Logistic Regression

Certification Project

edureka!

Version 2.0

About the data:

A Retail Company has a dataset with 10 variables and more than 1700 customers of which over 350 bought some merchandise.

recency	history_se	history	mens	womens	zip_code	newbie	channel	segment	visit	conversio	spend
10	2) $100 - $	142.44	1	0	Surburbar	0	Phone	Womens E	0	0	0
6	3) $200 - $	329.08	1	1	Rural	1	Web	No E-Mail	0	0	0
7	2) $100 - $	180.65	0	1	Surburbar	1	Web	Womens E	0	0	0
9	5) $500 - $	675.83	1	0	Rural	1	Web	Mens E-M	0	0	0
2	1) $0 - $10	45.34	1	0	Urban	0	Web	Womens E	0	0	0
6	2) $100 - $	134.83	0	1	Surburbar	0	Phone	Womens E	1	0	0

Variable Name	Description
Cust Id	Unique Id to identify each customer
Months Since Last Buy	Number of months that have gone by since last purchase was made
Spend Numeric	The amount spent by the customer during last year
Spend Category	The amount spent by the customer during last year, captured into buckets
Men's Merchandise	Whether the customer purchased men's merchandise
Women's Merchandise	Whether the customer purchased women's merchandise
Area	Does the customer belong Rural, Urban part of the city
New Customer	Whether the customer has been acquired newly or an existing customer
Purchase Channel	Whether the customer made a purchase through Phone or Website
Sale Made	(Dependent Variable) Whether the customer actually bought anything

A retailer, with all the promotional campaigns along with all the fancy discounts and attractions, would want to make a "Sale Conversion" at the end. Would you not agree to this? Any retailer's ultimate objective is to lure a certain customer into buying something. This is, in layman terms, known as "making a sale".

Making a sale has many aspects to it, such as marketing a product, pricing a product, tempting customers with high discounts, etc. However, the inevitable question is, how do we know which customer would buy and which customer would not!

Problem: Now the question what we would want you to answer at the end of this project is to identify potential buyers from the customer base of a retail chain company i.e. understand the dataset and design a model in a way to classify between the customers who would want to buy the product or not.

Hint: You can use any algorithms/ techniques to address this classification problem

You can use the below steps while working on this project (but it's not mandatory).

→ Loading Packages, Setting up Working Directory and Reading the Dataset
→ Exploratory Data Analysis (EDA)
→ Data Cleaning, Modification and Sampling
→ Model Iterations
→ Model Validation

Dataset Used: You can find the dataset in the below link:
http://blog.minethatdata.com/2008/03/minethatdata-e-mail-analytics-and-data.html

The dataset can also be found in your LMS.