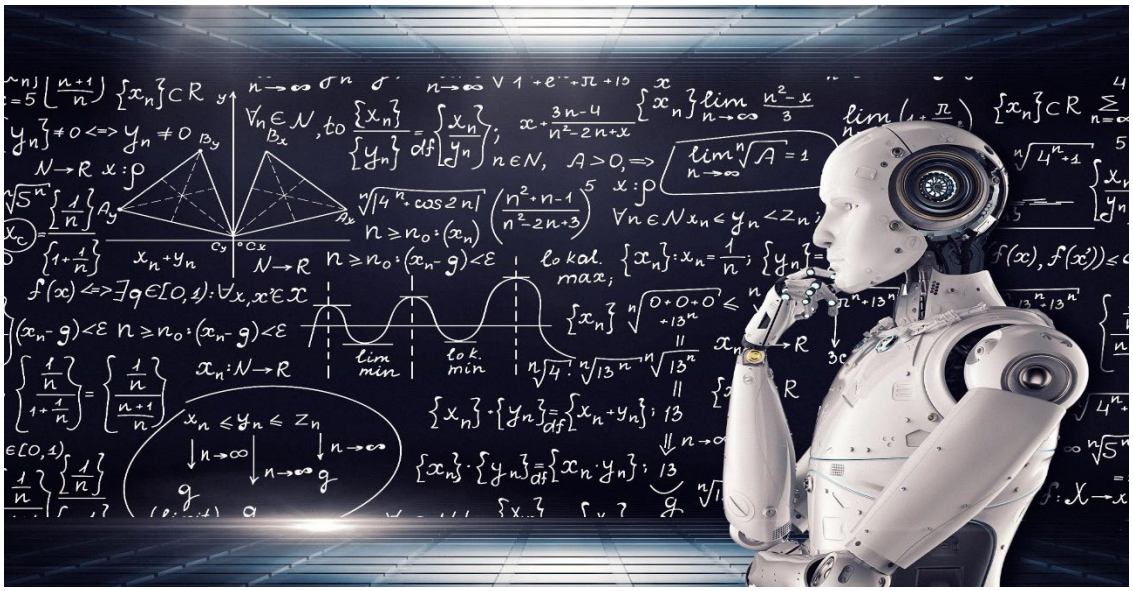

AAI-501

Adam Graves

The University of San Diego

Module 7

Final Project: Pay Per Click (PPC) Activity

(Digital Marketing)

Prof. Lin

12/11/2022

Documentation: https://github.com/kavrangraves/AAI501-M7-PPC-Project.git

Presentation: https://youtu.be/LAg6XSaT_Pw

Abstract

The Pay Per Click has been around for over 15 years, and today it is a major marketing service available mainly by Google. A serious amount of revenue for Google, "In 2016 alone, $79.38 billion of Google's revenue came from its pay per click advertising service Google Adwords." (Sam Carr, Luni 2021). For any business that has a digital presence, the PPC is a major service and expense in the business budget. I will examine factors that will provide answers to the PPC usage by gender and age, by utilizing Python's sklearn libraries (https://www.tutorialspoint.com/scikit_learn/scikit_learn_introduction.htm#:~:text=Scikit%2Dlearn%20(Sklearn)%20is,a%20consistence%20interface%20in%20Python), and statistical calculations.

Keywords Marketing Advertising, Digital Marketing, Internet Marketing, PayPerClick, Search Engine.

Introduction

Since the millennium marketing has been stepped up to go digital; with digital marketing, also called internet marketing, businesses utilizing digital technologies (Smith, 2012).

Digital Marketing is the main marketing tool used in today's businesses. Google is by far the leader in these services offering a platform for Websites, Maps, and advanced reporting services. The Pay Per Click (PPC) which is now part of their Ad Services, see: https://ads.google.com/home, and provides the services of "An online advertising solution that businesses use to promote their products and services on Google Search, YouTube, and other sites across the web. Google Ads also allows advertisers to choose specific goals for their ads, like driving phone calls or website visits. With a Google Ads account, advertisers can customize

their budgets and targeting, and start or stop their ads at any time." (https://ads.google.com/home/faq/)

This service is an expensive one, and can become much more so if the industry is in such of competition in which the costs Per Click go up. Many businesses look to evaluate their Return On Investment (ROI) for the PPC service as to compare to the common Search Engine Optimization (SEO) see:(https://developers.google.com/search/docs/fundamentals/do-i-need-seo). SEO can be free if the business wishes to do by themselves, but can take time to show results. (Average is between 30-90 days)

This project is going to enable my agency to help clients streamline their budgets correctly as they launch new websites for their business. The discussion of allocating funds between the efforts of building a natural Search Engine Optimization (SEO) vs. Pay Per Click (PPC or Ads).

I will be examining the relations between:

- Determine the ratio of web usage between male vs. female
- Determine the ratio of user Clicked on Ad between male vs. female, evaluate if there is a significant difference. This can help in determining on which campaigns it is worth utilizing Ad clicks.
- Determine the general percentage of "Clicked on Ad" (Ad Clicks) vs general web activity
- Identify patterns to Click Ad = 1 (True)
- Generate Visuals (Plots) and Statistics that display all the analytic results

Much of the results for the dependent variable (DV) will be based on the Boolean Logic, a logic developed by the 19th-century English mathematician George Boole, in which the

values are True (1), or False (0). In my case it's about if the user Clicked on the Ad (the PPC service), being a T or F, or 1 or 0.

The goal is to determine if Ad Clicks (PPC) are being used by internet surfers. This is very important in marketing strategies as this is a major expense in a business marketing plans i.e. Google Pay Per Click service. Each click on an Ad is a cost to a business deploying this service. Due to the high cost, there are business concerns as to the "worth" of such a service. For staring businesses, it is a very import to deploy Ad Clicks in the marketing campaign, but as a more established business that has a SEO presence, the question of the expensive Ad clicks or what is called Pay Per Click (PPC) becomes a legitimate question for the business.

This project will show the stakeholders the information about the percentages of Ad clicks as a general Internet activity.

Background

PPC is a sponsored search mechanism, originally introduced by Overture and formerly known as Goto.com back in 1996. Leading online multinationals, such as Google, Yahoo!, and others have made PPC a huge advertising business worth billions of dollars (Dellarocas, 2012).

Google's AdSense, ruling the keyword advertising business for the corporation, came into being in 2003 (Ratliff and Rubinfeld, 2010). Today it is easy to tell the PPC service on Google as they have a bold Ad by the link, any time one clicks on it, it creates a charge to the business based on the budget and bidding price the business had setup.

This process is basically an easy setting, it is about setting up Keywords and entering a bid price per word, meaning that for every word entered into the search engine by a user, your web link will potentially show up, and if the user clicks on it you, the web owner, will be

charged the fee you set in the bidding setup. The war of bidding is what determines the position of your website link as the keyword is entered, the highest bidder has top page space.

Search Engine Optimization on the other hand has been around since the mid-1990, Although it could be argued that SEO and all things search engine marketing began with the launch of the first website published in 1991, or perhaps when the first web search engine launched, the story of SEO "officially" begins a bit later, around 1997.(Loren Baker, 2021). Google's service include a smart algorithm that explores the Internet universe and "ranks" your site based on many factors, and that will eventually determine the page position of your website link as a natural search is made that has keywords in your site. This service is a key player in the Digital Marketing space, and can be free, or for sure much less than the PPC service.

Methodology

I will be using Python code to load a dataset in the format of a csv file with 1009 rows (Kaggle.com).

The goal is to determine if Ad Clicks are being used by internet surfers. This is very important in marketing strategies as this is a major expense in a business marketing plans i.e. Google Pay Per Click service. Each click on an Ad is a cost to a business deploying this service. Due to the high cost, there are business concerns as to the "worth" of such a service.

For staring businesses, it is a very import to deploy Ad Clicks in the marketing campaign, but as a more established business that has a SEO presence, the question of the expensive Ad clicks or what is called Pay Per Click (PPC) becomes a legitimate question for the business.

This project will show the stakeholders the information about the percentages of Ad clicks as a general Internet activity.

The strategy is for This project will look at several factors that will determine the activity of the Ad Clicks:

- Determine the general percentage of "Clicked on Ad" (Ad Clicks) vs general web activity
- Determine the ratio of web usage between male vs. female
- Determine the ratio of user Clicked on Ad between male vs. female, evaluate if there is a significant difference. This can help in determining on which campaigns it is worth utilizing Ad clicks.
- What is the Mean of Age for users
- What is the Mean of Age for users that Clicked on Ad, and by Gender
- What is the Mean of Age for users that did not Clicked on Ad, and by Gender
- The Mean of users the Spent Time on Site that Clicked on Ad
- The Mean of users the Spent Time on Site that did not Clicked on Ad

(1) Data Importing and Pre-processing

The process of the Extract, Transform, Load (ETL) for a csv file with 1009 rows, downloaded from Kaggle.com File name: advertising_ef

Fields of the dataset:

1. Daily Time Spent on Site: Use
2. Age: Use
3. Area Income: Not used
4. Daily Internet Usage: Use
5. Ad Topic Line: Use

6. City: Use
7. Gender: Use
8. Country: Use
9. Timestamp: Use – Need to separate date from time values.
10. Clicked on Ad: Use as T,F (1,0)

Importing Libraries

Importing the necessary libraries to read the dataset, doing the ETL

(See Python code): https://github.com/kavrangraves/AAI501-M7-PPC-Project.git

Importing Dataset

Actual Importing the dataset from the csv file and storing it in a dataframe

- advertising_ef.csv

Initial Data Summary

Review data field attributes and data types, ensure proper import and verify the field types

Organize Dataframes

Normalization of the data by removing unnecessary columns and replacing missing values

Data Cleaning and Normalization

Normalization of the data by removing unnecessary columns and replacing missing values

(2) Statistical Analysis and Visualization

Running the calculations to produce the required values:

- Determine the ratio of web usage between male vs. female
- Determine the ratio of user Clicked on Ad between male vs. female, evaluate if there is a significant difference. This can help in determining on which campaigns it is worth utilizing Ad clicks.
- Determine the general percentage of "Clicked on Ad" (Ad Clicks) vs general web activity
- Identify patterns to Click Ad = 1 (True)
- Generate Visuals (Plots) and Statistics that display all the analytic results

Final Dataframe Information

Filtering in on the final set of data used

Dataframe Description

Descriptive statistics of the dataframe, table of values

Dataframe Insights

Grouping Dataframe by Clicked on Ad and calculating the mean of each column sorted by Gender, Age: Show the ratio of the two as to Clicked on Ad

Dataframe Visualization (Grouped)

Multiple graph presentations to show the relations between all fields

Convert Categorical Data to Numerical Data

Using One Hot Encoding to convert categorical data to numerical data

Setting the Independent and Dependent Variables

Set the independent variables (X): Zip Codes, and dependent variable (y): Clicked on Ad

Results

Display all results

(3) Prepare Dataset for AI based calculations

My aim for this work is to examine the patterns, if at all, for the reason of a web surfer (User) to click on the Ad link. The examinations are directed towards the Users as in their Gender and Age as a primary value to show the correlation to the Clicked on Ad=1 or 0.

I will utilizing the sklearn libraries in Python, "Scikit-learn (formerly scikits.learn and also known as sklearn) is a free software machine learning library for the Python programming language.[3] It features various classification, regression and clustering algorithms including support-vector machines, random forests, gradient boosting, *k*-means and DBSCAN, and is designed to interoperate with the Python numerical and scientific libraries NumPy and SciPy." (Wikipedia)

- Load libraries
- Assign scaling to the df

Splitting the Dataframe

Splitting the dataframe into training and testing sets (75/25)

(4) Regression Model

Importing Libraries

Importing the necessary libraries to split the dataframe (sklearn)

Dataframe Modeling

Importing the necessary libraries to model the dataframe and Setting the model

Fitting the Model

Fit the model to the training set

SVM

Model Score

Scoring the model on the training set (R^2)

Model Score using OLS

Mean Absolute Error (MAE) and Mean Squared Error (MSE)

(5) Model Prediction Test (User Input)

Selecting a random row from the testing set and predicting the price

Run multiple sets

Results

Dataset rows: 1009

2. Male spend more time on the Web surfing in general

2. Male spend more time on a website they are on

3. The average age of the Male is younger than the Female surfers: Male 20-30 the majority while Female 35-45

4. Age 31 has the highest count for web surfers

5. The web surfing activity is almost 50-50 between the genders: F:%52, M:%48

5. The Clicked on Ad is also close to 50-50:M:501 F:508

6. The mean age for users that clicked on the ad is: 40

7. The mean age of users that did not click on the ad is: 32

8. the mean of the Daily Time Spent on Site is higher on the users that did not click on the ad: 76.8 vs. 53.27 that is a 23 point increase which is a %43 increase

9. the mean of the Daily Internet Usage is higher on the users that did not click on the ad: 214.56 vs. 145.73 that is a 68.83 point increase which is a %47 increase

10. Mean Age-General: 35.96 (36)

11. 25th Quantile Age: 29

12. Standard Deviation: 8.72

13. Margin of error: 0.54

14. %95 CI : (35.42, 36.50)

15. On a large scale of data the Test Predictions are >%50

Conclusion

The results show that the ratio of gender surfing the web in general is about an even 50-50 split, and that is the same result for the ratio of the user's gender that clicks on Ads. I did find it interesting that the average age of the males is younger than the female, but that is definitely a pattern for the results. The mean age is 36 which I found to be interesting as I assumed this age is the of the busier age group that would not have the time to surf the web and explore all the sites, but it seems I was wrong, another great pattern to use.

The results show that the PPC is still a powerful %50 of user's web surfing experience and that businesses should continue to use and budget into their marketing services. And as for directing their campaigns to a certain gender, that is, of course, based on the product, but in general it is a 50-50 split.

References

Sam Carr , *"Ad Words",* 2021, https://lunio.ai/blog/author/sam-carr/

Lisbeth Karltorp, *"Digital transformation strategies in small businesses",* 2017, Jonkoping University

Chrysanthos Dellarocas, *"Double Marginalization in Performance-Based Advertising: Implications and Solutions",* 2012, , Boston University

Loren Baker, "20+ Years of SEO: A Brief History of Search Engine Optimization", 2021, https://www.searchenginejournal.com/seo-guide/seo-history/#close

scikit-learn, https://en.wikipedia.org/wiki/Scikit-learn