PREDIXION ANALYTICS CLUB
INDIAN INSTITUTE OF MANAGEMENT NAGPUR

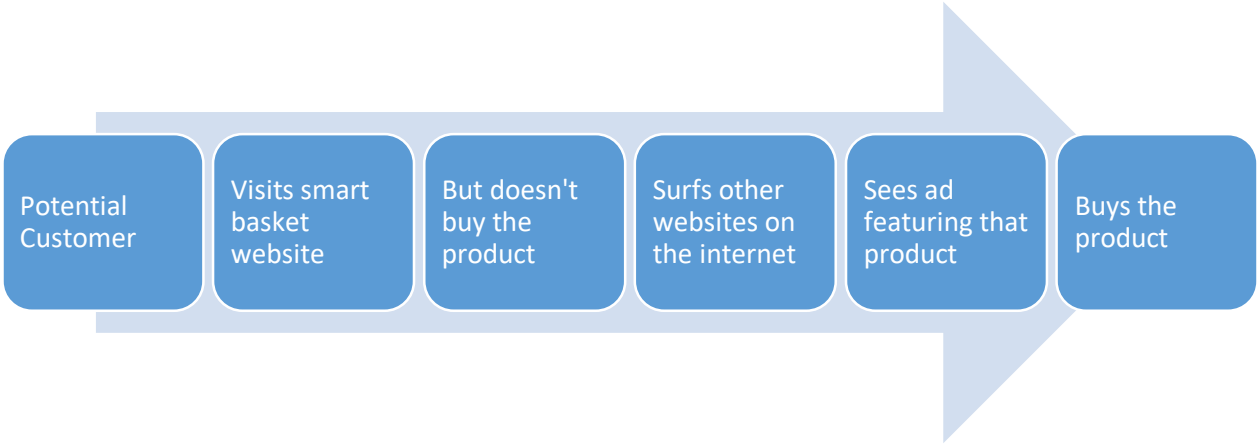
Smart Basket

On average, a user visits more than a dozen of websites in one sitting while he or she surfs the internet. This journey of customers from one channel to another forms a sequence that contains a lot of hidden information. The online customer journey can be leveraged in several ways by marketers to gain sales and revenue. This digital footprint of customers hence needs to be recorded and forms a crucial part of the big data. It is called clickstream data. It can be used to understand customer behavior by obtaining important information like the user's demographics, browsing history, interests, purchasing habits, and online activities. The commercial companies can reverse engineer this sequence to find answers to questions like which channels triggered the customer to start his/her journey and use predictive modeling techniques to forecast future behavior.

The E-commerce company, Smart Basket, offers a diverse product mix to its customers. It uses clickstream data to promote sales. Often customers view a product and do not buy. The company keeps a log of this data and uses this information to advertise its products on the other websites the customer visits. For Example, Customer A searches a given product on Smart Basket website but doesn't buy it. He then logs in to Facebook or other social media handles. The product starts to appear in the advertisement views to the customer. The potential customer then clicks back on this advertisement to be redirected to the Smart Basket website again, on the particular product page. Generally, this time, conversion happens, i.e., the customer buys the products.

Smart Basket hires another company, ABC Ltd., to advertise its products on different online platforms using this data. ABC builds the required system to show relevant advertisements to potential customers on its partner websites. Hence it is an important aspect of marketing to improve the sales performance of Smart Basket.

The following diagram explains the above-mentioned process:



Note:- This case was prepared by Predixion- The Analytics Club of IIM Nagpur solely for use in Brain-A-Lytics 2.0. This is a hypothetical case and is not intended to serve endorsements, sources of primary data, or illustrations of effective or ineffective management

Submit the following 2 documents on D2C by 20th November:

1. test.csv sheet with values for 'click_flag' for each impression_id
2. A PowerPoint presentation - maximum of eight slides briefing the following:
    a. Insights gained via exploratory data analysis
    b. Assumptions made
    c. Reasoning / logic for approach & method selected
    d. Validation methodology
    e. Results obtained from modeling
    f. A summary slide

Data Description

- The train data (train.csv) along with the label which specifies whether the ad is clicked or not is for the duration of 15th Nov 2029 to 6th Dec 2029
- Historical view log file for each session (view_log.csv) is from 15th Nov 2029 to 6th Dec 2029.
- Your model will be evaluated on Test data(test.csv) which is without label. Duration of test data is 7th Dec 2029 to 13th Dec 2019.

Table Name	Variable	Definition
item_data.csv	item_id	Item ID
	item_price	Price of the Item
	product_type	Anonymized item type
train.csv	impression_id	Ad impression ID
	time_stamp	Time of impression at partner website
	cust_id	Customer or user ID
	app_code	Application code for partner website where the ad was shown
	os_version	Operating System Version (3 levels)
	lte_flag	1 - Using 4G, 0 - Not using 4G
	click_flag	1 – User clicked the ad, 0 –User didn't click the ad
view_log.csv	server_time	Time Stamp of log
	device_type	Device type of use (3 levels)
	session_id	Browser session ID
	cust_id	Customer or user ID
	item_id	Item ID
test.csv	impression_id	Ad impression ID
	time_stamp	Time of impression at partner website
	cust_id	Customer or user ID
	app_code	Application code for partner website where the ad was shown
	os_version	Operating System Version (3 levels)
	lte_flag	1 - Using 4G, 0 - Not using 4G

Note:- This case was prepared by Predixion- The Analytics Club of IIM Nagpur solely for use in Brain-A-Lytics 2.0. This is a hypothetical case and is not intended to serve endorsements, sources of primary data, or illustrations of effective or ineffective management