
Analysis Of Product Price For Creative Economic Business Yogyakarta Special Region Using Data Scrapping

Astika Ayuningtyas^{1)*}, Yuliani Indrianingsih²⁾, Uyuunul Maudzhoh³⁾

¹⁾²⁾ Informatics Study Program, Faculty of Industrial Technology, Adisutjipto Institute of Aerospace Technology, Indonesia

³⁾ Industrial Engineering Study Program, Faculty of Industrial Technology, Adisutjipto Institute of Aerospace Technology, Indonesia

¹⁾astika@itda.ac.id, ²⁾yulistta@gmail.com, ³⁾farahzizy@gmail.com

ABSTRACT

The government in the National Economic Recovery (PEN) launched a strategy to increase business activity, namely by maximizing the potential for product sales through E-Commerce. E-Commerce sites in Indonesia display product information with different descriptions and prices. Through the pages on E-Commerce can collect data to form a useful information. Collection of data variations requires a method or system that can automatically collect the data provided on the Web page. In addition to data collection, another thing that needs to be considered is the meaning of knowledge from the collection of information that has been obtained. For these data to be useful, it is necessary to design a place for data management. Data must be selected and arranged properly so that it is easy to use so that it gives a good meaning. This research requires methods and tools in the process. The method used is Web Scrapping while the tools used are Tableau. There are two objectives in this research. First, obtain product price data for creative economy entrepreneurs in the Special Region of Yogyakarta through E-Commerce, especially Tokopedia and Bukalapak. Second, present product price data visually to facilitate the process of price distribution analysis so that it can provide information to local governments, especially product managers from creative economy business actors to support strategic policies in determining product prices from creative economy business actors in the Special Region of Yogyakarta to support promotion and tourism.

Keywords: Data Visualization; Promotion; Tourism; Web Scrapping

1. INTRODUCTION

The Covid-19 pandemic has had a major impact on the world economy, and Indonesia is no exception. Pakpahan (2020) as quoted by Nazuli et al., (2021) reveals that global economic growth will be minus 1.1 percent and the International Monetary Fund (IMF) projects that global economic conditions will drop to minus 3 %. One of the policies of the Indonesian government to restore the national economy as a short strategy includes increasing domestic consumption, increasing business activity and maintaining economic stability and monetary expansion. The government's efforts to move the business world are carried out through providing assistance to MSMEs and cooperatives. In addition, structural policies that focus on the introduction and use of digital technology for MSMEs as well as preparation for entering the Industry 4.0 era are needed as a solution to long-term problems (Pakpahan, 2020) in (Nazuli et al., 2021), (Yusroni et al., 2022).

Gunungkidul MSME is one of the UMKM that is currently developing and is receiving the attention of the government in Java Island, especially in the Special Region of Yogyakarta. The development of regional MSMEs has a very important role for economic growth for the community (Pujiastuti et al., 2021). Besides that, Gunungkidul as a part of the Special Region of Yogyakarta (DIY) is a destination for domestic and foreign tourists. The designation of DIY as a tourism city illustrates the potential of this province from a tourism perspective (Indrianingsih et al., 2020), (Dewi et al., 2021). One of these potentials is superior products produced by creative economy entrepreneurs in each region, for example in Gunungkidul to support tourism activities.

* Corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International (CC BY-SA 4.0).

In the second formulation of the National Economic Recovery (PEN) policy, namely increasing business activity, one of the media that can be utilized is E-Commerce (Ginting et al., 2021), (Tambunan et al., 2023). In line with what was conveyed by (Dinisari, 2020) in (Nazuli et al., 2021) that E-Commerce is one of the main drivers that makes Indonesia the country with the largest digital economic value in Southeast Asia reaching \$40 billion in 2019 and is predicted to increase to \$130 billion by 2025. E-Commerce activities have great potential in the pandemic era because buying and selling activities can take place without the limitations of a pandemic. What's more, government calls to limit activities outside the home make E-Commerce a desirable and appropriate option (Djufri, 2020; Nazuli et al., 2021). The 2021 E-Commerce Survey shows that 1,774,589 E-Commerce businesses (75.15 percent) of the total E-Commerce businesses in Indonesia (2,361,423 businesses) are still concentrated on the island of Java. In addition, the food, beverage and groceries group is the type of goods or services that are sold the most via the internet in 2020, the total businesses selling these goods or services are 40.86 percent of all businesses sampled E-Commerce (BPS, 2021).

E-Commerce provides product information sold by each seller. Through the pages on E-Commerce, data can be collected that is used to form useful information (Liu et al., 2020), (Miao et al., 2022). The large number of sellers and the variety of products found on the Web makes it difficult to collect the necessary data so that a method or system is needed that can automatically collect data from each page that displays or data provided on that web page (Husada et al., 2020). In order for the data set to be useful, it is necessary to design a place for data management (Ahmad et al., 2022). Data must be selected and arranged properly so that when used it can give good meaning and make it easier to analyze according to need (Li et al., 2022). One technique that can be used is Web Scraping. Web Scraping is also called Web Harvesting, Web Data Extraction or Web Mining, which is the activity of constructing agents to download, parse, and organize data from web pages automatically (Husada et al., 2020). This means that the user's activities in retrieving data from the Web screen are carried out by the Web Scraper with the aim that data retrieval can be carried out faster and more precisely.

LITERATURE REVIEW

Mauidzoh (2020) in his research entitled "Utilization of Augmented Reality to Support the Promotion of Superior Products and Tourism Objects in Patuk District, Gunungkidul Regency." The problems raised in this study are (1) How to develop superior technology-based product promotions in Patuk District as a tourist attraction (2) How to introduce the potential of superior products in Patuk District which can become tourism supporting products (3) How to design a promotion based on Augmented Reality (AR) technology for superior products in Patuk District in supporting Tourism in Gunungkidul. The specific objective is to develop existing website-based promotions with Augmented Reality technology to introduce superior products in Patuk District in support of Tourism in Gunungkidul. In line with its objectives, further research conducted by Akbar et al. (2022) and (Pujiastuti et al., 2021) in Gunungkidul to support tourism activities through developing the skills of creative economy entrepreneurs by utilizing digital promotions.

According to a study published by the Massachusetts Institute of Technology (MIT), the human brain can recognize images after viewing them within 13 milliseconds (Trafton, 2014). Referring to the results of this study, humans generally can be able to understand more information through existing images (Perdana, 2021). E-Commerce sites in Indonesia offer each product with different descriptions and prices. Descriptions with a large and varied number of words can take a very long time to manage them. This limitation in managing detailed product descriptions can be overcome by the process of retrieving information with product Web Scraping (Setiawan et al., 2020). Web Scraping Also Called Web Harvesting, Web Data Extraction Or Web Mining, is the activity of constructing agents to download, parse, and organize data from Web pages automatically. This means that the user's activities in retrieving data from the Web screen are carried out by the Web Scraper with the aim that data retrieval can be carried out more quickly and more precisely (Husada et al., 2020).

Business Intelligence (BI) is basically an evolution of a decision support system. BI's primary goal is to provide interactive access to data, enable data manipulation, and provide managers with the ability to perform business analysis (Oumkaltoum, 2019), (Rashid et al., 2022). With the ability to analyze data, performance, and business environment situations, managers can make better decisions. In general, BI is an umbrella term that combines various architectures, techniques, analytical tools, applications, and methodologies that allow easy access to data to help managers carry out business analysis. BI assists in the transformation of data, into information (and knowledge), then into decisions and finally into action (Valentinus, 2017).

* Corresponding author



In research conducted by (Husada et al., 2020), the data retrieval process with predetermined marketplace sources has been successful, and the data stored is in accordance with what is desired. The data taken has also been successfully visualized according to the original purpose of data collection (Wang et al., 2019), (Rosales et al., 2021). It is hoped that further work will be able to carry out the Web Scraping process by entering per page so that the selection of data to be retrieved is expected to be more complete compared to the list of data that can be retrieved in pages per specified item category so that the results of the visualization will be more varied depending on which data will be used as the object of research and comparison.

METHOD

In this research, the data collection method will be carried out by implementing the Web Scraping method. Data is taken from E-Commerce Shopee, Tokopedia and Lazada which are the largest E-Commerce and also have a lot of active users and use food and beverage product input as search input. can be used as research material to determine the distribution of product data of creative economy entrepreneurs in Yogyakarta and see whether there is important information from the distribution of the data that can support promotion and tourism activities.

Research Framework

Figure 1 shows the research framework. Starting from accessing E-Commerce which will be used as a source of information.

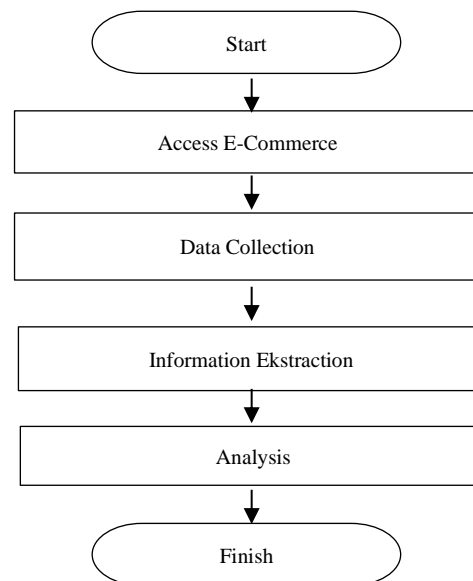


Fig 1. Research framework

At this stage, a search for products for creative economy entrepreneurs in DIY is carried out in the food and beverage category. The second stage is to collect data with Web Scraping Techniques. Web Scraping is the process of retrieving semi-structured document information from the internet in the form of web pages in the form of HTML or XHTML (Kurniawati et al., 2017), (Darmawiguna et al., 2019), (Henrys, 2021). This data collection was taken from E-Commerce Shopee, Tokopedia and Lazada pages based on E-Commerce URL id categories of food and beverages in the Yogyakarta area. The next stage is the extraction of information resulting from data collection by scraping then the unstructured data is made structured, to facilitate data processing. In the next stage, the data that has been obtained will be visualized using Tableau tools to simplify the data analysis process to produce a conclusion that can be used for other purposes. To support the results of the analysis, this study uses graphic visualization, charts which aim to visualize data to communicate information clearly and efficiently.

* Corresponding author



Method of collecting data

In this study the data used were <https://shopee.co.id>, <https://www.tokopedia.com>, <https://www.lazada.co.id/> especially in the food and beverage product category. The object of this research is all shops located in the Special Region of Yogyakarta which consists of Gunungkidul, Kulon Progo, Sleman and Yogyakarta. This research was conducted by applying the Web Scrapping Technique which was carried out by observing on page three of E-Commerce.

Data Visualization Methods

The data collected has visualization needs and certain goals from visualizing product data from the three E-Commerce. The method used for the development of applications used in this study is the Fry Method (can be seen in Figure 2).

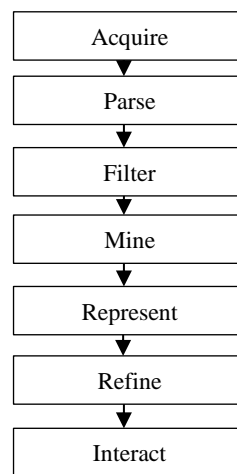


Fig 2. Process of data visualization fry method

Process Analysis

This research uses a case study method on open data in three e-commerce products for creative economy entrepreneurs in Yogyakarta related to the category of food and beverage products. This data was visualized using the fry method with the Tableau application and analyzed using Descriptive Statistics. In Figure 3.3 is a visualization of the research stages.

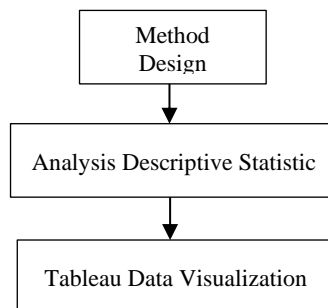


Fig 3. Research chart

The results of this research analysis are in the form of data visualization using Tableau to describe the information obtained clearly, effectively, and efficiently. The results of the analysis from this study are expected to assist buyers

* Corresponding author



or sellers in making decisions, especially for creative economy actors in the Yogyakarta region. The results of the analysis will show product information from the three E-Commerce categories for each category, you can see the distribution of prices for certain products, the best-selling products, the average price for a product.

RESULT

The results of this study are the application of Web Scrapping Techniques to E-Commerce Shopee, Tokopedia and Lazada by using Web Scrapper. The results of scrapping data are stored in Ms. Excel with .CSV format is then filtered to facilitate the process of analyzing product data for each category, including the distribution of prices for certain products, the best-selling products, the average price of a product.

The analysis process will be carried out using the Fry Method, including:

1. Acquire

At this stage data from the E-Commerce page is selected and downloaded according to the needs of the research (Figure 4 to 7). The data collected is product name, product price, amount sold, seller's area.

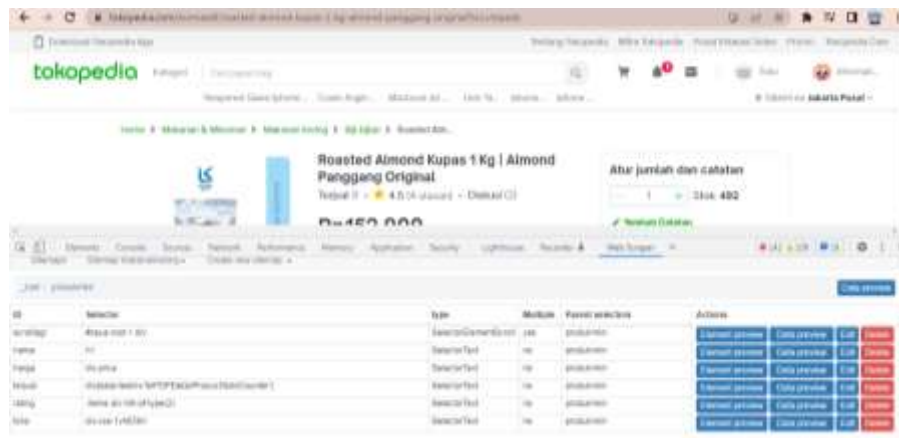


Fig 4. Acquire on the Tokopedia Website (a)

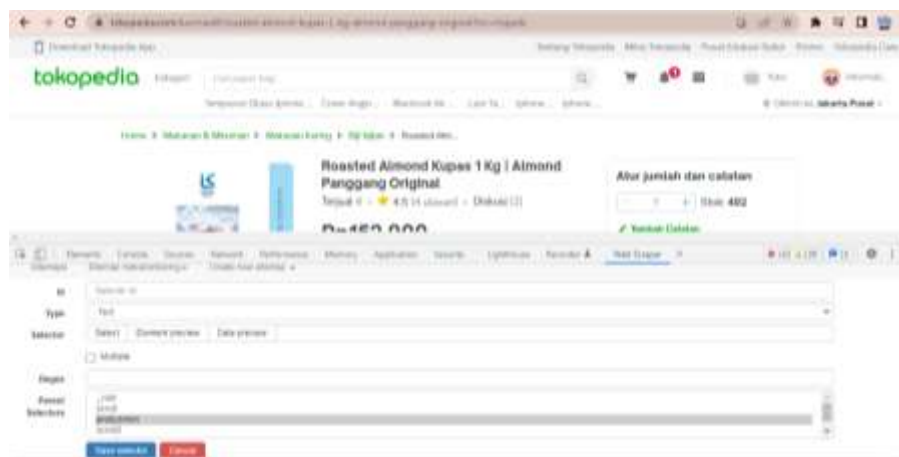


Fig 5. Acquire on the Tokopedia Website (b)

* Corresponding author



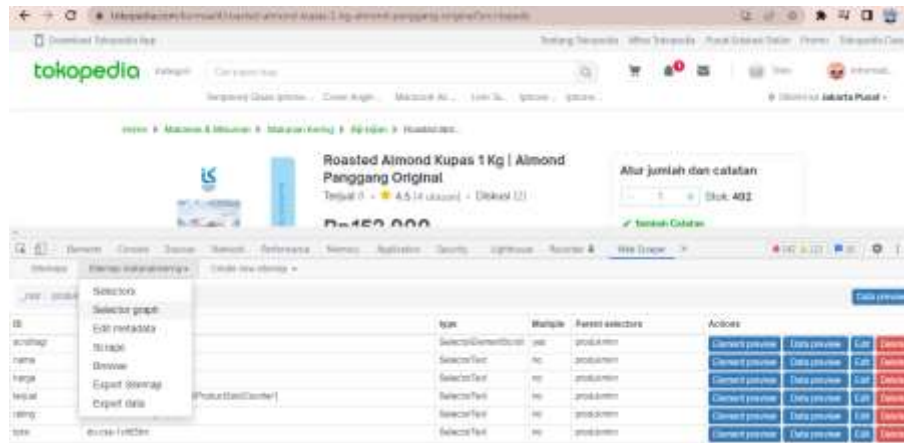


Fig 6. Acquire on the Tokopedia Website (c)

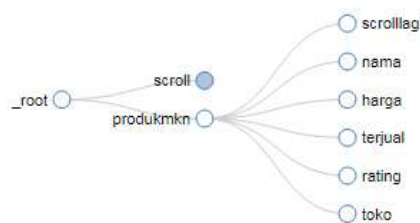


Fig 7. Product Data Sitemap on the Tokopedia Website (c)

2. Parse

The scrapped data is stored in .CSV format (Fig. 8 to 10).

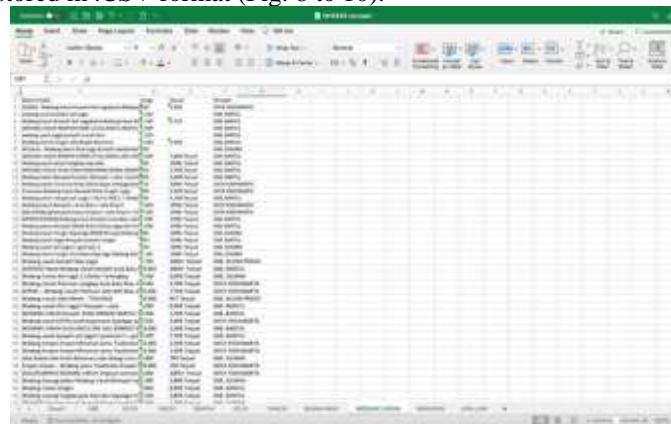


Fig 8. Shopee Website Scrapping Result Data

* Corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International (CC BY-SA 4.0).

Fig 9. Shopee Website Scrapping Result Data

Fig 10. Lazada Website Scrapping Result Data

3. Fliter

The data was sorted and the irrelevant data was deleted (Figure 11) with the use and purpose of the research. Filters are done with concern for the purpose of data visualization.

Fig 11. Filtering Result Data

4. Mine

* Corresponding author



A process that involves statistics using a descriptive method to see the distribution of visualized product data. Descriptive statistics are a way of collecting, labeling, describing, processing and analyzing numbers and interpreting them in graphical form to be analyzed and interpreted by drawing conclusions (Haden, 2019), (Silvia, 2020), (Cooksey, 2020). The data analyzed in this study are products sold in three E-Commerce (Shopee, Tokopedia and Lazada) by creative economy entrepreneurs in DIY.

5. Represent
 Specifies the product data set to display. The way to do this is to choose a basic visual model, such as a histogram/bar graph, list or tree.
6. Refine
 The process of improving or improving from the basic visual that has been carried out at the represent stage. At this stage improvements are made so that the resulting visual data is easier for users to understand because the data is clearer and more visually appealing.
7. Interact
 Methods for manipulating data about what appears or wants to be displayed as information. Figures 12 to 14 display information on the types of products and the number of products in each region. The visualization chart shows that the City of Yogyakarta is the area with the most creative economy business actors offering products in E-Commerce, while the least are in the Kulon Progo region.



Fig 12. Visualization of Data on the Number of Food Products in Bantul Regency

* Corresponding author



BhnMkn-GK-source1



Fig 13. Visualization of Data on the Number of Food Products in Gunungkidul Regency

BhnMkn-KP-source1

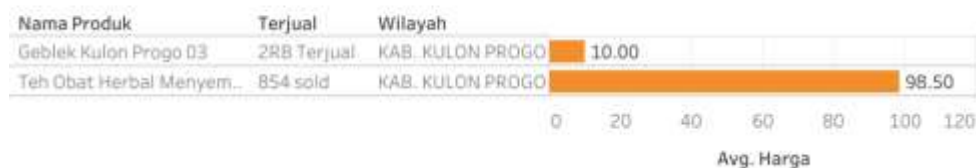


Fig14. Visualization of Data on the Number of Food Products in Kulon Progo Regency

One example of data visualization in Figure 15, the products that are widely offered in the Gunungkidul area are chips-type snacks. In contrast to those in the Kulon Progo region (Figure 14), the dominant products offered are regional specialties, namely geblek and herbal drinks.

BakpiaKulius-Source1



Fig 15. Data Visualization of Bakpia Product Names

One of the typical foods in DIY is bakpia, if you look at the data in Figure 15, business actors come from the Sleman and Yogyakarta City regions. Bakpia products that are widely sold in E-Commerce are Bakpia Tugu Jogja.

DISCUSSIONS

Data on the products of creative economy entrepreneurs in three E-Commerce which have been visualized so that the data can easily have a useful meaning in making business decisions for business actors or the local government. This research produces some information to help data users, especially business actors or the government. Making clusters according to the sources of information obtained, namely Shopee, Tokopedia and Lazada based on their respective regional groups (Figure 16) is useful for data users in determining marketing strategies, namely in terms of market segmentation and targeting.

* Corresponding author



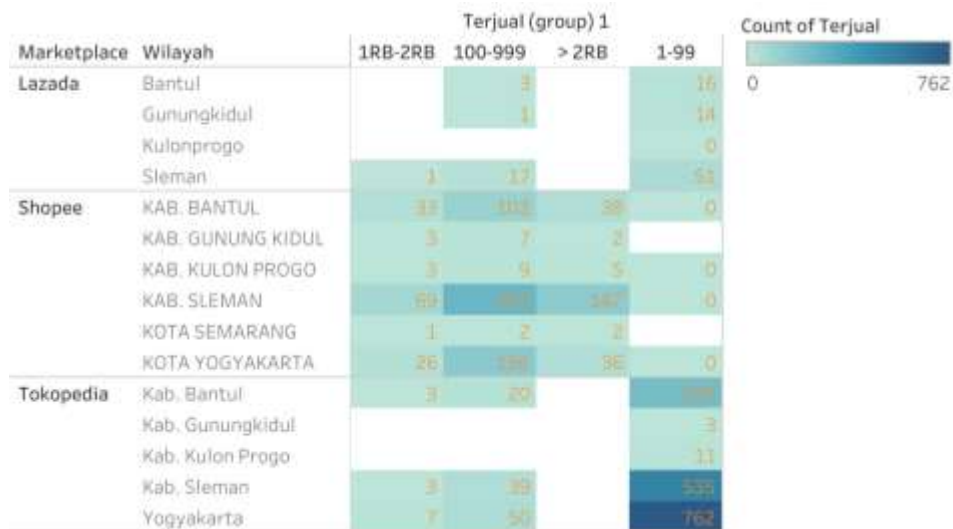


Fig 16. Product Data Tableau Based on 3 E-Commerce Clusters (sold)

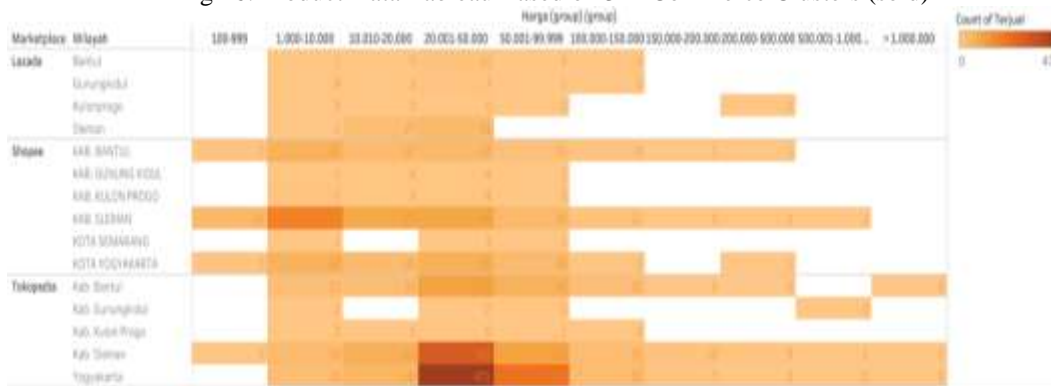


Fig 17. Product Data Tableau Based on 3 E-Commerce Clusters (price)

Figures 16 and 17 show the data distribution of products that have been sold in the three E-Commerce based on their respective regions. Grouped data visualization helps users optimize product marketing or make decisions. Shopee is the most widely used E-Commerce by businesses in DIY to market their products. Prices for several products offered by business actors in the three E-Commerce average range from 20,000 to 50,000.

This data visualization can be used as a decision-making tool, for example determining product price standards for the same type of product in E-Commerce that will be used by business actors to market their products. So that business actors can obtain valuable information from visualization results for business (Boldrini et al., 2021), (Martins et al., 2020). This information includes Shopee being the E-Commerce with the first order that is most often used by actors, the price range of products offered in the region, the products that are best selling in these three E-Commerce are DIY special food products, namely bakpia. If you look at the types of products that are mostly sold by business actors in the three E-Commerce products, the average is bakpia, snacks (snacks such as chips), wedang uwuh/spice and typical cakes such as (wingko and yanko).

CONCLUSION

Based on the results of the research above, it can be concluded that Web Scrapping can be implemented in three E-Commerce (Shopee, Tokopedia and Lazada) to collect product data for the food and beverage category in DIY. Data visualization requires planning in order to turn it into useful information. This product data was successfully visualized using Tableau software and is useful for businesses or local governments. The results of product data analysis from the scrapping process reveal information that the majority of products sold in the three E-Commerce are

* Corresponding author



bakpia, snacks (snacks such as chips), wedang uwuh or spices and special cakes such as (wingko and yanko). Bakpia is the best-selling product in all of the E-Commerce. Based on the visualization, it was found that the platform that is widely used by business actors in the DIY area is Shopee.

ACKNOWLEDGMENT

The author would like to thank the Adisutjipto Institute of Aerospace Technology Yogyakarta for providing financial support for this research.

REFERENCES

- Akbar, M. P. R., Ayuningtyas, A., Indrianingsih, Y. (2022). Pemanfaatan Media Komunikasi Visual Sebagai Alat Bantu Promosi Obyek Wisata Menggunakan Metode Markerless Augmented Reality (Studi Kasus: Kampung Baran, Desa Salam, Kap. Patuk Gunungkidul). *Seminar Nasional Teknologi Informasi dan Kedirgantaraan, 2021(7)*. <http://dx.doi.org/10.28989/senatik.v7i1.456>
- Ahmad, K., Maabreh, M., Ghaly, M., Khan, K., Qadir, J., & Al-Fuqaha, A. (2022). Developing future human-centered smart cities: Critical analysis of smart city security, Data management, and Ethical challenges. *Computer Science Review, 43*, 100452.
- BPS. (2021). *Statistik E-Commerce 2021*.
- Boldrini, J. C., & Antheaume, N. (2021). Designing and testing a new sustainable business model tool for multi-actor, multi-level, circular, and collaborative contexts. *Journal of Cleaner Production, 309*, 127209.
- Cooksey, R. W., & Cooksey, R. W. (2020). Descriptive statistics for summarising data. *Illustrating statistical procedures: Finding meaning in quantitative data*, 61-139.
- Darmawiguna, I. G. M., Pradnyana, G. A., & Santyadiputra, G. S. (2019, February). The Development of Integrated Bali Tourism Information Portal using Web Scrapping and Clustering Methods. In *Journal of Physics: Conference Series*(Vol. 1165, No. 1, p. 012010). IOP Publishing.
- Dewi, K. K., Solichin, A., Amalia, D. A. R., Wisudawati, T., & Saputro, W. A. (2021, April). The Role of Social Media Instagram and Identification of Factors in The Revisit Intention to Gunung Api Purba Nglangeran. In *Proceeding of International Conference on Science, Health, And Technology*(pp. 332-336).
- Djufri, M. (2020). Penerapan Teknik Web Scraping Untuk Penggalan Potensi Pajak (Studi Kasus Pada Online Market Place Tokopedia, Shopee Dan Bukalapak). *Jurnal BPPK : Badan Pendidikan Dan Pelatihan Keuangan, 13(2)*, 65–75. <https://doi.org/10.48108/jurnalbppk.v13i2.636>
- Ginting, K. R., & Pancawati, G. (2021, July). Post-Pandemic National Economic Recovery. In *2nd International Conference on Administration Science 2020 (ICAS 2020)* (pp. 25-29). Atlantis Press.
- Haden, P. (2019). Descriptive statistics. *The Cambridge Handbook of Computing Education Research*, 102-132.
- Husada, I. N., Fernando, E. H., Sagala, H., Budiman, A. E., & Toba, H. (2020). Ekstraksi dan Analisis Produk di Marketplace Secara Otomatis dengan Memanfaatkan Teknologi Web Crawling. *Jurnal Teknik Informatika Dan Sistem Informasi, 5(3)*, 350–359. <https://doi.org/10.28932/jutisi.v5i3.1977>
- Henry, K. (2021). Importance of web scraping in e-commerce and e-marketing. *Available at SSRN 3769593*.
- Indrianingsih, Y., Ayuningtyas, A., & Mauidzoh, U. (2020). Promotion System For Promotional Products and Tourist Object in Patu Gunungkidul District. *Angkasa, 12(1)*, 79–88.
- Kurniawati, D., & Triawan, D. (2017, November). Increased information retrieval capabilities on e-commerce websites using scraping techniques. In *2017 International Conference on Sustainable Information Engineering and Technology (SIET)* (pp. 226-229). IEEE.
- Liu, G., Fei, S., Yan, Z., Wu, C. H., Tsai, S. B., & Zhang, J. (2020). An empirical study on response to online customer reviews and E-commerce sales: from the mobile information system perspective. *Mobile Information Systems, 2020*, 1-12.
- Li, V. Q., Ma, L., & Wu, X. (2022). COVID-19, policy change, and post-pandemic data governance: a case analysis of contact tracing applications in East Asia. *Policy and Society, 41(1)*, 129-142.
- Martins, A., Martins, P., Caldeira, F., & Sá, F. (2020). An evaluation of how big-data and data warehouses improve business intelligence decision making. *Trends and Innovations in Information Systems and Technologies: Volume 1 8*, 609-619.

* Corresponding author



- Miao, M., Jalees, T., Zaman, S. I., Khan, S., Hanif, N. U. A., & Javed, M. K. (2022). The influence of e-customer satisfaction, e-trust and perceived value on consumer's repurchase intention in B2C e-commerce segment. *Asia Pacific Journal of Marketing and Logistics*, 34(10), 2184-2206.
- Nazuli, M. F., Utami, N. A., & Putri, S. R. (2021). Information Dashboard Penjualan Produk Fashion Muslim di DKI Jakarta Sebagai Pendukung Strategi Pemulihan Ekonomi Nasional. *Seminar Nasional Official Statistics*, 2021(1), 1025–1035. <https://doi.org/10.34123/semnasoffstat.v2021i1.1009>
- Oumkaltoum, B. (2019, April). Toward A Business Intelligence Model for challenges of interoperability in egov system: Transparency, Scalability and Genericity. In *2019 International Conference on Wireless Technologies, Embedded and Intelligent Systems (WITS)* (pp. 1-6). IEEE.
- Pujiastuti, A., Ayuningtyas, A., Kusumaningrum, A., & Saputra, A. D. (2021). Nawani and Nukoni Applications for Supporting Small and Medium Micro Enterprises in Gunungkidul District. *Angkasa: Jurnal Ilmiah Bidang Teknologi*, 13(1), 15–26. <https://doi.org/10.28989/angkasa.v13i1.807>
- Rashid, A., & Khurshid, M. M. (2022, July). A Descriptive Literature Review and Classification of Business Intelligence and Big Data Research. In *Intelligent Computing: Proceedings of the 2022 Computing Conference, Volume 1* (pp. 865-879). Cham: Springer International Publishing.
- Rosales, J., Deshpande, S., & Anand, S. (2021). IIoT based augmented reality for factory data collection and visualization. *Procedia Manufacturing*, 53, 618-627.
- Setiawan, D. F., Tristiyanto, T., & Hijriani, A. (2020). Aplikasi Web Scraping Deskripsi Produk. *Jurnal Teknoinfo*, 14(1), 41. <https://doi.org/10.33365/jti.v14i1.498>
- Tambunan, N., Gumilar, R. T. P., & Karimah, L. N. (2023). Post-Pandemic Economic System Changes. *Journal of Social Research*, 2(2), 289-299.
- Wang, C. S., Lin, S. L., Chou, T. H., & Li, B. Y. (2019). An integrated data analytics process to optimize data governance of non-profit organization. *Computers in Human Behavior*, 101, 495-505.
- Yusrani, N., & Pratiwi, R. (2022). Impact Of The Covid 19 Pandemic On Micro, Small And Medium Enterprises (Msmes) Performance: A Study In Indonesia. *Journal of Positive School Psychology*, 6(7), 1610-1615.

* Corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International (CC BY-SA 4.0).