Available online at:
https://jurnal.integrasisainsmedia.co.id/index.php/JIMS
Journal Integration of Management Studies
Volume 1 Number 2, 2023:176-183
DOI: 10.58229/jims.v1i2.97

Unveiling Post Engagement Rate Using Predictive Analytics Model for Instagram Accounts of Coffee Shops in West Jakarta

Sachio Dariell Sutanto[*1]
School of Business and Management, Institut Teknologi Bandung[1,2]
Email: sachio_dariell@sbm-itb.ac.id

Abstract

The rapid growth of the coffee shop industry is currently happening in Indonesia. The rapid growth induces fierce competition between the two, and one of the areas in Indonesia is affected by the growing number of coffee shops in West Jakarta. When competing, a coffee shop must establish a positive relationship with its customers to promote its products. One of the popular methods of establishing relationships with customers is by using Instagram. The social interaction between the coffee shop and its customers happens through liking or commenting on Instagram posts. The application will then quantify those interactions on Instagram by the number of likes and comments on the posts. Then, using both quantified metrics, with the addition of the count of followers of the accounts, we are intrigued to create predictive analytics for the engagement metrics in Instagram to predict the engagement rate of future posts. In this study, we will analyze the Instagram engagement metrics, likes, comments, and followers to predict the engagement rate on an Instagram post. With multiple linear regression, we developed a model to forecast the outcome of future Instagram posts. The prediction variables include the factors we can input when posting on Instagram, such as captions, location, audio, hashtags, and post type. Although we have analyzed several variables that we may modify when posting on Instagram, we found that not all variables significantly impact the engagement rate. Only two variables significantly impact the predicted value of an Instagram post. Consequently, this study is a foundation for predicting engagement rates of coffee shop Instagram posts, helping coffee shop owners create a plan for their Instagram posts.

Keywords: Predictive analytics, Coffee shop, Instagram engagement rate, Multiple linear regression

A. INTRODUCTION

Over the past decade, there has been a new and emerging industry that gradually changes people's behavior into consuming more and more of their products and services: the Coffee Shop Industry. The global coffee shop industry has grown rapidly and seems to continue growing, especially in recent years when the COVID-19 pandemic has become endemic. The rapid growth of this industry creates fierce competition that forces coffee shop owners to constantly seek innovation or strategies to differentiate themselves from their competitors, allowing them to attract loyal customers and drive sustainable growth. This global trend is not limited to certain regions; we can see a similar phenomenon in countries like Indonesia.

The number of coffees consumed by Indonesian people is increasing yearly, even when the country is still affected by the COVID-19 pandemic, as reflected in coffee consumption from 2019/2020 to 2020/2021. At the same time, the number of coffee shops in Indonesia also grew from 1083 in 2016 to 2937 in 2019 (Toffin and Mix, 2020:9). The growing coffee consumption and the flourishing coffee shop industry in Indonesia have had a noticeable impact on various regions, and West Jakarta is one of the areas that has experienced the rapid growth of coffee shops. West Jakarta was quite famous for several colonial relics that were still preserved and maintained by the government, attracting people from outside West Jakarta to visit this area. In addition to the colonial relics, many private universities in Jakarta are in West Jakarta, e.g., Bina Nusantara University, Trisakti University, Universitas Kristen Krida Wacana, and Tarumanegara University. These universities in West Jakarta have the biggest student density at 5252 students/km2 (BPS, 2021).

The tourist attractions and universities allow coffee shops to be built in West Jakarta to accommodate the need for places to rest and hang out with friends or families. As coffee shops emerge in this thriving environment, it becomes essential to recognize that a coffee shop venture's successful establishment and growth go beyond physical infrastructure alone. Building a coffee shop includes marketing to create a consumer base and grow the

* Corresponding author

business. When we want to grow a business, one of the most frequently used methods to market its business is by using social media.

Social media is an internet-based channel of mass personal communication that facilitates interactions between users that have value primarily from user-generated content (Carr and Hayes, 2015:49), and it has changed how information spreads around the world. Social media is now the top choice for businesses to engage with their customers and consumer base. It is also the best media to boost the number of customers that come to their coffee shops by creating social media campaigns. Social media offers unparalleled opportunities to boost the number of customers who visit coffee shops. With a well-executed social media campaign, a coffee shop can leverage the power of viral marketing, potentially attracting dozens of customers to the coffee shop. However, attracting customers on social media is more complex than it may seem, so the posts should be visually appealing to the netizens who see them. Using social media correctly could be a fantastic tool to generate customer engagement and brand awareness for all kinds of businesses (Sledge, 2014). However, it is important to note that simply posting what we believe is the best may not align with what consumers think is attractive. Nowadays, companies need to have the ability to understand the targeted customers' value and the reason for their attraction to the company (Thompson, 1998:21) to create a better social media interaction that aligns our product with our customers' value.

In selecting specific social media platforms, coffee shops must consider several factors that contribute to effective customer engagement, and two of the main factors are the number of users and the engagement rate of the platform. In January 2023, the number of social media users in Indonesia was 167.0 million, equivalent to 60.4% of the total population (DataReportal, 2023). In Indonesia, WhatsApp is the most used social media platform, followed by Instagram and Facebook (DataReportal, 2023). Since WhatsApp is not a platform we can use to engage with followers and strangers on the internet, this research decided to use Instagram because of its ability to interact with followers, with the possibility of counting the engagement rate for each post. The other reason we chose Instagram is that the average engagement rate of all posts is better on Instagram than on Facebook.

Engagement Rate is a way to measure how much influence an Instagram account has on its followers. It is one of the metrics we can use to know how to optimize an endorsement on Instagram influencers (Arman and Sidik, 2019:5). The variable engagement rate will be crucial for coffee shops to understand whether their Instagram post is doing great in giving exposure to the coffee shop or not. Instagram is one of the world's most used social media platforms, reaching 1.628 billion users on advertising reach in April 2023 (DataReportal, 2023). We can use this platform to capture, edit, and share photos or videos as Instagram posts. The post may accompany captions, hashtags, location tags, and other features. These companions on the post are the data that we can collect to predict the engagement rate for an Instagram post.

In this research, we want to understand the interaction between the variables. Understanding the interaction means we can predict what type of Instagram posts could generate the most engagement with the consumer base. If we can predict the most effective post for Instagram's engagement rate, we can create an effective and efficient Instagram marketing plan. To understand the relationship between the variables, we will use linear regression, with a little tweak to accommodate the independent variables, which are mostly categorical. Linear regression can be defined as a formula or a method to understand the correlation of a linear relationship between dependent and independent variables (Groß, 2003:4). The regression method uses a continuous variable as the dependent variable, and the independent variable may vary between categorical or continuous variables.

The information from social media can be used to forecast future outcomes, specifically by using the Linear Regression model (Asur and Huberman, 2010:499). We can use linear regression to create a predictive model for social media posts. In that case, we can forecast the outcomes of our engagement rate, creating the possibility of a better Instagram marketing plan using the given variables. This research develops a linear regression predictive analytics model to predict the expected engagement rate of instagram posts using RStudio. RStudio is an application that could be used for R programming, typically for regression or other machine learning techniques. RStudio uses all functions available in R, and it uses a more user-friendly interface to allow the users to analyze data comfortably and generate the graphical outcome for the data (Kronthaler and Zöllner, 2021:1-2).

This research aims to understand the interaction between the engagement rate and other observable variables. Suppose we can understand the correlation between those variables. In that case, we can predict what type of Instagram posts could generate the most engagement with the consumer base.

B. RESEARCH METHODS

In this study, we aim to investigate the impact of independent variables on the value of engagement rate on the Instagram posts of several West Jakarta Coffee Shop's Instagram Accounts. Based on our theoretical exploration, we portray the following theoretical conceptual model in Figure 1.

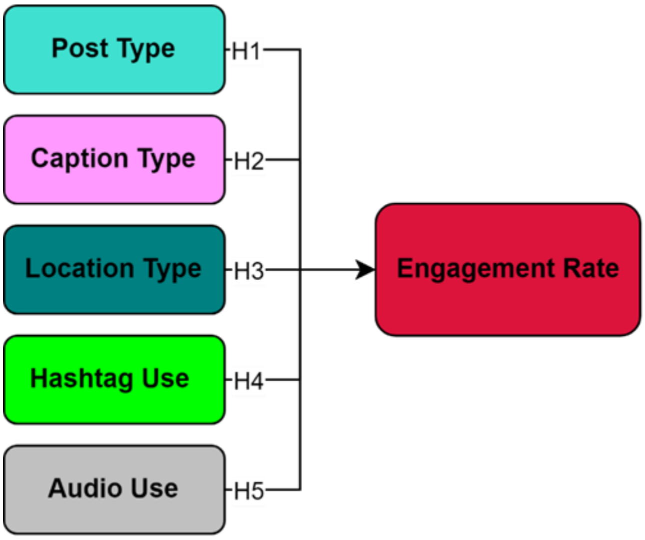

Figure 1. Theoretical Framework

In line with the theoretical conceptual model definition, we propose the following hypotheses to guide our research process and address the objectives of this research:
H1: Post Content Type has a significant impact on the Engagement Rate.
H2: Caption Type has a significant impact on the Engagement Rate.
H3: Location type has a significant impact on the engagement rate.
H4: Hashtag use has a significant impact on the engagement rate.
H5: Audio use has a significant impact on the engagement rate.

In this research, we will use Web Data Extraction, a data extraction method that uses the data stored from a website source as the main source of data (Ferrara et al., 2014:301). We will use Apify to create a data extraction task on the Instagram Scraper API to do that. In this task, we should define the input parameter for the Instagram account data that will be collected. The input parameter here is the Instagram Account's URL, time frame, and max number of posts that will be extracted.

In collecting the data, we will use the time frame between the 1st of January 2022 and the 31st of December 2022. The maximum number of posts will be 100 posts per Instagram Account. After giving the parameters or limitations on the Apify Task, we will use the Start button to start the scraping task on the API. Then, after collecting the data for this research, we cleaned, transformed, and validated the data. The total amount of collected clean data is 417. In addition to collecting the data, there will be several manually categorized data: Post Type, Location Type, Audio Use, Caption Type, and the Engagement Rate calculated using the formula.

$$Engagement\ Rate = \frac{Likes + Comments}{Total\ Followers}$$

After processing the data, we generate a multiple linear regression model representing the relationship between the dependent and independent variables. We will divide the data into training and test datasets; then, the training dataset will generate an equation that will serve as a model for predictions or draw conclusions based on the independent variable. These models will be developed by using RStudio.

The division will be conducted by separating the collected data into 70% and 30% using a "set.seed" function to ensure that any iteration of the division will be the same when we clean the workspace in RStudio.

Then, we evaluate the predictive analytics model by using the equation to calculate predictions and testing it on the test dataset.

To create the prediction model, we will use the function "lm()" from RStudio to create the linear regression model using the training dataset using the following code:

linearregression <- lm(engagement_rate ~ caption_type + hashtag_use + location_type + post_type + audio_use, data = trainSet)

By executing the defined code, Rstudio has created a linear regression equation, represented by the variable "linearregression". To further comprehend the equation, we will use the function "summary()" with "linear regression" as the variable for the function to return the result of the regression and other statistical insights.

Then, to validate the reliability of the machine learning model, we will use RMSE and MAE as the metrics to measure the model's reliability. Both can be done in RStudio by using the respective function of "mean((modelEval$Actual)- (modelEval$Predicted)^2)" for predicting RMSE, and we can then use the function "mean(abs(modelEval$Actual - modelEval$Predicted))" for predicting MAE.

RMSE is a measure to understand the overall difference between the predicted and actual values, indicating how well the prediction aligns with the actual data. A lower RMSE means the model is quite accurate since the overall difference between the predicted value and the actual value is low.

MAE is a measure of understanding the average absolute difference between a dataset's predicted and actual values. MAE indicates the average error of the prediction model in comparison to the actual value from the dataset. Like the RMSE, the lower value MAE generates means better prediction accuracy.

The research aims to understand the interaction between the engagement rate and other observable variables using a linear regression prediction model. Suppose we can understand the correlations, which means we can predict what type of Instagram posts could be generated to maintain a high engagement rate. The generated model will then answer the hypotheses and research question, thus fulfilling the research objective.

C. RESULTS AND ANALYSIS

From the summary function in RStudio, we analyzed each factor from the analyzed variables using a training dataset to determine the significance of all variables, specifically each factor within the variables toward the dependent variable. In this research, we will use the significance level of 0.05 to understand the significance of each variable using the P Values of each factor to answer the proposed hypotheses. The equation summary shows that only two specific factors significantly impact the engagement rate value: an Endorsement Post from the Post Type variable and the City of the Coffee Shop from the Location Type variable.

To further understand the result of the linear regression equation from RStudio, the researchers converted the estimated value from the summary into a Multiple Linear Regression equation that can be seen in the formula below:

$$\begin{aligned} y = {} & 0.0137009x_1 + 0.0031128x_2 + 0.0042954\ x_3 + 0.0063035x_4 + 0.0042047x_5 + 0.0109943x_6 \\ & + 0.0071867x_7 + 0.0060929x_8 + 0.0020010x_9 + 0.0001449x_{10} + 0.0021084x_{11} \\ & + 0.0062277x_{12} + 0.0081009x_{13} - 0.0020412x_{14} + 0.0049077x_{15} + 0.0008522x_{16} \\ & + 0.0041713x_{17} - 0.0018726x_{18} - 0.0039268x_{19} - 0.0020377x_{20} \end{aligned}$$

Where:

y = Engagement Rate

x_1 = post_type1 (endorsement post)

x_2 = post_type2 (events within the coffee shop or celebration post)

x_3 = post_type3 (photo of the menus)

x_4 =post_type4 (ambiance of the coffee shop)

x_5 =post_type5 (discount or promotion poster)

x_6 = post_type6 (video about the coffee shop)

x_7 = post_type7 (video of the menus)
x_8 = post_type8 (video of events within the coffee shop)
x_9 = caption_type1(grateful or playful interaction with the customers)
x_{10} = caption_type2 (promoting purchase packages)
x_{11} = caption_type3 (promoting specific menus)
x_{12} = caption_type4 (promoting the ambiance)
x_{13} = caption_type5 (promoting the coffee shop or the coffee shop location)
x_{14} = caption_type6 (promoting or explaining the services)
x_{15} = caption_type7 (celebrating special day)
x_{16} = hashtag_use1 (presence of hashtag within the caption)
x_{17} = location_type1 (city of the coffee shop)
x_{18} = location_type2 (coffee shop location)
x_{19} = location_type3 (every other location)
x_{20} = audio_use1 (presence of audio)

In that equation, the coefficients represent the effect of the predictor variables on the dependent variable, the engagement rate. Since the regression variables are categorical, the coefficient value will only be triggered if the specific categorical value is present. We tweaked the equation to show the first variable for post type, endorsement post type, at the expense of the intercept value.

For example, when the caption is caption_type1, only the variable x1 has the value of 1; it increases the value of engagement_rate by 0.0020010. Meanwhile, all other caption_type variables have a 0 value, which does not affect the value of the Engagement Rate. Similarly, other variables contribute to the dependent variable based on their respective coefficients and other variables in the same categorical variable type.

From the same summary function, we can further analyze the significance of each factor of the variables by using the P Values of each respective variable in the equation, and then we reject or not reject each variable according to the given hypotheses. The result of analyzing the P Values can be seen in the table below.

Table 1. Hypotheses Testing

	Hypotheses	T Statistics	P Values	Result
$H_{1.1}$	Post Content Type 1 has a significant impact on the Engagement Rate.	3.395	0.001	Not Rejected
$H_{1.2}$	Post Content Type 2 has a significant impact on the Engagement Rate.	0.863	0.389	Rejected
$H_{1.3}$	Post Content Type 3 has a significant impact on the engagement rate.	1.321	0.187	Rejected
$H_{1.4}$	Post Content Type 4 has a significant impact on the Engagement Rate.	1.908	0.057	Rejected
$H_{1.5}$	Post Content Type 5 has a significant impact on the engagement rate.	1.040	0.299	Rejected
$H_{1.6}$	Post Content Type 6 has a significant impact on the Engagement Rate.	1.619	0.107	Rejected
$H_{1.7}$	Post Content Type 7 has a significant impact on the engagement rate.	1.110	0.268	Rejected
$H_{1.8}$	Post Content Type 8 has a significant impact on the engagement rate.	0.930	0.353	Rejected
$H_{2.1}$	There is a significant impact of Caption Type 1 on the Engagement Rate.	0.620	0.536	Rejected
$H_{2.2}$	There is a significant impact of Caption Type 2 on the Engagement Rate.	0.039	0.969	Rejected
$H_{2.3}$	There is a significant impact of Caption Type 3 on the Engagement Rate.	0.636	0.525	Rejected
$H_{2.4}$	There is a significant impact of Caption Type 4 on the Engagement Rate.	1.685	0.093	Rejected
$H_{2.5}$	There is a significant impact of Caption Type 5 on the Engagement Rate.	1.887	0.060	Rejected
$H_{2.6}$	There is a significant impact of Caption Type 6 on the Engagement Rate.	-0.480	0.632	Rejected
$H_{2.7}$	There is a significant impact of Caption Type 7 on the Engagement Rate.	1.407	0.161	Rejected
H_3	Hashtag Type has a significant impact on the engagement rate.	0.624	0.533	Rejected
$H_{4.1}$	There is a significant impact of Location Type 1 on the Engagement Rate.	2.150	0.032	Not Rejected
$H_{4.2}$	There is a significant impact of Location Type 2 on the Engagement Rate.	-1.742	0.083	Rejected
$H_{4.3}$	There is a significant impact of Location Type 3 on the Engagement Rate.	-1.360	0.175	Rejected
H_5	Audio type has a significant impact on the engagement rate.	-0.362	0.718	Rejected

Source: research data, 2023

From the equation, we will then evaluate the machine learning model; we will use RMSE and MAE as the evaluation metrics for the prediction performance.

Table 2. The Performance Evaluation Of The Machine Learning Model

Model	Predictors	Evaluation Metric	
		RMSE	MAE
Multiple Linear Regression	Post Content Type + Caption Type + Hashtag Type + Location Type + Audio Type	0.09779162	0.005533414

Source: research data, 2023

D. CONCLUSIONS

In this research, we concluded that not all variables significantly impact the engagement rate of an Instagram post. Although insignificant, the machine learning model may still use all variables from the equation to make predictions for the engagement rate of future Instagram posts. When we mention that not all variables have a significant impact, two factors significantly impact the engagement rate: Endorsement Post Type and City of the Coffee Shop Location Type. It is important to note that being significant here means the P values are lower than the level of significance, and the significance is tested using the training dataset, thus making a limitation from the already limited number of samples from the dataset. Then, we evaluated the model and gained an insight that the model has the value RMSE and MAE of 0.09779162 and 0.005533414, respectively, which means the model is quite accurate in predicting the engagement rate for Instagram posts, having an MAE value of 0.55% from the engagement rate value.

The researcher then concluded several points according to the research result: 1) The predictive analytics model can predict several Instagram posts simultaneously. If we want to grow the number of customers engaging with our coffee shop business, we can use this prediction model to create an Instagram marketing post plan, avoiding the combination that hurts the potential engagement rate for the Instagram post; and 2) when a coffee shop wants to post on their Instagram account, they should consider using the endorsement post and use the city of their coffee shop to be put in the location input of the Instagram post for better average value of engagement rate.

In addition, the researchers want to recommend future research as an evaluation for this research, such as: 1) Using other machine learning methods like artificial neural networks or other statistic models for the predictive analytics to improve the accuracy of the prediction. The limitation of Multiple Linear Regression is that we must do statistical training and comprehension of the statistics model, and there could be a missed detection of complex relationships between variables when creating a prediction model (Tu, 1996:1229); 2) Using a photograph or image classifier or other Machine Learning/Artificial Intelligence method further to reduce the subjectivity of the image within the post. The subjectivity here can come from different people interpreting the guidelines of classification differently, instead of using a machine with one perspective toward the given guidelines to classify the images; 3) using Naive Bayes or other text classifier types to reduce the subjectivity for the caption type further. The text classification can be inconsistent without clear keywords or limitations toward what kind of word strings directly classify the captions of the Instagram post; and 4) the future researchers or business owners may expand the scope of the research to make it more sophisticated or be ready to be used as a predictive model for a larger area, not limited to West Jakarta. Making the dataset larger can also create a more sophisticated machine-learning model, increasing the accuracy of the prediction.

REFERENCES

Alhabash, S., & Ma, M. (2017). A Tale of Four Platforms: Motivations and Uses of Facebook, Twitter, Instagram, and Snapchat Among College Students. Social Media + Society, 3(1), 205630511769154. https://doi.org/10.1177/2056305117691544

Arman, A. A., & Pahrul Sidik, A. (2019). Measurement of engagement rate in Instagram (case study: Instagram Indonesian Government Ministry and Institutions). 2019 International Conference on ICT for Smart Society (ICISS) [Preprint]. https://doi.org/10.1109/iciss48059.2019.8969826

Asur, S., & Huberman, B. A. (2010). Predicting the Future with Social Media. In 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (pp. 492–499). https://doi.org/10.1109/WI-IAT.2010.63

Azmi, A. F., & Budi, I. (2018). Exploring practices and engagement of Instagram by Indonesia Government Ministries. 2018 10th International Conference on Information Technology and Electrical Engineering (ICITEE) (pp. 18–21). https://doi.org/10.1109/ICITEED.2018.8534799

Bakhshi, S., Shamma, D. A., & Gilbert, E. (2014). Faces engage us. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 965–974). https://doi.org/10.1145/2556288.2557403

Brettel, M., Strese, S., & Flatten, T. C. (2012). Improving the performance of business models with relationship marketing efforts – An entrepreneurial perspective. European Management Journal, 30(2), 85–98. https://doi.org/10.1016/j.emj.2011.11.003

Carr, C. T., & Hayes, R. A. (2015). Social Media: Defining, developing, and divining. Atlantic Journal of Communication, 23(1), 46–65. https://doi.org/10.1080/15456870.2015.972282

Dhisasmito, P. P., & Kumar, S. (2020). Understanding customer loyalty in the coffee shop industry (A survey in Jakarta, Indonesia). British Food Journal, 122(7), 2253–2271. https://doi.org/10.1108/BFJ-10-2019-0763

Ferrara, E., de Meo, P., Fiumara, G., & Baumgartner, R. (2014). Web data extraction, applications, and techniques: A survey. Knowledge-Based Systems, 70, 301–323. https://doi.org/10.1016/j.knosys.2014.07.007

Ferreira, J., Ferreira, C., & Bos, E. (2021). Spaces of consumption, connection, and community: Exploring the role of the coffee shop in urban lives. Geoforum, 119, 21–29. https://doi.org/10.1016/j.geoforum.2020.12.024

Giannoulakis, S., & Tsapatsoulis, N. (2017). Defining and Identifying Stophashtags in Instagram (pp. 304–313). https://doi.org/10.1007/978-3-319-47898-2_31

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. International Journal of Forecasting, 22(4), 679–688. https://doi.org/10.1016/j.ijforecast.2006.03.001

IBM. (n.d.). About linear regression. Retrieved July 4, 2023, from https://www.ibm.com/topics/linear-regression

Kurniawan, Y., Setiawan, S., Bhutkar, G., Johan, & Cabezas, D. (2020). Instagram Engagement for University. 2020 International Conference on Information Management and Technology (ICIMTech) (pp. 887–892). https://doi.org/10.1109/ICIMTech50083.2020.9211134

Kronthaler, F., & Zöllner, S. (2021). Data Analysis with RStudio. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-62518-7

Mayfield, A. (2008). What is social media? iCrossing. Retrieved July 17, 2015.

Nurhadian, N. (2023). Indonesia: Most used social media platforms 2022. Statista. Retrieved June 10, 2023, from https://www.statista.com/statistics/1343136/indonesia-most-used-social-media-platforms/

Patma, T. S., Wardana, L. W., Wibowo, A., Narmaditya, B. S., & Akbarina, F. (2021). The impact of social media marketing for Indonesian SMEs sustainability: Lesson from Covid-19 pandemic. Cogent Business & Management, 8(1). https://doi.org/10.1080/23311975.2021.1953679

Pilgrim, K., & Bohnet-Joschko, S. (2019). Selling health and happiness how influencers communicate about dieting and exercise on Instagram: mixed methods research. BMC Public Health, 19(1), 1054. https://doi.org/10.1186/s12889-019-7387-8

SI, S. (2015). Social Media and Its Role in Marketing. Business and Economics Journal, 07(01). https://doi.org/10.4172/2151-6219.1000203

Singh, A., & Masuku, M. (2014). Sampling Techniques and Determination of Sample Size in Applied Statistics Research: An Overview. International Journal of Commerce and Management, 2, 1–22.

Sledge, E. (2014, March 8). Using social media to grow a small business. CBS News. CBS Interactive. Retrieved from https://www.cbsnews.com/newyork/news/using-social-media-to-grow-a-small-business/

Stockwell, P. (2017). Ambiance. In The Language of Surrealism (pp. 133–148). Macmillan Education UK. https://doi.org/10.1057/978-1-137-39219-0_8

Stroebele, N., & de Castro, J. M. (2004). Effect of ambiance on food intake and food choice. Nutrition, 20(9), 821–838. https://doi.org/10.1016/j.nut.2004.05.012

Thompson, H. (1998). Marketing strategies. Journal of Business Strategy, 19(4), 16–21. https://doi.org/10.1108/eb039946

Toffin. (2020). BREWING IN INDONESIA: Insights for Successful Coffee Shop Business. Jakarta: Toffin.

Tu, J. v. (1996). The advantages and disadvantages of using artificial neural networks versus logistic regression to predict medical outcomes. Journal of Clinical Epidemiology, 49(11), 1225–1231. https://doi.org/10.1016/S0895-4356(96)00002-9