ABSTRACT:

In this project, we are analysing user behavior and optimizing user work-flow for an ecommerce website using a suitable machine learning algorithm.

INTRODUCTION:

With latest trends more and more people are indulging in online shopping and with time these numbers are bound to increase.
Understanding how customers act as they navigate to and through an ecommerce site is the first step in building an e-commerce company that will thrive. One needs to supplement their own business intuition with careful analysis of customer/user behavior to come up with a solid data-informed strategy that will instill confidence in the team to take the company to the next level and overcome the shortcomings that the company is facing by optimising the site according to user behavior.
Google Analytics is a web analytics service offered by Google that tracks and reports website traffic.
In this project to understand the user behavior we first had to create an ecommerce website of our choice. After creating the website, in order to collect some user data we integrated a web analytics tool with the website. In our case, we chose Google analytics to collect user data for our website. After collecting the user data of about 125 people, we prepared a spreadsheet of all the collected data and processed it for according to our requirement. We then researched for a machine learning algorithm which can help us analyze this clickstream data, and zeroed in on using Markov chain algorithm. With the use of Markov Chain algorithm on our data, we were able to gain significant insights.

THE WEBSITE:

My website - www.forthefit.shop.
Forthefit is an ecommerce website for purchase of fashion products for women like dresses, shoes, t-shirts, and accessories.

Steps I took to create the website:

- I got the domain name forthefit.shop.
- I started building the website using free website building tool from wix.com. I selected their ecommerce theme and customized the background, layout and product arrangement.
- The website contains following pages :-

- Home : Consists of information about the website.
- Shop : Consisting of all the products on the website, filters.
- Cart : Consists of the products that are added to the cart by the users.
- Account : Consists of account details of the user like address, wallet information and order details.

My website has the following features:

- View a list of products
- View product details
- Search products
- Use filters to change the product list (eg. Category, price range, etc.)
- Add a product to the cart.
- Contact Us

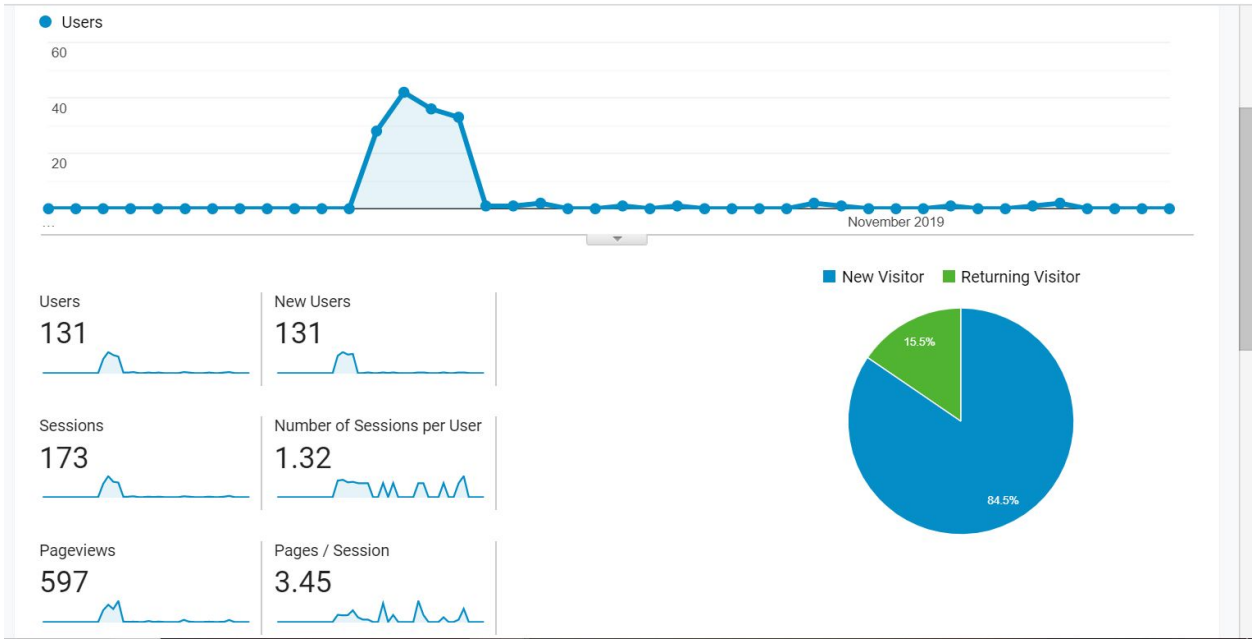
User Guidance : I used the light box tool provided by Wix.com to create instructions for the users on how to navigate the site, the instructions were created for each page of the website in order to help users to navigate the website easily.

ANALYTICS:

I used Google Analytics as our web analytics tool. It helped us gain insights into aspects like:
- during what days the maximum number of people visited the website
- browser used by the user
- location of the user
- how much time user spent on a particular page
- which were the exit pages i.e. where the user left the website
- bounce rate - percentage of users who left the website after visiting only one page
- active users at a particular time

I collected the data of over 125 users.

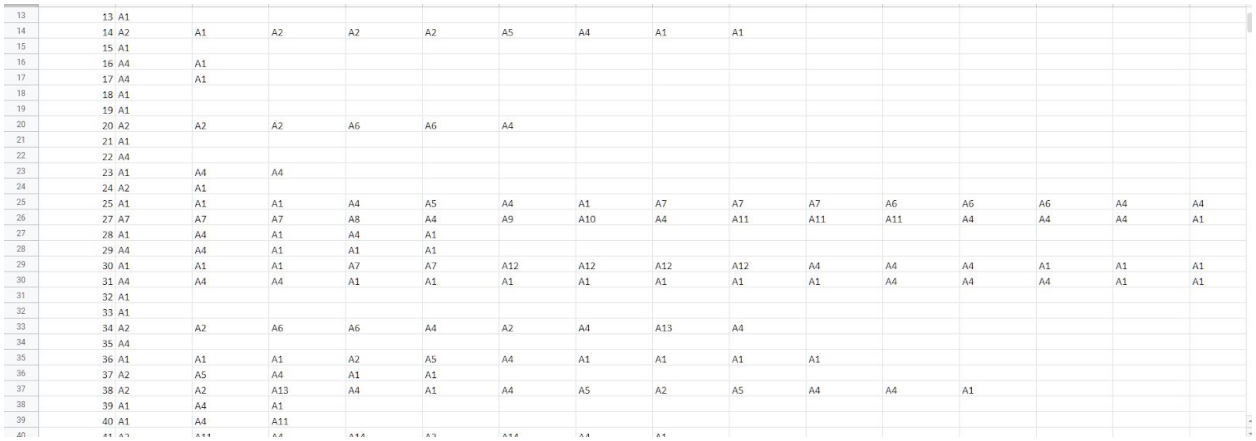
Below is the overview of user behavior:



Below is the spreadsheet of user data collected from the website:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	date	USER_ID	hasGoal	hasRevenue	sessionCount	sessions__durati	sessions__devic	sessions__chan nel	sessions__activi tySummary__E VENT	sessions__activi tySummary__P AGEVIEW	sessions__activi ties__time	sessions__activi ties__type	sessions__activities__det ails__Page title__-	sessions__activi ties__details__ Page URL__-	sessions__activi ties__pageTitle
2	Oct 13, 2019	1	FALSE	FALSE	1	0:00	mobile	Direct	1	8:02 PM	PAGEVIEW	Home | FortheFit	/	Home | FortheFit	
3		1													
4	Oct 14, 2019	2	FALSE	FALSE	1	1:24	mobile	Direct	4	7	3:40 PM	PAGEVIEW			
5		2									3:40 PM	PAGEVIEW			
6		2									3:40 PM	PAGEVIEW			
7		2									3:40 PM	EVENT	Enhanced Ecommerce - Stores	Add to Cart	Off-Shoulder white Dress
8		2									3:40 PM	PAGEVIEW			
9		2									3:40 PM	EVENT	Enhanced Ecommerce - Stores	View Content	Off-Shoulder white Dress
10		2									3:40 PM	PAGEVIEW			
11		2									3:40 PM	EVENT	Enhanced Ecommerce - Stores	Product Click	Off-Shoulder white Dress
12		2									3:39 PM	EVENT	Enhanced Ecommerce - Stores	Product Impressions	(not set)

Below is a sample of the data preparation to implement the algorithm:



13	A1														
14	A2	A1	A2	A2	A2	A5	A4	A1	A1						
15	A1														
16	A4	A1													
17	A4	A1													
18	A1														
19	A1														
20	A2	A2	A2	A6	A6	A4									
21	A1														
22	A4														
23	A1	A4	A4												
24	A2	A1													
25	A1	A1	A1	A4	A5	A4	A1	A7	A7	A7	A6	A6	A6	A4	A4
26	A7	A7	A7	A8	A4	A9	A10	A4	A11	A11	A11	A4	A4	A4	A1
27	A1	A4	A1	A4	A1										
28	A4	A4	A1	A1	A1										
29	A1	A1	A1	A7	A7	A12	A12	A12	A12	A4	A4	A4	A1	A1	A1
30	A4	A4	A4	A1	A1	A1	A1	A1	A1	A1	A4	A4	A4	A1	A1
31	A1														
32	A1														
33	A2	A2	A6	A6	A4	A2	A4	A13	A4						
34	A4														
35	A1	A1	A1	A2	A5	A4	A1	A1	A1	A1					
36	A2	A5	A4	A1	A1										
37	A2	A2	A13	A4	A1	A4	A5	A2	A5	A4	A4	A1			
38	A1	A4	A1												
39	A1	A4	A11												

ALGORITHM:

The elusive clickstream data. Many platforms, like Facebook, rely on these generated data from what a user clicks and what doesn't. To start analyzing clickstream data, we need first to be able to capture step by step a user's activity across a web page or application. And that is of great value in the hands of any Internet marketer.
I used Markov chain Algorithm for our clickstream data.

Markov Chains work with sequential data, which is the type of data we're dealing with in this post.

The Markov process is a stochastic process that satisfies the Markov property of memorylessness. A Markov chain is a Markov process in either discrete or continuous time with a countable state space.

Markov Chain is a stochastic process containing random variables, transitioning from one state to another depending on certain assumptions and definite probabilistic rules.

These random variables transition from one to state to the other, based on an important mathematical property called Markov Property.

What Is The Markov Property?

Discrete Time Markov Property states that the calculated probability of a random process transitioning to the next possible state is only dependent on the current state and time and it is independent of the series of states that preceded it.

The fact that the next possible action/ state of a random process does not depend on the sequence of prior states, renders Markov chains as a memory-less process that solely depends on the current state/action of a variable.

Let’s derive this mathematically:

Let the random process be, {Xm, m=0,1,2,⋯}.

This process is a Markov chain only if,

$$P(X_{m+1} = j|X_m = i, X_{m-1} = i_{m-1}, \dots, X_0 = i_0) = P(X_{m+1} = j|X_m = i)$$

Markov Chain – Introduction To Markov Chains – Edureka

for all m, j, i, i0, i1, ⋯ im−1

For a finite number of states, S={0, 1, 2, ⋯, r}, this is called a finite Markov chain.

P(Xm+1 = j|Xm = i) here represents the transition probabilities to transition from one state to the other. Here, we’re assuming that the transition probabilities are independent of time.

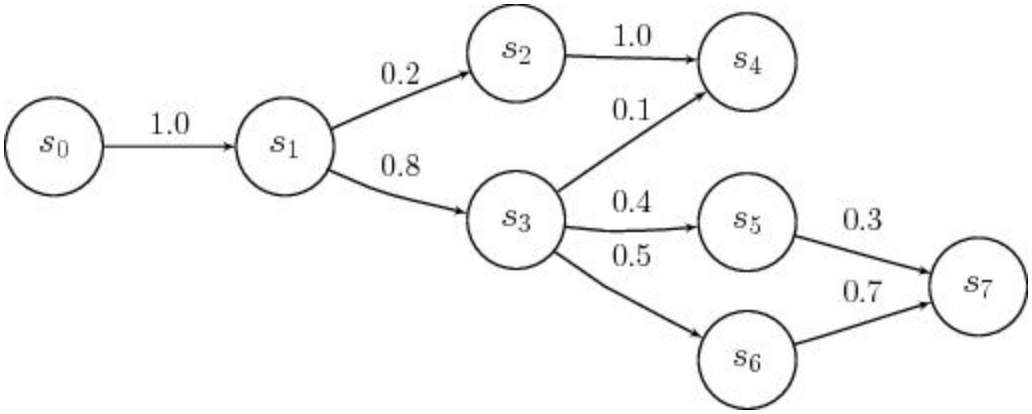
Which means that P(Xm+1 = j|Xm = i) does not depend on the value of ‘m’. Therefore, we can summarise,

$$P_{ij} = P(X_{m+1} = j|X_m = i)$$

Markov Chain Formula – Introduction To Markov Chains – Edureka

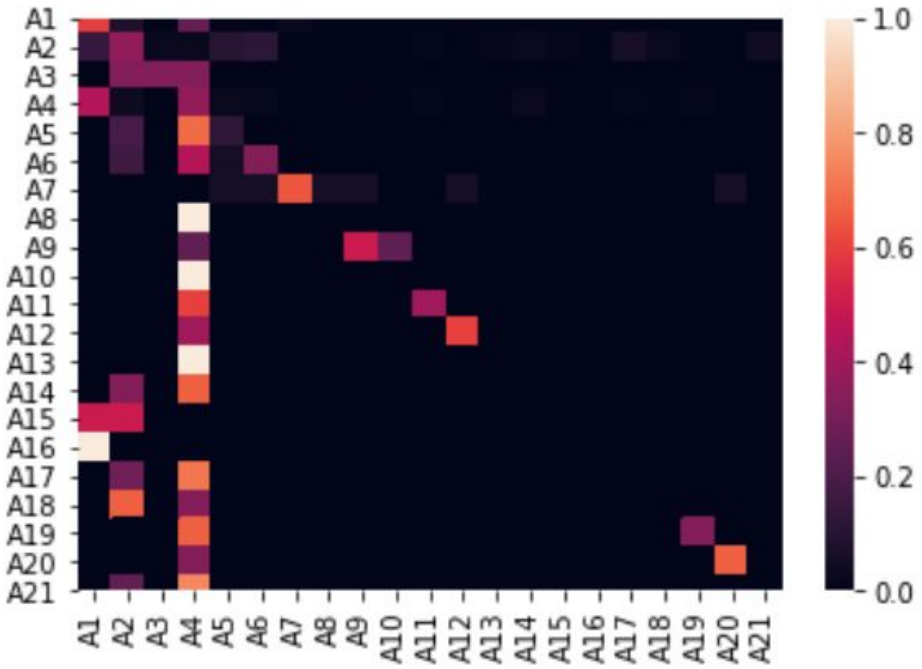
So this equation represents the Markov chain.

It can be graphically represented as a transition diagram along with the corresponding probabilities:



In clickstream analysis, we usually utilize these Markov Chains. The order of a Markov Chain is derived from the number of recent states on which the current state, we assume, depends. Based on this, zero-order chains imply that the probability of being in a state in the next step is independent of all previous states.

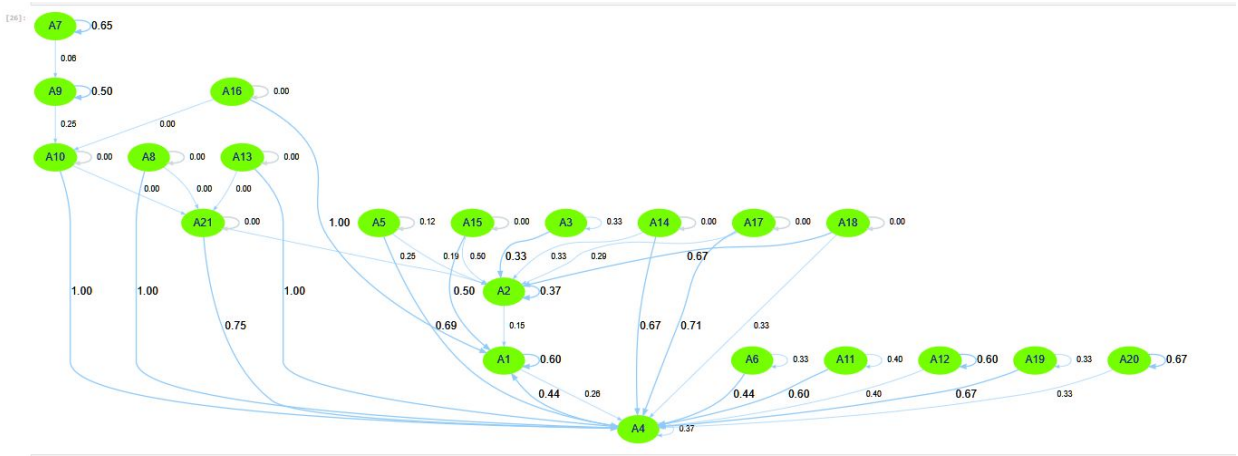
The transition probability matrix can be represented as a heat map with the y-axis representing the current state and x-axis the next one. The more bullish the colour, the more probable the indicated transition will occur.



Heat map:

The heatmap indicates the probability matrix,the y- axis represents the current page and the x- axis represents the next page. The lighter the color the more the probability of user to go from one page to another.

And finally, we got a transaction map too, that shows the probability of user going from one page to another:



#The nodes representing the individual pages are shown in green, and up to 3 edges from each node are rendered.

#The first edge is in a thick blue arrow, depicting the most likely transition from this page / state to the next page / state.

#The second edge depicted by a thinner blue arrow, depicts the second most likely transition from this state.

#Finally, a third edge is shown that depicts the transition from this page / state back to itself (light grey).

#This edge is only shown if the the two most likely transitions are not already to itself. For all transitions, the probability is shown next to the edge (arrow).

SUGGESTIONS FOR THE UI/UX DESIGNER:

1. Since we want users to follow this path :

   Home > Shop > Product > Cart Page > Payment Page, the crop tops page which is disrupting this path and has more probability of users going to the home page instead of the cart page needs to be dealt with by placing the filter at the very end in the filters list.

2. Since, the shop page was the most visited page, we can make it even more easily accessible to the users by making the shop page icon bigger or moving it to a position where it is right in the sight of the user.
3. According to the transaction diagram, there are a hundred percent chance that user selected the blue lace semi formal dress, or the brown strappy flats or the brown wedge heels are going back to the home page. So, the designer can place these products on the top of the product list or add others in suggestions when user opens one of these. This will allow the users to select these products easily without going back a page and thus, will increase sale of the products.
4. Since, many users are going back from the products to the shop page, we should make the add to cart or payment button larger for the user to not go back and also, include suggestions below the main product.
5. There is a hundred percent chance that a user viewing the “white crop top” will go back to the home page instead of the payment page or the shop page, we can avoid that by adding more attractive instructions on this product page navigating the user towards the cart page or the payment page.
6. Or by placing the “white crop top” at the very bottom of the list so that users visit the products with lesser probability of going back to the home page. Because if users keep going back to the home page it will negatively affect the business since the user is neither buying this nor visiting other products of the website.
7. We can also observe that there is a 67 percent chance that a user visiting the “Yellow Graphic Tee” will go to the cart page, so we should place this t-shirt at the top of the product list.
8. For some products like the “Grey graphic tee” and “Antique Gold Necklace” the probability of reloading this page is more than 60 percent and also have greater probability of going back to the Shop page, this indicates that users are having trouble going forward to cart page/payment page from these two product pages and thus we should check the “Add to cart” button on these pages.