Analyzing the Impact of Internet User Characteristics on Ad Clicks

Xiwen Ran 1004158154

2020/12/22

Abstract

The objectives of this report are to examine the impact of Internet user characteristics on ad clicks, and to analyze the associations between "Daily Internet Usage", "Daily Time Spent on a Site" and "Area Income", with ad clicks among a group of 1,000 people from age 19 to 61 all over the world. The dataset specifies whether or not a certain Internet user clicks on an advertisement. A logistic regression model of ad clicks with 1,000 adults in different regions was used to test for associations. Histograms of various Internet user characteristics were used to illustrate the significant distribution of the major factors affecting ad clicks.

Keywords

Observational Study, Logistic Regression Model, Online Advertising, Ad Clicks, Internet Users Characteristics

Introduction

In statistical thinking for marketing and advertising, statistics have an essential role in understanding consumers' online purchase intention, and in exploring the dominant factors affecting the online purchase intention. The statistical analysis provided in this study aimed to determine the factors which influence ad clicks. Observational data is used in the analysis of ad clicks. Compared to experimental design data, observational data is more effective and reliable for identifying the characteristics of the dataset.

Advertising plays a central role in modern society. Online advertising shapes how we consume in the digital world, while ad clicks can influence individuals' consumption to a certain extent. Ad clicks measure the behavioral responses to online ads at the individual level, which in turn lead to intentions to purchase and spread positive word of mouth (Zhang & Mao, 2016). The goal of this report is to demonstrate trends in Internet user characteristics of "Daily Internet Usage", "Daily Time Spent on a Site", and "Area Income", and analyze the statistically significant relationships of these predictor variables with the outcome variable "Clicked on Ad". The results of the report are determined by the multiple logistic regression model. The report is based on the dataset from the Kaggle website published by Fayomi (2017). The dataset contains 10 variables for 1,000 Internet user information.

The chosen dataset will be used to investigate how the number of ad clicks changes when the predictor variables change. The predictor variables include different characteristics of the user. In the Methodology section (Section 2), I describe the observational study, the data, and the model that was used to conduct the ad clicks analysis. Results of the ad clicks analysis are provided in the Results section (Section 3), while conclusions, areas of weakness, and the next steps are identified in the Discussion section (Section 4).

Methodology

	Estimate	Std. Error	z value	P-Value
Intercept	2.866e+01	2.355e+00	12.169	< 2e-16
Daily Internet Usage	-6.185e-02	5.714e-03	-10.824	< 2e-16
Daily Time Spent on Site	-1.676e-01	1.688e-02	-9.928	< 2e-16
Area Income	-1.032e-04	1.527e-05	-6.759	1.39e-11

Table 1: Logistic Regression Model of Ad Clicks

Data: The advertising dataset indicates whether or not a particular Internet user clicks on an advertisement. I have created a multiple logistic regression model that will predict whether or not they will click on an advertisement based on the features of the user. The features that I choose to predict ad clicks are: “Daily Internet Usage”, “Daily Time Spent on a Site”, and “Area Income”. “Daily Internet Usage” represents the average time a day the individual is on the Internet in minutes. “Daily Time Spent on a Site” represents the average time a day the individual is on the website in minutes. “Area Income” represents the average income in dollars of the individual in a particular geographical area.

Model: For the analysis, I used a model that included the three numerical variables influencing ad clicks. The binomial distribution was used when there were exactly two mutually exclusive outcomes of a single event. Since there was one outcome variable

"Clicked on Ad" with two states of the variable, either a "0" or a "1", it was a natural choice to use the logistic regression model. The logistic regression model was built using the R studio statistical software and the glm command. The multiple logistic regression model was used to investigate relationships between independent variables "Daily Internet Usage", "Daily Time Spent on a Site", and "Area Income", with ad clicks. The following model is considered to analyze whether or not a particular Internet user clicks on an advertisement:

$$log(\frac{p}{1-p}) = \beta0 + \beta1 \times Daily\ Internet\ Usage + \beta2 \times Daily\ Time\ Spent\ on\ Site + \beta3 \times Area\ Income$$

From the model, p is the probability of people clicking an advertisement. $\beta1,\ \beta2,\ \beta3$ coefficients represent the change in log odds for every one unit increase in $Daily\ Internet\ Usage,\ Daily\ Time\ Spent\ on\ Site,\ Area\ Income$ respectively. From Table 1: Logistic Regression Model of Ad Clicks, it can be observed that $\beta0 = 2.866e+01,\ \beta1 =- 6.185e-02,\ \beta2 =- 1.676e-01,\ \beta3 =- 1.032e-04$. It should also be pointed out that the variables "Daily Internet Usage", "Daily Time Spent on a Site", and "Area Income" are statistically significant because the corresponding p-values are less than the significance level of 0.05. That being said, all three predictor variables have significant relationships with the response variable.

Results

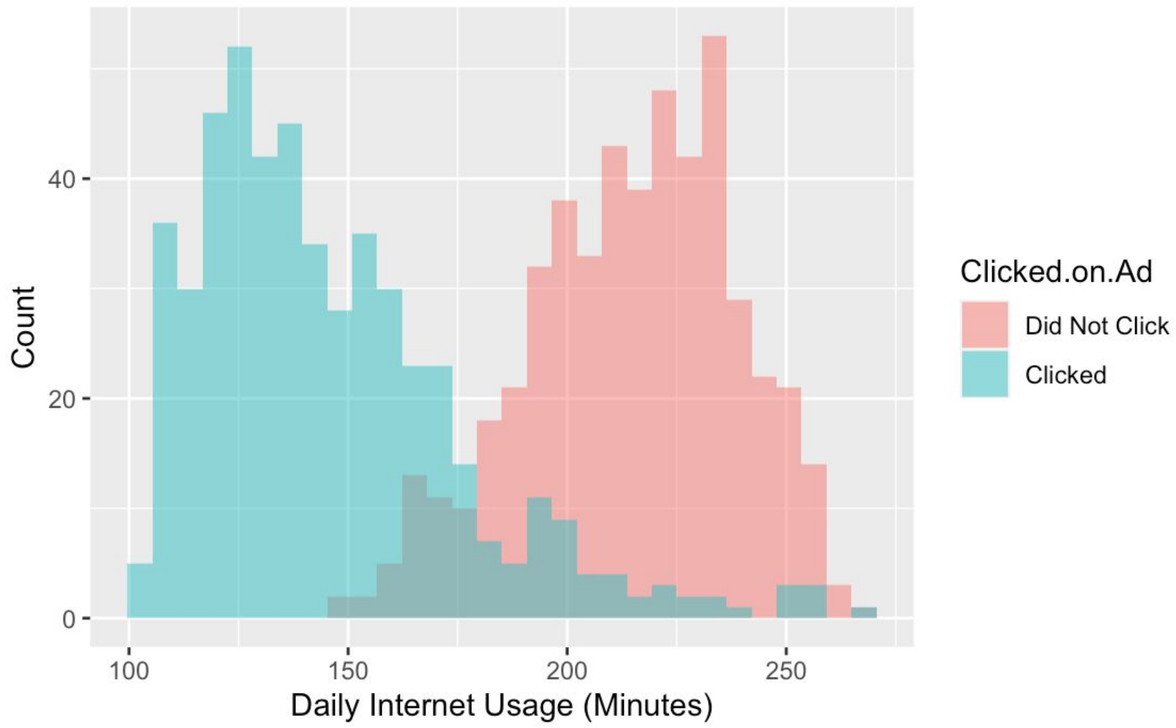

Figure 1: Histogram of Daily Internet Usage

Figure 2: Histogram of Daily Time Spent on a Site

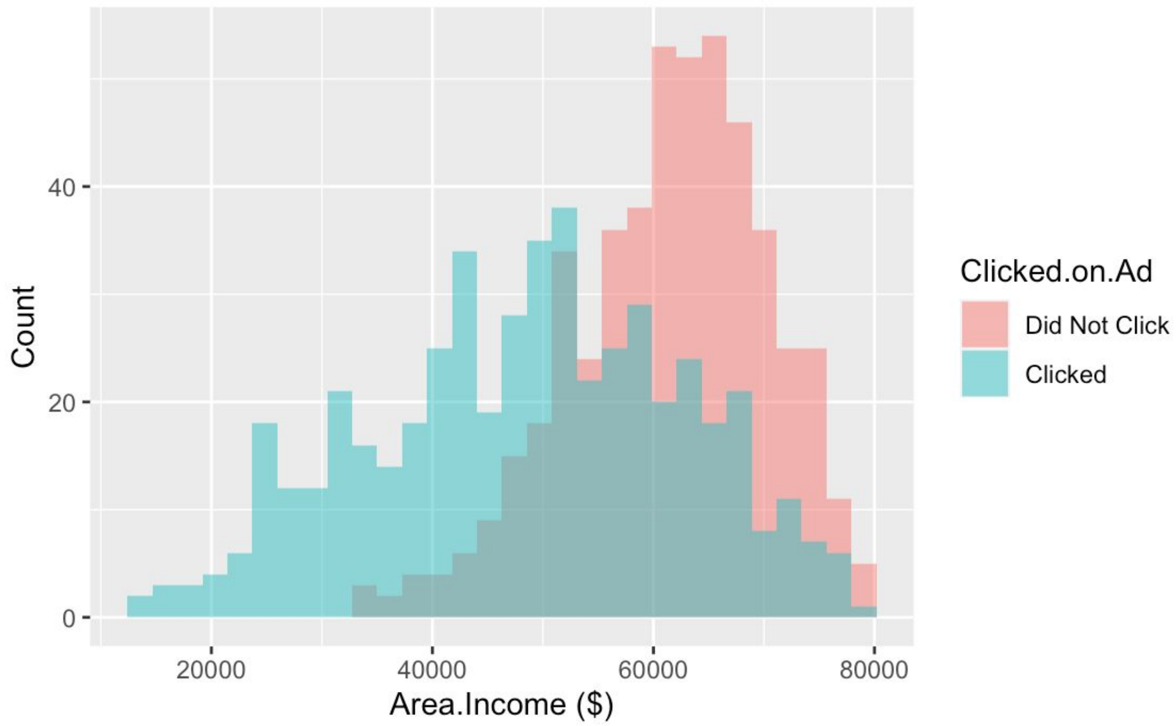

Figure 3: Histogram of Area Income

Logistic regression is used for binary outcomes having two possible values; the value is either a “0” or a “1”. For this analysis, the outcomes were appropriately labeled “Clicked” and “Did Not Click”. The selected individual could either choose to click the ad, or choose not to click the ad. From Figure 1: Histogram of Daily Internet Usage, the X-axis shows the intervals for the average time a day the individual is on the Internet in minutes, while the Y-axis shows the number of ad clicks within the intervals set by the X-axis. Figure 1 displays that the number of ad clicks is associated with less time of daily Internet usage. In other words, individuals who spend less time on the Internet will be more likely to click on the ads.

From Figure 2: Histogram of Daily Time Spent on a Site, the X-axis shows the intervals for the average time a day the individual is on the website in minutes, while the Y-axis shows the number of ad clicks within the intervals set by the X-axis. Figure 2 illustrates that the number of ad clicks is associated with less time spending on the website. To put it another way, individuals who spend less amount of time browsing the website will be more likely to click on the ads.

From Figure 3: Histogram of Area Income, the X-axis shows the intervals for the average income in dollars of the individual in a particular geographical area, while the Y-axis shows the number of ad clicks within the intervals set by the X-axis. Figure 3 represents that the number of ad clicks is associated with lower area income. Specifically, individuals who have a lower income in a particular area will be more likely to click on the ads.

Discussion

Summary: In summary, the study is focused on evaluating the impact of Internet user characteristics on ad clicks. A multiple logistic regression model is developed to identify whether the three independent variables affect individuals' decision to click on a random advertisement. Generally speaking, factors including "Daily Internet Usage", "Daily Time Spent on a Site" and "Area Income" have been shown to be associated with the number of ad clicks. These factors are essential in demonstrating the characteristics of Internet users and predicting if a particular Internet user will click on an advertisement.

Conclusions: All in all, the general purpose of the study is to observe and analyze how Internet user characteristics can result in different outcomes of clicking an advertisement Results from this study indicated that less time of daily Internet usage, less time spending on the website, and lower area income are associated with the increased number of ad clicks. In addition, the obtained results clearly state that individuals who spend less total time on the Internet, or spend less total time on the website, or have a lower income in a particular geographical area will be very likely to click on a random advertisement.

As a matter of fact, ad clicks are not only driven by individuals' responses to a specific ad, but also their general attitudes toward online advertising (Zhang & Mao, 2016). This indicates that individuals who tend to click on the ads will have more interest in online advertisements. Moreover, individuals who tend to click on the ads are more likely to make online purchases. It is undoubted that online advertisements can be displayed and reach different Internet users. Therefore, it is important for advertisers to think about who the target audience should be, and what characteristics or features they have reflected.

Weakness & Next Steps: One weakness of this study is that the "Area Income" variable is too broad, which only represents the average income based on a particular geographical area of people. The "Area Income" variable might not be as effective as personal income. Personal income statistics will measure the income that an individual receives from all sources, including salary, wages, bonuses, income from self-employment, dividends from investments, and receipts from real estate investments (Duffin, 2020). All things considered, it would be helpful if the predictor variable is more representative for the analysis.

Future studies using a set of more precise predictor variables would strengthen the observed findings. Additionally, future studies can obtain further insight into the various types of online advertising. The different online advertising types include social media advertising on Facebook, Instagram, LinkedIn, Twitter, YouTube, or Snapchat, native advertising, and display advertising (Braccialini, 2020). The different types of online advertising should be considered as significant factors that can affect the number of ad clicks. Further research will be conducted to better understand the influential factors leading to ad clicks.

References

Braccialini, C. (2020, May 12). Online Advertising: Everything You Need to Know in 2020. Retrieved December 22, 2020, from https://blog.hubspot.com/marketing/online-advertising

Duffin, E. (2020, March 27). U.S. per capita personal income, by state 2019. Retrieved December 22, 2020, from https://www.statista.com/statistics/303555/us-per-capita-personal-income/

Fayomi. (2017, December 14). Advertising. Retrieved December 22, 2020, from https://www.kaggle.com/fayomi/advertising

RStudio [Computer software]. Retrieved December 22, 2020, from https://rstudio.cloud/projects

Zhang, J., & Mao, E. (2016). From Online Motivations to Ad Clicks and to Behavioral Intentions: An Empirical Study of Consumer Response to Social Media Advertising. Psychology & Marketing, 33(3), 155-164. doi:10.1002/mar.20862

Appendix

This file was made using the R markdown package. All code used in this paper can be accessed from within the code blocks of the markdown document. All the methodology I have developed are learned in STA304H1F of University of Toronto (St. George campus).