

# CLICKSTREAM PROJECT

Final Capstone Project  
Morteza Moghaddam

# Capstone Project Milestone Report - Intro

## Capstone Project Milestone Report - Intro

- Clickstream data related to the path the user takes when navigating through the site. During a user journey, a lot of info, articles, and product details will be available to help a user in making a decision. What are the most viewed pages before users take some actions on the page? Is there any relationship between the pages users view and actions they take on the website? Is there any relationship between the pages users view and products they choose on the website?

## Goal and Questions

- The website is available 24/7 and there are 5 main products available to users. There are also other pages such as articles, product details and services available to help a user choose a product. We are trying to answer some questions such as:
- 1- Is there any difference between the user behavior in 1st visit versus later visits?
- 2- Does a number of visit impact on the behaviour? For example, a user with 100 visits vs a user with 20 visits
- 3- Do we have a path that gives us 100% click rate on the products?
- 4- Can we profile the user based on the behavior, metrics or dimensions?
- 5- Is the type of product important?
- 6- What are the characteristics of users who did not click?
- 7- What are the characteristics of users who viewed the “result” page and click on the product vs users who did not? Clicks and pageviews before result matter?

# Structure of the Project

## Structure of the Project

The project is structured in 4 big analysis parts.

**First**, we wrangled the data that we were going to use by cleaning and converting the variables in the dataset. The source of the data is clickstream data for 3 month from Apr 2016 to Jul 2016. We started with 41 variables and 1,031,246 rows. By doing feature engineering, we could have 86 variables and the number of rows went down to 140,336 after cleaning the data.

**Second**, we explored the data we work on and perform a detailed descriptive analysis that splits in two different parts. In the first one, the reader can understand what each variable is adding to the dataset and how it is structured. In the second one, we see how each variable is related to our dependent variable, which is “clicking on the product” that helps the reader figure out its relation and helps us understand the importance of each variable for our following analysis. We prepared our final dataset based on this exploratory data analysis and we started the next step with 27 variables and 4555 rows.

**Third**, the reader can see market basket analysis (association rules) that help us to find some business recommendations because we can see the influence of each variable towards the dependent variable. We fitted our model with 19 variables as a result of previous steps

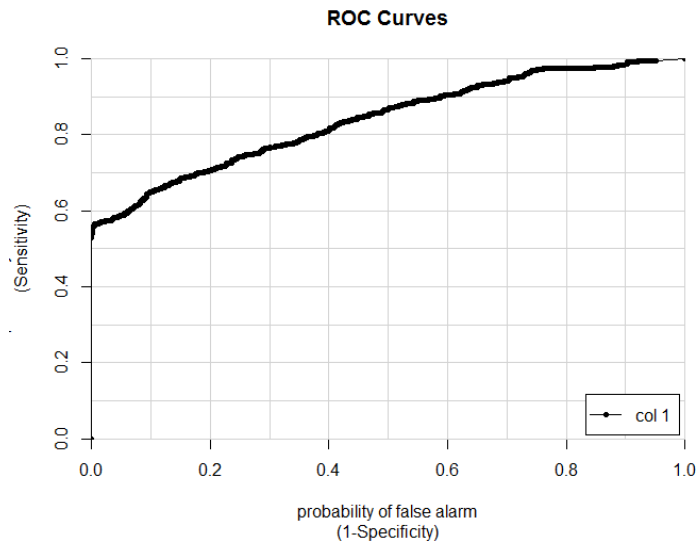
**Fourth**, we perform a logistic regression using a binary target variable `product_clicked`

We would like to mention that, a priori, we expect getting similar conclusions in our different analysis. We believe that since we want to find out which variables are the most important ones in predicting, all the analysis that we are performing tends to give us similar results.

# Logistic Regression

## Logistic Regression

- The goal of logistic regression is to correctly predict the category of outcome for individual cases using the most parsimonious model. To accomplish this goal, a model is created that includes all predictor variables that are useful in predicting the response variable. So, we want to know which variables can help us to predict about our target variable `product_clicked`.
- Our model has the accuracy of 79.5% on the training dataset and the evaluation of the test data shows the accuracy of %75. The sensitivity of our model is 89% and the threshold we chose based on the ROC was 0.65. The AUC was also 0.839 which is a good AUC.



# Answers to our Questions

## Answers to our Questions

We could find answers to some of our questions.

### **1- Is there any difference between the user behavior in the 1st visit versus later visits?**

- Our exploratory data analysis shows that visitors who are visiting the website for the first time account for 65% of total visitors. They tend to spend more time, view more pages, take more actions.
- We noticed that there is positive correlations between the pageviews and taking actions. The more users view pages the more they take actions.
- 75% of mobile users visited the website for the first time which was 14% more than what was expected
- 70% of no clicks on the products were done by the first timers

### **2- Does a number of visit impact on the behavior? For example, a user with 100 visits vs a user with 20 visits?**

- By grouping the visit frequency number to different categories, we realized that users with medium/high frequency of visit, spend less time on the site, view fewer pages and take fewer actions than first-time users.

### **3- Do we have a path that gives us 100% click rate on the products?**

- Both market basket analysis and exploratory data analysis shows that there are a specific group of users that took specific paths and got 100% conversion rates. Our regression model also shows their characteristics are unique and statistically significant. For example, some of them started their journey from a specific path and some of them referred by the specific sources. (Because of confidentiality and privacy policy we can't share more findings on this point)

### **4- Can we profile the user based on the behavior, metrics or dimensions?**

- Both market basket analysis and exploratory data analysis shows that there are a specific group of users that took specific paths and got 100% conversion rates and also there are some unique characteristics such as source and clicking on some features on the page that impact the user behavior in terms of not clicking on the product.

# Answers to our Questions

## **5- Is the type of product important?**

- In order to answer this question, we will need further analysis specific to the type of products. This can be part of our future analysis and recommendations.

## **6- What are the characteristics of users who did not click?**

- Our market basket analysis shows that there are some dimensions such as source, device browser, visit frequency and device category that impact the clicks on the product. (Because of confidentiality and privacy policy we can't share more findings on this point)

## **7- What are the characteristics of users who viewed the “result” page and click on the product vs users who did not? Do clicks and pageviews before result matter?**

- We already answered to some part of this question in the previous questions. Viewing some section of the website such as articles and product detail pages impact the later behavior of the users in terms of clicking. One of the key findings was if we direct some of the incoming traffic to specific pages and users start their journey from those pages, we will have better conversion rates and higher clicks on the product. Based on our logistic regression, assuming all the coefficients are equal, if we increase the number of visits to 3 specific pages by 1, we will increase the number of clicks to 4 for each entrance.

# Recommendations

## Recommendations

- 1- For paid traffic and paid ads, it is better to link the traffic to specific product pages as they will increase the engagement and product familiarity which will lead to more clicks on the products.
- 2- There are specific blogs, referrers, and websites that send visitors to the pages. There is a good opportunity to invest money in this segment as they have significant impact on the conversions
- 3- There is no need to change the flow in other sections such as articles, tools and contact us as they have no statistically significant impact on the conversions.
- 4- Traffic that comes from the emails have high statistically significant conversions and further investment may be a good choice on targeting the customers via email.
- 5- New visitors spend more time on the pages and there is an opportunity for the company to increase their conversions by showing them the product detail pages and educate them. This will lead to high-quality conversions and clicks on the products.

**(Because of confidentiality and privacy policy we can't share more findings on this point)**

## Further Research

- There should be a further study on the visitors who visit the website more than 6 times but less than 20. This segment has no interests on the product itself but the fact that they are visiting the pages, again and again, shows they are looking for other info. Understanding these visitors may have some good impacts on their user experience.
- Any analysis to measure ROI, any marketing activities such as online ads, social media posts and marketing automation process such as weekly, daily and monthly emails should be targeted based on web analytics modeling to increase the conversions and to measure the effectiveness of marketing channels.
-

# Further Research

## **Further Research**

- There should be a further study on the visitors who visit the website more than 6 times but less than 20. This segment has no interests on the product itself but the fact that they are visiting the pages, again and again, shows they are looking for other info. Understanding these visitors may have some good impacts on their user experience.
- Any analysis to measure ROI, any marketing activities such as online ads, social media posts and marketing automation process such as weekly, daily and monthly emails should be targeted based on web analytics modeling to increase the conversions and to measure the effectiveness of marketing channels.