AAI-501: Marketing: Pay Per Click Analytics

12.3.2022

Adam Graves

Github Link to Model: https://github.com/kavrangraves/AAI501-M7-PPC-Project.git

YouTube Presentation: https://youtu.be/LAg6XSaT_Pw

Table of Contents

Overview

Using a dataset of 1009 rows of data about female and male internet time spent and Ad clicks as T,F (1,0) International data. (kaggle.com)

Goals/Strategize

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

Business Need

FirstMedia.AI is a Southern California Marketing Agency specializing in digital media and building websites. This project is going to enable the 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) vs. Business Mapping is a big issue especially with new Start-Up businesses as the Ad Click (PPC) is very expensive and a close look needs to be taken to justify the return on investments (ROI).

Design

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)

(1) Importing Libraries

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

(See Python code)

- 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

(1.1) Organize Dataframes

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

(1.2) 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

2.2 Results

Display all results

3. Prepare Dataset for AI based calculations

- Load libraries
- Assign scaling to the df

(3.1) Splitting the Dataframe

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

4. Regression Model

(4.1) Importing Libraries

Importing the necessary libraries to split the dataframe (sklearn)

(4.2) Dataframe Modeling

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

(4.3) Fitting the Model

Fit the model to the training set

SVM

(4.4) Model Score

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

Model Score using OLS

(4.5) 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

(5.1) Run multiple sets

6. Conclusion

1. 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

Scope of Work

11/20/2022:

- Select the Dataset type for the project:
    - Selected the PPC dataset from Kaggle.com with 1009 rows
    - Submitted the Final Project outline

12/1/2022:

- Began the project plan build out:
    - Define the resources required
    - Start a Project Plan documentation
    - Identify the tasks and potential obstacles
- Download the Dataset: kaggle.com: advertising_ef.csv (1009 rows)
- Setup a shared GitHub repository: https://github.com/kavrangraves/AAI501-M7-PPC-Project
- Setup a Google Docs: Start the Project Management and documentation

12/2/2022:

- Setup Jupyter/Python environment
- Start project
    - Strategize the project scope
    - ETL, Organize df
    - Data Normalization

12/4/2022:

- Work on Python code
- Work on Project Scope documentation
- R&D work

12/5/2022:

- Continue code
    - Review current code progress
    - Evaluate alignment with the business requirements
    - Evaluate obstacles

12/6/2022:

- Continue code
    - Review current code progress

- Refine the reports: Highlight the IV and DV sections
- Verify the results

Update reports and documentation

12/7/2022:

- Finalize code
  - Review current code progress
  - Approve the graphs
  - Approve the numbers
  - Research outlier if any