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

Astika Ayuningtyas^{1)*}, Yuliani Indrianingsih²⁾, Uyuunul Mauidzhoh³⁾

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 to support strategic policies in determining product prices from creative economy business actors 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



Volume 5, Number 1, January 2023 https://doi.org/10.47709/cnahpc.v5i1.2184 **Submitted** : March 7, 2023 **Accepted** : March 8, 2023 **Published** : March 8, 2023

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 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



Volume 5, Number 1, January 2023 https://doi.org/10.47709/cnahpc.v5i1.2184 **Submitted** : March 7, 2023 **Accepted** : March 8, 2023 **Published** : March 8, 2023

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.

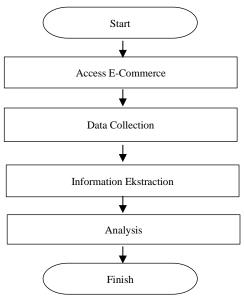


Fig 1. Research framework

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



Volume 5, Number 1, January 2023 https://doi.org/10.47709/cnahpc.v5i1.2184 Submitted : March 7, 2023 Accepted : March 8, 2023 Published : March 8, 2023

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).

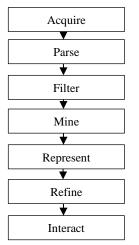


Fig 2. Process of data visualization fry method

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.

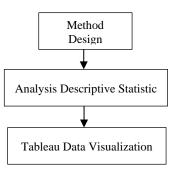


Fig 3. Research chart

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



Submitted : March 7, 2023 **Accepted** : March 8, 2023 **Published** : March 8, 2023

Volume 5, Number 1, January 2023 https://doi.org/10.47709/cnahpc.v5i1.2184

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.

+ + 1	C B Dispatation to all the second to	on (la division (seiger	- intransitionen					3. H. M.	* 17 18	21
0.00	and and the second is later			THEFT	(hereast)	-	in poore		rei Taiped	aller i
tok	opedia				14	W 80		the last		e.
	Heapened Game (physics)	Santage, Malan	al. and a		door -				- Industry Protor	é.
	Inche & Managine & Managine & Managine	ning 1 million 1 from	et dan.							
	Ĕ	Rossted Almond Panggang Origin Termini I - 4 4 104	at			itur jumlat	dan catab • Disk 40	107 F		
	N.S. av A	D=450 000				· home for	Latury .			
	Density Courts Borns, Named, Antonious, Division of Court	Names Approxime	here upon	e. 160	- 4 -	at hope		#141 m		5 I X
2010	inter /									
	bracker.		1.00	Molton .	Person and a	215	Action			
arring .	4191-0-1021-022		Salar price media set	140	point-ten		E Street	Address Contra	ILINY LIK	
1919			Darter Tank	-	-		1.000	man Cart	street Life	Canal of Lot
1994	di pri a		Selected	19	pane			erres Calls	Automa (Car)	(been
1010-00	distant with STITICS and Distillation (Banariar Taxà	10	-		1	Second Calls		(Thereas)
18814	June als 100-04 (pec)		description (see	10	100.000		a series	second Carls		
lake .	sta con treblan		instanting in the	14	100.000		1	and the second second	the second second	10000

Fig 4. Acquire on the Tokopedia Website (a)

D Description Print Print Print Print Print		Terrary Ter	service, Miles	Incode (1	 4.770		inter (lan
tokopedia		Santago Manada Larta pina pina		4 ⁰ m	 an Januaria	-	
tions 2 Main	arar-B.Mitamar-P. Mitamati-I	tartes & Sprinker & Researching.					
1	<u>к</u>	Roasted Almond Kapes 1 Kg Almond Panggang Orlginal Separt 0 - # 4111 statement - Dakas [1] Des 150,000		amilah dan 9 +) 0 nat Calatan			
E Dente Louis los	an America Netternets a Transition America a	Barry Applates Sector Systems Resided	A constitution	1.00	101-0.178	(RI)	
El terrer tren terrer End terrer Ed terrer Ed terrer Ed terrer Ed terrer Editore End	er - Tonani Ane atomar e	, Morey , Aprildian, Steam, , Lydnaw, Annale (A	100 - 1	CHE ALT	-	0
All Second Secon	er - Tonani Ane atomar e	, Morey , Application , Starting , Spectrum , Applied &	A nertis	Ann I		-	0

Fig 5. Acquire on the Tokopedia Website (b)

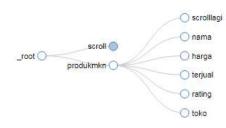


Submitted : March 7, 2023 **Accepted** : March 8, 2023 **Published** : March 8, 2023

Volume 5, Number 1, January 2023 https://doi.org/10.47709/cnahpc.v5i1.2184

	· Menufacercho											-
() Promis	of Dynamics in a				heir	physics of		61 Mar	Contract 1	beine P	termi To	in section (Co
toko	pedia	Carportion				10. 9		-	-	100	-	streat.
	And showing the state of the st	Report Data press	Contage.	Maning Mr., Sonth, S.					1	4 (main)	in select	a Planet +
	correct & Administ	a. 6 Million & Makazari	And & Spinso	ri Frankrider.								
		۲ 🚺	Panggan	Almond Kupas 1 Kg A g Original		Ab	a jumlah					
		C-STREET	Terrifield (I = 1	4.6.14 ukaccel - Ulekasi US			18	+ Disk	492			
	1		Ded 6				taménéh Gal		402			
fil ne			n-460				taménéh Ché	tat an				. 0
fi) the		Banch federate	n-460	000			taménéh Ché	tat an		410 L	in (*	•
States	en Gran Juan Tura Menerativas Silicito	Banch federate	n-460	000			taménéh Ché	tat an		*10 S	in (#1	•
anninger Anninger	nen Green Jaco Factor Green Jaco Senctor Senctor page	Banch federate	n-460	000	a . baa		handadi Cod	tat an		* in 1	in (e)	and period pro-
inings reg prob-	Sector page Sector page Est renativo	Banch federate	n-460	anno larta arre	e ter		handadi Cod	Acti				and period pro-
inings reg prob-	Sector Sector Sector Sector Sector Sector prot Solution Sciences Sector	Banch federate	n-460	National Sector Sector	e ter	ne à mi	handadi Cod	Acti	048	-	1	and period pro-
anninger rerr genten reg	Sectors Sectors Sectors Sectors prod Edit refactors Strap Drave	Banch federate	n-460	National Devis operation	n fan	Autor and a	handadi Cod	Action Action	in the factor	e Um	100100	and period pro-
Contractions Co	en Gren Jane Secto Secto Secto Estretato Estretato Tange Estretato Bowe Estretato	Banch federate	n-460	Non Ministra Davis, ripress Non Securiture Securiture	n fan	Anton and the positions positions	handadi Cod	Action Action	su su sultas	e Um		and period pro-

Fig 6. Acquire on the Tokopedia Website (c)



$\mathbf{E} \sim 7$	Duaduat	Data	Citorna	and the	Talaa	li a	Walacita	(a)
FI2 /.	Product	Data	Sueman	on the	ΙΟΚΟΙ	eaia	websile	(0)

2. Parse

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

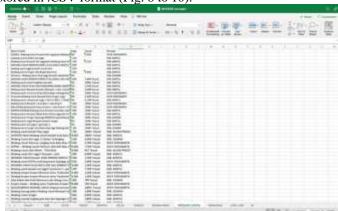


Fig 8. Shopee Website Scrapping Result Data



Submitted : March 7, 2023 **Accepted** : March 8, 2023 **Published** : March 8, 2023

Volume 5, Number 1, January 2023 https://doi.org/10.47709/cnahpc.v5i1.2184

	•• Q	5.0 0.1		als in an	
	fault for	e Antiney House but Now the 2.5	4-u		The local
-	-	the sea has been been been			The second s
		and the second se	haden in		
-	P.C.A.	the second	and states of the	the second second	termine over any dash dash many dark that, dash
	1000	ALCONTO A CARL OF A CARL OF A			through a loss when
	10.00	C			
	CONTRACT OF	The foreigneet one	and the second	10.000	
2.1	ALC: NO.	And the second second second second	100.00	THE REAL	
	side seen	internal manager in the set of the	10000	10.000	
2	COLUMN TWO IS NOT	Printed and control of the second sec	Land and	ingentation of	
2.	continues.	Print or type index for inits \$10.0 being the init \$10.0 being	Sec. at	THE PARTY	
2.1	plant Service	August and Aug	1000		
2.1	and the	Martine and a the for affiliate lang they i Contractly	factor and	100.0000	
	2161 Adap	The second	ALC: NO.		
	one have	built index and the section in the sec-	100.00		
20	PERSONAL PROPERTY AND INC.	the story offer a lot from the lot of	Tax Do to		
	LOUI BORNEY	Annual Control and Stational Institute Institute	Tanking .		
	+101-018-00	Disk for any one and the part of the part of the later.	1401.000	- manufactor	
	1010	All of the Rest of the other states of the state of the local states of	And Contemporate	Appendix	
20	101 586	And Any contents	And other	in terms	
2.1	and a second	Station & Balances Trades Instation	table last	and Autor	
2.1	and states	and any present prior and the	August -	and Assess	
	the second	- The second sec	August day	100 Harris	
	10100	Statistics, Second - 1752 All address of Strategy - Tanada - Statement of Strategy	Autom:	ind bollet	
	101.000	The second s	10100	on examine	
	arter are off	21 - In the Advisory Concession of Statements and Advisory of Statements	1415.05		
	alair and	ter ang, da barrar	2212	the strengt	
	and the state	the story of a fact that has	1000	the literate	
	trans ages	then had dealt	tank in	ing ballet	
2.	101 84	12.00	takine .	the Annual	
2.1		the first them we have a meaning of a first strike.	Automatical Statement	the same	
	the local in	the other which is an on the design of the property limit in the party	Accessory.	-	
	10114440	contraction (the balance late deschooled) \$122.02 (Str.)	the second	and beauty	
	And a state	tion arrived have steep white saved first	And in case	and Annual	
	and the second	The Largence that allowing the factors	Sectors.	100000	
	allow states	The same size and a start fight the set from the	10000	- Contract of Cont	
	and and	An other lands with	And in	The local	
	or the design of	Approximate to Traine Hereign (Strengt	tracing.	the Association	
	and the	this bail there is an in an include a second second second	No. of Concession, Name	Takenter.	
2.1	101100	Transfer transporter to an or the set of the Advancement of the set of	to the second	Taken of the local division of the local div	
2	and and	Annual age of the second of the second		the land	
2.	and a state of the	Annual Cold, Annual Cold Lines	1000		
2	order march.	Appropriate Carl & Southern State St	tartist.	Transmitter of	
	and a second	The first of the second s	tailed.	10.000	
	1101 March 1	And and the state of the state		tab. Kimu	
	10.00	tion by beauty and	Address of the	the Annual	
	222	San and the second second	and the second	the second	
				- Longe	
		deal had			

Fig 9. Shopee Website Scrapping Result Data

ties:	-	-	Frage 1.8	****	-			-	-		-												1.00	**	-
fer.	1.1	-		- 11	- 4	1	×	10	**	14	-	÷			10	1.18	- 10	1	-	100	142	123	39.4	0.	提
14	\$1. 1		- 12		14		1.1	π.	11.1	18	-	-	1.00	•	 -	:23	81		1000	-		10.	122	222	
4	Ŧ.	- 4																							
																	-								
	100								104	164 C	+++														
	(along									200															
	1000								2021																
	Channel .	Taxa ta																							
	i land								ar) an 5		in the second														
	1.000										Sec.														
	144								ene hait t		Remark.														
	i'mme.										ine i														
	inte.																								
	t interiori								un he s		Terms.														
	0.000	inter in									10.00														
	100								222																
	1100																								
	-	ingeneration of the second							and don't																
	i badai '								the last		-														
	1000							-		8															
	Colored T	-							ومدمي		-														
	i anije	-									-														
	i kising	-							60 BC		-														
	198	1000							week,		-														
	in the second	Auguan a							-114																
	1.000																								
	i Antonio	Annual State																							
	A Autom								2123		-														
	Antonio	An opposite the second									-														
C . 4	6 Alanaga (14440		-				41.11		K	-														
	Chinksi-	Manuality									hanse.														
	0.000	1000									-														
	1.00	- 8-3 LBV									-			_											
	1.000	(second second s							spin filer																
	-	Contraction of the local division of the loc			-				tern and T	<u>.</u>	100														
	- 22	1000																							
	-																								
	1000	diam'r									-														
	1000	-					تو التق				-														
	1000	100-101							some?	e	-														
	1.000								com)		-														
	Part .		414						1.1.1.1		-														
	All and a second	Appalent					-				-														
0.01	Canal of Can	10.00								T	- e														
		- 14		and it																					
-	- Hinter	-	-																	- 1	100.00	1000		-	
																				_					-

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.

	Dec fu	minghal forman data darang tina ji fafina				C 944	
田子				新一部で	반전만	影響量	西
enter 2	3.46	PRIAMA AND INSULTS					
-	14						
-	and a second	Second and the second state of the second stat	100	100.00	diame.		
in the second se	-	the start she is service that and it is as to determine the	5	See	And contents		
the state	-	Restor of the street income and the line of the street of	5.00		and some		
-	-	amonto the react	Gii	1000 1000	and industry		
and the	-	Number of States	5	A dist balance	and industry.		
diam'r.	-	NUM .	See	a late inclusion	And and the set		
in the second se	-	Insulating channel the channel state	C.	A STATE STATES	And industry		
and a		Name the party of the party of the	See	and the second	And the second		
-	had -	security from the complete day day	-	an long	And residently		
and and	-	Private 110 state / Westmann Berlandt	C	A 1984 Million	NO COMO		
1000	-	stations because as universation based on the	See	A AND INCOME.	And Designation		
And a local diversion of the local diversion	-	search of the local to the shear have	Sec.	27 Subar	and the second s		
-	-	Particular (Inc.) 1980	See.	and home	And the other description.		
-	-	angled over the	Cont.	All Tribel	and taxable		
	-	Links had being	Sau	Land Instant	And induced		
- Anne	2	started big bings by a set big	San.	141 Tuber	THE COMPANY		
		No. on Address of the start of the start and a 1 de ferror manual de starte free fille fille distant	5	1 APRIL TOTAL	And include		
-	-	No. on Address of the second state of the seco	100	All Inder	ATTA DOLLARS		
22		Nor having investment from Hig- context lagest	200	Address .	And Address		
	-	Not having theme that the upper lagent	200		And material		
-	-		Sam		And Linkson		
inger .		rain time file	100				
inger .	-	No. of the Tagle Dank Upp			AND DESIGN		
-	-	200301 (1993) 3 5/81 We can straig tariful (1914) her meritaling has the	0.00	100,400	ANI LOND		
	-	rear to forme	3.40		NYT-TOLASER.		
	100	Michights High Statement Inscrimental Reveal Bill Ann Description and	2.44	01101	. All 141800		
10.00	-	HINAR CONTRACTORS AND ADDRESS	3446		X00.12.0000		
1000	-	NUMBER OF THE OWNER	244	304.004	All COST TIME		
The state	-	servers the sector had	244	24 144	10.0070		
	100	he behalvestermine	2-0	Tales .	10.000		
-	100	processing and the second se	2.00	Contractory	10,000		
Train .	22	herebulus with a fail of stational hassing for	200	-1 800 Testini	10.000		
-	-	the set of the last of the las	210	CORPORATE AND A DESCRIPTION OF A DESCRIP	100.000		
		and and the attract of the second sec		100100	10.1100		
	-	The second of the second	-	- 100 King	And Address		
100	-	THERE IS NOT THE TAXABLE PARTY AND A DESCRIPTION OF THE PARTY	-	Table Table	VID DURING		
			No.	A MER PROPERTY.	And the second s		
-	-	Annual of the second state		A ME TONE	And Designed		
-					And instants		
-	-	Addated Laws Burling, ME King	200	1.000 August	And Address		
- The second	-	(Relief) and two-burble taken a stream momentary		A DEST TAQUES			
- Annal -	-	Sector the cost (Marrishal publ)	Same -	Admit August	And address		
1.00 million and	1000	server a children was a server of the server	1000	- VMP Talasia	- Physical Street Stree		
							

Fig 11. Filtering Result Data

4. Mine

* Corresponding author



Volume 5, Number 1, January 2023 https://doi.org/10.47709/cnahpc.v5i1.2184

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.

Rama Produk	Terjual	Wileyah	
(III) Gettial Barry Curs R	5.2990 Termont	ALC: BARTIN.	40.0
ABON AVAM SUPERR SHI	101 word	WAR BARTLE	117.0
Agarinea	1,040 Terpan	ARE BARTIN.	30
AMEX POTATO FIETHOLE	TA3 acrit	MAR BANTUL	7.5
Autor Uniatied Bottler M.	4,400 Terrise	AM, BANTUL	23.0
Bana keingenyehaayy	103.4001	AAB. BANTID	6.0
Disert Victor Article Vice.	SAT Twipsel	KAB. BANTUL	63.5
Bryongrovery, Sarder Par.	SDM Tempani	1848, \$548710.	6.0
Nation Colder Royal 3 Hz	A27 Fergiél	KAR BANTILL	50.0
BUNING RACIE INDOFOOD	SILL Temposi	NAR BARTIN	10.0
Company of Next Institution	809 years	NAR SANTUL	10.0
Enging Wolves Memall 0.	013 Petpant	KNB, BANTUL	23.0
CARRAN HEADALAYA PRIMA.	249-uon0	10.649: \$1.636/TULL	25.0
Repart Durit, Narsalo Gay.	Links Depart	ANE BARTIN	22.0
Barling the Tyrtro Kendel Lau.	3.2ND Terpus	INAR BHINTLE	29.0
Rump Harn Anglethy We.	177 Sprpad	SAB BARTIN	52.5
DU AHE DUE Bats date 5	265 2001	KAR BAHTID.	21.5
Hala was / Gala Longata ? .	647 Terson	AAR BOORTIN	15.0
Gula bebo petate ito 10 pert.	Apello-Terrold	WHE BANTLE	85.0
Kata Baba fersitat Haring d.	men Torgant	AAR SANTIR	16.9
Bala Bathy Percentage Addit 2.1	548 uoto	HOME BANKTLD	23.0
DULA JAWA JAHE MERAH.	242 mild	AAR BANTIN	135
UULA INDER CURAH KOMA	355.5010	INAS BANTLE	0.0
Herball Portievit Herball A.	2.488 Terrorit	AND BANTIN	tekd
Jantos Murrito Mantocari	1.649 Terror	ANE, BANTID	24.0
Kacara Nethile Organii M.	100 wind	KAR BANTIN	42
Kalula Jamar Manang Liba.	Sittlin Good	NAR BANTUL	17.0
Kalifu John Weining Nort :	Robbin Terlant	KAR BARTIN	7.8
Katiki, Jurnur Maratty Dask	EE1 yours	NAME BAAYTLE	E8-0
Kalita Jarriar Totole 80 gra.	1.648 Terpet	WAR, BANTUR.	ES.0
KATL Parget Gerray Bulls	ALC: TOPJARI	NAR BARTLE	
LARONTE PALTA PETTUEL	134 wind	KNB. BANTIN.	10.3
Langear Organik - Servic Or.	TLS error	4348 85419713	215
Maxiela Catella (14/ Oare D)	ST2 Tereser	AND BURNETH	47.5
Magginerig interrespirate to	Add and	WAR BARTIN	1.0
blue perment / blue through st.	805 Telbari	AGE SANTUS	100
Nazzel Ayler Hertato 500	DAT sold	KAR BANTH	111 IN 8
Naturgal Discourses 10 gram	Mail	RAD, BANTID	1.9
Chahaet (Kart Isertiak Supres)	E13 with	AAR BANTIN	14.5
DRIFLAKES LOW/G DUARE	410 hergial	AAB, BANTLIL	es.u
Britlanes Samuel Strein Bar	2003 orrel	#3/8: 8/4/97131	0.429
pinial-barri hullak marri	MAR works	KONGE BERBETTILL	100000 47.0
barribal sellar e cubin bara c	1.310 1011	848 849710	42.5
teblak Instan Kompit	404 Televiti	ADAR BANTON	1.4
TEPHNICLARD KUNING DE.	#0.7 amit	AND, SANTLE	0.08
			0 AD 300 300 20

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

* Corresponding author



Submitted : March 7, 2023 Accepted : March 8, 2023 Published : March 8, 2023

Volume 5, Number 1, January 2023 https://doi.org/10.47709/cnahpc.v5i1.2184

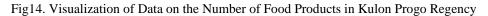
BhnMkn-GK-source1

Nama Produk	Terjual	Wilayah				
Kentang Mustofa Full Ke.	2,4RB Terjual	KAB. GUNUNG KIDUL		11.50		
Kering Kentang Full Kenta	291 sold	KAB. GUNUNG KIDUL			29.00	
Kering Kentang Mustofa K.	328 sold	KAB. GUNUNG KIDUL			29.00	
Kering Kentang+ Kacang+	1,1RB Terjual	KAB, GUNUNG KIDUL		14.00		
KERUPUK RENYAH kerupu	884 sold	KAB. GUNUNG KIDUL	5.80)		
PATHILO/PATILO/Pathol.	886 Terjual	KAB. GUNUNG KIDUL	3.75			
TELUR maggot BSF per Gr	511 sold	KAB. GUNUNG KIDUL				50.00
TERHEMAT !! KERING KEN	335 sold	KAB, GUNUNG KIDUL			24.50	
TERHEMAT !! Kering Kent	1,6RB Terjual	KAB. GUNUNG KIDUL				56.00
TERHEMATY III (COD)Keri	4,5RB Terjual	KAB. GUNUNG KIDUL	1	1.00		
			0	20	40	60
					Avg. Harga	

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

Rhn	Mkn-K	P-sour	co1
DIIII	1014/11-17	r-suur	CET

Nama Produk	Terjual	Wilayah							
Geblek Kulon Progo 03	2R8 Terjual	KAB. KULON PROGO	10.0	0					
Teh Obat Herbal Menyem		KAB. KULON PROGO						98	50
		0	20		40	60	80	100	120
					А	vg. Harg	a		



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

BalipiaKulius-Source1

	17				7000	1,528 1,528 1,528 1,528	Organi	216	120	Kong Private Kong Private Kong Private Private All Private Private Particular Private Pri
Harrish Probab Release Release Tongging Cough French Stationer Raine, The Proof to	111111	111110	0.0000	91,849,000	9,0009					 Respire Annual Topic Huger Huger Tomas Tarat Respire Topics Topic Huger Dynamic Database Minist Facel
BARMA KARATTA KARATTA (D. 1994) Matalak Kalono Traga Janga Brownian Datarta - Hind Papa Manga Kalono Traga Janga Di Yinga Katalan Matang Mahan Kaloni Progo Janga Hagam (Prawa) Pash		3						•		Temper Facher Frage anges Fragman Konserny Frager (Facher) Facher Frager (Facher) (Specific Frager (Facher)) Facher Facher (Facher) (Specific Frager (Facher)) Facher Facher (Facher)
Annual Anime Taga, taga Dingkot Chanan, Man Pare Bahara Anima Taga, taga Dingkot Koomp Pijan, Travel Part Bahara Anima, Taga, taga Dingkot Davimer Tagan Bahara Anima, Taga, taga Dingkot Davimer Tagan Mat	•									

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



Submitted : March 7, 2023 Accepted : March 8, 2023 Published : March 8, 2023

Volume 5, Number 1, January 2023 https://doi.org/10.47709/cnahpc.v5i1.2184

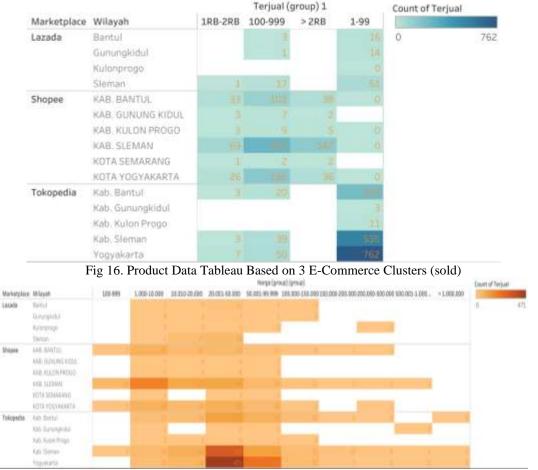


Fig 17. Product Data Tableau Based on 3 E-Commerce Clusters (price)

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



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). <u>http://dx.doi.org/10.28989/senatik.v7i1.456</u>
- 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 Nglanggeran. In Proceeding of International Conference on Science, Health, And Technology(pp. 332-336).
- Djufri, M. (2020). Penerapan Teknik Web Scraping Untuk Penggalian Potensi Pajak (Studi Kasus Pada Online Market Place Tokopedia, Shopee Dan Bukalapak). *Jurnal BPPK : Badan Pendidikan Dan Pelatihan Keuangan*, *13*(2), 65–75. <u>https://doi.org/10.48108/jurnalbppk.v13i2.636</u>
- 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. <u>https://doi.org/10.28932/jutisi.v5i3.1977</u>
- Henrys, 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.



Volume 5, Number 1, January 2023 https://doi.org/10.47709/cnahpc.v5i1.2184

- 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. <u>https://doi.org/10.34123/semnasoffstat.v2021i1.1009</u>
- 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. <u>https://doi.org/10.28989/angkasa.v13i1.807</u>
- 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. <u>https://doi.org/10.33365/jti.v14i1.498</u>
- 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.
- Yusroni, 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.

