# **Website Ad-Click Assignment**

# Content

Once the company launch their website, they’ll need to spread the word to potential customers. While there are several ways to do this, one of the most effective method is online advertising. Though advertising is often considered expensive, it’s actually quite feasible for companies on a budget. In fact, plenty of platforms let you advertise at zero cost, including online directories, social platforms and more. So, this company do advertisement on their website and they provided with historical website data by users with different countries located in different cities.

# **Objective**

## **Problem description:**

Ranging from high-profile platforms to local listings, online directories cover a range of audiences as well as additional details about your business or organization. This helps promote your website by boosting its search engine ranking, increasing the chances that they’ll get paid by Google by drawing traffic to google ad on the website. The goal of the case study is to Predict who is likely going to click on the Advertisement so it can contribute to the more revenue generation to the organization.

## **Recommended Project Steps & Guidelines:**

1. Identify the Problem Statement - what are you trying to solve?
2. Import the dataset and identify the Target variable in the data.

Identifying the type of variables: Identifying the nature of different columns (Continuous / Categorical ), removing garbage columns (if any) and conversion of categorical variables to factors if they are not in factors.

1. Data pre-processing: Checking and treating the missing values with appropriate measures. Checking the presence of outliers by creating boxplots and treating the outliers (if any).
2. Univariate and Bivariate Analysis: Explore each "Potential" predictor for distribution (visual analysis –histogram/bar plot) and also explore their relationship with the target variable (visual analysis –boxplot/grouped bar chart and statistical tests – ANOVA/Chi-square).
3. Feature selection: Finalize the set of potential predictors to be used in the logistic regression algorithm.
4. Splitting the data into train & test: Divide the data into two parts: Train sample (70%) and test sample (30%). The machine learning algorithm will be applied on the Train set and the model will be validated on the test set.
5. Model Building: Form the logistic regression model with the set of potential predictors identified from Exploratory data analysis and obtain the significant predictors.
6. Multicollinearity check: Check the presence of Multicollinearity in your final model and remove the variables with high multicollinearity one by one from your final model to arrive at the model which will be used to generate predictions on the test data.
7. Predictions: Using the obtained model, generate the prediction probabilities on the test data. Considering the threshold, obtain the predictions on the test data set.
8. Accuracy measures: Obtain the confusion matrix. Obtain the “overall” value of accuracy, Precision, Recall/Sensitivity, Specificity, Balanced accuracy, F1-score.

***Deliverables***

1. Codes file.
2. Approach, steps, results found and their interpretation, significance of the variables and their business meaning etc. in Word doc.
3. PowerPoint containing Objective, significant variables and results obtained in laymen language, business recommendations (if any).

# Variable Description

This data set contains the following features:

**'Time\_Spent'**: Average time spent by user on site in minutes

**'Age'**: User age in years

**'Area\_Income'**: Average Income of geographical area of user

**'Internet\_Usage'**: Average minutes a day user spent on the internet

**'Ad\_Topic’**: Headline of the advertisement

**'Country\_Name'**: Country of user

**'City\_Code'**: City of user

**'Male'**: Whether or not user was male

**‘Time\_Period’**: Time at which consumer clicked on Ad

**‘Weekday’**: Name of the day

**‘Month’**: Name of the months

**‘Year’**: Which year the data is collected

**'Clicked'**: 0 means not clicked and 1 means that user clicked the Ad.