



Article

New RFI Model for Behavioral Audience Segmentation in Wi-Fi Advertising System

Shueh-Ting Lim, Lee-Yeng Ong * and Meng-Chew Leow

Faculty of Information Science and Technology, Multimedia University, Jalan Ayer Keroh Lama, Melaka 75450, Malaysia; 1201300175@student.mmu.edu.my (S.-T.L.); mcleow@mmu.edu.my (M.-C.L.)

* Correspondence: lyong@mmu.edu.my

Abstract: In this technological era, businesses tend to place advertisements via the medium of Wi-Fi advertising to expose their brands and products to the public. Wi-Fi advertising offers a platform for businesses to leverage their marketing strategies to achieve desired goals, provided they have a thorough understanding of their audience's behaviors. This paper aims to formulate a new RFI (recency, frequency, and interest) model that is able to analyze the behavior of the audience towards the advertisement. The audience's interest is measured based on the relationship between their total view duration on an advertisement and its corresponding overall click received. With the help of a clustering algorithm to perform the dynamic segmentation, the patterns of the audience behaviors are then being interpreted by segmenting the audience based on their engagement behaviors. In the experiments, two different Wi-Fi advertising attributes are tested to prove the new RFI model is applicable to effectively interpret the audience engagement behaviors with the proposed dynamic characteristics range table. The weak and strongly engaged behavioral characteristics of the segmented behavioral patterns of the audience, such as in a one-time audience, are interpreted successfully with the dynamic-characteristics range table.

Keywords: behavioral model; engagement; behavioral audience segmentation; clustering algorithms; behavioral characteristics and patterns



Citation: Lim, S.-T.; Ong, L.-Y.; Leow, M.-C. New RFI Model for Behavioral Audience Segmentation in Wi-Fi Advertising System. *Future Internet* **2023**, *15*, 351. <https://doi.org/10.3390/fi15110351>

Academic Editors: Francesca Fallucchi and Ernesto William De Luca

Received: 30 September 2023
Revised: 20 October 2023
Accepted: 24 October 2023
Published: 26 October 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The Wi-Fi advertising system is a popular advertising platform for businesses to expose their advertisement to the public in a more effective way. The public Wi-Fi advertising systems share indistinguishable general procedures, as shown in Figure 1. The procedure involves users logging in, watching a non-skippable advertisement, and skipping to connect to free Wi-Fi or visit the advertisement website. A most common example of the Wi-Fi advertising system in Malaysia consists of almost 8000 public hotspots distributed nationwide [1]. When the users connect to the public hotspots for free Wi-Fi, a login with personal information is required. After that, there will be a non-skippable advertisement playing for 10 s before providing two options for selection. In the first option, the users are able to visit the website related to the advertisement, whereas the second option will redirect the users to proceed to connecting to the free Wi-Fi. The other example is the Boingo hotspot that is sponsored by the Boingo Wireless American company [2]. There are almost 1 billion hotspots available in airports, hotels, transits, and shopping malls of the United States of America. When the users connect to the Boingo hotspot, a login with personal information and watching a non-skippable advertisement for 17 s are the compulsory processes. After that, similar to the Malaysian public hotspot, Boingo also provides two options to redirect the user to the advertisement page or to connect them to the Wi-Fi.

As the name implies, Wi-Fi advertising is a marketing strategy that reaches the public through Wi-Fi networks. It appears on a smartphone, tablet, laptop, or other devices, as

opposed to traditional advertising marketing, where ads are placed in a magazine or a newspaper. Wi-Fi advertising is a hot topic in recent years due to the increasing number of Internet users. According to [3], among the 32.98 million Malaysian citizens, a total of 29.55 million are Internet users, accounting for 89.5% of Malaysian citizens. The public relies on the Internet for a wide variety of purposes, including socializing, e-learning, amusement, online banking, and e-wallet transactions. In light of this, mobile users connect their devices to Wi-Fi in their daily routine. The Wi-Fi advertising system that provides free Wi-Fi access is a great choice for the public to connect to. Hence, it is an effective marketing medium to deliver the advertisement to the public. The purpose of the Wi-Fi advertising industry is to advertise and implant a brand and its products into the subconscious mind of the public as much as possible. Therefore, the engagement of the audience towards the advertisement is a crucial element to investigate. The engagement of the audience towards the advertisement is defined as how well the advertisement can retain the audience to continue watching them or how well the audience is attracted to the advertisement.

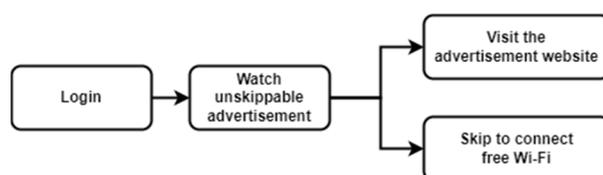


Figure 1. General procedure of the public Wi-Fi advertising system.

In fact, there are existing customer behavioral models developed for businesses to analyze customer behaviors to guide the decision-making process on the marketing strategies [4–11]. However, those models are solely developed for the e-Commerce, banking, and insurance industries instead of the Wi-Fi advertising business. The existing customer behavioral models investigate customer behaviors in terms of recency, frequency, and the purchase's monetary values since these industries are more concerned with the purchases made by customers [4–11]. However, the Wi-Fi advertising industry is more concerned of the audience engagement level towards the advertisement instead of their purchases. To the best of the authors' knowledge, a customer behavioral model is yet to be created for the Wi-Fi advertising industry. Therefore, a new RFI model is proposed to investigate the audience behaviors particularly in evaluating the audience engagement level within the Wi-Fi advertising industry.

From the review of the existing behavioral segmentation applications, it is found that both clustering-based [4–7,9,11] and the score-based [8,10] segmentation approaches are widely used to interpret the behaviors of the audience. In clustering-based segmentation, clustering algorithms are utilized to identify natural groupings and patterns in the data without pre-established rules. On the other hand, score-based segmentation is a rigid approach that requires pre-established rules, such as using the quartiles concept as the rules to score the criterion [8,10]. It will generate a huge number of group combinations that makes it exponentially more complicated when interpreting the patterns of the audience's behaviors. Therefore, clustering-based segmentation was chosen as the audience segmentation approach in this study to investigate and interpret the behavioral patterns of the audience.

In this study, investigations have been conducted to evaluate the influence of the newly proposed RFI model in segmenting the behaviors of the audience that are interacting with the Wi-Fi advertising system. The Wi-Fi advertising system dataset that was collected in Malaysia was tested with the proposed RFI model to evaluate the engagement of the audience towards the advertisements. In addition to that, three types of clustering algorithms are tested to segment the audience according to their behaviors to give a better insight in understanding the audience behaviors in terms of their recency, frequency, as well as interest. Lastly, the segmented audience groups are interpreted with their representative

behavioral characteristics based on the dynamic-characteristics range table. The main contributions of this paper are summarized as follows:

- To propose a new RFI model that is generally applicable to measure the audience behaviors in Wi-Fi advertising system.
- To segment the audience behaviors into well-defined groups based on the RFI model.
- To create a dynamic-characteristics range table to interpret the segmented behavioral characteristics of the audience based on their respective RFI values.

The introductory section is followed by Section 2, which describes the related analysis methods and clustering algorithms in the existing publications. After that, Section 3 presents the procedures of the experiment’s setup with the experimental results included in Section 4. Section 5 illustrates the discussion of this study, whereas the last section encapsulates the conclusions and summary of the whole study as well as the future work.

2. Related Works

2.1. Behavioral Analysis

Behavioral analysis has been the focus of numerous research investigations in recent years. Many of these have significant practical uses, particularly from a marketing standpoint. There are many indicators that serve as metrics to examine the behaviors of the customers or users. In existing publications, the RFM model, which refers to the recency, frequency, and monetary values, is a method of behavioral analysis that is widely used in different domains ranging from the e-Commerce, banking, to insurance industries. Table 1 illustrates the application of behavioral segmentation in these domains.

Table 1. Applications of behavioral segmentation in different domains.

Domain	Year	Behavioral Analysis	Behavioral Segmentation	Purpose of Application
e-Commerce	2021 [4]	RFM model	K-means clustering	Purchasing behavioral segmentation - Segment the customers according to their purchasing behaviors as the marketing reference.
	2021 [5]	RFMT model	Agglomerative hierarchical clustering (AHC)	Online shopping behavioral segmentation - Segment the customers based on their shopping behaviors to discover their online shopping patterns.
	2022 [6]	RFM model	K-means clustering	Customer classification - Segment the customers according to their behaviors for improving sales.
	2022 [7]	RFM model	Gaussian Mixture Model (GMM)	Supermarket customer segmentation - Segment the customers according to their purchasing behaviors.
Banking	2021 [8]	RFM model	RFM score	Customer segmentation - Segment the customer behaviors in bank activities in increasing bank average balances.
	2022 [9]	RFMT model	K-means clustering, Agglomerative hierarchical clustering (AHC)	Mobile banking behavioral segmentation - Segment the customers to discover the customer’s transaction patterns in banking.
Insurance	2020 [10]	RFM model	RFM score	Policyholder segmentation - Segment the policyholders according to their claiming patterns.
	2022 [11]	RFM model	K-means clustering	Client segmentation - Segment the clients according to their behaviors for the needs of the policy.

From Table 1, most of the behavioral analysis studies are using the RFM model, which is a cost-effective behavioral analysis model. The RFM model has been popularized by Arthur Hughes since 1994 [12,13]. The metrics of RFM are vital indicators in measuring the customer lifetime value (CLV) in the customer's behaviors [14]. It was based on the marketing axiom of the Pareto Principle that "80% of your sales come from 20% of your customers" [15]. The usage of RFM analysis is for ranking the customers in terms of three quantifiable factors—recency (R), frequency (F), and monetary value (M)—and thus classifying them into homogeneous groups. It is a basic marketing strategy for the customer relationship management (CRM) process in most of the business [16].

Each component in the RFM model corresponds to distinct and significant characteristics of the customer. The most important component of the three RFM models is the recency of the customer, which is represented by the letter R in the RFM analytic alphabet [16]. The recency component refers to the gap of time since the customer's most recent interaction or transaction, which measures the time between the customers' analyzing and purchasing phases. The gap in time is typically measured in days or months. The greater the gap period, the lower the recency number. This can be explained by the extended interval between customer purchases, which suggests that the customer is not particularly drawn to it and does not have a strong wish to make another purchase [16]. The general rule that almost all market sectors follow stated that the more recently a customer transacts or interacts with a product, the more likely it is that they will react to the next promotion offer [17].

The frequency component of the RFM is typically determined by the total number of transactions customers have made or the average amount of time between their most recent interaction or transaction and their most recent transaction [16]. The frequency score can be rated higher to show the longer times of transactions that a client makes. The frequency score is used to quantify customer steadfastness with a brand. Customer loyalty, product demand, and market size all improve with frequency score, which, consequently, increases with product repurchases [16]. It is important to note that first-time customers can be viewed as potential customers because they represent a good target market for the company to follow-up with and entice with different products in order to turn them into a group of regular customers [18].

There are two methods to investigate the monetary component in the RFM model. The former indicates the average transaction amount made by the customer over the course of a specified period, whereas the latter indicates the overall accumulated transaction amount made by the customer over the course of a defined period. The former approach of average transaction amount is strongly advised over the latter approach for decreasing the collinearity of the frequency and monetary components [16]. This monetary component will portray the customer's identity as a light or hefty spender to the business. The more the customer spends, the more likely is the customer to respond to the new deal, and the more likely they are to become a repeat customer and make purchases [18].

However, the RFM model is not suitable for interpreting all the customer behaviors. Therefore, an improvement or variation is performed to the RFM model in the existing works. From Table 1, there is a new proposed RFMT model for the e-Commerce domain [5] and banking domain [9]. The RFM of the RFMT still represents the recency, frequency, and monetary value. The only difference is the addition of another component to the RFM model. The alphabet T of the RFMT model represents the time component, which is another significant component that should be considered for achieving the model's purpose. In [5], the time component refers to the inter-purchase time, which is calculated by the period between two continuous customer transactions made in the same location or on the same website, to understand their pattern of online shopping. While, in [9], the time component represents the transaction time, which is measured by the median of the gap time between the current and subsequent transactions, for discovering the customer's transaction pattern in the banking domain. However, the RFM or RFMT model is only suitable for interpreting the audience behaviors that are related with monetary, particularly in the Wi-Fi advertising

industry. This can be explained by the main purpose of Wi-Fi advertising, which is to advertise and implant the brand and the products into the subconscious mind of the public as much as possible (involving engagement) instead of promoting the products to let the audience purchase (involving monetary) directly.

Up to 95% of customer purchasing decisions are influenced by the subconscious impression generated from advertising impact [19]. When compared to other brands, customers are more likely to purchase from brands that exist subconsciously. Therefore, the customer's interest that evaluates the impression and action is a significant factor to customer's subconscious perceptions of a product or brands. To the best of the authors' knowledge, customer behavioral models are yet to be created for the Wi-Fi advertising industry. Hence, investigations have been conducted in this study to evaluate the influence of the newly proposed RFI model in segmenting the engagement behaviors of the audience in the Wi-Fi advertising industry. The new RFI model is designed to discover the pattern of the audience's behavior towards the advertisement. The RFI model is replacing the monetary value with the interest of the audience towards the advertisement, which is represented as the *I*. Although the recency component remains in this proposed model, the definition of recency is different. In the RFI model, the recency is defined as the duration of when the audience is being impressed by the advertisement. A shorter interval of time results in a stronger impression of the advertisement remaining in the audience's mind. The frequency is defined as how many times the audience is being impressed by the advertisement. The higher the frequency, the stronger impression implanted in the subconscious mind of the audience, whereas the interest level of the audience is used for indicating the impression and attraction of the advertisement to the audience [20].

In the study [21], the interest and engagement of users towards social media posts are determined by the multiple types of attributes, such as the actions of clicking, viewing, liking, and commenting, as well as sharing the advertisement post. These attributes are considered the impressions of the social media users towards the advertisement posts. In the Wi-Fi advertising system, the available advertising attributes are limited to viewing, clicking for accessing the advertisement details, or skipping the advertisement. The advertisement video will continuously play if there is no action either to view the advertisement details or skip the advertisement. Therefore, the view duration of the audience towards the advertisement is a basic advertising attribute to evaluate the interest level of the audience towards the advertisement. The view duration of the audience is recorded from the initial moment when they begin watching the advertisement until they skip it or access the advertisement details. According to [22], the click attribute is an important indicator for evaluating the effectiveness of an advertisement to attract and engage its viewers. In the study [23], the click motivations are divided into four categories: literal interpretation (an evaluative signal of liking particular content), acknowledgement of viewing (demonstrating that they have seen and recognized a specific post), social support and grooming (with a goal of maintaining or deepening social relationships), and utilitarian (as a personal archive of content curated for future reference). The audience will only click on advertisement details wherever they are well engaged and attracted by the advertisement. This is a crucial component that can determine the interest level of the target audience in the advertisement.

2.2. Audience Segmentation

Audience segmentation is the process of partitioning the target audiences into smaller segments based on certain criteria or traits, such as demographics, behaviors, hobbies, or requirements [24]. It is widely used in the marketing and advertising domains to discover more specialized and focused audience groups for businesses. In existing papers, audience segmentation is usually performed with two different segmentation approaches, which are the clustering-based [4–7,9,11] and score-based [8,10] approaches to segment the audience's behaviors.

The clustering-based segmentation utilizes the unsupervised machine learning technique, which is the clustering algorithm. The clustering algorithm aims to generate audi-

ences that are related to each other in the same group and distinct from the audiences in other groups. The audience is grouped based on distance or similarity functions, allowing for the formation of clusters from the audiences that are closely related to each other in the same group [25]. It is able to identify the natural groupings and patterns in the data without establishing any rules or criteria beforehand and discovering the patterns that might not be apparent from prior knowledge. Therefore, with the help of the clustering algorithms, the underlying patterns and trends for marketing and advertising motives with the huge amount of data can be used to classify the audience into meaningful categories [26].

Score-based segmentation is a rigid approach that relies on pre-established rules or criteria to assign the scores. To take an example from the existing paper [10], the pre-established rules to assign each criterion of RFM model are from score 1 to score 4. The upper limit of each criterion's score is determined by three quartiles, which are 25%, 50%, and 75%. The larger the score assigned to the specific criterion, the better the performance of the specific criterion. By assigning four different scores (score 1 to score 4) to three criteria (recency, frequency, and monetary), there might be a total of 72 different groups interpreted. It is hard to gain insight into the huge number of groups for interpreting the behavioral patterns. Therefore, clustering-based segmentation was chosen as the audience segmentation approach in this study to investigate and interpret the behavior patterns of the audience for the businesses.

Broadly, there are three types of clustering algorithms that are being applied in the existing papers [4–7,9,11], which are centroid-based, hierarchical-based, as well as model-based clustering [27,28]. Table 2 illustrates the categories of clustering algorithms that are stated in Table 1. The key characteristics of each category are different. Therefore, each category will be included in the experiments of this study.

Table 2. Clustering algorithms in different categories.

Categories	Algorithm	Key Characteristic
Centroid based	K-means [29]	Partitioning the data into k clusters based on the centroid of cluster
Hierarchical based	Agglomerative Hierarchical Clustering [30]	Recursively merging the nearest pair of data or clusters to generate a hierarchy of clusters
Model based	Gaussian Mixture Model (GMM) [31]	Estimating the probabilities of each data belonging to each cluster

2.3. Performance Evaluation Metrics

As clustering is an unsupervised machine learning algorithm, the assessment of similarity or dissimilarity is used to evaluate the performance of clustering [32]. By locating disparate data points and combining related data points, the clustering algorithms can perform well. The elbow method, silhouette score, Calinski–Harabasz Index, as well as the Dunn Index are four widely used evaluation metrics for assessing the quality of clusters.

The elbow method's underlying theory says that the mean of the entire within-cluster sum of square error (SSE) should be as small as possible. The distortion score, which is an average of SSE, is calculated as the sum of the squared distances between each data point and the cluster centroid. Each cluster's compactness serves as a gauge for the SSE [33]. The cluster is more compact and has less fluctuation when the SSE value is lower. The elbow point, as seen in the graph below, serves as the graph's pivot point and is the definition of this method's arm structure. The elbow point that was discovered shows the correct number of clusters (k).

According to [34], the silhouette score determines the consistency within the data clusters. The silhouette score is employed to quantify the separation of clusters in terms of their spatial separation. It shows how closely connected each data point in a cluster is to those in the nearby clusters. The value range for evaluating the performance of the

silhouette score is from -1 to 1 . Incorrect cluster assignment is denoted by -1 , whereas 0 denotes no discernible distance between clusters, making it difficult to distinguish between them. A large distance between clusters is denoted by 1 , indicating well-defined clusters. As a result, one of the iterations of the silhouette coefficient closest to 1 will be used to determine how many clusters, which is the *optimal cluster* (k) to be used.

The Calinski–Harabasz Index (CH Index) is a metric that is utilized to measure the similarity of cohesion compared to separation [35]. Cohesion refers to the distance of a data point to its own cluster. The cohesion is measured by summing the distances between each data point of the cluster and its own centroid. The separation is the measure of how far the cluster centroids are from the global centroid and is measured by summing up the distances from each centroid of clusters to the global centroid of the dataset. The CH Index is measured by the separation divided by cohesion. The higher the score of the CH Index, the stronger the cohesiveness of the cluster.

Another measurement for assessing the robustness of the clustering performance is the Dunn Index. The Dunn Index is evaluated using an internal evaluation schema, and the clustered data are used to obtain the final result [36]. The purpose of the Dunn Index is to determine the compaction and variation in the clusters, similar to the other evaluation metrics. It will be able to distinguish between clusters with a high inter-cluster distance from those with a low intra-cluster distance. The value of the smallest intra-cluster distance divided by the largest inter-cluster distance produces the Dunn Index. When the distance between the data points in a cluster is small, and the distance between clusters is large, leading to a higher value of Dunn Index, the clustering works better.

3. Proposed Framework

In this section, descriptions of the dataset and research methodology with the proposed framework are provided. The process flow of this Wi-Fi advertising system implements similar procedures, as illustrated in Figure 1. The principal objective of this Wi-Fi advertising system is to show the advertisement video to the audience before granting them free Wi-Fi on their digital devices. A login to the system requires supplying an email address, age range, and gender for getting access to the free Wi-Fi. After login, there will be a non-skippable advertisement played for 10 s. After the 10th second, there are two option buttons to either visit the advertisement website to access more information about the advertisement or skip the rest of the advertisement to directly access the free Wi-Fi. If the audience has selected to visit the advertisement website, then the audience is directed to the advertisement page.

The process of the proposed framework is illustrated in Figure 2. Data cleaning is first executed. After that, data transformation is performed. The three factors of the RFI model, which are the recency, frequency, and interest values, are calculated. Lastly, clustering is applied as the behavioral segmentation approach to summarize the characteristics of the audience.

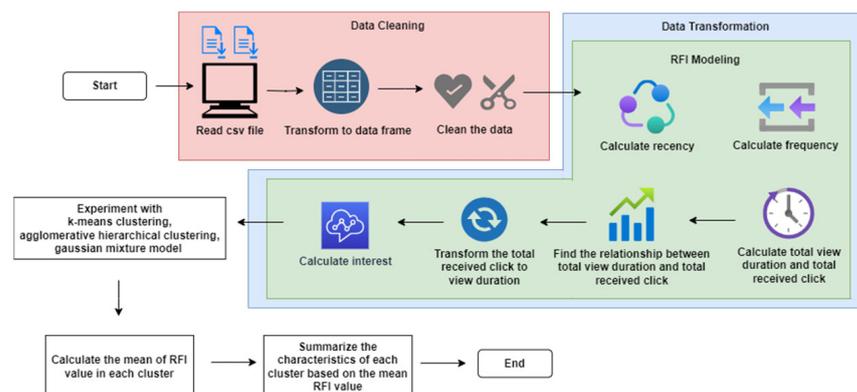


Figure 2. The proposed framework.

3.1. Dataset

The dataset was collected from a Wi-Fi advertising system in Malaysia. The Wi-Fi service subscribers that participated in the Wi-Fi advertising system are referred to as the “audience” in the subsequent sections. There are multiple advertising campaigns in the dataset. Each campaign is different from one another. In order to fairly investigate the audience behavioral patterns towards the advertising campaign, the data modeling and clustering process were processed based on the criteria stated in the following sections.

In the dataset, there was a total number of 154,531 audiences and a total of 580,229 records of audience interactions gathered in January 2023. Campaign 764, which was the most active campaign, had a total of 28,676 records. It was followed by Campaign 776, which recorded a total of 28,141 records. The more active the campaign, the larger the volume of the data recorded, offering a relatively richer dataset for analysis. By having a rich dataset, the variety of the patterns could be explored, offering insights into the audience behavioral patterns that influenced the success or failure of a campaign in achieving its objectives. Therefore, the two most active campaigns, which were Campaign 764 and Campaign 776, were selected for the investigation of the behavioral patterns of the audience.

Figure 3 illustrates a snippet of the dataset. The *campaign_id* refers to the id of the advertisement, *play_time* refers to the broadcast time of the advertisement, *click_time* refers to the time-of-click on the advertisement, and the *audience_id* refers to the id of the audience.

	campaign_id	play_time	click_time	audience_id
1283	764	2022-12-20 19:37:53	NaT	263
1289	764	2022-12-20 19:50:01	2022-12-20 19:50:17	286
1295	764	2022-12-20 20:12:23	NaT	293
1301	764	2022-12-20 20:24:08	2022-12-20 20:24:29	123
1307	764	2022-12-20 20:37:08	NaT	129

Figure 3. Snippet of the dataset.

3.2. Data Cleaning

Data cleaning was firstly executed by removing unnecessary data columns, dealing with missing values and converting the data type into a suitable format. As the missing value on *click_time* in the dataset indicated that the audience did not click to the advertisement, it remained unchanged. The data type of *play_time* and *click_time* were in a string format when the data were initially read. Hence, they were converted into date–time format to calculate the view duration of the advertisement.

3.3. Data Transformation

After data cleaning, the dataset was then processed to data transformation and transforming those meaningless data into meaningful data, which represented the behavior of each audience in the recency, the frequency, and the interest level towards the advertisement. Therefore, the first step of the data transformation was the RFI modeling. The computations of the recency, frequency, as well as the interest metrics were measured based on the engagement behaviors of an individual audience towards common advertisement campaigns.

The recency metric was defined as the time interval between when the audience returned to the Wi-Fi advertising system. The shorter interval allowed for a stronger impression to be retained in the audience’s mind. The recency of the specific audience in a campaign was calculated by using the difference between the latest broadcast time $Broadcast_t$ and the previous broadcast time $Broadcast_{t-1}$ in seconds. If the audience was a one-time audience, the latest broadcast time was the last active time of accessing the specific campaign. As the recency metric was measured in days, the difference of broadcast

time was divided, with 86,400 s in a day. The recency metric of a specific audience in a campaign is shown in Equation (1).

$$\text{Recency (day)} = \frac{\text{Broadcast}_t - \text{Broadcast}_{t-1}}{\text{Total Seconds in 1 Day}}. \quad (1)$$

The frequency metric was defined as the total number of advertisement occurrences recorded by the specific audience in a campaign. The higher occurrence of an advertisement allowed for a stronger impression to be implanted in the subconscious mind of the audience.

As the click action of audience to visit the advertisement webpage indicated the audience showing interest towards the advertisement, the interest metric in previous research was measured by the number of click actions [37]. Whenever the audiences were interested, they would respond to the advertisement on its first occurrence; otherwise, they would not respond to the advertisement. The previous research [37] highlighted the limitations of a recorded click action without specifying the view duration of the advertisement. As stated in Section 2.1, both the view duration of the advertisement and the click to visit the advertisement webpage action are crucial components that are used as the measurement criteria of the interest level towards the advertisement. Prior to interest metric formulation, this study must investigate the relationship between the total view duration and the overall number of clicks received within the Wi-Fi advertising system.

In this study, a relationship between the total view duration and the overall number of clicks received was found. As the interest metric was formulated to assess the interest level expressed towards the advertisement, it was recommended to exclude the mandatory 10 s non-skippable advertisement display time. Consequently, the view duration was calculated by excluding the mandatory 10 s non-skippable advertisement display time *Mandatory* from the broadcast time Broadcast_t of a specific audience in a campaign. Hence, the total view duration *View* of a specific audience in a campaign is deduced in Equation (2).

$$\text{View(seconds)} = \sum_{t=1}^n \text{Broadcast}_t - \text{Mandatory}. \quad (2)$$

The relationship between the total view duration and the overall number of clicks received are extracted to determine their correlation. Figure 4 depicts the correlation between total view duration and the overall number of clicks received for the two most active campaigns, which were Campaign 764 and Campaign 776. In order to evaluate the goodness of fit of the linear regression model to the data, the R^2 values of Campaign 764 and Campaign 776 were found to be 0.704 and 0.710, respectively. These values indicated that both campaigns exhibited a strong and positive linear relationship between the total view duration and the overall number of clicks received. The positive linear relationship implies that as the total view duration increased, the overall number of clicks received also increased. Therefore, it can be inferred that the total view duration of the audience is indeed related with the overall number of clicks received from the audience, signifying their overall interest towards the advertisement. Given the best-fit line equation, as expressed in Equation (3), the relative view advertisement duration x of a specific audience in a campaign from overall number of clicks received y of a specific audience can be transformed and formulated into the interest metric shown in Equation (4).

$$y = c + mx. \quad (3)$$

$$\text{Interest(seconds)} = \text{View} + \frac{y - c}{m}. \quad (4)$$

Once the recency, frequency, and interest metrics have been computed, these three derived factors are ready to be utilized in the clustering algorithms to segment the similar audience behavioral patterns.

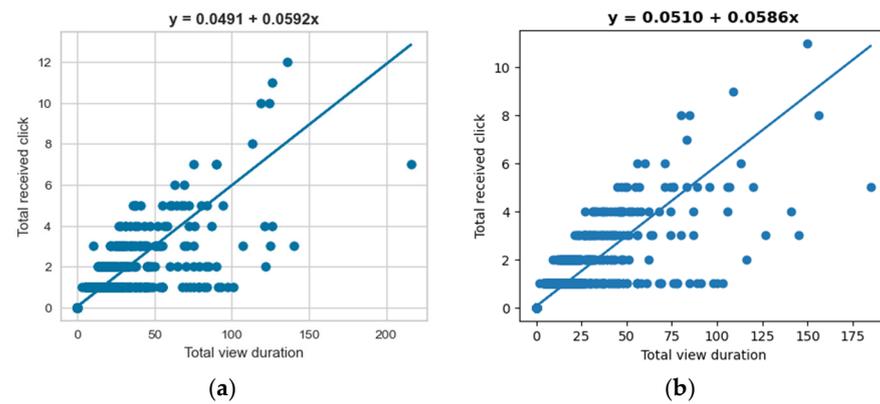


Figure 4. Relationship between total view duration and overall number of clicks received in (a) Campaign 764 and (b) Campaign 776.

4. Experiments and Results

To investigate the audience behavioral patterns towards a single campaign rather than all campaigns collectively, clustering algorithms were applied to analyze audience behavioral patterns within a specific campaign. In this section, both the normal case and the special case of the formulated RFI model are being tested to prove the new RFI model is generally applicable to different advertising attributes. In normal case, the interest metric (I) relies solely on the number of click actions, as suggested in existing research [37]. In contrast, the formulated relationship of total view duration and overall number of clicks received in Equation (4) are utilized in the special case. The recency (R) and frequency (F) metrics remained the same in both cases.

Both cases are applied to the same dataset and same campaign for performance comparison. The two most active campaigns of the dataset, which is Campaign 764 and Campaign 776, are selected to demonstrate audience segmentation with three clustering algorithms. The three clustering algorithms are k-means clustering, agglomerative hierarchical clustering as well as Gaussian Mixture Model. The characteristics of each cluster and the performance evaluation of each algorithm are elaborated in this section. To perform clustering, identifying the optimal number of clusters is necessary. Four performance evaluation metrics, which are the elbow method, silhouette score, CH Index and Dunn Index are used to determine the optimal number of clusters. Table 3 highlighted the optimal number of clusters for each algorithm in both cases.

It is not surprising that those clustering algorithms produced different cluster assignments for the same dataset. This could be explained based on the nature of the methodology of clustering algorithms. K-means minimizes the sum of squared distances between the data points and their assigned cluster centroids. Agglomerative hierarchical clustering is a bottom-up approach that starts with each data point as a separate cluster and iteratively merges the closest pairs of clusters until a single cluster containing all data points is obtained. Gaussian Mixture Model is a type of probabilistic model that characterizes data as a combination of multiple Gaussian distributions, where each Gaussian value represents a specific cluster within the data.

To identify the best-performing clustering algorithm, performance evaluation metrics, which included the silhouette score, CH Index, and Dunn index, were used for comparison, as listed in Table 4. In both cases, k-means outperformed the other two metrics by achieving the lowest score in the elbow method and the highest scores in both silhouette score and CH Index. It was followed by the agglomerative hierarchical and Gaussian Mixture models. For the Dunn Index, the Gaussian Mixture Model obtained the highest score, followed by the agglomerative hierarchical and k-means models. This was because the Gaussian Mixture Model tended to produce more compact and less-separated clusters. It is worth noting that the performance of k-means in the special case was slightly lower than the Gaussian Mixture Model, with a difference of 0.152 in Campaign 764 and 0.150 in Campaign 776.

Therefore, k-means clustering is the top-performing algorithm in the overall performance, followed by agglomerative hierarchical, whereas the Gaussian Mixture Model showed the least favorable results. Consequently, the groups segmented by k-means clustering were further analyzed to interpret the segmented audience behaviors based on the RFI model.

Table 3. Result of optimal cluster evaluation for Campaign 764 and 776.

Case	Campaign	Algorithm	Elbow Method	Silhouette Score	CH Index	Dunn Index
Normal case	764	K-means	5	5	5	3
		Agglomerative hierarchical	5	5	5	4
		Gaussian Mixture Model	3	5	3	3
	776	K-means	5	5	5	3
		Agglomerative hierarchical	5	4	5	3
		Gaussian Mixture Model	3	4	3	3
Special case	764	K-means	5	5	5	3
		Agglomerative hierarchical	4	4	4	4
		Gaussian Mixture Model	3	5	3	3
	776	K-means	5	5	5	3
		Agglomerative hierarchical	5	5	5	3
		Gaussian Mixture Model	3	3	3	3

The bold value in blue indicates the mode of the optimal cluster number.

Table 4. Performance comparison of the clustering algorithms for Campaign 764 and 776.

Case	Campaign	Clustering Algorithm	Elbow Method ↓	Silhouette Score ↑	CH Index ↑	Dunn Index ↑
Normal case	764	K-means	17,179.932	0.512	12,127.681	0.337
		Agglomerative hierarchical	23,695.748	0.459	9911.987	0.734
		Gaussian Mixture Model	34,949.082	0.143	6897.579	0.771
	776	K-means	16,297.157	0.511	12,658.076	0.393
		Agglomerative hierarchical	17,555.738	0.495	11,400.780	0.375
		Gaussian Mixture Model	34,077.653	0.419	7015.656	0.785
Special case	764	K-means	17,737.915	0.501	11,590.687	0.320
		Agglomerative hierarchical	19,004.798	0.490	10,488.531	0.399
		Gaussian Mixture Model	38,536.754	0.367	5750.592	0.472
	776	K-means	17,198.727	0.496	11,738.731	0.349
		Agglomerative hierarchical	19,000.083	0.480	10,163.169	0.400
		Gaussian Mixture Model	37,381.868	0.379	5901.983	0.499

The downward arrows (↓) indicate that the lower the score, the better the performance, whereas the upward arrows (↑) indicate that the higher the score, the better the performance. The blue-colored values indicate the best performance among all clustering algorithms.

5. Discussion

The RFI model was used to understand the behavior of the audience towards the effect of advertising. Table 5 shows the average values of recency, frequency, and interest level in each cluster of the two most active campaigns for the normal case, whereas Table 6 illustrates the special case. To provide a more comprehensive representation of the audience behavioral patterns, a dynamic characteristics range is created in Table 7, elaborating on the behavior of the audience groups mentioned in Tables 5 and 6. The dynamic characteristics range table is established based on similar practices from [38,39], by dividing all the RFI values for a campaign into four quartiles. In Table 7, five characteristics are defined based on the non-negative RFI values, ranging from zero to positive infinity, with the initial characteristics for each metric set to zero. The subsequent characteristics of each metric are defined according to different quartiles, where the upper limit of each quartile is determined by 25%, 50%, and 75%. As the characteristic range table is generated dynamically from

the RFI values of each campaign, the actual range will vary depending on the RFI values associated with different campaigns. The actual ranges of Campaign 764 and Campaign 776 in the special case are shown in Table 7.

Table 5. Mean of RFI value in each cluster of two most active campaigns for the normal case.

Campaign	Audience Number	Audience Percentage	Cluster	Recency ↓ (Day)	Frequency ↑ (Time)	Interest ↑ (Second)
764 (Total audience 19,777)	1693	8.56%	1	2.3156	4.1784	0.1110
	8457	42.76%	2	3.1211	1.2157	0.0000
	7511	37.98%	3	10.1687	1.0570	0.0000
	122	0.62%	4	2.5039	5.3934	3.6148
	1994	10.08%	5	6.4750	1.3661	1.0587
776 (Total audience 19,524)	7483	38.33%	1	10.2028	1.0588	0.0000
	1601	8.20%	2	2.3031	4.1537	0.0906
	8268	42.35%	3	3.0987	1.2124	0.0000
	131	0.67%	4	2.4740	5.2595	3.7023
	2041	10.45%	5	6.3460	1.3988	1.0642

The downward arrows (↓) indicate that the lower the score, the better the behavior of the audience towards the advertisement, whereas the upward arrows (↑) indicate that the higher the score, the better the behavior of the audience towards the advertisement. The blue-colored values indicate the strongly engaged behavior of the audience towards the advertisement, whereas the red-colored values indicate the weakly engaged behavior of the audience towards the advertisement.

Table 6. Mean of RFI value in each cluster of two most active campaigns for the special case.

Campaign	Audience Number	Audience Percentage	Cluster	Recency ↓ (Day)	Frequency ↑ (Time)	Interest ↑ (Second)
764 (Total audience 19,777)	1764	8.92%	1	2.2631	4.1508	3.9372
	8458	42.77%	2	3.1215	1.2158	0.0000
	7514	37.99%	3	10.1718	1.0568	0.0112
	127	0.64%	4	2.7980	4.8110	115.7868
	1914	9.68%	5	6.6500	1.3161	27.4169
776 (Total audience 19,524)	7570	38.77%	1	10.2471	1.0577	0.2763
	1647	8.44%	2	2.2784	4.1299	3.0429
	8274	42.38%	3	3.1014	1.2127	0.0000
	138	0.71%	4	2.8428	4.8696	115.9700
	1895	9.70%	5	6.0973	1.3858	28.2536

The downward arrows (↓) indicate that the lower the score, the better the behavior of the audience towards the advertisement, whereas the upward arrows (↑) indicate that the higher the score, the better the behavior of the audience towards the advertisement. The blue-colored values indicate the strongly engaged behavior of the audience towards the advertisement, whereas the red-colored values indicate the weakly engaged behavior of the audience towards the advertisement.

Referring to the special case in Table 7, the audience of Campaign 764 in Cluster 1 is interpreted as the audience that portrayed the shortest gap time from the last engagement (2.2631), high frequency (4.1508), and low interest in the advertisement (3.9372). In contrast, the largest proportion of audience from Cluster 2 portrays the shortest gap time from the last engagement (3.1215), low frequency (1.2158), and no interest in the advertisement (0.0000). Cluster 3 demonstrates the least favorable outcome, with the audience displaying the longest gap time from the last engagement (**10.1718**), low frequency (**1.0568**), and low interest in the advertisement (**0.0112**). On the other hand, the most favorable outcome is shown in Cluster 4, featuring an audience group with the shortest gap time from the last engagement (**2.7980**), highest frequency (**4.8110**), and high interest in the advertisement (**115.7868**). Lastly, the audience group in Cluster 5 exhibits a long gap time from the last engagement (6.6500), low frequency (1.3161), and medium interest in the advertisement (27.4169). To interpret the audience characteristics of each cluster in the two most ac-

tive campaigns, Table 8 presents the characteristics of each cluster associated with their actual range.

Table 7. Dynamic characteristics range table of two most active campaigns for the special case.

Criteria	Quartile Range	Actual Range		Characteristics
		Campaign 764	Campaign 776	
Recency (R)	R = 0	value = 0.0000	value = 0.0000	Audience with no gap time from the last engagement
	0 < R ≤ 25%	0 < value ≤ 3.2551	0 < value ≤ 3.2566	Audience with shortest gap time from the last engagement
	25% < R ≤ 50%	3.2551 < value ≤ 6.0630	3.2566 < value ≤ 6.0915	Audience with medium gap time from the last engagement
	50% < R ≤ 75%	6.0630 < value ≤ 8.8710	6.0915 < value ≤ 8.9265	Audience with long gap time from the last engagement
	R > 75%	value > 8.8710	value > 8.9265	Audience with longest gap time from the last engagement
Frequency (F)	F < 0	value = 0.0000	value = 0.0000	Audience with no frequency
	0 < F ≤ 25%	0 < value ≤ 2.2271	0 < value ≤ 2.2105	Audience with low frequency
	25% < F ≤ 50%	2.2271 < value ≤ 3.0900	2.2105 < value ≤ 3.0348	Audience with medium frequency
	50% < F ≤ 75%	3.0900 < value ≤ 3.9530	3.0348 < value ≤ 3.8592	Audience with high frequency
	F > 75%	value > 3.9530	value > 3.8592	Audience with highest frequency
Interest (I)	I = 0	value = 0.0000	value = 0.0000	Audience with no interest in the advertisement
	0 < I ≤ 25%	0 < value ≤ 18.0680	0 < value ≤ 29.2051	Audience with low interest in the advertisement
	25% < I ≤ 50%	18.0680 < value ≤ 54.2039	29.2051 < value ≤ 58.4103	Audience with medium interest in the advertisement
	50% < I ≤ 75%	54.2039 < value ≤ 176.1219	58.4103 < value ≤ 188.3255	Audience with high interest in the advertisement
	I > 75%	value > 176.1219	value > 188.3255	Audience with highest interest in the advertisement

Based on the observation in Table 8, it is evident that the characteristics revealed distinct behavioral patterns among the groups of audiences. In both cases, audience group exhibiting strongly engaged behaviors covered the smallest portion, whereas the group displaying weakly engaged behavior comprised the second largest portion across all the campaigns. The strongly engaged behavior group could be interpreted as individuals who expressed the shortest gap time from the last engagement and the highest frequency of engagement. This behavior signified a strong interest in the advertisement. Hence, this group of audience is more inclined to revisit the Wi-Fi advertisement service within a short period (approximately within 3 days). As a result, the audience retained a more lasting impression of the content and derived greater enjoyment from watching the advertisements.

On the other hand, the weakly engaged behavior audience group could be defined as one-time users who watched the advertisement just once. This behavior was reflected in the longest gap time from the last engagement and their low frequency, typically only occurring once in Table 6. In both cases, this audience group directly exhibited very minimal to no interest in the advertisement (exactly 0 s in Table 5 and less than half a second in Table 6). The difference arose because of the interest metric in the normal case was solely considered on the overall number of click actions to visit the advertisement webpage, lacking the insight provided by view duration, which was considered in the special case. Therefore, the special case offered a comprehensive understanding of audience behavioral patterns in the Wi-Fi advertising system. It is noteworthy from Table 8 that the largest proportion of the audience in both cases demonstrated the shortest gap time from the last engagement and had low frequency and no interest in the advertisement.

Table 8. Audience behavioral characteristics in each cluster of the two most active campaigns.

Campaign	Case	Cluster	Audience Percentage	Characteristics
764	Normal case	1	8.56%	Audience with shortest gap time from the last engagement, highest frequency, and low interest in the advertisement
		2	42.76%	Audience with shortest gap time from the last engagement, low frequency, and no interest in the advertisement
		3	37.98%	Audience with longest gap time from the last engagement, low frequency, and no interest in the advertisement
		4	0.62%	Audience with shortest gap time from the last engagement, highest frequency, and highest interest in the advertisement
		5	10.08%	Audience with long gap time from the last engagement, low frequency, and medium interest in the advertisement
	Special case	1	8.92%	Audience with shortest gap time from the last engagement, highest frequency, and low interest in the advertisement
		2	42.77%	Audience with shortest gap time from the last engagement, low frequency, and no interest in the advertisement
		3	37.99%	Audience with longest gap time from the last engagement, low frequency, and low interest in the advertisement
		4	0.64%	Audience with shortest gap time from the last engagement, highest frequency, and high interest in the advertisement
		5	9.68%	Audience with long gap time from the last engagement, low frequency, and medium interest in the advertisement
776	Normal case	1	38.33%	Audience with longest gap time from the last engagement, low frequency, and no interest in the advertisement
		2	8.20%	Audience with shortest gap time from the last engagement, highest frequency, and low interest in the advertisement
		3	42.35%	Audience with shortest gap time from the last engagement, low frequency, and no interest in the advertisement
		4	0.67%	Audience with shortest gap time from the last engagement, highest frequency, and highest interest in the advertisement
		5	10.45%	Audience with long gap time from the last engagement, low frequency, and medium interest in the advertisement
	Special case	1	38.77%	Audience with longest gap time from the last engagement, low frequency, and low interest in the advertisement
		2	8.44%	Audience with shortest gap time from the last engagement, highest frequency, and low interest in the advertisement
		3	42.38%	Audience with shortest gap time from the last engagement, low frequency, and no interest in the advertisement
		4	0.71%	Audience with shortest gap time from the last engagement, highest frequency, and high interest in the advertisement
		5	9.70%	Audience with long gap time from the last engagement, low frequency, and medium interest in the advertisement

The strongly engaged behavior is indicated in blue, whereas the weakly engaged behavior is indicated in red.

6. Conclusions

As the number of internet users continues to grow annually, businesses are increasingly using Wi-Fi services to reach a wider audience, leading to an increasing demand for audience behavioral analysis. This study aimed to introduce a new RFI model designed to analyze the audience behaviors in the Wi-Fi advertising system. With the help of clustering algorithms in segmenting the audience, this model successfully summarized and presented behavioral characteristics and patterns in distinct groups based on audience recency, frequency, and interest metrics.

Through experiments, the unique characteristics of each cluster were revealed and demonstrated. With the aid of clustering and the dynamic characteristic range table, the audience behavioral patterns in both normal case and special case could be successfully interpreted. Consequently, it could be inferred that the new RFI model could be broadly

applied to different Wi-Fi advertising attributes that exhibited different advertising effectiveness in terms of audience engagement. Furthermore, the RFI values of each segmented audience group were interpreted into meaningful characteristics through the utilization of the dynamic characteristic range table. Therefore, it could be concluded that the dynamic characteristic range table offered a viable approach for audience behavioral segmentation based on their respective RFI values. With the knowledge of the audience's behavioral patterns, businesses gain a better insight into their engagement and thus leverage the characteristics for designing and implementing more effective marketing strategies to boost their sales.

In the future, exploring the engagement of advertisements among different genders or age groups using the RFI model shows a promising direction of research. Additionally, segmenting audience behaviors based on their engagement times is significantly important for the implementation of targeted advertising.

Author Contributions: Funding acquisition, L.-Y.O. and M.-C.L.; Investigation, S.-T.L.; Project administration, L.-Y.O.; Supervision, L.-Y.O.; Visualization, S.-T.L.; Writing—original draft, S.-T.L.; Writing—review and editing, L.-Y.O. and M.-C.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Telekom Malaysia Research & Development, RDTC/221073 (MMUE/230002).

Data Availability Statement: Not Applicable, the study does not report any data.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Lim, J. TM Offers Free WiFi at 5800 Hotspots Nationwide with Newly Launched UniFi App. *SAYS*, 26 September 2017. Available online: <https://says.com/my/tech/tm-offers-free-wi-fi-at-5-800-hotspots-nationwide-with-newly-launched-unifi-app> (accessed on 29 September 2023).
2. Stellin, S. Free Wi-Fi, but Speed Costs. *The New York Times*, 4 September 2012. Available online: <https://www.nytimes.com/2012/06/05/business/airports-and-hotels-look-at-tiered-pricing-for-internet-access.html> (accessed on 29 September 2023).
3. Simon, K. Digital 2022: Malaysia—DataReportal—Global Digital Insights. *DataReportal*. 15 February 2022. Available online: <https://datareportal.com/reports/digital-2022-malaysia> (accessed on 29 September 2023).
4. Shirole, R.; Salokhe, L.; Jadhav, S. Customer Segmentation using RFM Model and K-Means Clustering. *Int. J. Sci. Res. Sci. Technol.* **2021**, *8*, 591–597. [[CrossRef](#)]
5. Zhou, J.; Wei, J.; Xu, B. Customer segmentation by web content mining. *J. Retail. Consum. Serv.* **2021**, *61*, 102588. [[CrossRef](#)]
6. Wang, T.; Li, N.; Wang, H.; Xian, J.; Guo, J. Visual Analysis of E-Commerce User Behavior Based on Log Mining. *Adv. Multimed.* **2022**, *2022*, e4291978. [[CrossRef](#)]
7. Oliveira, W.V.; Araújo, D.S.A.; Bezerra, L.C.T. Supermarket customer segmentation: A case study in a large Brazilian retail chain. In Proceedings of the 2022 IEEE 24th Conference on Business Informatics (CBI), Amsterdam, The Netherlands, 15–17 June 2022; Volume 1, pp. 70–79. [[CrossRef](#)]
8. Heikal, J.; Rialialie, V.; Rivelino; Supriyono, I.A. Hybrid Model of Structural Equation Modeling PLS and RFM (Recency, Frequency and Monetary) Model to Improve Bank Average Balance. *Aptisi Trans. Technopreneurship* **2021**, *4*, 1–8. [[CrossRef](#)]
9. Mamashli, Z.; Zolfani, S.H. Customer Segmentation Based on Mobile Banking User's Behavior. *Int. J. Mechatron. Electr. Comput. Technol.* **2022**, *12*, 5267–5276.
10. Nandapala, E.Y.L.; Jayasena, K.P.N.; Rathnayaka, R.M.K.T. Behavior Segmentation based Micro-Segmentation Approach for Health Insurance Industry. In Proceedings of the 2020 2nd International Conference on Advancements in Computing (ICAC), Malabe, Sri Lanka, 10–11 December 2020; Volume 1, pp. 333–338. [[CrossRef](#)]
11. Kumar, S.J.; Oommen Philip, A. Achieving Market Segmentation from B2B Insurance Client Data Using RFM & K-Means Algorithm. In Proceedings of the 2022 IEEE International Conference on Signal Processing, Informatics, Communication and Energy Systems (SPICES), Thiruvananthapuram, India, 10–12 March 2022; Volume 1, pp. 463–469. [[CrossRef](#)]
12. RFM Migration Analysis: A New Approach to a Proven Technique. Available online: <http://www.dbmarketing.com/articles/Art123.htm> (accessed on 31 March 2023).
13. Kohavi, R.; Parekh, R. Visualizing RFM Segmentation. In Proceedings of the 2004 SIAM International Conference on Data Mining (SDM), Lake Buena Vista, FL, USA, 22–24 April 2004. [[CrossRef](#)]
14. Khajvand, M.; Zolfaghar, K.; Ashoori, S.; Alizadeh, S. Estimating customer lifetime value based on RFM analysis of customer purchase behavior: Case study. *Procedia Comput. Sci.* **2011**, *3*, 57–63. [[CrossRef](#)]
15. Rajeev, S. Pareto principle and compulsive buying disorder—An analysis. *J. Educ. Soc. Res.* **2022**, *8*, 44–59.

16. Wei, J.-T.; Lin, S.-Y.; Wu, H.-H. A review of the application of RFM model. *Afr. J. Bus. Manag. Dec. Spec. Rev.* **2010**, *4*, 4199–4206.
17. Giesen, C.G.; Schmidt, J.R.; Rothermund, K. The Law of Recency: An Episodic Stimulus-Response Retrieval Account of Habit Acquisition. *Front. Psychol.* **2020**, *10*, 2927. Available online: <https://www.frontiersin.org/articles/10.3389/fpsyg.2019.02927> (accessed on 29 September 2023). [[CrossRef](#)]
18. Christy, A.J.; Umamakeswari, A.; Priyatharsini, L.; Neyaa, A. RFM ranking—An effective approach to customer segmentation. *J. King Saud Univ. Comput. Inf. Sci.* **2021**, *33*, 1251–1257. [[CrossRef](#)]
19. Zaltman, G. *How Customers Think: Essential Insights Into the Mind of the Market*; Harvard Business Press: Boston, MA, USA, 2003.
20. Johnston, K.A.; Taylor, M. (Eds.) *The Handbook of Communication Engagement*, 1st ed.; Wiley-Blackwell: Hoboken, NJ, USA, 2018.
21. Dolan, R.; Conduit, J.; Fahy, J.; Goodman, S. Social media: Communication strategies, engagement and future research directions. *Int. J. Wine Bus. Res.* **2017**, *29*, 2–19. [[CrossRef](#)]
22. Jaisinghani, M.R.; Lundwani, C.; Mukherjee, O.; Nagori, N.; Solanke, P. CTR Prediction of Advertisements using Decision Trees based Algorithms. In Proceedings of the 2022 International Seminar on Application for Technology of Information and Communication (ISemantic), Semarang, Indonesia, 17–18 September 2022; pp. 107–112. [[CrossRef](#)]
23. Hayes, R.A.; Carr, C.T.; Wohn, D.Y. One Click, Many Meanings: Interpreting Paralinguistic Digital Affordances in Social Media. *J. Broadcast. Electron. Media* **2016**, *60*, 171–187. [[CrossRef](#)]
24. Peelen, E.; Beltman, R. *Customer Relationship Management*, 2nd ed.; Pearson: London, UK, 2013.
25. Rodrigues, F.; Ferreira, B. Product Recommendation based on Shared Customer's Behaviour. *Procedia Comput. Sci.* **2016**, *100*, 136–146. [[CrossRef](#)]
26. Mishra, R.K.; Raj, H.; Urolagin, S.; Jothi, J.A.A.; Nawaz, N. Cluster-Based Knowledge Graph and Entity-Relation Representation on Tourism Economical Sentiments. *Appl. Sci.* **2022**, *12*, 8105. [[CrossRef](#)]
27. Xu, D.; Tian, Y. A Comprehensive Survey of Clustering Algorithms. *Ann. Data Sci.* **2015**, *2*, 165–193. [[CrossRef](#)]
28. Lim, Z.-Y.; Ong, L.-Y.; Leow, M.-C. A Review on Clustering Techniques: Creating Better User Experience for Online Roadshow. *Future Internet* **2021**, *13*, 233. [[CrossRef](#)]
29. Ahmed, M.; Seraj, R.; Islam, S.M.S. The k-means Algorithm: A Comprehensive Survey and Performance Evaluation. *Electronics* **2020**, *9*, 1295. [[CrossRef](#)]
30. Tan, P.-N.; Steinbach, M.; Karpatne, A.; Kumar, V. *Introduction to Data Mining*; Pearson India: Bengaluru, India, 2016.
31. Wan, H.; Wang, H.; Scotney, B.; Liu, J. A Novel Gaussian Mixture Model for Classification. In Proceedings of the 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC), Bari, Italy, 6–9 October 2019; pp. 3298–3303. [[CrossRef](#)]
32. Syakur, M.A.; Khotimah, B.K.; Rochman, E.M.S.; Satoto, B.D. Integration K-Means Clustering Method and Elbow Method for Identification of The Best Customer Profile Cluster. *IOP Conf. Ser. Mater. Sci. Eng.* **2018**, *336*, 012017. [[CrossRef](#)]
33. Liu, Y.; Li, Z.; Xiong, H.; Gao, X.; Wu, J. Understanding of Internal Clustering Validation Measures. In Proceedings of the 2010 IEEE International Conference on Data Mining, Sydney, NSW, Australia, 13–17 December 2010; pp. 911–916. [[CrossRef](#)]
34. Shi, C.; Wei, B.; Wei, S.; Wang, W.; Liu, H.; Liu, J. A quantitative discriminant method of elbow point for the optimal number of clusters in clustering algorithm. *EURASIP J. Wirel. Commun. Netw.* **2021**, *2021*, 31. [[CrossRef](#)]
35. Lima, S.; Cruz, M. A genetic algorithm using Calinski-Harabasz Index for automatic clustering problem. *Rev. Bras. Comput. Apl.* **2020**, *12*, 97–106. [[CrossRef](#)]
36. Luna-Romera, J.M.; del Mar Martínez-Ballesteros, M.; García-Gutiérrez, J.; Riquelme-Santos, J.C. An Approach to Silhouette and Dunn Clustering Indices Applied to Big Data in Spark. In *Advances in Artificial Intelligence*; Luaces, O., Gámez, J.A., Barrenechea, E., Troncoso, A., Galar, M., Quintián, H., Corchado, E., Eds.; Springer International Publishing: Berlin/Heidelberg, Germany, 2016; Volume 9868, pp. 160–169. [[CrossRef](#)]
37. Försch, S.; de Haan, E. Targeting online display ads: Choosing their frequency and spacing. *Int. J. Res. Mark.* **2018**, *35*, 661–672. [[CrossRef](#)]
38. Lydersen, S. Mean and standard deviation or median and quartiles? *Tidsskr. Den Nor. Lægeforening.* **2020**, *140*, 1–3. [[CrossRef](#)]
39. Wan, X.; Wang, W.; Liu, J.; Tong, T. Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. *BMC Med. Res. Methodol.* **2014**, *14*, 135. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.