Analysis of user behaviour & optimization of the user workflow using a machine learning algorithm

G. Yashwanth

Abstract

For an online shopping site, it is very important to identify the various pages which are responsible for most of the revenue and analyse how the users are surfing through the website. In order to obtain the maximum revenue, the web site needs to be improved and changed accordingly. It was done so by analysing the user behaviour using a machine learning algorithm known as ‘MarkovClick’. The site was linked to Google analytics and the required data was extracted, modified to the desired format to perform analysis and observe various insights.

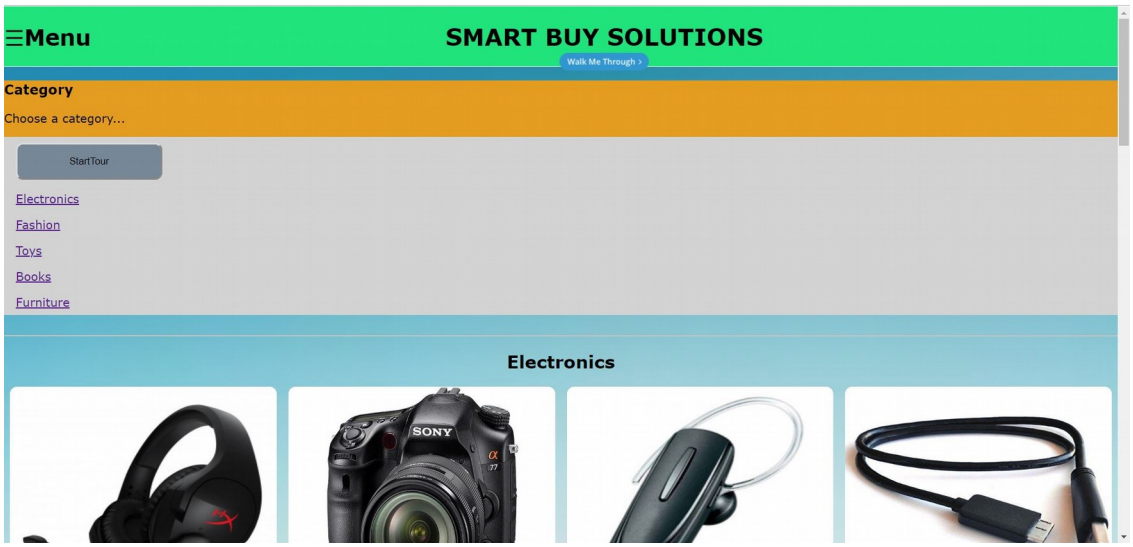
Steps:

- Creating a website
- Data collection
- Data preparation
- Applying MarkovClick algorithm to the data
- Analyse the plots and suggest the changes required to be made

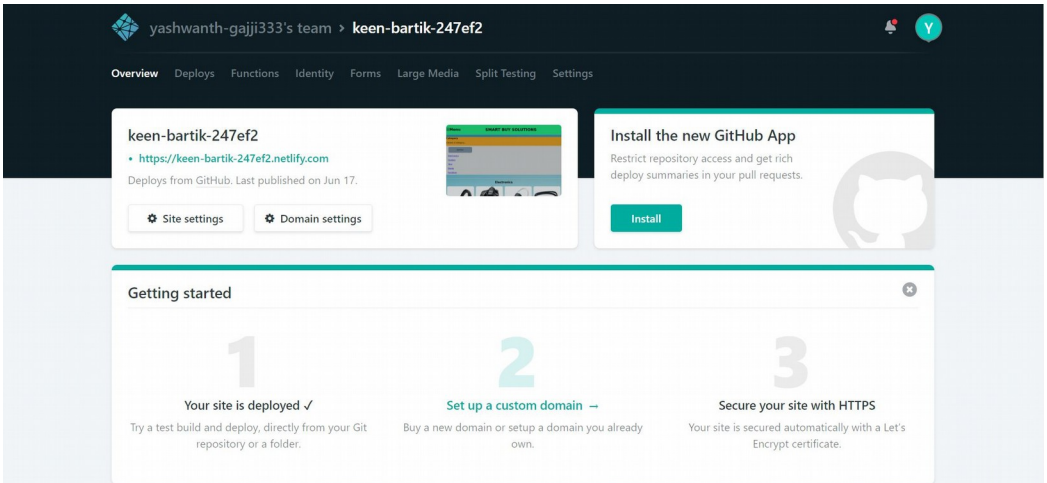
Website Creation

Creating a website is the base of the project.

- A domain can be chosen to deploy and host the website for free.
- In this project I created a website from scratch in one page. We can instead choose the existing things from word press or bootstrap.
- All the products for the website are added from scratch.



- All the pages are connected in a single page using Javascript.
- When the website is ready by any means, it is to be deployed online.
- For hosting the website use Netlify.com where we can directly add our source code of our website to deploy.
- We can add source code from Github or directly from the computer.
- Netlify will give a url address for the website added which we can access it from anywhere.



- Now the next part is to add a walkthrough for the website so that the user can understand clearly what he can do in the website.
- For doing this we can use Walkme, Appcues etc.
- Walkme is the best option and create a walkthough to the website and deploy it to the site so that whenever a new user opens the website, walkthrough should start automatically.

Now the website is ready for the users to use. The next part is to connect the website with Google analytics so that we can track the behaviour of the users.

- Register to the Google analytics and add your website to it and deploy the website.
- Now it is ready for the users to surf.
- Now spread and share the website url to various people to use.
- Users have to browse the the website and spend some time on it and do various tasks on it.

MarkovClick Algorithm

Markovclick allows you to model clickstream data from websites as Markov chains, which can then be used to predict the next likely click on a website for a user, given their history and current state.

Clickstream is the list of clicks made by each user. This clickstream data is used to model Markov chains with the help of a probability matrix that is formed from the data that has been provided.

Markov chains are mathematical systems that hop from one "state" (a situation or set of values) to another.

The type of data Markov Chains work with are sequential data. 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 the right balance between these two.

Installation:

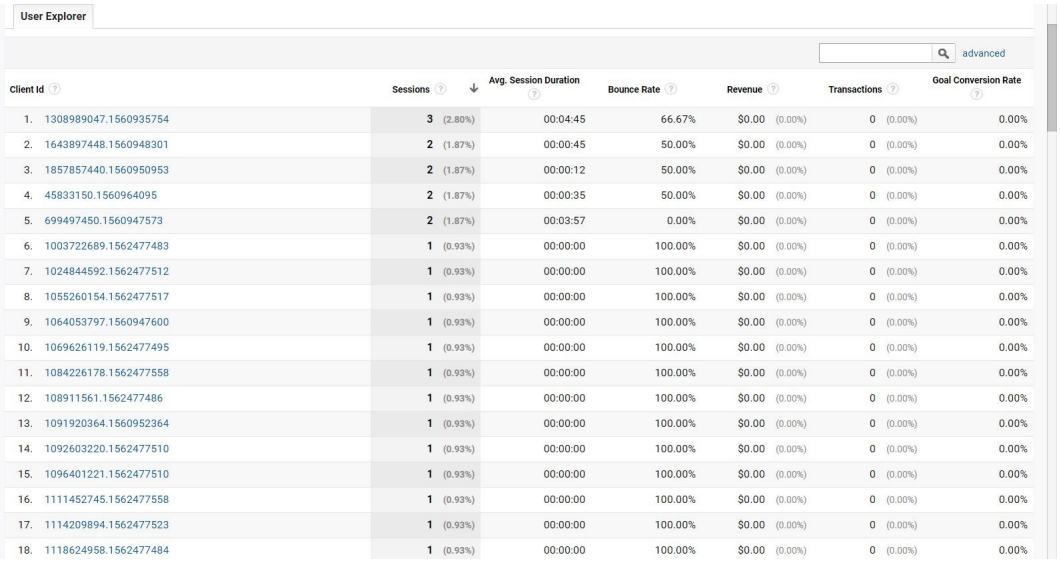
- pip installs morkovclick
- pip install Graphviz

Requirements:

- Python 3.X
- numpy
- matplotlib
- seaborn (Recommended)
- pandas

Data Preparation

- All the data for training the algorithm gets from Google Analytics.
- In Google analytics goto Audience and User explorer. There we can see the data of all the users surfed in the website.

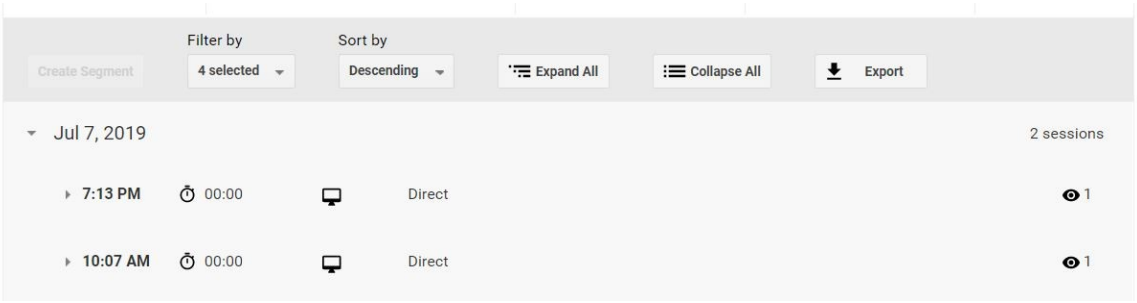


User Explorer

advanced

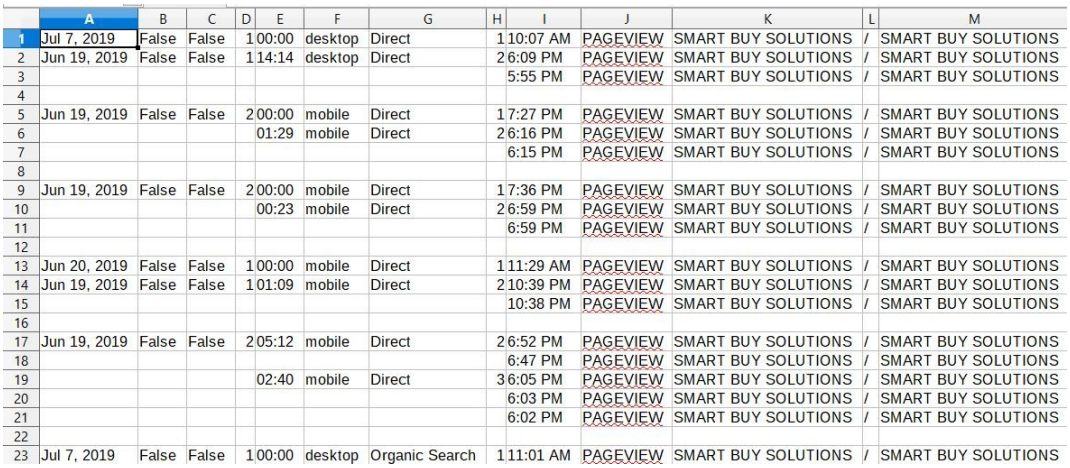
	Client Id	Sessions	Avg. Session Duration	Bounce Rate	Revenue	Transactions	Goal Conversion Rate
1.	1308989047.1560935754	3 (2.80%)	00:04:45	66.67%	$0.00 (0.00%)	0 (0.00%)	0.00%
2.	1643897448.1560948301	2 (1.87%)	00:00:45	50.00%	$0.00 (0.00%)	0 (0.00%)	0.00%
3.	1857857440.1560950953	2 (1.87%)	00:00:12	50.00%	$0.00 (0.00%)	0 (0.00%)	0.00%
4.	45833150.1560964095	2 (1.87%)	00:00:35	50.00%	$0.00 (0.00%)	0 (0.00%)	0.00%
5.	699497450.1560947573	2 (1.87%)	00:03:57	0.00%	$0.00 (0.00%)	0 (0.00%)	0.00%
6.	1003722689.1562477483	1 (0.93%)	00:00:00	100.00%	$0.00 (0.00%)	0 (0.00%)	0.00%
7.	1024844592.1562477512	1 (0.93%)	00:00:00	100.00%	$0.00 (0.00%)	0 (0.00%)	0.00%
8.	1055260154.1562477517	1 (0.93%)	00:00:00	100.00%	$0.00 (0.00%)	0 (0.00%)	0.00%
9.	1064053797.1560947600	1 (0.93%)	00:00:00	100.00%	$0.00 (0.00%)	0 (0.00%)	0.00%
10.	1069626119.1562477495	1 (0.93%)	00:00:00	100.00%	$0.00 (0.00%)	0 (0.00%)	0.00%
11.	1084226178.1562477558	1 (0.93%)	00:00:00	100.00%	$0.00 (0.00%)	0 (0.00%)	0.00%
12.	108911561.1562477486	1 (0.93%)	00:00:00	100.00%	$0.00 (0.00%)	0 (0.00%)	0.00%
13.	1091920364.1560952364	1 (0.93%)	00:00:00	100.00%	$0.00 (0.00%)	0 (0.00%)	0.00%
14.	1092603220.1562477510	1 (0.93%)	00:00:00	100.00%	$0.00 (0.00%)	0 (0.00%)	0.00%
15.	1096401221.1562477510	1 (0.93%)	00:00:00	100.00%	$0.00 (0.00%)	0 (0.00%)	0.00%
16.	1111452745.1562477558	1 (0.93%)	00:00:00	100.00%	$0.00 (0.00%)	0 (0.00%)	0.00%
17.	1114209894.1562477523	1 (0.93%)	00:00:00	100.00%	$0.00 (0.00%)	0 (0.00%)	0.00%
18.	1118624958.1562477484	1 (0.93%)	00:00:00	100.00%	$0.00 (0.00%)	0 (0.00%)	0.00%

- Now open each user and download the JSON file of each user by clicking the export button.



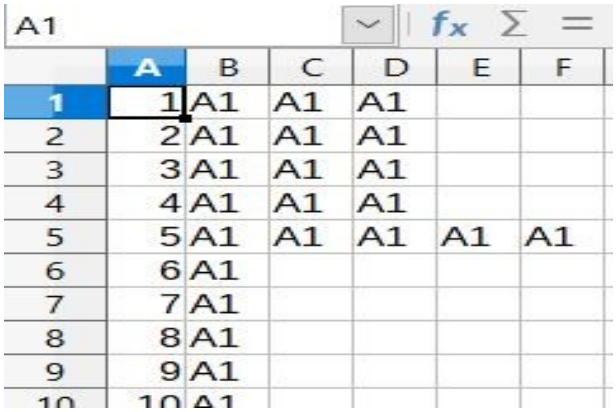
- Now use some online converter to convert the JSON files to CSV files.
- Now arrange data of all the users into a single CSV file.

- The final data may look like this.



	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Jul 7, 2019	False	False	1	00:00	desktop	Direct	1	10:07 AM	PAGEVIEW	SMART BUY SOLUTIONS	/	SMART BUY SOLUTIONS
2	Jun 19, 2019	False	False	1	14:14	desktop	Direct	2	6:09 PM	PAGEVIEW	SMART BUY SOLUTIONS	/	SMART BUY SOLUTIONS
3									5:55 PM	PAGEVIEW	SMART BUY SOLUTIONS	/	SMART BUY SOLUTIONS
4													
5	Jun 19, 2019	False	False	2	00:00	mobile	Direct	1	7:27 PM	PAGEVIEW	SMART BUY SOLUTIONS	/	SMART BUY SOLUTIONS
6					01:29	mobile	Direct	2	6:16 PM	PAGEVIEW	SMART BUY SOLUTIONS	/	SMART BUY SOLUTIONS
7									6:15 PM	PAGEVIEW	SMART BUY SOLUTIONS	/	SMART BUY SOLUTIONS
8													
9	Jun 19, 2019	False	False	2	00:00	mobile	Direct	1	7:36 PM	PAGEVIEW	SMART BUY SOLUTIONS	/	SMART BUY SOLUTIONS
10					00:23	mobile	Direct	2	6:59 PM	PAGEVIEW	SMART BUY SOLUTIONS	/	SMART BUY SOLUTIONS
11									6:59 PM	PAGEVIEW	SMART BUY SOLUTIONS	/	SMART BUY SOLUTIONS
12													
13	Jun 20, 2019	False	False	1	00:00	mobile	Direct	1	11:29 AM	PAGEVIEW	SMART BUY SOLUTIONS	/	SMART BUY SOLUTIONS
14	Jun 19, 2019	False	False	1	01:09	mobile	Direct	2	10:39 PM	PAGEVIEW	SMART BUY SOLUTIONS	/	SMART BUY SOLUTIONS
15									10:38 PM	PAGEVIEW	SMART BUY SOLUTIONS	/	SMART BUY SOLUTIONS
16													
17	Jun 19, 2019	False	False	2	05:12	mobile	Direct	2	6:52 PM	PAGEVIEW	SMART BUY SOLUTIONS	/	SMART BUY SOLUTIONS
18									6:47 PM	PAGEVIEW	SMART BUY SOLUTIONS	/	SMART BUY SOLUTIONS
19					02:40	mobile	Direct	3	6:05 PM	PAGEVIEW	SMART BUY SOLUTIONS	/	SMART BUY SOLUTIONS
20									6:03 PM	PAGEVIEW	SMART BUY SOLUTIONS	/	SMART BUY SOLUTIONS
21									6:02 PM	PAGEVIEW	SMART BUY SOLUTIONS	/	SMART BUY SOLUTIONS
22													
23	Jul 7, 2019	False	False	1	00:00	desktop	Organic Search	1	11:01 AM	PAGEVIEW	SMART BUY SOLUTIONS	/	SMART BUY SOLUTIONS

- Now label the data as A1,A2. Modified data may look like this.

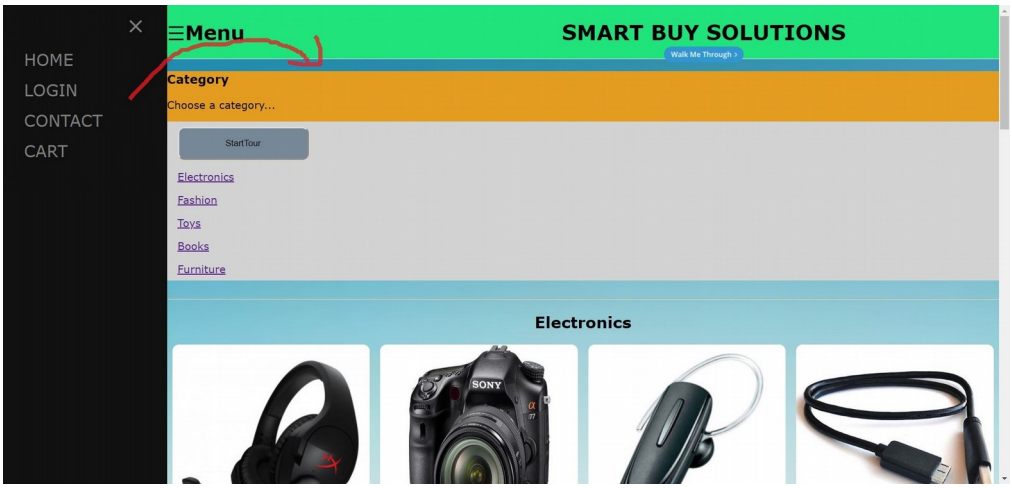


A1

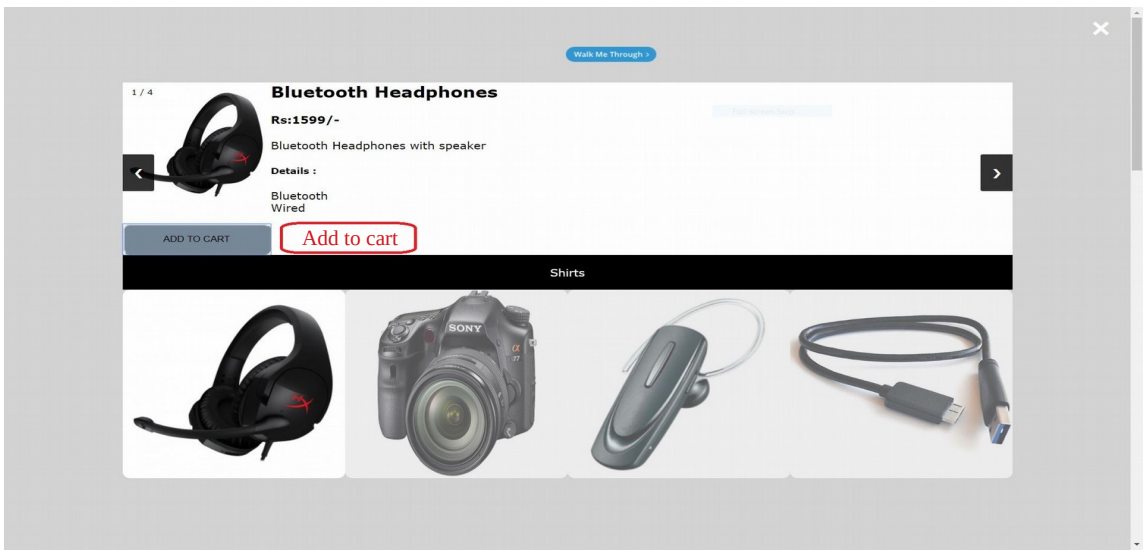
	A	B	C	D	E	F
1	1	A1	A1	A1		
2	2	A1	A1	A1		
3	3	A1	A1	A1		
4	4	A1	A1	A1		
5	5	A1	A1	A1	A1	A1
6	6	A1				
7	7	A1				
8	8	A1				
9	9	A1				
10	10	A1				

SUGGESTIONS AND IMPROVEMENTS

- A Top Navigation bar is more helpful than a side navigation bar as the users can directly see where the login and cart are. A new user may confuse finding the login and cart.

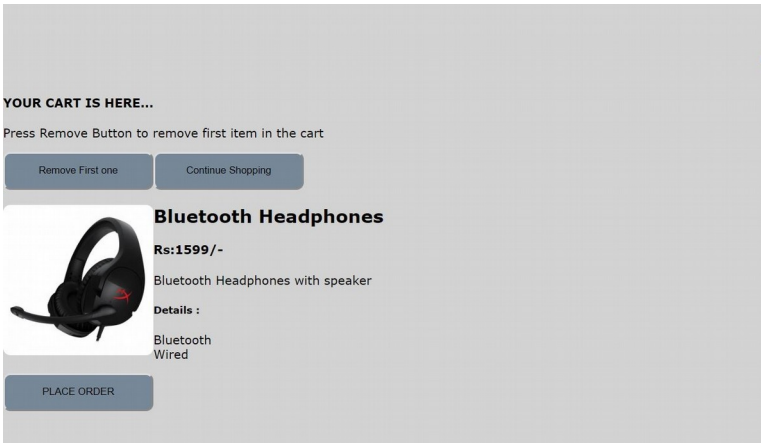


- After add a product to cart, user has to go back to home page and open cart to see the items he added. It is better to add a cart button on the products page so that he can directly open and check his products.

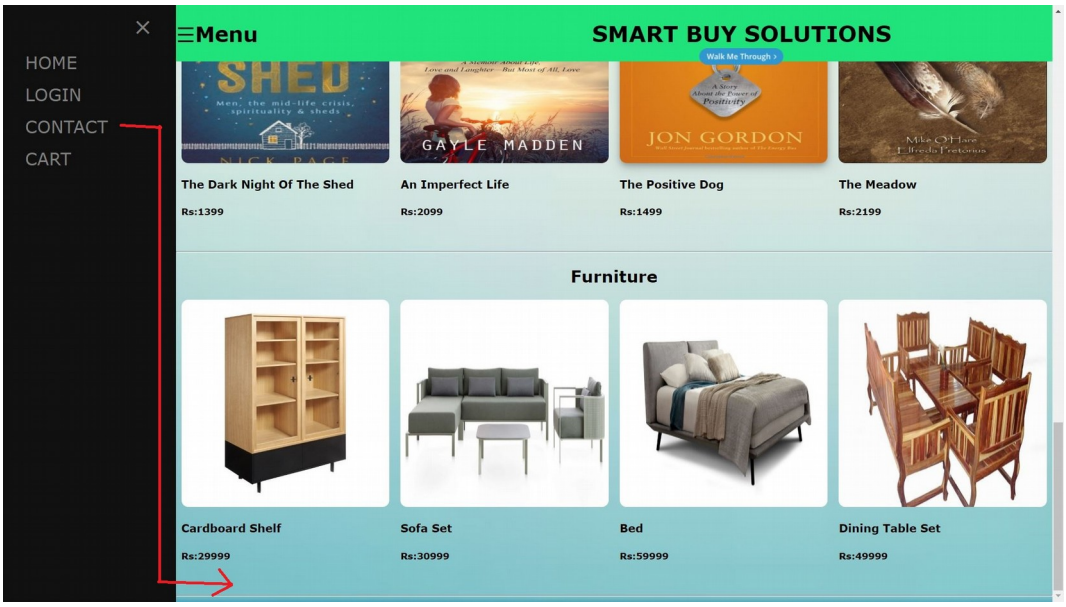


- Product display is too clumsy. It is better to show all the products and on clicking a product only that product's details are to be displayed.

- Show similar products in cart as user can easily buy another product on the same page without going to home page and and searching for the product.



- Get the contact details down the page as it is the least thing user will click on. It can be moved to bottom of the page.



- Change the design of website, instead of using some fixed colors and backgrounds, change the UI to more good looking.

Thank you