information about their accommodation (see Figure 3).

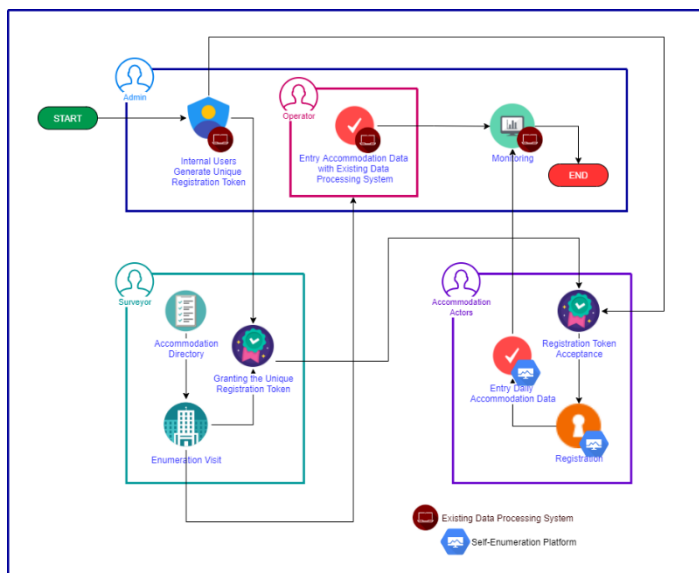


Figure 3. Self-enumeration accommodation survey platform user registration scheme



Figure 4. Official cover letter of granting registration token

The registration token will be distributed through an official cover letter presented by surveyors during the visit, as shown in the schematic in Figure 3. It is suggested that this cover letter be physically handed over by the surveyor in order to also provide an explanation of the processing flow as well as the concept and definition of the questionnaire's content.



The creation of this official cover letter is automatically generated along with the registration token generation process. A registration token will be included in the official cover letter, which may then be used to register for the self-enumeration platform. One registration token can only be used for one registration account, but when an accommodation requires more than one account to handle, it only requires requesting a new token. The letter also contains the token's expiration date, which has been configured to be available no later than one month from the moment it was generated (see Figure 4).

3.2. Data entry

In general, the self-enumeration entry scheme slightly differs from the regular entry scheme. The data entry processing system is typically constructed to adapt the original questionnaire as much as feasible, and the question structure in the questionnaire is typically numerical, which is aimed to make it easier for both field surveyors and data entry operators. Choice questions, for example, will typically not be marked with a specific mark as an answer marker, but will instead require the surveyor to write down the numerical results of the choice for the question in the provided column, allowing the entry operator to easily enter the answer code through the system later. However, if this optional question is then provided to be filled in independently by the respondent, of course, the display with the radio button option will make the respondent much more comfortable in filling it out. Nonetheless, this conversion adjustment is not implemented in this data input process on the self-enumeration accommodation survey since the content of the questionnaire is mostly about quantity type.

The data entry system on this self-enumeration platform was designed to still adapt the original questionnaire, which is nothing much different than the existing data processing system, considering that actually, this questionnaire should have become something that the accommodation side usually encounters every month. Figure 5 shows a comparison of the questionnaire view with the data entry displays.

2. Kabupaten/Kota *	KABUPATEN SEKADAU - [09]	6. Alamat	JL. MERDEKA TIMUR KEC. SEKADAU HULU
3. Kecamatan	SEKADAU HILIR - [040]	7. Jenis Akomodasi	NON-BINTANG
4. Kelurahan/Desa *	MUNGGIK - [003]	8. Kelas Akomodasi	2

1. HARGA/ TARIF KAMAR PER HARI								
Tarif	1. Rp	2. USD	Hari Kerja	Hari Libur	Suite	Hari Kerja	Hari Libur	
Non Suite	Standard					Junior Suite		
	Superior					Suite		
	Deluxe					President Suite		

2. JUMLAH KAMAR, TEMPAT TIDUR, DAN TAMU												
Tanggal	Jumlah Kamar Tersedia	Kapasitas Tempat Tidur Tersedia	Banyaknya Kamar			Banyaknya Tamu Menginap						
			Digunakan Kamar	Check In	Check Out	Kamarin		Masuk Hari Ini		Keluar Hari Ini		
						Asing	Indonesia	Asing	Indonesia	Asing	Indonesia	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	
01 Juli 2021												
02 Juli 2021												
03 Juli 2021												

Questionnaire View

The screenshot shows a data entry form with fields for Kabupaten/Kota, Kecamatan, and Kelurahan/Desa. Below the form is a table titled 'JUMLAH KAMAR, TEMPAT TIDUR, DAN TAMU' with columns for Tanggal, Jumlah Kamar Tersedia, Kapasitas Tempat Tidur Tersedia, and columns for Digunakan Kamar, Check In, Check Out, and columns for Kamarin, Masuk Hari Ini, and Keluar Hari Ini.

Existing data processing system view

The screenshot shows a self-enumeration platform interface with a data entry form and a table view. The table is titled 'Banyaknya Tamu Menginap' and has columns for Tanggal, Jumlah Kamar Tersedia, Kapasitas Tempat Tidur Tersedia, and columns for Digunakan Kamar, Check In, Check Out, and columns for Kamarin, Masuk Hari Ini, and Keluar Hari Ini.

Self-enumeration platform view

Figure 5. Comparison of the questionnaire view with the data entry displays



Data validation is the activity to review and validate data against predefined rules to verify the value and to identify potential problems such as missing data, inconsistency, and inappropriate editing/imputation [6, 7]. Each document will receive two sets of data as a result of the validation process: an error or a clean status. Documents with an error status still have incorrect details, but those with a clean status have been verified as clean and complete. This validation procedure is, of course, also implemented on the self-enumeration platform, ensuring the integrity and quality of the data submitted by the accommodation actors. However, the validation process is being implemented differently between the existing data processing system and the self-enumeration platform: on the self-enumeration platform, validation occurs immediately after the user inputs, whereas in the existing data processing system, validation occurs after the document has been triggered to be saved. Also, when the user enters, it also generates live daily accommodation statistics indicators based on the input data, namely *occupancy rates* and *night guest stay* (see Figure 6). In addition, this self-enumeration data entry is designed to be filled in available on a daily basis, whereas the existing data processing system requires waiting for the current month to end before entering the data on a monthly basis.

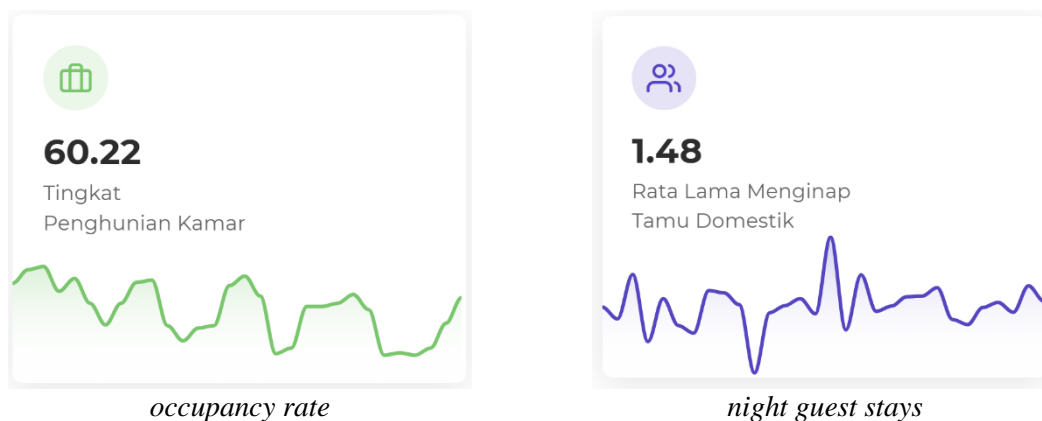


Figure 6. Live daily accommodation statistics indicators based on users input

3.3. Special features

Profit is the primary goal of any company, including accommodation. Profit is the lifeblood of any business; without it, no company can compete in today's market. It is common for accommodations to keep an eye on each other's competitors, especially if they are in the same neighborhood. Price is currently the most common technique for assessing competitors, albeit this can be problematic due to price discounting. Mathews recommended network analysis, and in particular affiliation matrices, as a systematic technique to build the foundation for developing a competitive strategy [8]. Meanwhile, Olmo added that it is important to study the market and including competitor's data in the research, as well as to identify which core elements and additional services that generate more value to the target market [9]. Hesford et al. additionally found that monitoring more contenders was related to better revenue, attributable to a solid, beneficial effect on prices. Accommodation actors can enhance pricing decisions by acquiring more information about market competitors, even allowing them to raise prices without causing a drop in occupancy [10]. As a response, rather than providing merely self-enumeration data entry, this platform is also designed to provide competitors monitoring, particularly competition in the nearby area location. This special feature is also present as a reward for accommodations who are willing to participate in surveys.

The monitoring feature is designed not to freely monitor accommodation specifically, because it is clear that this activity is kept confidential under the Law of The Republic of Indonesia no. 16 of 1997 concerning Statistics [11]. Nonetheless, this particular feature will allow the users to compare their occupancy rates with other nearby accommodation competitors in the same class category (see Figure 7). Meanwhile, if the nearby location is too specific with only a few accommodations available; the



comparison will automatically calculate the results from expanding to a wider level and area. This calculations algorithm has been designed to dynamically adjust to all circumstances, ensuring that no one is exposed to the comparison outcomes for a specific accommodation.

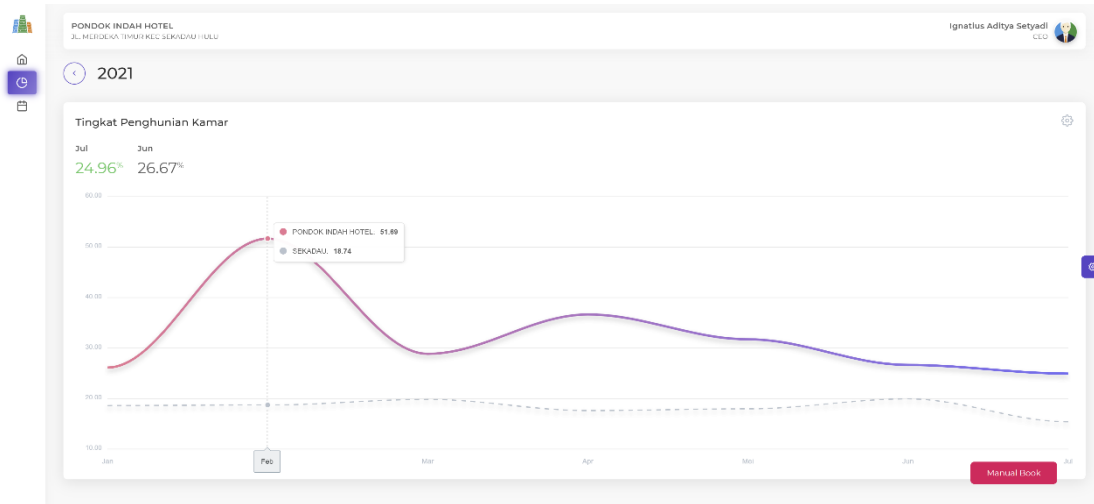


Figure 7. Monitoring feature to compare occupancy rates between nearby accommodations in the same class category

4. Discussion and findings

The self-enumeration platform was initially made available in October 2020, and the campaign was heavily promoted by the end of the year. Figure 8 shows that the number of tokens generated increased significantly from the end of 2020 to the first quarter of 2021, followed by the number of tokens used for registration, despite the fact that some tokens were reported as being deactivated to terminate the registration. It has an activation percentage of 10.81%, with a total of 2,036 tokens generated and about 220 utilized for the registration process. The majority of the deactivations were due to user cancellations that were deemed excessive, though it was later discovered that several cancellations were made because they were eventually willing to accept surveyors' visits after the COVID-19 pandemic ease.

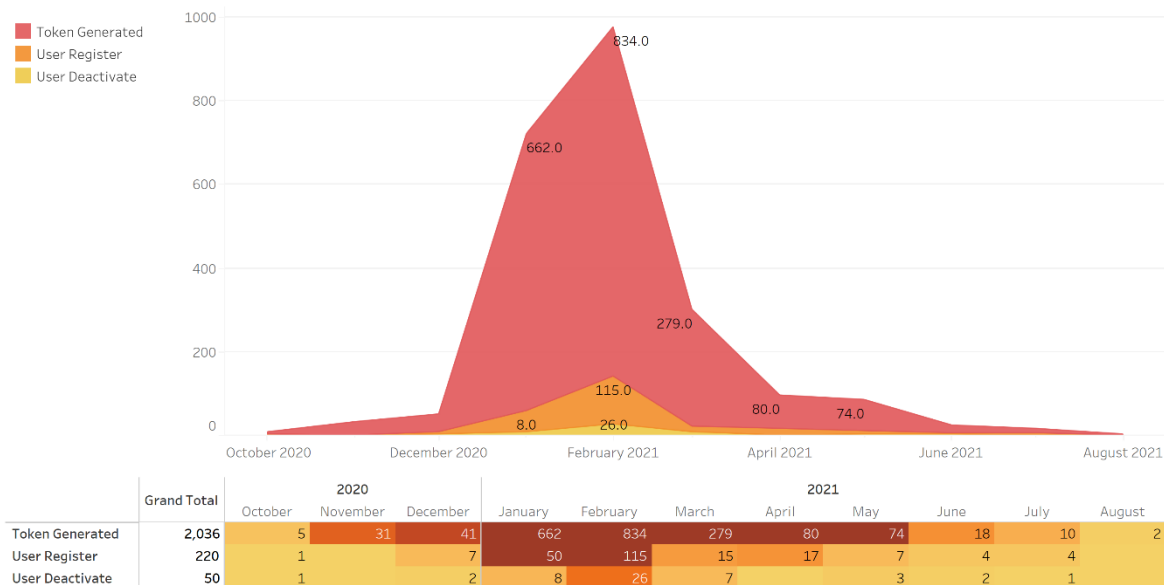


Figure 8. The number of tokens generated, user registration, and user deactivation by month



There are currently 170 registered users from 142 different accommodations. With a user ratio of 1 to 2 for each accommodation, 3-star class accommodations have the most users, followed by 4-star and 2-star, while 1-star, non-star, and 5-star classes are among the bottom three (see Table 2). Meanwhile, the number of registered users is in line with the number of available accommodations, and the ratio of registered users compared to the number of distinct accommodations is a 4-star class, which is 1.27.

Table 2. Number of user register and distinct accommodations by accommodation class category since October 2020

Class Category	KBLI	Number of User Register	Number of Distinct Accommodations
5-star	55111	5	4
4-star	55112	37	29
3-star	55113	65	52
2-star	55114	27	26
1-star	55115	14	12
non-star	55120	15	14
	55130	5	3
	55199	2	2
Grand Total		170	142

The map of the spread of accommodation sites registered on the platform shows that accommodation has practically nearly reached every area of Indonesia, with the distribution center in the northern area of Sumatra Island and the western parts of Java Island. It's interesting that some of the accommodation's users are on island territory and that numerous accommodations in Indonesia's eastern region are already using this platform (see Figure 9).

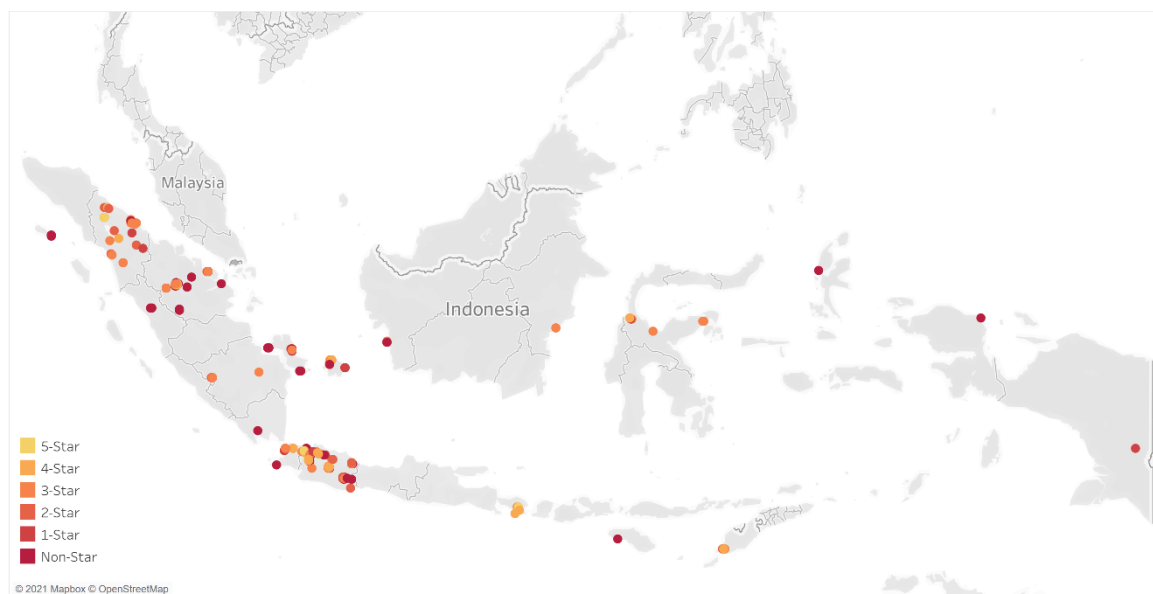


Figure 9. Distribution of accommodation locations based on the map of Indonesia by accommodation class category

The presence of the self-enumeration platform has also been found to boost response rates; as seen in Figure 10, the response rate has increased significantly since its release. As is well known, the COVID-19 pandemic first occurred in Indonesia in February 2020 and is directly proportional to the decline in the response rate since that month, which peaked in May 2020. However, after the first self-



enumeration platform was released in October 2020, the response rate effects gradually improved. The response rate climbed again in February-July 2021, nearly matching the percentage seen before the COVID-19 pandemic spread (January-February 2020), which was around 70%. Inter-month trends on response rate have also improved in the last two years, with an average of 57.17% in 2020 rising to roughly 68.35% in 2021.

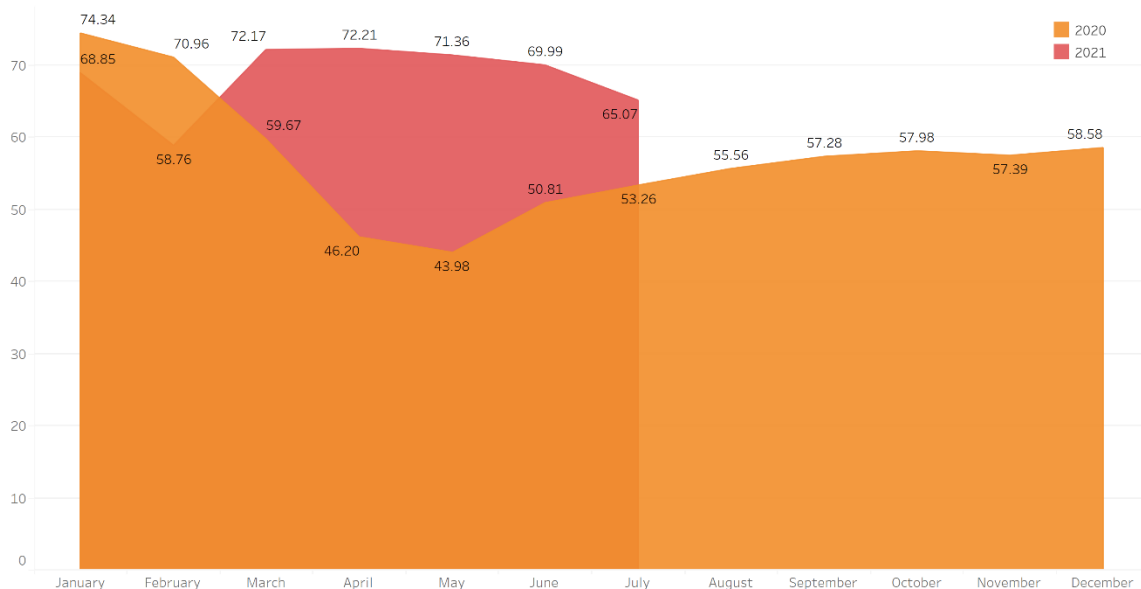


Figure 10. Response rate comparison by month

5. Conclusion

In general, having this self-enumeration platform as an option for the accommodation survey has been shown to substantially benefit the survey's implementation. The survey's major problem is its location, as well as respondents' confidence in it, which has resulted in a low response rate thus far. Nonetheless, the presence of this platform has been found to mitigate the impact of this problem by achieving a higher response rate than before it was publicly released. The distance issue is thought to have been somewhat handled, as seen by the spread of several accommodations registered on the platform are even located on island territories.

The special feature to monitor nearby competitors is expected to be the main attraction to the users in the platform, while hopefully might also attract other accommodation parties to participate independently in this survey. Even so, this element is still thought to be lacking in variety and in further development need to add another special feature to draw greater interest from other accommodation parties. It is also expected that there will be a redesign of the questionnaire that is not only oriented to the needs of the BPS - Statistics Indonesia in particular, but also a design that focuses on user convenience.

Furthermore, the support and appeal of the local government will be tremendously helpful in getting other accommodation actors to participate in the survey through this platform. Besides, since the data is obtained on a daily basis, it will be valuable to the government in a timely manner, both nationally and regionally.

Reference

- [1] Ministry of Tourism and Creative Economy of The Republic of Indonesia 2013 Regulation No. PM.53/HM.001/MPEK/2013 concerning Hotel Business Standards
- [2] BPS - Statistics Indonesia 2015 Regulation of the Chief No. 95 of 2015 on Indonesian Standard Business Classification



- [3] Adetia A, Budi I and Setiadi F 2020 Identification and Analysis of Factors Affecting E-survey Response Rate at Central Bureau of Statistics *In 2020 International Conference on Information Management and Technology (ICIMTech)* pp 560-565 IEEE doi:10.1109/ICIMTech50083.2020.9211236
- [4] Statistics Division of the United Nations Economic and Social Commission for Asia and the Pacific (UNESCAP) 2020 Surveys Under Lockdown; a pandemic lesson *Stats Brief* April 2020 — Issue No. 23
- [5] Bogue D J 1965 The Pros and Cons of “Self-Enumeration” *Demography* vol 2 pp 600–626 doi:10.2307/2060143
- [6] United Nations Statistics Division Johannesburg 2017 Data Validation and Analysis of Performance of Imputation <<https://unstats.un.org/unsd/demographic-social/meetings/2017/johannesburg--regional-ws-census-data/docs/s03-01-UNSD.ppt>> accessed 8 October 2021
- [7] Di Zio M, Fursova N, Gelsema T, Gießing S, Guarnera U, Petrauskienė J, Quensel-von Kalben L, Scanu M, ten Bosch KO, van der Loo M, and Walsdorfer K 2016 Methodology for data validation 1.0. *Essnet Validat Foundation, Brussels, Belgium* pp 1-76
- [8] Mathews V E 2000 Competition in the international hotel industry *International Journal of Contemporary Hospitality Management* vol 12 no 2 pp 114–118 doi:10.1108/09596110010309916
- [9] Olmos G 2012 Hospitality Competitiveness Measurement System *Journal of Global Business and Technology* 2(8) pp 29–37.
- [10] Hesford J W, Turner M J, Mangin N, Thomas C R and Hoffmann K 2019 Competitor Monitoring and Revenue Performance: Evidence from the Hospitality Industry *Advances in Management Accounting* vol 31 pp 1–22 doi:10.1108/S1474-787120190000031002
- [11] Law of The Republic of Indonesia No. 16 of 1997 on Statistics