ABSTRACT

Trading, via barter exchange or purchasing and selling the goods has been there for centuries. Since the demand and supply of goods and services are increasing at an enormous rate, it has been really hard to keep the track hold of many transactions. As these transactions have always existed in trading for centuries. But in the current growing digital world the level of handling such data is a difficult task. Thus , E-commerce came into existence , it is simply defined as the sale and purchase of services and goods over an electronic platform, such as the internet. It comprises digital data and funds between several parties. To put it in a simple way, we call it online shopping. It all began in the 1960s when organizations began to use Electronic Data Interchange (EDI) to transfer documents of their business back and forth. The 1990s saw the emergence of online shopping businesses, which is quite a phenomenon today. The first ever online purchase was a Sting CD, sold by US retailer, NetMarket on 11 August 1994. Since then E-commerce is growing day by day. E-Commerce websites are online portals that facilitate online transactions of goods and services through means of the transfer of information and funds over the Internet. In the early days, e-commerce was done partially through emails and phone calls. Now, with a single website, anything and everything that a transaction needs, can be executed online. The popular E-commerce websites are Amazon, Flipkart, Ebay, Alibaba, Walmart etc. The user's behaviour on these websites are analysed by the recommendation engine, and the data that is collected from the user is put into data analytics. This is done to analyse the workflow and behaviour of the user, based on which user's interests are determined. Every order that is placed, every product that is visited, every part the user explores gives us the digital interests and behaviour of the user, for the site engines to predict what will be the best to suggest the user to buy. Since Machine Learning makes it easier to analyse these data, we make a model based on the data. We train that model to predict the output , we create a best suited algorithm for the machine to implement and we run the program in the backend. Producing the reports and the data set for analysis.

INTRODUCTION

E-Commerce websites are online portals that facilitate online transactions of goods and services through means of the transfer of information and funds over the Internet. In the early days, e-commerce was done partially through emails and phone calls. Now, with a single website, anything and everything that a transaction needs, can be executed online.

"We see our customers as invited guests to a party, and we are the hosts. It's our job every day to make every important aspect of the customer experience a little bit better." - Jeff Bezos

A website user is a person who is accessing, browsing or interacting with a website, and user behavior refers to how people use a website. Behaviors include everything from the journey they take through the site to interactions such as clicks. When it comes to optimising a website, simply monitoring behavior can only get you so far. The real value comes from analysing users' actions to get to the bottom of what makes them behave as they do.

Behaviour is complex and varies across different websites depending on the target audience. This means you need to learn specifically about your users. Who are they? What are their needs? Which browsers and devices do they prefer? How often do they purchase? Answering these questions is crucial if you want to have a competitive edge, meet consumer needs and retain your customers. By researching online behaviours, you can get an idea of what users are trying to achieve, the factors driving certain behaviours, where they experience friction and areas where user experience can be better. Ultimately, learning how visitors behave on your website allows you to provide an enhanced experience that's in line with user needs, which in turn will ensure your business continues to grow.

One of the main tools we use to carry out research is Google Analytics. This is free and fairly easy to set up. Once installed, you'll have access to valuable data about how users behave on your site, including where they land and go next, where they drop off and what they interact with. You can also use it to discover overall trends and patterns and source opportunities for growth.

Web Application

E-commerce websites are online portals that facilitate online transactions of goods and services through means of the transfer of information and funds over the Internet. In the early days, e-Commerce was done partially through emails and phone calls. Now, with a single website, anything and everything that a transaction needs, can be executed online.

My website is only a dummy site, thus some features don't work due to the lack of Database in it. This site was made for research purposes only.

I have used following steps to create my E-commerce website:

- ➔ Firstly, I used HTML to create a structure of my webpages, and connected all of them together.
- ➔ Second, I used CSS via Bootstrap to add style to my webpages, in order to make the website look more attractive.
- ➔ Then, I used JavaScript to add actions and functionality to my website to work accordingly.
- ➔ My website contains following webpages:
  - ◆ a) Home- It is the homepage of my website.
  - ◆ b) Store- It contains various products categorized into different categories.
  - ◆ c) Account- It contains account details of the user.
  - ◆ d) Contact Us- It contains the contact details of the company.
  - ◆ e) About- It contains team details.
- ➔ Website contains various features:
  - ◆ a) View a list of products
  - ◆ b) View product details
  - ◆ c) Search products
  - ◆ d) Use filters to change the product list (eg. Category, price range, etc.)
  - ◆ e) Add a product to the cart.
  - ◆ f) Payment Gateway.
- ➔ I used netlify to host my website as “proxp.netlify.app”. Hosting on netlify is free of cost and it’s just drag and drop of project folder.

Data Analytics

User Behavior Analytics or UBA focuses on what the user is doing: apps launched, network activity, and, most critically, files accessed (when the file or email was touched, who touched it, what was done with it and how frequently).

UBA technology searches for patterns of usage that indicate unusual or anomalous behavior — regardless of whether the activities are coming from a hacker, insider, or even malware or other processes. While UBA won't prevent hackers or insiders from getting into your system, it can quickly spot their work and minimize damage.

I have integrated my website to an analytics platform named Google Analytics. This tool is able to track the users action and the time spent by the users on the various actions of the website. I have accumulated the analytics data from various users and stored it.

I have collected data of around 102 users with various data points such as:

- Time spent by users on the different pages on website,
- Filters being used by the users ,
- Time spent by users on a particular product page, etc.

Audience Overview

Page	Pageviews	Unique Pageviews	Avg. Time on Page	Entrances	Bounce Rate	% Exit
/	161	126	0:02:33	125	56.80%	59.01%
/cart.html	57	27	0:00:08	2	100.00%	12.28%
/checkout.html	50	24	0:00:11	3	100.00%	16.00%
/shop.html	48	21	0:01:23	0	0.00%	10.42%
/login.html	44	21	0:00:38	1	0.00%	13.64%
/contact-us.html	18	11	0:00:05	0	0.00%	16.67%
/index.html	17	11	0:00:24	0	0.00%	23.53%
/404.html	16	13	0:00:16	0	0.00%	6.25%
/product-details.html	15	13	0:00:05	1	100.00%	13.33%
/blog-single.html	9	8	0:00:07	0	0.00%	11.11%
	436	276	0:00:55	133	58.65%	30.50%

Analytics All Website Data Pages

1	A1	A1																	
2	A1	A1	A1	A1	A1	A1	A1												
3	A2	A2	A2	A1	A3	A1													
4	A1	A1	A1	A1	A1	A1													
5	A1	A1	A1	A4	A5	A3	A4	A5	A6	A7	A2	A8	A9	A10	A1				
6	A1	A7	A7	A2	A8	A1													
7	A1	A1																	
8	A1	A1																	
9	A3	A4	A5	A6	A7	A2	A8	A9	A10	A1	A1								
10	A1	A7	A2	A2	A2	A2	A8	A10	A10	A10	A10	A10	A10	A10	A10	A10	A10	A10	A1
11	A1	A1	A10	A10	A10	A1													
12	A1	A1																	
13	A8	A8	A8	A8	A1														
14	A1	A1																	
15	A1	A1	A1																
16	A7	A2	A8	A9	A10	A7	A8	A8	A2	A1	A1								
17	A1	A1	A1																
18	A1	A1	A1																
19	A1	A7	A7	A2	A2	A2	A1	A5	A1	A1									
20	A1																		
21	A1	A1																	
22	A1																		
23	A1	A10	A1																
24	A1																		
25	A5	A3	A6	A7	A2	A8	A9	A10	A1										
26	A1	A11	A1	A1	A1														
27	A10	A10	A10	A1															
28	A1																		
29	A1																		

Action Perform By Users

ML Model

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. Getting a 360-degree view of a customer by knowing what he is clicking and what he is not can get you a huge improvement in both your products and your customers' experience.

Data Collection

Either you have your data in your data warehouse, or you need to enrich it with more data sources you need to have a way to collect and store data consistently into a database.

Data Preparation

Raw data is like a rough diamond; It requires some refinement before being truly valuable.

In the data world, refinement includes data processing, cleaning, and transformation of the initial data into something convenient for the analysis you are going to carry out.

In this case, we would like to have our data grouped into users. It would be good too, we could arrange the events of each user in time order before moving to actual analysis.

In contrast to other data sequences, clickstream data can have varying length for every different user.

In order to transform the initially collected event log into clickstream data we need to:

- ★ Identify events/actions performed by the same user and group them together
- ★ Split them further into subgroups of events based on which of those were performed during the same session according to the session's definition given above.

At this point the dataset we are going to use for the rest of the analysis should look like this:

2	A1	A1	A1	A1	A1	A1	A1												
3	A2	A2	A2	A1	A3	A1													
4	A1	A1	A1	A1	A1	A1													
5	A1	A1	A1	A4	A5	A3	A4	A5	A6	A7	A2	A8	A9	A10	A1				
6	A1	A7	A7	A2	A8	A1													
7	A1	A1																	
8	A1	A1																	
9	A3	A4	A5	A6	A7	A2	A8	A9	A10	A1	A1								
10	A1	A7	A2	A2	A2	A2	A8	A10	A10	A10	A10	A10	A10	A10	A10	A10	A10	A10	A1
11	A1	A1	A10	A10	A10	A1													
12	A1	A1																	
13	A8	A8	A8	A8	A1														

In this representation, each line corresponds to a user. The first field is the user's name while the next fields the actions performed by the user during this session.

Model Construction

As in most cases, the methods we can deploy for solving this problem are many.

Markov Chains

The type of data Markov Chains work with are sequential data.

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

In clickstream analysis, we usually utilize these Markov Chains. The process takes the state from a finite set at each time . 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.

Higher order Markov Chain introduced by the Raftery (1985) will lead to more realistic models. At the same time, the parameters needed for the representation increase exponentially and so it is important to find a right balance between these two.

Fitting a Markov Chain

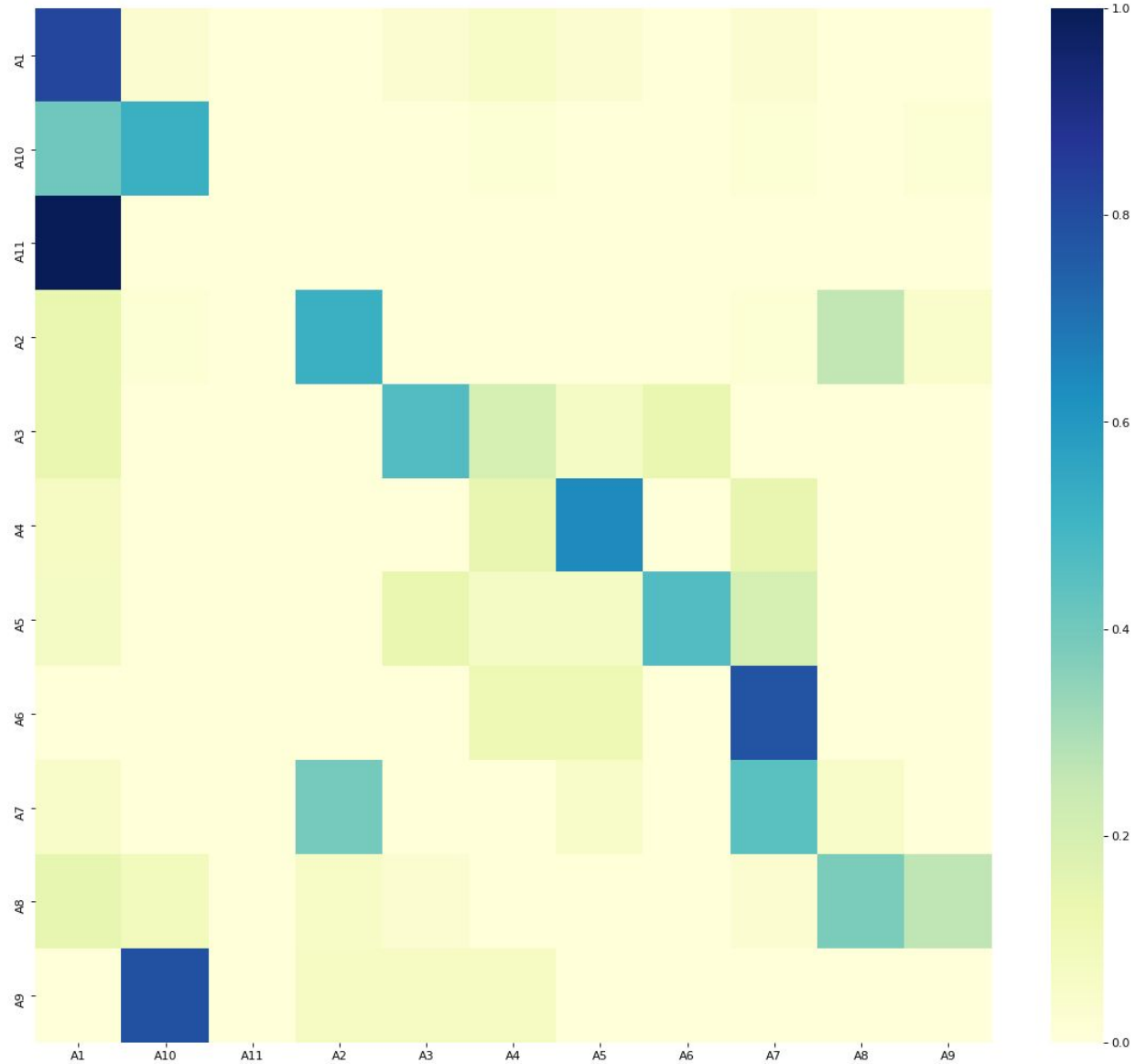
As mentioned before at this point our dataset looks like:

4	A1	A1	A1	A1	A1	A1													
5	A1	A1	A1	A4	A5	A3	A4	A5	A6	A7	A2	A8	A9	A10	A1				
6	A1	A7	A7	A2	A8	A1													
7	A1	A1																	
8	A1	A1																	
9	A3	A4	A5	A6	A7	A2	A8	A9	A10	A1	A1								
10	A1	A7	A2	A2	A2	A2	A8	A10	A10	A10	A10	A10	A10	A10	A10	A10	A10	A10	A1
11	A1	A1	A10	A10	A10	A1													

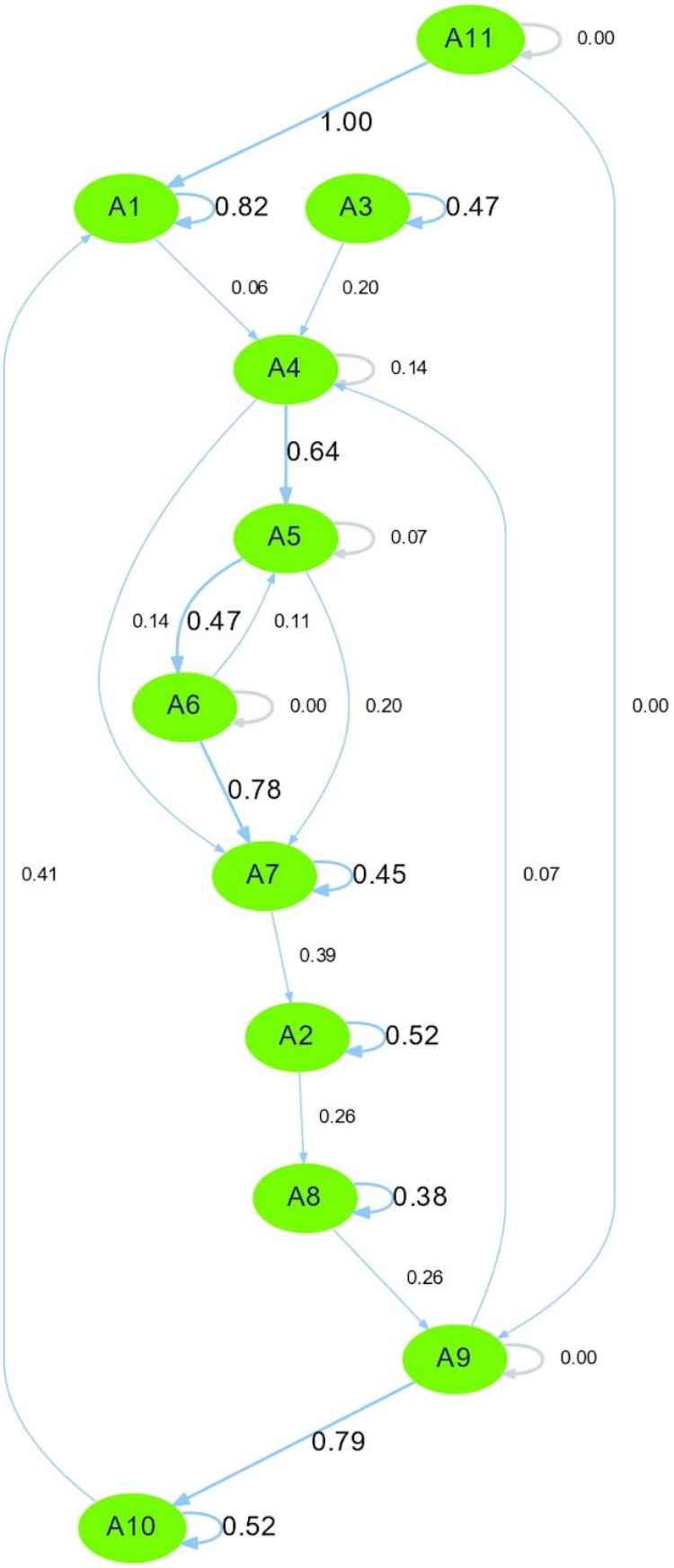
Fitting the Markov Chain model gives us the transition probabilities matrices and the lambda parameters of the chain for each one of the three lags along with the Start and End Probabilities.

Start and End probabilities correspond to the probability that a clickstream will start or end with this specific event.

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 blueish the color, the more probable the indicated transition will occur.



Heat Map



Transaction diagram

Actions Details

Home	A1
Cart	A2
contact-us	A3
Index	A4
404	A5
Blog-single	A6
Login	A7
Checkout	A8
Product-details	A9
Shop	A10
Sessions__activities__details__Page	A11
Payment	A12

Suggestion to UX Designer

The following are the insights from the Heat Map generated by the Markov chain model:

The transaction from Action 1 to Action 12 is more correlated to each other. Suggestion taken from transaction diagram along with the corresponding probability are:

- All the users must start from A1 (homepage).
- The users from action A2, A8, A7, A4 (all are products related) are occasionally not going to A2 (cart). They should go to A2 (cart) so the UX designer should guide the user to go to the A2 from these actions.
- The UX designer should guide the user to go to the A8 (Checkout) from A2 (cart).
- Users are either going back to A2 (cart) or close the application at A6 (Checkout) without proceeding to payment, so the UX designer should guide the user to place an order and make payment. So the UX designer should guide the user to go to next actions to place orders.