

Prepared for

[Prof. Mahshid Fardadi]

Prepared by

[Mr. Sudarshan Powar]

[sudarshan.powar01@student.csulb.edu]

Introduction

The problem is about the challenge that merchants face in distinguishing between one-time buyers and potential loyal customers who may return for future purchases, especially after promotional events. The impact of promotions on long-term sales may be limited as many customers are only interested in one-time deals. The goal of the project is to design a system that can predict the probability of a customer making future purchases within 6 months, reduce promotional costs, and identify one-time buyers. This will help the merchants maximize their ROI and focus their efforts on converting potential loyal customers.

Dataset

The data set contains anonymized users' shopping logs in the past 6 months before and on the "Double 11" day (*Promotional Event*). In summary, there are two datasets, TrainData-group7.csv and TestData-group7.csv, which are subsets of a larger dataset consisting of data points numbered from 961 to 1120. The datasets have been concatenated for training purposes, although it is unclear what kind of model is to be trained. Sample outputs are provided below.

	user_id	age_range	gender	merchant_id	label	activity_log
0	173184	3	0	4519	-1.0	426158:1047:5625:0728:0#426158:1047:5625:0728:...
1	173184	3	0	4785	-1.0	890128:1271:1422:0525:0
2	173184	3	0	2489	-1.0	760996:451:814:1104:0#760996:451:814:1104:0#42...
3	173184	3	0	3070	-1.0	520330:673:3929:1104:0
4	173184	3	0	2607	-1.0	1091667:1472:3270:0728:0
5	173184	3	0	2894	-1.0	93415:662:4098:1001:0#612617:662:4098:1001:0#4...

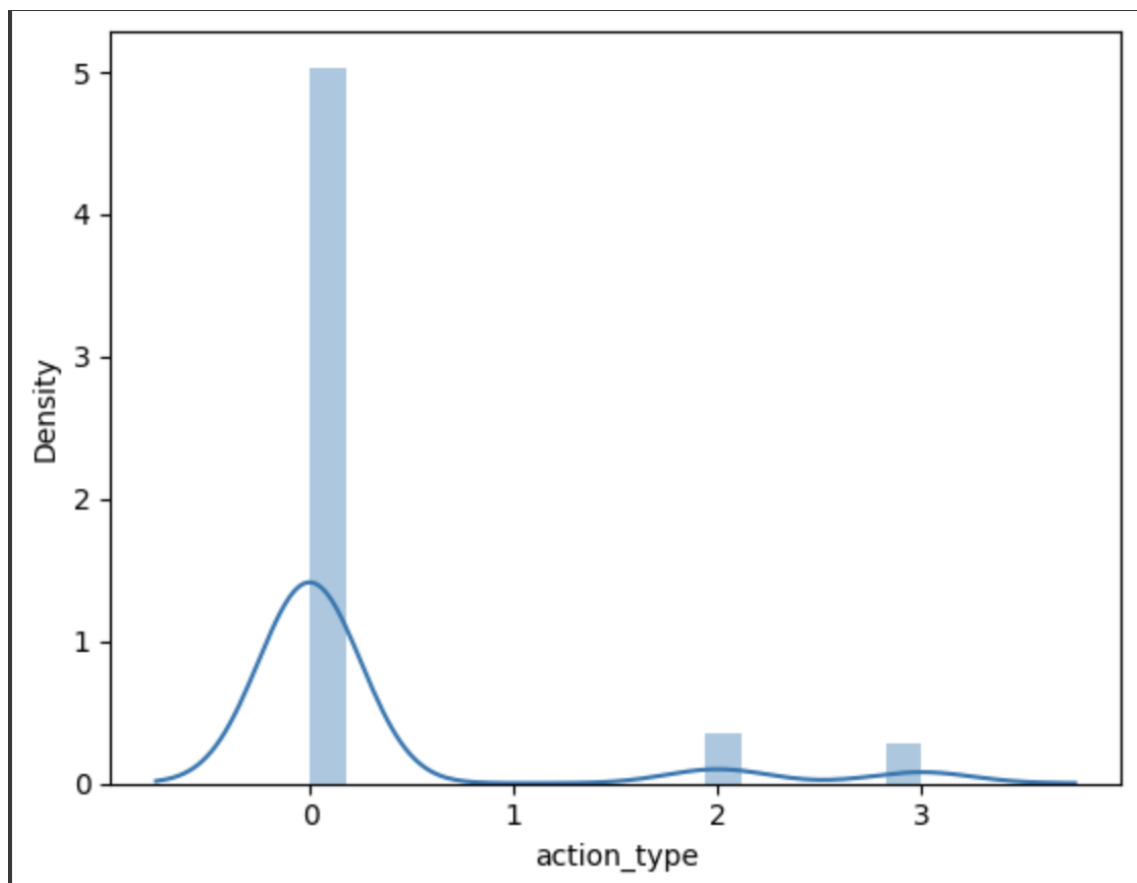
Current Progress

****based on table, we break it down individually so that we can correlate

1. We have converted labels from float to integers.
2. Next, we have converted activity logs into 5 sections such as **item_id**, **category_id**, **brand_id**, **time_stamp** and **action_type**.

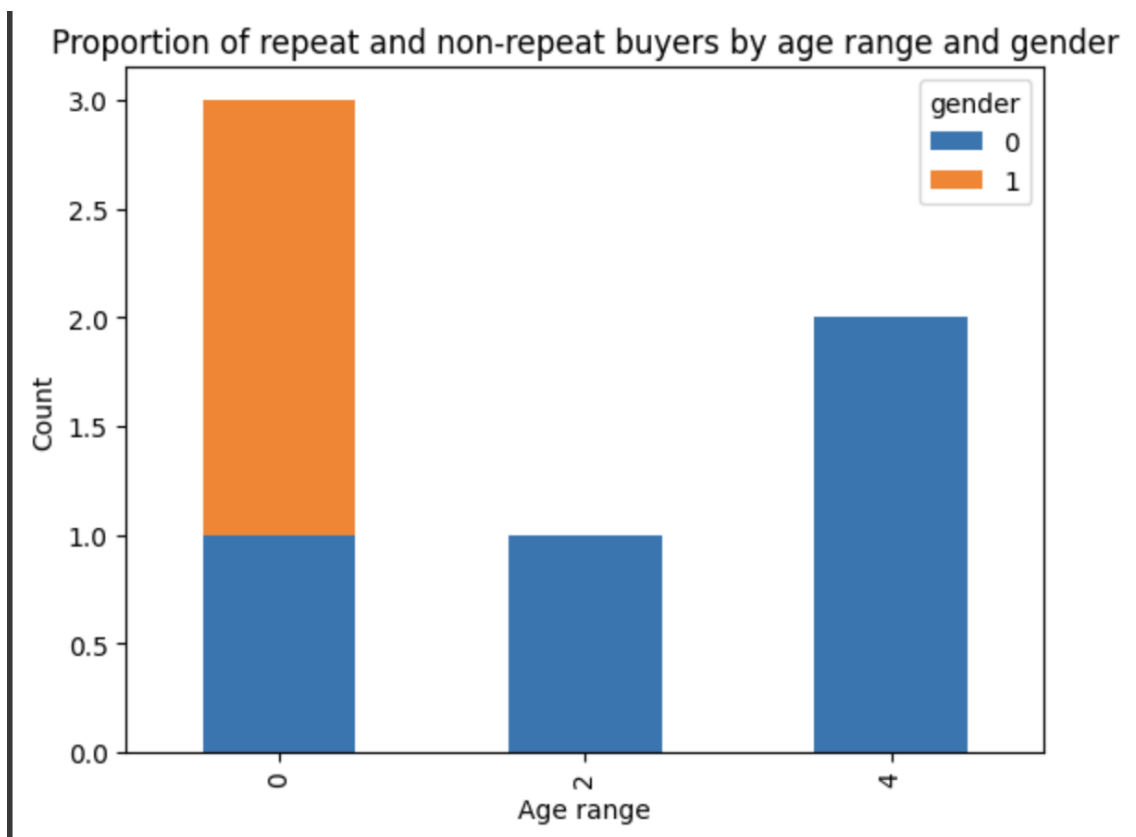
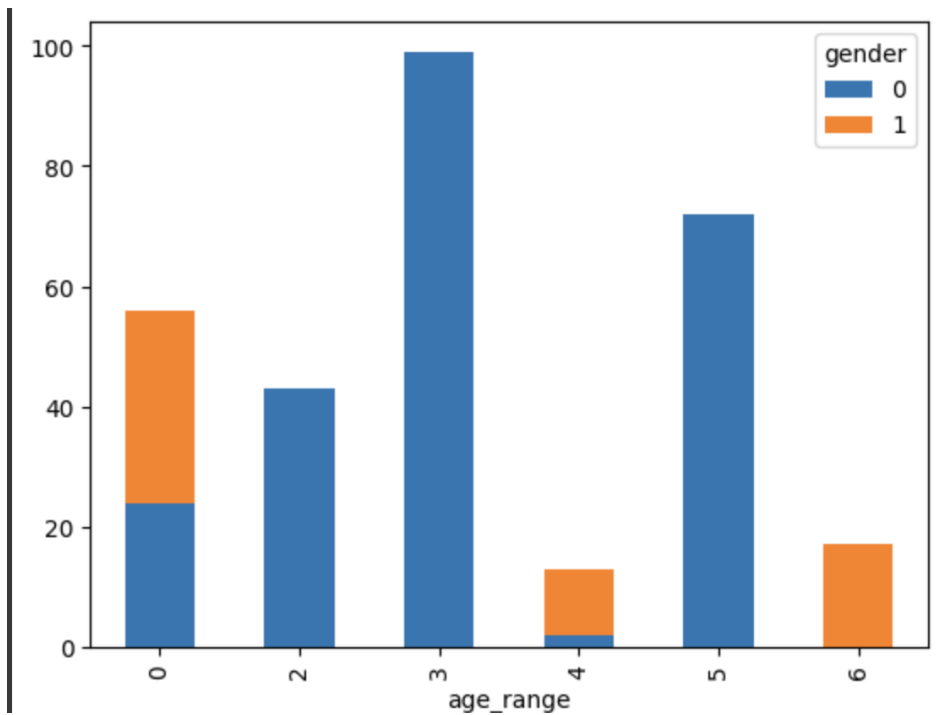
	act	item_id	category_id	brand_id	time_stamp	action_type
[426158:1047:5625:0728:0, 426158:1047:5625:072...		426158	1047	5625	0728	0
[890128:1271:1422:0525:0]		890128	1271	1422	0525	0
[760996:451:814:1104:0, 760996:451:814:1104:0,...		760996	451	814	1104	0
[520330:673:3929:1104:0]		520330	673	3929	1104	0
[1091667:1472:3270:0728:0]		1091667	1472	3270	0728	0

3. Next, we have plotted a histogram for visualization of **action_type**.



Where 0 represents, no. of clicks.

4. We have plotted Proportion of repeat and non-repeat buyers by age range and gender'





Next Steps

1. Maintain current customers.
2. Find the preferences of brands that are being viewed and not purchased.