Available online at https://ejournal.corespub.com/index.php/ijmsc/index

International Journal of Mathematics, Statistics, and Computing

e-ISSN 3025-0803

Vol. 2, No. 2, pp. 76-84, 2024

Analysis and Visualization of Advertising Sales Data Using Python Software Through an Internship Program

Muhammad Naufal Ardiansyah[1*], Kalfin[2]

[1,2] *Statistics Study Program, Faculty of Science Technology and Mathematics, Matana University, Banten 15810, Indonesia*

**Corresponding author email: muhammad.ardiansyah@student.matanauniversity.ac.id*

Abstract

This research discusses the use of data visualization as a tool for analyzing and presenting information effectively. The main goal of this research is visualization that allows better understanding of complex data. In this context, research explores various data visualization techniques, including the use of bar charts, correlation heatmaps, and interactive technology to present information intuitively. In improving decision-making abilities, identifying patterns or trends, and exploring relationships and insights contained in large and heterogeneous datasets. The research methodology includes comparative analysis of various visualization methods, user experiments, and the application of new techniques to evaluate the effectiveness and usefulness of data visualization in advertising video sales data at PT XXX. Based on the research results, it was found that data visualization involves presenting data information in graphic or image form to facilitate understanding. This helps in explaining the facts and determining the steps that need to be taken. The research results are expected to provide in-depth insight into the development of data visualization techniques. The results of this data visualization can be widely applied in various fields, especially creative, marketing and sales management.

Keywords: Data visualization, Bar charts, Patterns, Trends

1. Introduction

As a form of practical learning in the world of work, internships play an important role in combining theory and direct experience. Internships reflect the needs of the ever-evolving labor market, the desire to develop practical skills, and efforts to produce work-ready graduates (Rothschild, et.al., 2020; Hora, et.al., 2020). In an ever-changing business environment, practical experience is becoming increasingly important. Internships provide students with the opportunity to apply theoretical knowledge gained at school to real situations in the workplace. Due to a lack of work experience, graduates often face difficulties when entering the job market. Internships are a solution that bridges this gap by providing direct access to the world of work and helping students and trainees develop relevant skills. Competition in the labor market is getting tighter (Tarmazi, et.al., 2017; Dewi, et.al., 2018). Employers typically look for candidates who not only have an academic background, but also work experience that can demonstrate the ability to adapt and apply concepts to everyday work situations. Rapid changes in business dynamics require graduates who are able to adapt quickly. Internships help students develop the skills of adaptability, problem solving, and initiative, which are important qualities in a dynamic work environment. The campus internship program is one of the main initiatives that connects the academic world and the world of work (Putera, et.al., 2020). The internship program planned by the campus reflects efforts to enhance students' learning experience, equip them with practical skills, and motivate students to face the challenges of the world of work. The internship program planned by the campus aims to adapt formal training to industry needs and requirements. Internships allow students to combine academic knowledge with practical application in the workplace, thereby increasing the relevance of the curriculum to the ever-changing world of work.

In support of the study program chosen during college, namely Statistics, the position obtained during the internship at PT XXX was Data Entry and Analysis Section Intern. Tasked with analyzing, monitoring, making reports, checking everything related to advertising video references, which are obtained through the creative team to increase sales value. The research was carried out as a form of understanding the theories that have been studied so far in campus learning, as well as the application of learning theories in real work environments. Therefore, this research focuses on the analysis and visualization of advertising sales data. Where, using a dataset from advertising video sales. The dataset was then prepared for analysis and visualization using the Python platform (Raschka, et.al., 2020; Jalolov,

2023). Data visualization involves presenting data information in the form of graphics or images to facilitate understanding. This helps in explaining the facts and determining the steps that need to be taken. The benefits can be felt in various fields of study that require new ways of presenting large and complex information (Cao, et.al., 2021). Visualization is the use of computer-supported visual representations of data. Interactive data visualization allows users to determine the format used to display data. The center of the data visualization process is the transformation of data into information. Understanding the difference between the concepts of data and information, as well as the relationship between the two, is very important for understanding the data visualization process (Li, 2020; Qin, et.al., 2020; Dai, et. al., 2022). Most visualization designs aim to aid decision making and act as tools that enhance cognitive processes. When designing and prototyping data visualizations, it is important to consider how the visualization will be used. Data visualization is more than just the representation of numbers, it involves selecting and re-analyzing the numbers that form the basis of the visualization (Sadiku, et.al., 2016). The visualizations produced in the research are in the form of bar charts, scatter plots and correlation heatmaps. Based on the description of the problem, this research aims to analyze and visualize advertising sales data at PT XXX. This visualization provides the information the company needs to increase sales.

2. Materials and Methods

2.1. Materials

In conducting data analysis, the author used data from sales of reference video advertisements at PT XXX. The data used is in the form of advertising video references obtained from social media such as Facebook and Instagram. Then the next data uses data on sales of advertising videos that have been produced. In processing data, we use Python software to analyze and visualize datasets into information that suits the company's needs.

2.2. Methods

This research uses qualitative methods in data analysis. This method focuses on in-depth observation of references for making and selling advertising videos. The use of this qualitative method is intended to examine a more comprehensive reference phenomenon of making and selling advertising videos. This is done to provide detailed information on the data used in accordance with the company's needs.

3. Results and Discussion

3.1. Preparation of Data to be Analyzed

The first step before carrying out data analysis is to prepare the data itself. Here, the author prepares data using Python and explains it in detail so that it is easy to understand. The first data preparation carried out by the author was to import the data into Python using the coding in Figure 1.

```
import pandas as pd
Content = pd.read_excel('/content/drive/MyDrive/Magang.xlsx')
Content
```

	Platform	Submit Date	Category	Duration	Posted	Objective	Sales Data	Purchased	Price
0	Instagram	2023-11-30	Properti	15s-60s	Tesar	Others	Property Hack	Terjual	250000
1	Instagram	2023-11-30	Properti	>60s	Tesar	Others	Property Hack	Terjual	300000
2	Instagram	2023-11-30	Properti	15s-60s	Tesar	Others	Property Hack	Terjual	300000
3	Instagram	2023-11-30	Properti	>60s	Tesar	Others	Property Hack	Terjual	250000
4	Facbook	2023-11-30	Kursus	<15s	Sendy	Others	Gold Hack	Terjual	400000
5	Facbook	2023-11-30	Properti	15s-60s	Sendy	Others	Gold Hack	Terjual	200000
6	Facbook	2023-11-30	Kuliner	>60s	Sendy	Others	Kuliner Mastery	Terjual	500000
7	Facbook	2023-11-30	Trading	15s-60s	Sendy	Others	Buku Jago Trading	Terjual	600000
8	Facbook	2023-11-30	Trading	15s-60s	Sendy	Others	Buku Jago Trading	Terjual	700000
9	Facbook	2023-11-30	Properti	15s-60s	Sendy	Others	Property Hack	Terjual	400000
10	Facbook	2023-11-30	Kursus	15s-60s	Sendy	Others	Lelang Mastery	Terjual	450000
11	Facbook	2023-11-30	Properti	15s-60s	Sendy	Others	Lelang Mastery	Terjual	750000

Figure 1: Import Data into Python

In Figure 1 it can be seen that the data import process has been successfully carried out. The next step is to find information from the data you have. This aims to see the type of data you have and also to check the total number of rows in the data.

```
df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 49 entries, 0 to 48
Data columns (total 9 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   Platform     49 non-null     object
 1   Submit Date  49 non-null     datetime64[ns]
 2   Category     49 non-null     object
 3   Duration     49 non-null     object
 4   Posted       49 non-null     object
 5   Objective    49 non-null     object
 6   Sales Data   49 non-null     object
 7   Purchased    49 non-null     object
 8   Price        49 non-null     int64
dtypes: datetime64[ns](1), int64(1), object(7)
memory usage: 3.6+ KB
```

Figure 2: Description of Research Data

Furthermore, in Figure 2 above is a description of the research data input into Python by the author. In the data description, it can be seen that almost all of the data types are object data types, except for sales data, which means that the research will focus on qualitative research. Then, from the description of the data created, the author analyzes whether there is missing or damaged data from the data set used. The results of the analysis can be seen in Figure 3.

```
df.isnull().sum()

Platform       0
Submit Date    0
Category       0
Duration       0
Posted         0
Objective      0
Sales Data     0
Purchased      0
Price          0
dtype: int64
```

Figure 3: Dataset analysis results show that the dataset is in good condition

The analysis process continues because the dataset is in good condition, where the next step is to carry out analysis on the advertising video sales data variable, where this data is the only research data which is the int64 data type. Where the analysis carried out is an analysis of calculating the amount of data, mean, standard deviation, minimum and maximum values, and values in quartiles 1, 2, and 3. The results of the analysis can be seen in Figure 4.

```
df.describe()
```

	Price
count	49.000000
mean	396428.571429
std	219924.229376
min	0.000000
25%	250000.000000
50%	400000.000000
75%	550000.000000
max	800000.000000

Figure 4 : Quantitative Data Analysis of Advertising Video Sales

After the entire process of preparation and analysis of the dataset has been carried out, the next research step is to visualize the data from the research dataset used.

3.2. Data Visualization Results from Reference Data and Advertising Video Sales

The data visualization process was carried out using Python software using all the datasets that had been prepared. Data visualization was carried out first for data on video advertisement reference uploaders. Where the data visualization shows the highest number of video uploaders, the video uploaders in question are creative staff and HRD staff, namely Tesar, Sendy, Revin, Joey, and Jeremy. In the data visualization, it can be seen that Sendy is the staff who uploaded the most advertising video references with 18 videos, followed by Revin with 12 videos, then Joey with 8 videos, then Jeremy with 7 videos, and finally Tesar with 4 videos. This shows that Sendy is the most productive staff in finding references for advertising videos that are made and then sold. Where, Sendy is one of the creative team staff. Data visualization of video advertisement reference uploader data can be seen in Figure 5.

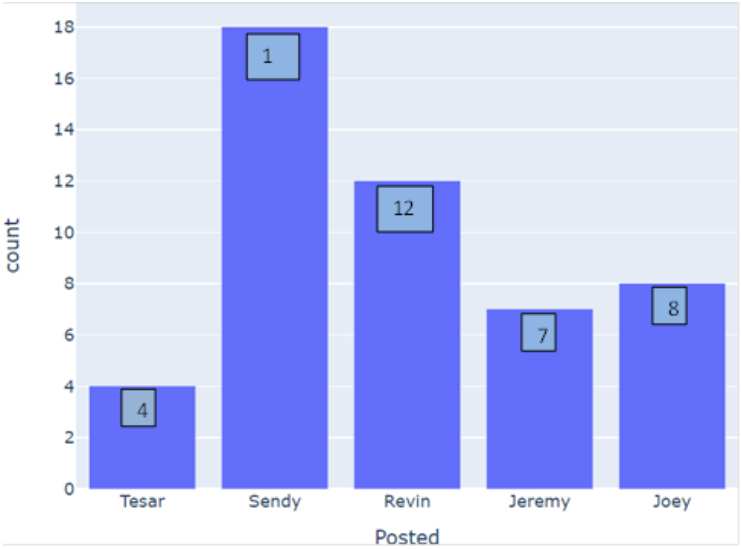

Figure 5 : Data Visualization Results for Ad Reference Video Uploaders

Next, the data that is visualized is the platform data from which the advertising video reference was obtained. There are only two platforms in the data, namely Instagram and Facebook. According to the results of the data visualization, Facebook is the platform that has the most references for advertising videos, namely 29 advertising reference videos

taken from Facebook. Instagram contributed 20 advertising videos which were taken as references for advertising videos. Facebook is the main destination for taking advertising video references where this platform has lots of video references that are useful in creating new advertising videos. Even though Instagram is a platform that is more popular among young people, in the data visualization results it can be seen that Facebook contributes more video references that are better than Instagram. The results of the platform data visualization can be seen in Figure 6.

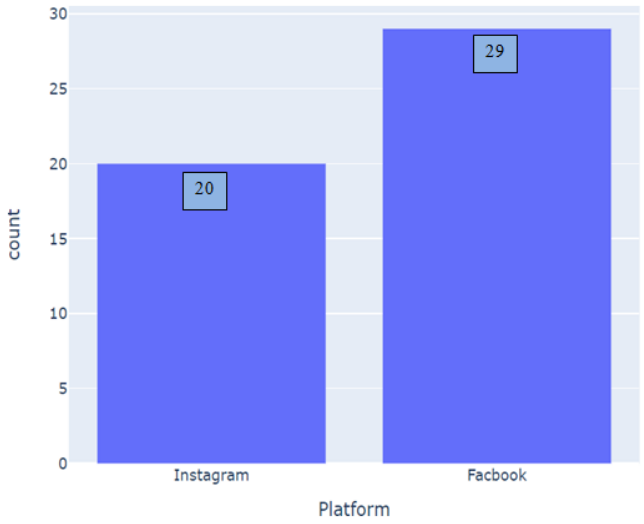

Figure 6 : Data Visualization of Advertising Video Referral Platform Data

The next data that is visualized is data from the company to which the advertising video created by the creative team is sold. The main companies aimed at selling video advertisements include Property Hack, Gold Hack, Culinary Mastery, Buku Jago Trading, and Auction Master. Visualization of sales data shows that the Buku Jago Trading and Auction Master companies are the companies with the highest sales level of 15 videos sold, followed by Property Hack which has a sales level of 14 videos, Kuliner Mastery has a sales level of 3 videos, and finally the company Gold Hack a total of 2 videos. This shows that sales of video advertisements are highest at the Buku Jago Trading and Auction Master companies. The results of the data visualization can be seen in Figure 7.

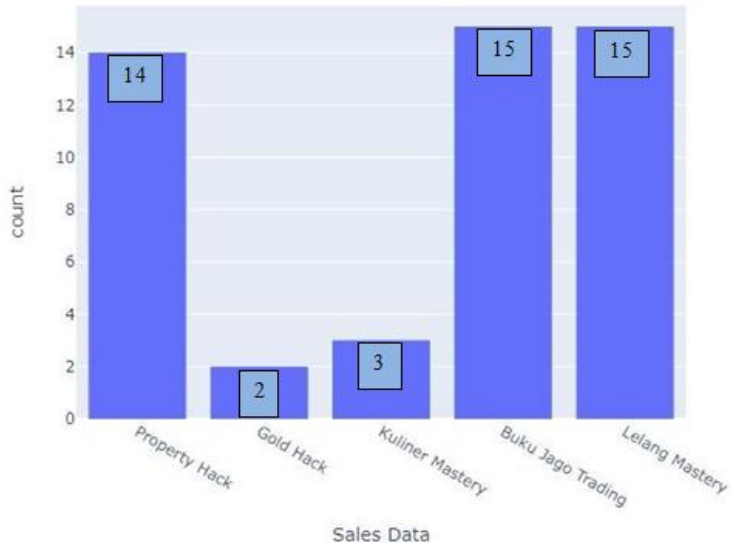

Figure 7 : Visualization of Sales Data on Target Companies

Next, the data that is visualized is the sales price data for advertising videos. Where sales price data tends to vary starting with the price ranges presented in the plot being IDR 0 – 90,000, IDR 100,000 – 190,000, IDR 200,000 – 290,000, IDR 300,000 – 390,000, IDR 400,000 – 490,000, IDR 500,000 – 59 0,000, IDR 600,000 – 690,000, and IDR 800,000 – 890,000. In the data, there are advertising videos that are sold for IDR 0 rupiah, meaning that the videos are not sold, so the data visualization plot of IDR 0 – 90,000 shows the number of unsold advertising videos, namely 6 videos. Then in the IDR 100,000 – 190,000 range there was 1 video sold, then at IDR 200,000 – 290,000 there were 6 videos sold, at the price of IDR 300,000 – 390,000 8 videos were sold, then in the IDR 400,000 – 490,000 price range the highest video sales price data was, namely a total of 12 videos, then in the IDR 500,000 – 590,000 range 7 videos

were sold, and for the IDR 600,000 – 690,000 and IDR 700,000 – 790,000 ranges there were the same number of videos sold, namely 3 videos. It can be concluded that the creative team sells many advertising videos to target companies in the price range of IDR 400,000 – 490,000. The difference in selling prices for advertising videos is based on the quality of the video, the idea and concept of the video, and the duration of the advertising video to be sold. The results of data visualization for the sales price of advertising videos can be seen in Figure 8.

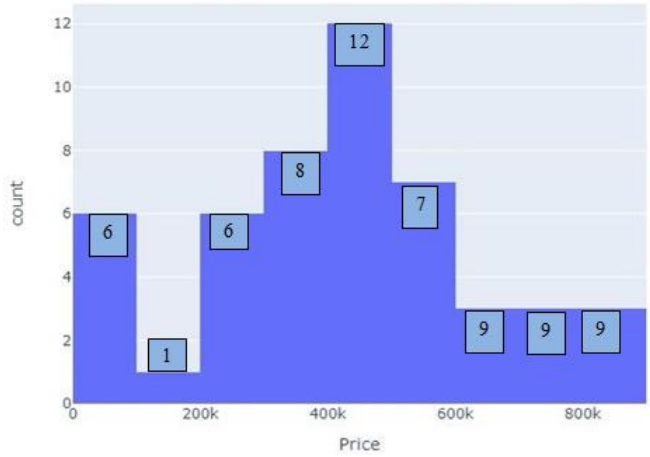

Figure 8 : Data Visualization Results of Advertising Video Sales Prices

The next data visualization is a scatter plot or distribution plot of research data which will be connected to each other. Where this research links data from the target sales company and the price at which the advertising video was sold at the target company. The scatter plot results show that the Property Hack company has a purchasing range for advertising videos at prices below IDR 600,000. Furthermore, the Gold Hack company has a video purchase range at a maximum price of IDR 400,000. The Culinary Mastery company has a maximum video purchase price range of IDR 500,000. The Jago Trading Book Company with the highest spread, has a purchasing range for advertising videos with the highest maximum price, namely IDR 800,000. Then finally, the Mastery Auction company has a maximum purchase price range of IDR 750,000. This shows a correlation between the sales target company data and the sales price data from the advertising video, the data visualization of which can be seen in Figure 9.

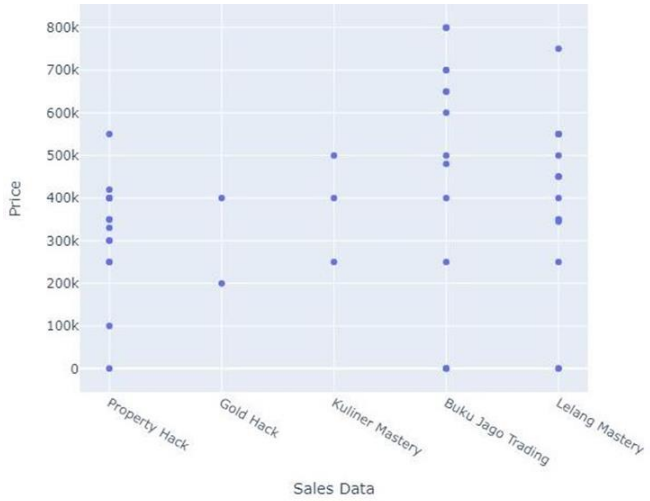

Figure 9 : Correlation Data Visualization Results using scatter plots for Target Company Data and Video Advertising Sales Prices

Furthermore, data visualization was also carried out to see the correlation between the data of the staff uploading the reference video and the sales price of the advertising video made based on the reference video. The plot shows that the distribution of reference video uploaders and the largest advertising video creators is Sendy with a total of 18 videos made, and a sales price range with a maximum of IDR 800,000. Then, there was Revin who uploaded 12 advertising videos with a maximum sales price range of IDR 500,000. Next, there is Joey who uploaded 8 videos with a maximum sales price range of IDR 800,000. Jeremy uploaded 7 videos with a maximum sales price range of IDR 550,000. Lastly, there is Tesar who uploaded 4 videos with a maximum sales price range of IDR 300,000. In the scatter plot it can also be seen that Sendy produces advertising videos with the highest sales and price ranges, so it can

be concluded that Sendy has a large contribution to sales of advertising videos. The results of the data visualization can be seen in Figure 10.

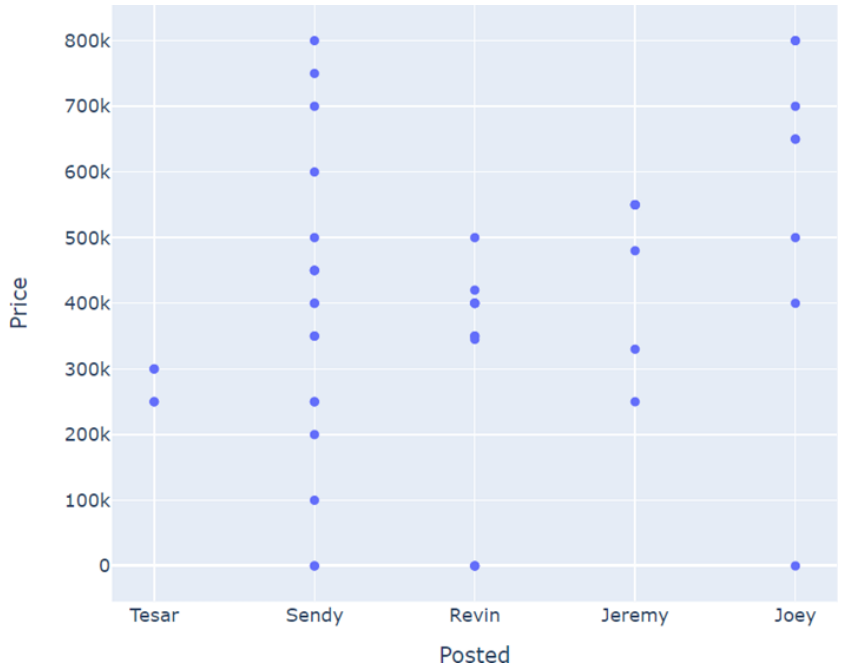

Figure 10 : Scatter plot visualization of the correlation between uploaders and sales prices of video advertisements made

Furthermore, data visualization is also carried out to see the correlation between data variables, whether there is a relationship between one variable and other variables. This correlation is explained in Correlation Heatmap in Python, where Correlation Heatmap is a plot that shows a discrete 2D correlation matrix that uses colored cells to represent data from a monochromatic scale (Rupji, et.al., 2019; Tiessen, et.al., 2017; Buyrukoğlu, et.al., 2022). The first dimension values appear as table rows while the second dimension appears as columns. The cell color is proportional to the number of measurements corresponding to the dimension value. The dataset used in the research is almost entirely object data, where carrying out testing requires changing the object dataset to an int64 dataset. Therefore, converting object data into int64 data uses coding. The results of the int64 data created for each dataset are then poured into a Correlation Heatmap which shows the correlation for each variable. In the Correlation Heatmap there are various colors in each column, where each color has a different value. The brighter the color in the column has a value closer to 1, which means the correlation shows a perfect positive relationship where the two variables move together linearly. Slightly darker colors have a value close to 0, which means the correlation shows that there is no significant linear relationship between the two variables. And if the column has a dark color, it has a value close to -1, which means the correlation shows a perfect negative relationship where one variable increases as the other variable decreases.

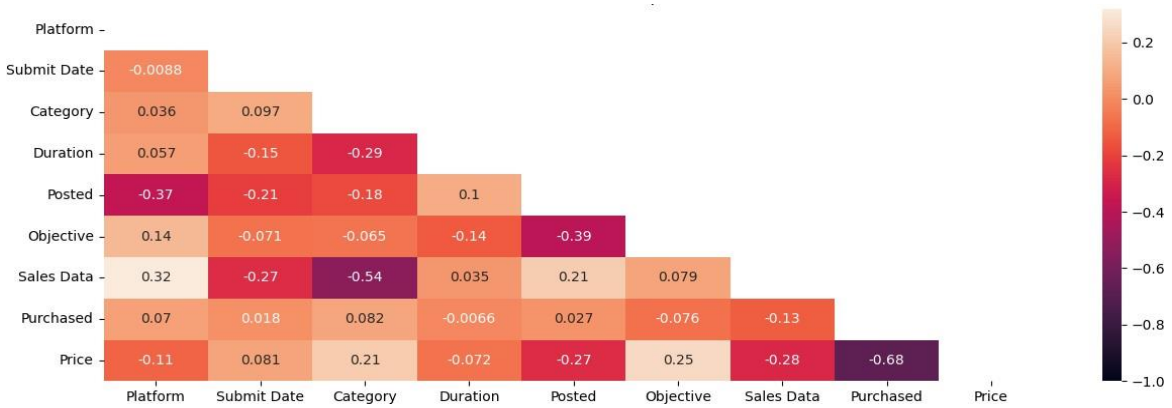

Figure 11 : Correlation Heatmap

For example, in the Correlation Heatmap shown in Figure 11, the brightest column color is shown in the correlation between the Sales Data and Platform variables which has a value of 0.32, where it can be concluded that the

correlation between these two variables looks weak, but is still positive. Meanwhile, the Purchased and Price variables in the column have the darkest color and have a correlation value of -0.68, which means that there is a fairly strong negative relationship between these two variables. A negative correlation value indicates that one variable tends to increase while the other variable tends to decrease. In this case, the higher the value of one variable, the lower the value of the other variable, and vice versa. And the Purchased and Submit Date variables have a value of 0.018, which means the correlation value shows that the linear relationship between these two variables is very weak. A correlation value close to zero indicates that there is no significant linear relationship between these variables.

4. Conclusion

Data visualization involves presenting data information in the form of graphics or images to facilitate understanding. This helps in explaining the facts and determining the steps that need to be taken. The benefits can be felt in various fields of study that require new ways of presenting large and complex information. Visualization is the use of computer-supported visual representations of data. Interactive data visualization allows users to determine the format used to display data. The center of the data visualization process is the transformation of data into information. Understanding the difference between the concepts of data and information, as well as the relationship between the two, is critical to understanding the data visualization process. The results of data visualization show that data visualization can make it easier to draw conclusions and analyze data. Based on the data visualization carried out, several things can be concluded from the dataset used in this research. The largest uploader of advertising video references according to the data is Sendy with 18 uploaded reference videos. The advertising video reference platform that is most often used as a reference is Facebook with 29 reference videos taken. Furthermore, the largest sales company target data is the Jago Trading Book and Master Auction companies with a total of 15 videos. Then, the price range for most advertising video sales is in the IDR 400,000 – 490,000 price range, with a mean sales price of IDR 396,428. Visualization of correlation data between Sales Data (sales target company) and Price (sales price of advertising videos), shows that Buku Jago Trading has the highest maximum sales price range at IDR 800,000. Furthermore, the correlation between uploaders and the sales price of advertising videos shows that Sendy has the highest number of sales with the highest maximum sales price, namely IDR 800,000. Finally, data visualization also produces a correlation heatmap which shows that the variable from the research data that has the highest positive correlation value is Sales Data with Platform even though the correlation is still relatively weak. On the other hand, the variables whose correlation has the highest negative value which is quite strong are Price and Purchased.

References

Cao, S., Zeng, Y., Yang, S., & Cao, S. (2021). Research on Python data visualization technology. In *Journal of Physics: Conference Series* (Vol. 1757, No. 1, p. 012122). IOP Publishing.

Li Q. (2020). Overview of Data Visualization. *Embodying Data: Chinese Aesthetics, Interactive Visualization and Gaming Technologies*, 17–47.

Sadiku, M.N.O. (2016) 'Data Visualization', *International Journal of Engineering Research And Advanced Technology(IJERAT), 02*(12), pp. 11–16.

Rupji, M., Dwivedi, B., & Kowalski, J. (2019). NOJAH: NOt Just Another Heatmap for genome-wide cluster analysis. *PloS one, 14*(3), e0204542. 1-14.

Tiessen, A., Cubedo-Ruiz, E. A., & Winkler, R. (2017). Improved representation of biological information by using correlation as distance function for heatmap cluster analysis. *American Journal of Plant Sciences, 8*(3), 502-516.

Buyrukoğlu, S., & Akbaş, A. Y. H. A. N. (2022). Machine Learning based early prediction of type 2 diabetes: a new hybrid feature selection approach using correlation matrix with heatmap and SFS. *Balkan Journal of Electrical and Computer Engineering, 10*(2), 110-117.

Tarmazi, S. A. A., Jumain, M. A. M., Idris, A. K. M., & Tan, Z. M. A. (2017). Internship program: A study on students' satisfaction. *Journal of Tourism, Hospitality & Culinary Arts, 9*(2), 371-382.

Dewi, L. R., & Kartowagiran, B. (2018). An evaluation of internship program by using Kirkpatrick evaluation model. *REID (Research and Evaluation in Education), 4*(2), 155-163.

Putera, L. J., & Sugianto, R. (2020). Perception and optimism about two-semester off-campus internship program of the kampus merdeka-merdeka belajar (freedom campus-freedom to learn) policy among university students. *Journal of Languages and Language Teaching, 8*(3), 264-275.

Qin, X., Luo, Y., Tang, N., & Li, G. (2020). Making data visualization more efficient and effective: a survey. *The VLDB Journal, 29*, 93-117.

Dai, C., Jiang, Y., Yin, C., Su, R., Zeng, X., Zou, Q., ... & Wei, L. (2022). scIMC: a platform for benchmarking comparison and visualization analysis of scRNA-seq data imputation methods. *Nucleic Acids Research*, *50*(9), 4877-4899.

Raschka, S., Patterson, J., & Nolet, C. (2020). Machine learning in python: Main developments and technology trends in data science, machine learning, and artificial intelligence. *Information*, *11*(4), 193.

Rothschild, P. C., & Rothschild, C. L. (2020). The unpaid internship: Benefits, drawbacks, and legal issues. *Administrative Issues Journal*, *10*(2), 5.

Hora, M. T., Parrott, E., & Her, P. (2020). How do students conceptualise the college internship experience? Towards a student-centred approach to designing and implementing internships. *Journal of Education and Work*, *33*(1), 48-66.

Jalolov, T. S. (2023). TEACHING THE BASICS OF PYTHON PROGRAMMING. *International Multidisciplinary Journal for Research & Development*, *10*(11).