

What Datasets Do We Have?

	Dataset 1	Dataset 2	Dataset 3	Dataset 4	Dataset 5
Name	E-Commerce Behavior Data from Multi-Category Store	E-Commerce Purchase History from Electronics Store	E-Commerce Events History in Cosmetics Shop	E-Commerce Purchase History From Jewelry Store	E-Commerce Events History in Electronics Store
Size	14.68 GB	298.65 MB	2.43 GB	13.62 MB	105.76 MB
Time	Oct 2019 to Nov 2019	Apr 2020 to Nov 2020	Oct 2019 to Feb 2020	Dec 2018 to Dec 2021	Sep 2020 to Feb 2021
Columns	event_time, event_type, product_id, category_id, category_code, brand, price, user_id, user_session	event_time, event_type, product_id, category_id, category_code, brand, price, user_id, user_session	event_time, event_type, product_id, category_id, category_code, brand, price, user_id, user_session	event_time, event_type, product_id, category_id, category_code, brand, price, user_id, user_session	event_time, event_type, product_id, category_id, category_code, brand, price, user_id, user_session
Comments	Using this dataset for building Recommendation System and performing Customer Segmentation	Each user had only one transaction A lot of fluctuation in purchases over time	This dataset has a high %age of nulls (78% user ids are null values) thus we discard it for any further analysis.	This dataset has a high %age of nulls (64% user ids are null values) thus we discard it for any further analysis.	Utilizing this dataset for extensive EDA and analysing consumer journey from view to cart to purchase

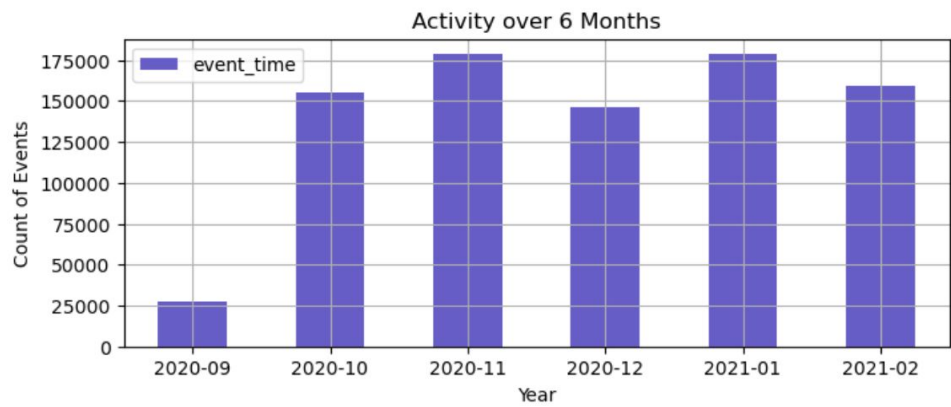
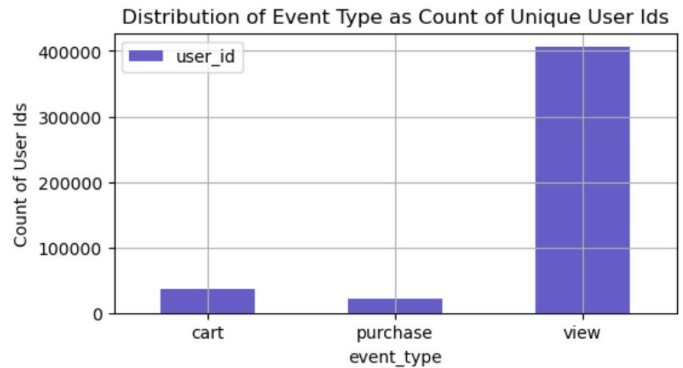
Dataset 5 - Electronics Store (Events History): Deep Dive

Understanding Interaction (View/Cart/Purchase)

Key Takeaways:

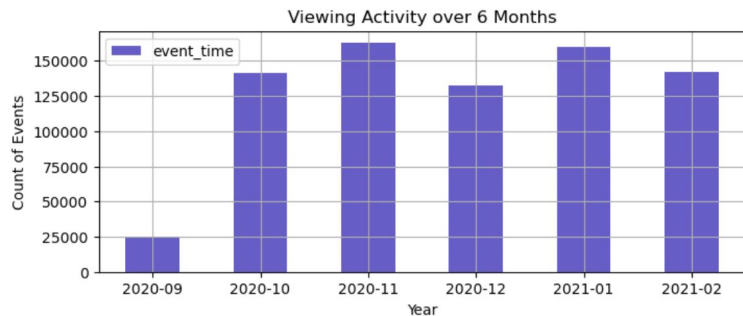
- 1. 87% records indicate viewing activity
- 2. There appears to be an alternating trend from odd to even months but the difference is not too big

user_id	
event_type	
cart	36952
purchase	21304
view	406863



event_time	
year_month	
2020-09	27059
2020-10	154758
2020-11	178957
2020-12	146462
2021-01	178730
2021-02	159075

Understanding Interaction (View/Cart/Purchase)

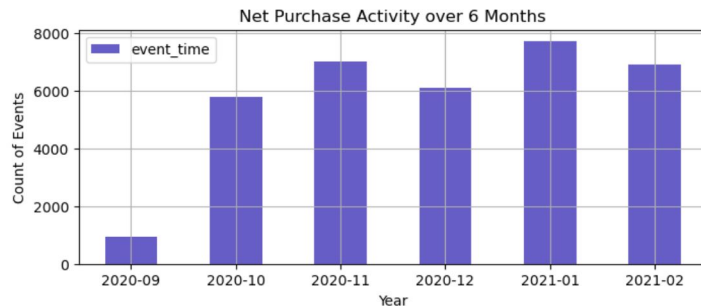


event_time	
year_month	
2020-09	24864
2020-10	141267
2020-11	163019
2020-12	132090
2021-01	159883
2021-02	142140

Key Takeaways:

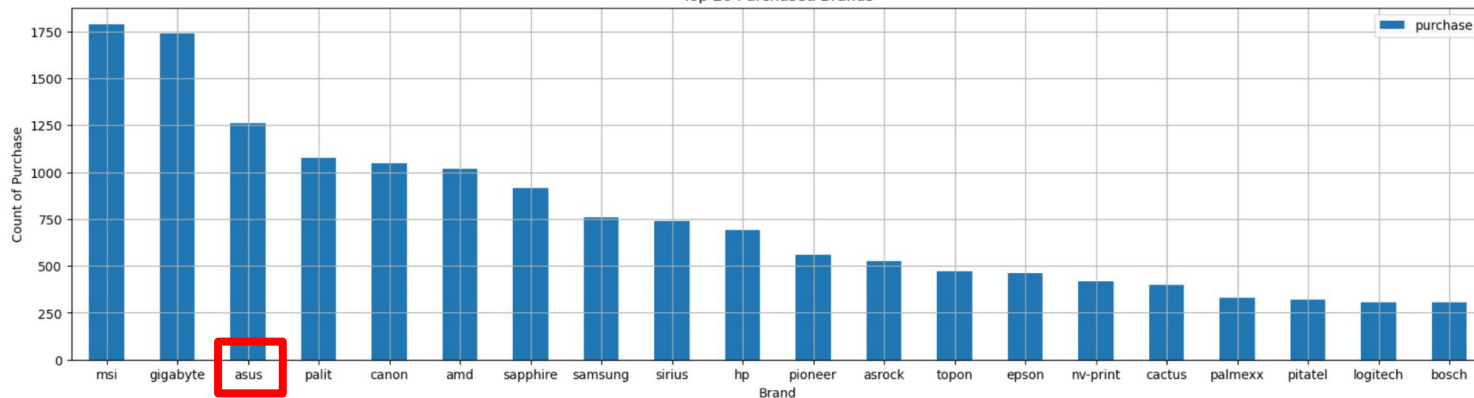
1. Trend of viewing activity over the 6 months is similar to previous plot of net activity
2. Jan 2021 has the highest Purchase activity; Nov 2020 has the highest viewing activity

event_time	
year_month	
2020-09	928
2020-10	5784
2020-11	6996
2020-12	6092
2021-01	7725
2021-02	6898

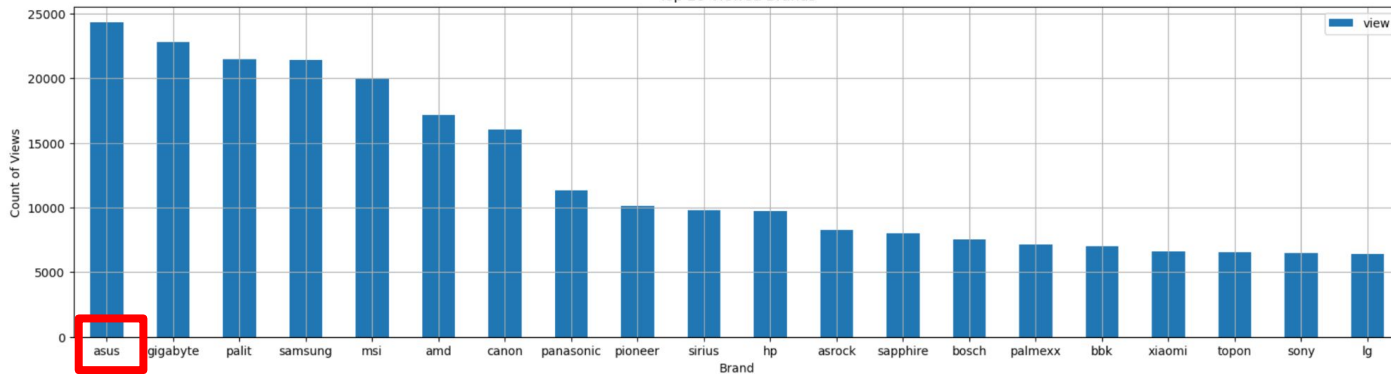


Understanding Brand Popularity

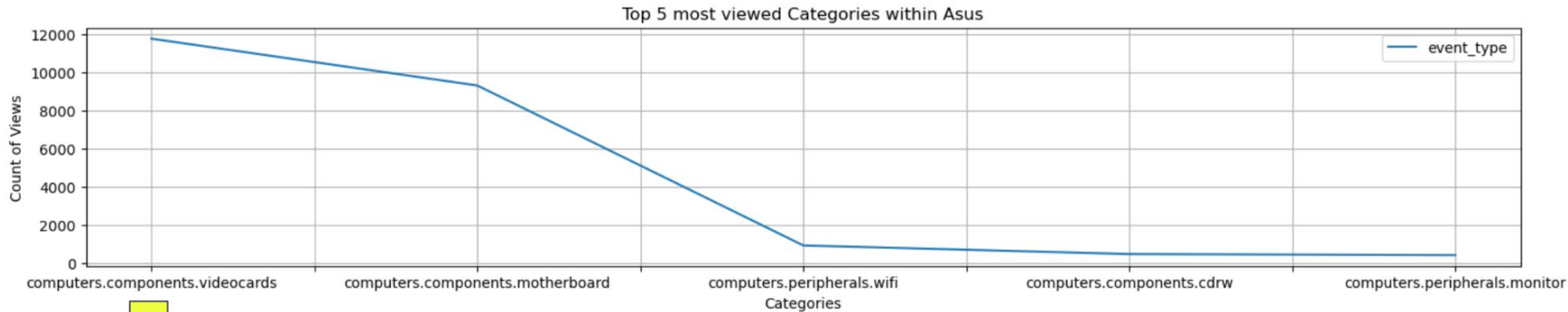
Top 20 Purchased Brands



Top 20 Viewed Brands



Exploring the most popular brand: Asus



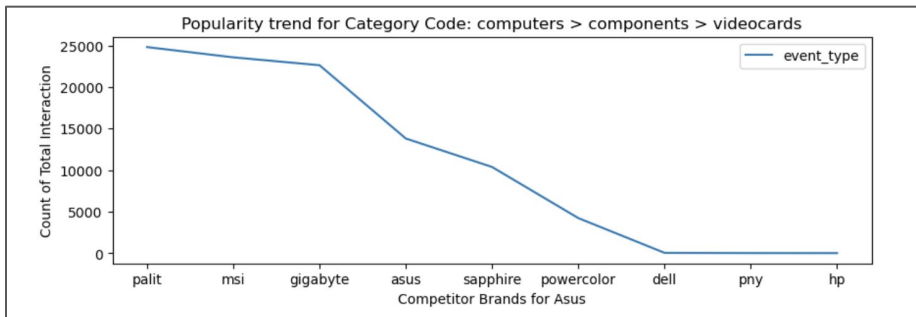
asus_vc['event_type'].value_counts():[:3]		
view	11782	
cart	1311	
purchase	716	

asus_vc['user_id'].value_counts():[:3]		
1515915625596734979	222	
1515915625593948004	96	
1515915625594475859	76	

asus_vc['product_id'].value_counts():[:3]		
387956	1483	
3961719	1409	
4101569	935	

asus_vc['user_session'].value_counts():[:3]		
651c519c-a50d-4ad1-b95e-c7867751dd51	49	
qkVkgX6mDz	44	
6lnJx9TUzD	39	

Bonus Competitor Comparison:



User Journey for highest interacting user_id (1)

User 01:

'651c519c-a50d-4ad1-b95e-c7867751dd51'

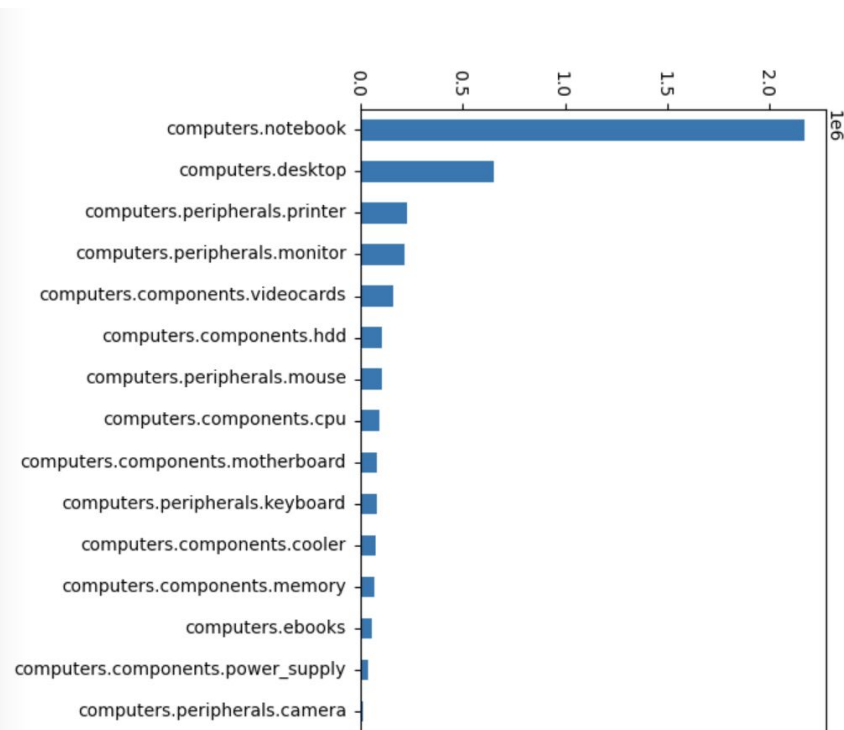
- User interacting with other brand products
 - Multiple Product IDs explored for competitive brand
 - doesn't compare different products for Asus
-
- User takes about 2 months for view → cart → purchase action

Takeaway: For this dataset, it was observed that each user_id is linked with an unique user_session, which is not natural. Thus we would not be furthering with this dataset for further analysis.

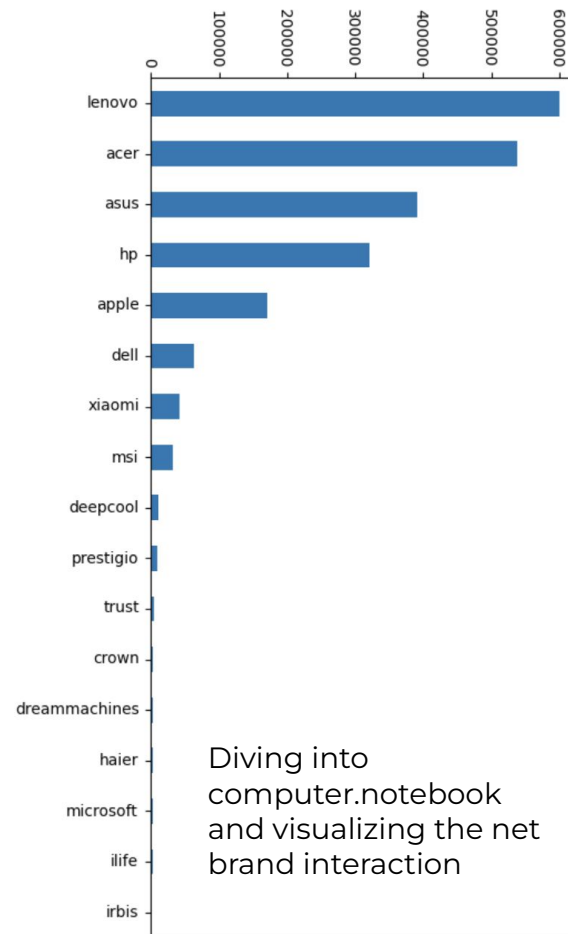
	event_time	event_type	product_id	brand	price
1	2020-09-26 05:46:25 UTC	view	4100254	asus	187.29
2	2020-09-28 15:17:31 UTC	view	4100254	asus	187.29
3	2020-09-28 16:11:55 UTC	view	4100254	asus	187.29
4	2020-09-29 15:22:58 UTC	view	4100254	asus	187.29
5	2020-09-30 10:42:01 UTC	view	4100254	asus	187.29
6	2020-10-01 19:17:18 UTC	view	4100254	asus	187.29
7	2020-10-02 14:37:21 UTC	view	4100254	asus	187.29
8	2020-10-03 06:00:50 UTC	view	4100254	asus	187.29
9	2020-10-04 16:33:12 UTC	view	4100254	asus	187.29
10	2020-10-07 18:30:37 UTC	view	4100254	asus	187.29
11	2020-10-08 10:01:52 UTC	view	4100254	asus	187.29
12	2020-10-08 16:38:29 UTC	view	4100254	asus	187.29
13	2020-10-08 18:24:50 UTC	view	4100254	asus	187.29
14	2020-10-08 18:25:36 UTC	view	4100254	asus	187.29
15	2020-10-11 06:53:49 UTC	view	4100254	asus	187.29
16	2020-10-12 04:23:57 UTC	view	4100254	asus	187.29
17	2020-10-12 19:02:58 UTC	view	4100254	asus	187.29
18	2020-10-13 08:34:32 UTC	view	4100254	asus	187.29
19	2020-10-15 18:12:24 UTC	view	4100254	asus	187.29
20	2020-10-15 18:12:33 UTC	view	4100254	asus	187.29
21	2020-10-17 09:40:39 UTC	view	4100254	asus	187.29
22	2020-10-19 04:21:13 UTC	view	4100254	asus	187.29
23	2020-10-20 16:18:02 UTC	view	4100254	asus	187.29
24	2020-10-20 18:14:42 UTC	view	4100254	asus	187.29
25	2020-10-22 04:31:00 UTC	view	4100254	asus	187.29
26	2020-10-23 04:31:31 UTC	view	4100254	asus	187.29
27	2020-10-25 13:13:55 UTC	view	4100254	asus	187.29
28	2020-10-30 04:21:56 UTC	view	4100254	asus	187.29
29	2020-10-31 09:15:08 UTC	view	4100254	asus	187.29
30	2020-11-01 09:28:54 UTC	view	4100254	asus	187.29
31	2020-11-01 09:56:16 UTC	view	4100254	asus	187.29
32	2020-11-07 09:22:14 UTC	view	4100254	asus	187.29
33	2020-11-15 16:40:25 UTC	view	4100254	asus	187.29
34	2020-11-16 18:49:04 UTC	view	4100254	asus	187.29
35	2020-11-19 19:27:07 UTC	view	4100254	asus	187.29
36	2020-11-19 19:27:16 UTC	view	4100254	asus	187.29
37	2020-11-20 17:53:46 UTC	view	4100254	asus	187.29
38	2020-11-20 21:40:45 UTC	view	4100254	asus	187.29
39	2020-11-20 21:45:20 UTC	view	4078916	sapphire	415.54
40	2020-11-21 18:01:27 UTC	view	4078916	sapphire	415.54
41	2020-11-21 18:02:03 UTC	view	4100254	asus	187.29
42	2020-11-21 18:02:42 UTC	view	893192	asus	253.38
43	2020-11-21 18:03:49 UTC	view	4100254	asus	187.29
44	2020-11-21 18:04:32 UTC	view	919786	sapphire	265.05
45	2020-11-21 20:17:52 UTC	view	4078916	sapphire	415.54
46	2020-11-27 19:31:29 UTC	view	4100254	asus	187.29
47	2020-11-27 19:32:12 UTC	view	4100254	asus	187.29
48	2020-11-27 19:48:43 UTC	view	4078916	sapphire	415.54
49	2020-11-27 19:49:59 UTC	view	4100254	asus	187.29
50	2020-11-27 19:50:59 UTC	cart	4100254	asus	187.29
51	2020-11-27 19:51:54 UTC	purchase	4100254	asus	187.29
52	2020-11-27 19:53:08 UTC	view	4100254	asus	187.29
53	2020-11-28 03:12:58 UTC	view	4100254	asus	187.29
54	2021-01-08 08:50:17 UTC	view	4100254	asus	187.29

Dataset 1 - Multi Category Store (Events History): Deep Dive

For the purpose of our project, we consider only a subset of the actual dataset containing **“computers”** as the first level **category code**.

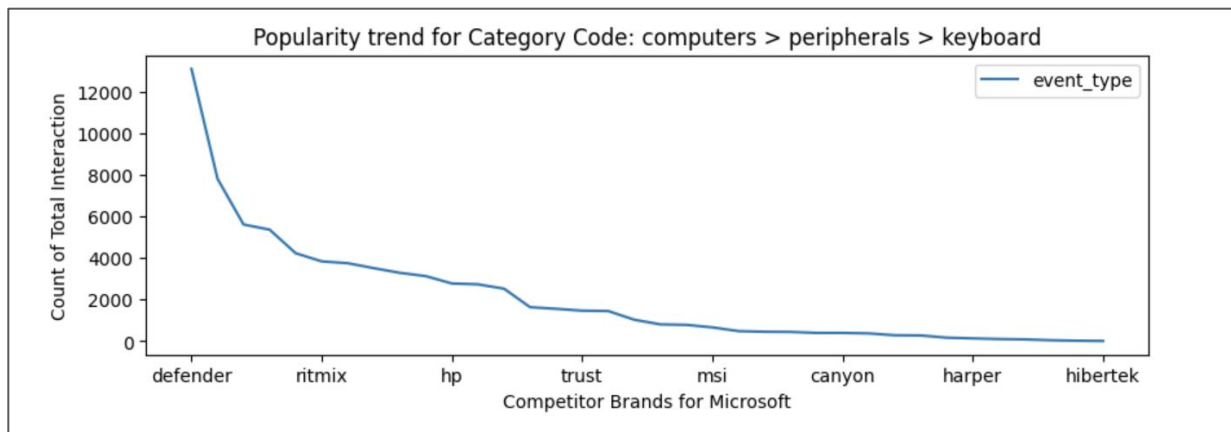
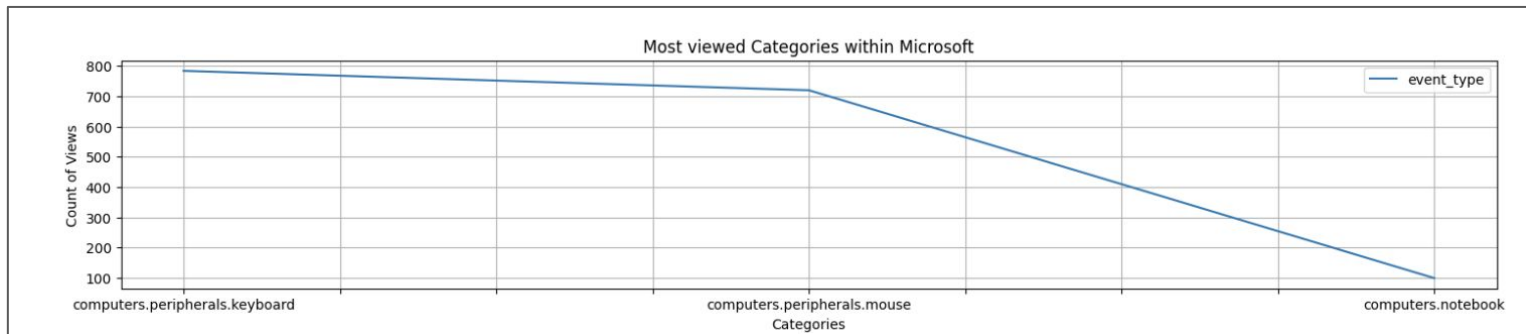


The category_code that start with “computers”



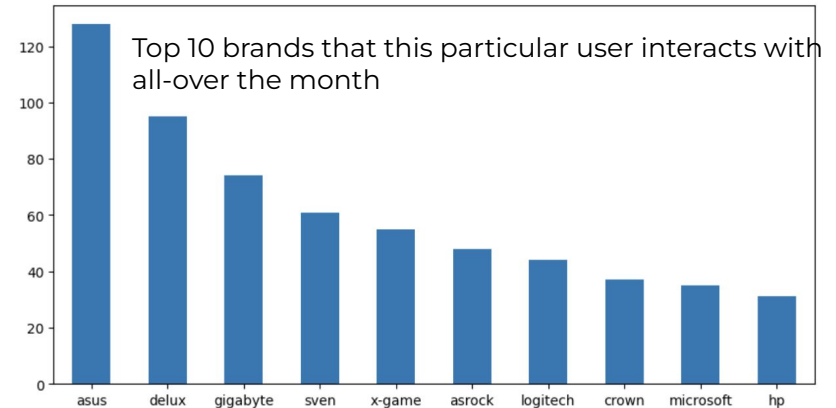
Diving into computer.notebook and visualizing the net brand interaction

Now we narrow down our analysis to discover some details about the brand: "Microsoft"



Pulling out the most active session for Microsoft: “e369b283-e714-474c-82d8-1afd85d17d6b”
Associated user_id: “546309572”

	event_type	product_id	category_code	brand	price	date
743233	view	9200606	computers.peripherals.keyboard	genius	19.06	2019-11-08
744023	view	9200600	computers.peripherals.keyboard	microsoft	20.57	2019-11-08
744398	cart	9200600	computers.peripherals.keyboard	microsoft	20.57	2019-11-08
744412	view	9200600	computers.peripherals.keyboard	microsoft	20.57	2019-11-08
744442	view	9200452	computers.peripherals.keyboard	canyon	17.99	2019-11-08
744563	view	9200452	computers.peripherals.keyboard	canyon	17.99	2019-11-08
744870	view	9200452	computers.peripherals.keyboard	canyon	17.99	2019-11-08
744993	view	9200312	computers.peripherals.keyboard	crown	10.09	2019-11-08
745254	view	9200406	computers.peripherals.keyboard	hp	17.99	2019-11-08
745344	view	9200600	computers.peripherals.keyboard	microsoft	20.57	2019-11-08
745818	view	9200638	computers.peripherals.keyboard	microsoft	41.16	2019-11-08
745842	view	9200588	computers.peripherals.keyboard	microsoft	42.45	2019-11-08
745896	view	9200625	computers.peripherals.keyboard	microsoft	79.56	2019-11-08
746885	view	9200600	computers.peripherals.keyboard	microsoft	20.57	2019-11-08
746895	cart	9200600	computers.peripherals.keyboard	microsoft	20.57	2019-11-08
746900	view	9200600	computers.peripherals.keyboard	microsoft	20.57	2019-11-08
746918	cart	9200600	computers.peripherals.keyboard	microsoft	20.57	2019-11-08
746945	view	9200600	computers.peripherals.keyboard	microsoft	20.57	2019-11-08



Final Purchase History of selected

User	product_id	category_code	brand	price	date
	747927	9200600	computers.peripherals.keyboard	microsoft	20.57 2019-11-08
	906218	9101252	computers.peripherals.mouse	crown	3.83 2019-11-10
	1087861	9700217	computers.components.power_supply	gigabyte	70.79 2019-11-11
	1087983	9700217	computers.components.power_supply	gigabyte	70.79 2019-11-11
	1088017	9700217	computers.components.power_supply	gigabyte	70.79 2019-11-11
	1088066	9700217	computers.components.power_supply	gigabyte	70.79 2019-11-11
	1088181	9700217	computers.components.power_supply	gigabyte	70.79 2019-11-11
	1090277	9700217	computers.components.power_supply	gigabyte	70.79 2019-11-11
	1090521	9700217	computers.components.power_supply	gigabyte	70.79 2019-11-11
	1090684	9700140	computers.components.power_supply	huntkey	50.66 2019-11-11
	1414918	9200600	computers.peripherals.keyboard	microsoft	20.57 2019-11-14
	2230408	9200564	computers.peripherals.keyboard	sven	19.07 2019-11-16
	3160008	9200564	computers.peripherals.keyboard	sven	19.07 2019-11-21

Key insights:

1. User interacts with different brands in one session
2. Cart action is associated only with microsoft
3. Products belong to same category_code

Data Analysis – Summary

With our EDA we explored 2 of the 5 datasets we had, and deep-dived into the following levels of analysis and interaction:

- 1. Brand**
- 2. Price**
- 3. User**
- 4. Category**

This extensive analysis is our map for the next steps of this journey where we perform customer segmentation and build a recommendation system.