Online store analysis: events of “remove from cart”: a concern or a hidden value?

Sandra Morkūnienė

Table of Contents

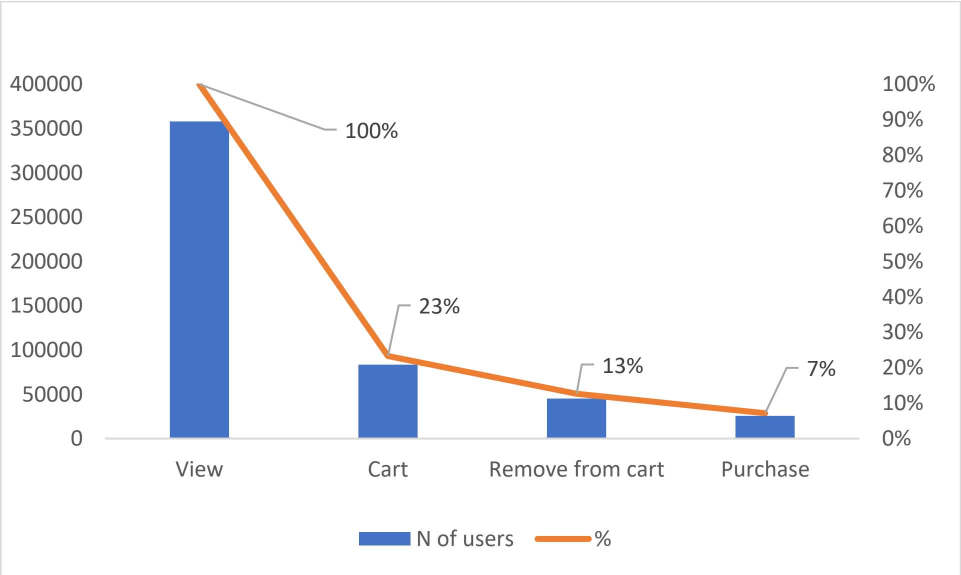
Executive Summary

Activity of the cosmetics e-commerce shop was reviewed. Key performance indicators were established for further monitoring in the Product analyst dashboard. The concerning number of “remove from cart” events were identified thus deeper analysis was made with an intent to get to know consumers’ behavior. Investigation revealed differences on web engagement and potential value per consumer for groups of purchasers and cart abandoners, on web engagement and value gained per consumer for groups of purchasers with removals and without removals and the groups of different volume purchasers. Price range of products which were dropped mostly was identified and the most removed from products were evaluated against top sellers, top viewed, top added to the cart and bought together products. Analysis results let us to make a conclusion that “remove from cart” effect was rather positive than negative for this online business in a way that removals from a shopping cart were rather the part of overall engagement with a web product than a concern this online business should deal with.

Problem

Funnel analysis was applied for the period from 1st of December 2019 to 31st of December 2019. It was noticed that only 7% of users who started an interaction with this online store made a purchase.

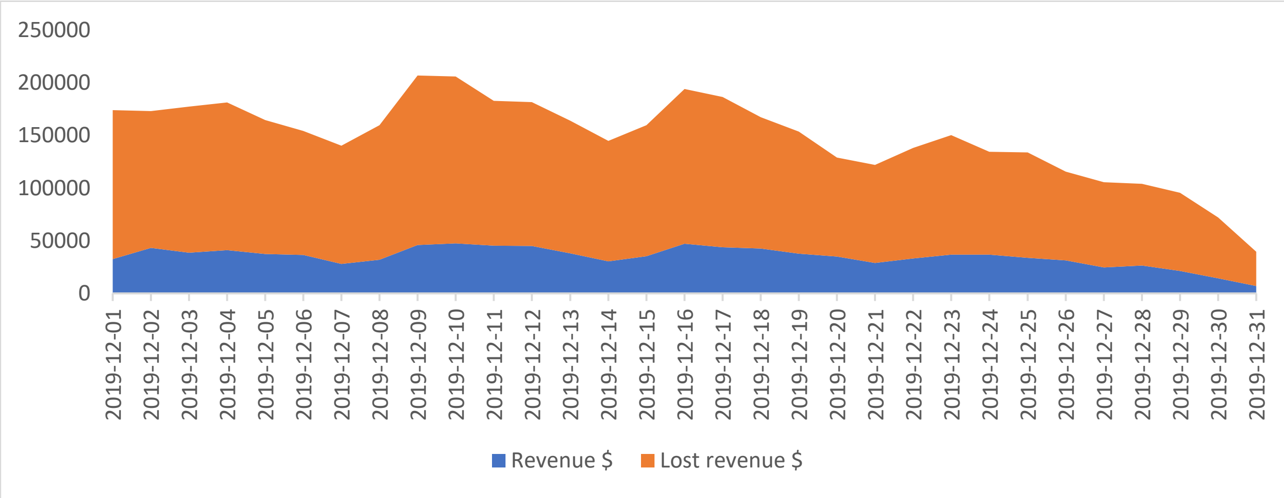
Chart 1. Funnel analysis for December 2019



For one month period 83458 users added at least one item to the cart, while 45217 users removed at least one item from the cart. 54% of those who had an intention to buy something removed at least one item from the cart and it appeared concerning. This led us to another question: how big was the potential revenue lost as a consequence of removed items from the cart?

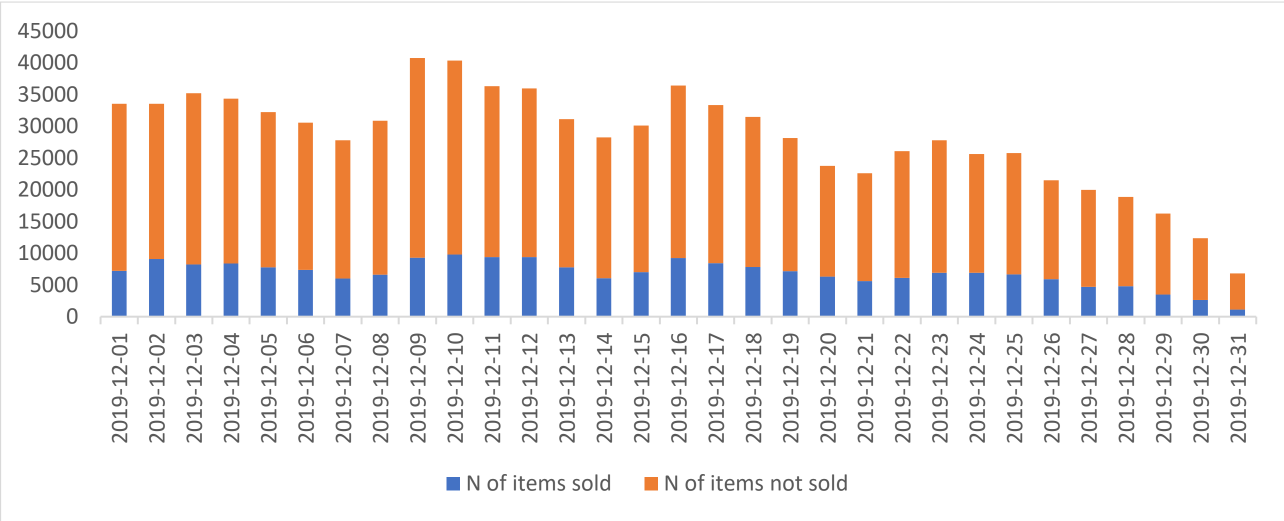
To answer this question, the potential revenue of the goods which have been removed from the cart were analyzed. In December of 2019 the customers put products to the cart with value of $4877332 but removed products from a cart with a value of $3538369, only $1338963 of revenue was actually gained.

Chart 2. Daily potential revenue $



It seemed that Pareto principle could be applied: on average, 77% of potential card value was removed, and only 23% of potential card value was actually purchased. For December of 2019 the average daily revenue was $34780, while the average daily lost revenue was $114141, so 3,3 times more revenue was lost than gained due to removed items from cart.

Chart 3. Daily potential number of items sold



Pareto principle could be applied for potential number of items to be sold: on average, 76% of potential items sold were removed, only 24% of potential items sold were actually purchased. Average daily number of items purchased was 6876, while average daily number of items removed from a cart was 21430, so 3,12 times more items were removed than purchased.

Analysis

With all that was discussed earlier an event of "remove from cart" might have some impact for online shop business, but it is usually hard to evaluate missed opportunities. If at first sight the damage from the potentially missed revenue might seem huge, we cannot be sure that the possibility to remove products and intensive usage of this option could have only negative effect – in case the users decide to replace one product with other product this could lead business to even greater revenue if the value and quantity of newly add products is greater.

It is valuable to make a distinction of a cart abandonment and "remove from cart" actions, since the removal of the item from a shopping cart not necessarily means that the client had not purchased anything. Some interpretations of consumer behavior could be discussed.

We can argue that even those who removed at least one item from a cart still processed to the purchase, i.e.: customers simply found it difficult to choose one product over the other. It was also possible that customers who removed at least one item did not purchase anything at the same day, but purchased later. Some customers removed at least one item from a cart and did not purchase anything. Finally, there might have been some cases when the customers abandoned their shopping cart without removing any item from a cart. Business should be most worried about those users who didn't make it to the act of purchase: either abandoned cart without any removals or removed at least one item and still didn't purchase anything.

First of all, it would be useful to identify the differences if any between those who made a purchase and those who abandoned shopping cart (either with or without product removals). Then we will analyze purchasers with and without removals to know how the same day purchasers differs in the number of removals, assuming that other day purchasers tend to remove more items from a cart and how this so called "purchase postpone to other day" might affect other important aspects for our business: revenue and engagement with our page in general. Removals might be related to the volume of items bought, so it might be beneficial to do a segmentation of customers by purchase volume and identify differences of engagement. Finally, which products customers tend to drop the most and evaluating them against top sellers, top viewed, top added to the cart and bought together products might give us an answer to the question: to worry or not to worry about online shop's goods assortment. Questions to be answered with further analysis:

- How purchasers of this online shop differ from cart abandoners?
- What part of these customers who removed at least one item still process a purchase and how they differed from those who purchased without any removals?
- Did the other day purchasers remove from a cart more often than the same day purchasers and how their web engagement differed from the same day purchasers?
- Which of those: one item, low/medium/high volume buyers drop products from cart mostly?
- Are more expensive products dropped more often?
- Which products are mostly dropped: do they match top products per other events and top products bought together?

How purchasers of online store differed from cart abandoners?

In December of 2019, 58147 customers abandoned the cart (no "purchase" event was identified for these users for reviewed time period). Their initial cart value (before any removals) was $2592351 (53% of total cart value: $4877333, of total number of users who had "cart" event), average order value $44,58. 23444 abandoned a cart but made at least one removal (their total value of initial cart was $1718388, after removals $70015, an average order value before removals was $73,30, after removals was $2,98). 34703 users abandoned a cart without removals, their total value of cart was $873965, an average order value was $25,18.

A comparison between group of purchasers and cart abandoners was made. Average number of views, average number of puts in the cart, average number of items removed, average initial cart value was chosen as the metrics.

Table 1. Comparison of purchasers and cart abandoners

Customer type	**Average N of Views**	**Average N of Cart**	**Average N of Removals**	**Average initial cart value**
Purchasers (N=25613)	22,10 (STD: 44,73)	17,91 (STD: 28,89)	13,41 (STD: 39,33)	$301,94 (STD: 322,38)
Cart abandoners (N=58147)	8,86 (STD: 30,70)	8,03 (STD: 22,06)	5,26 (STD: 24,37)	$44,58 (STD: 127,83)

Mean comparison with Student t test was made for these two types of consumers, statistically significant difference was identified ($p<0.01$) for all metrics. Cart abandoners viewed, put in the cart, removed from cart less items than purchasers. Average initial cart value for purchasers was greater than for cart abandoners.

What part of these customers who remove at least one item still process a purchase and how they differ from those who purchase without any removals?

In December of 2019, out of 45207 customers who once removed an item from a cart, 19650 made a purchase (43% of total number of customers who removed at least one item). The total number of customers who bought at least one item (N=25613) consisted of those who once removed at least one item from a cart 77% (N=19650) and those who bought straight without any removals – 33% (N=5963).

A comparison between group of purchasers with removals and purchasers without removals was made. Average number of views, average number of puts in the cart, average number of items purchased, average revenue per customer was chosen as the metrics.

Table 2. Comparison of purchasers with and without removals

Type of purchasers	Average N of Views	Average N of Cart	Average N items purchased	Average revenue per customer
Purchasers with removals (N=19650)	26,92 (STD: 49,66)	21,91 (STD: 31,75)	9,64 (STD: 11,70)	$45,83 (STD: 57,12)
Purchasers without removals (N=5963)	6,21 (STD: 11,77)	4,73 (STD: 5,85)	3,99 (STD: 4,42)	$29,78 (STD: 30,12)

Mean comparison with Student t test was made for these two types of consumers, statistically significant difference was identified ($p<0.01$). Purchasers without removals viewed, put in the cart, purchased less items than purchasers with removals. More revenue per consumer was gained from purchasers with removals.

Do the other day purchasers remove from a cart more often than the same day purchasers and how their web engagement differs from the same day purchasers?

During December of 2019 25613 users of this online shop made a purchase. 59% (N= 15162) of them were the same day purchasers and 41% (N= 10451) other day purchasers - those who started session one day and finalized a purchase other day. The other day purchasers accounted for 238088 (69%) items removed from a cart, the same day purchasers accounted for 105402 (31%) items removed from a cart.

Behavior of these two types of purchasers was analyzed comparing the average number of removals from a cart. Mean comparison with Student t test was made for these two types of consumers, statistically significant difference was identified ($p<0.01$). The same day purchasers on average removed smaller number of items from a cart compared to the other day purchasers.

To get the broader view the average number of events such as “view”, “cart”, “purchase” and revenue gained per customer was also taken into the consideration. For these two groups average revenue per customer, average number of “views”, “cart” events and number of items purchased statistically significantly differed ($p<0.01$) as well.

Table 3. *Comparison of other day purchasers and the same day purchasers*

Customer type	Average N of items removed per customer	Average N Views	Average N Cart	Average N items purchased	Average revenue per customer
Other day purchaser (N=10451)	22,78 (STD: 56,25)	36,58 (STD: 61,02)	25,57 (STD: 37,24)	9,99 (STD: 11,92)	$47,82 (STD: 57,32)
Same day purchaser (N=15162)	6,95 (STD: 18,15)	12,11 (STD: 23,85)	12,63 (STD: 19,62)	7,17 (STD: 9,68)	$38,14 (STD: 48,68)

Even though the same day purchasers were accounted for $578334, 54% of total revenue (compare to the other day purchasers $499831, 46% of total revenue gained), on average they viewed 3 times less items, 2 times less added items to the cart, gained less revenue per customer, purchased smaller number of the goods than the other day purchasers. The other day purchasers removed items from a cart more often, but had greater average number of all events and gained more revenue per customer.

Which of these: one item, low/medium/high volume buyers drop products from a cart mostly?

To answer this question all purchasers were divided in 4 groups by the volume of items purchased, and majority of purchasers belonged to low volume: 2 to 10 items bought group (see table 4).

Table 4. Revenue by purchaser group

Purchaser group	**Revenue**	**Revenue % by groups**	**Average revenue per customer**
Mid volume:11-30 items bought (N=5298)	$350174	32%	$66,09
Low volume:2-10 items bought (N=16603)	$491195	46%	$29,58
High volume:31 & more items bought (N=802)	$159690	15%	$199,11
One item purchasers (N=2910)	$77105	7%	$26,50

46% of total revenue gained was from low volume purchasers, the second group was mid volume purchasers accountable for 32% of total revenue. The group of one item purchasers was accountable for 7% of total revenue, but this group was greater in number than high volume purchasers accountable for 15% of total revenue.

The p-value corresponding to the F-statistic of one-way ANOVA was lower than 0.05, suggesting that the one or more purchasers' groups were significantly different from other. Post hoc Tukey test was applied to identify which groups differed from each other. All combinations of groups showed statistically significant differences ($p<0.01$) on all metrics.

Table 5. Comparison of purchaser groups by volume

Purchaser group	**Average N of Views**	**Average N of Cart**	**Average N of items removed**	**Average N items purchased**
Mid volume:11-30 items bought (N=5298)	39,71	34,49	26,78	16,55
Low volume:2-10 items bought (N=16603)	17,19	11,68	8,67	4,99
High volume:31 & more items bought (N=802)	68,82	90,69	65,57	49,62
One item purchasers (N=2910)	5,16	3,21	1,73	1,00

On average, high volume purchasers viewed, added to cart, removed and purchased the greatest number of items compare to other segmented groups. One item purchasers were least engaged with web product. In general, these user groups which were more engaged with the product bought greater number of products.

Are more expensive products dropped more often?

In this online shop products could cost in the price range from $0,05 to $327,28. 81,26% of all products belongs to the price range from $0.01 to $10. 85,10% of total removed items belonged to the price range from $0.01 to $10 (see table 6).

Table 6. Number of items removed within price range

Price range	N products with price	% with price	N products removed	% of total removed
less_than_$1	3085	6,33%	2569	6,93%
$1_to_9.99	36494	74,92%	28997	78,18%
$10_to_19.99	5711	11,72%	3686	9,94%
$20_to_29.99	1668	3,42%	924	2,49%
$30_to_39.99	744	1,53%	396	1,07%
$40_to_49.99	351	0,72%	190	0,51%
$50_to_59.99	241	0,49%	109	0,29%
$60_to_69.99	128	0,26%	70	0,19%
$70_to_79.99	75	0,15%	42	0,11%
$80_to_89.99	47	0,10%	21	0,06%
$90_to_99.99	46	0,09%	26	0,07%
$100_and_more	119	0,24%	62	0,17%

It can be stated that the cheaper products were dropped from a cart more often, but in general cheap price range of goods was the most abundant.

Which products are mostly dropped: do they match top products per other events and top products bought together?

Firstly, correlation analysis was made on the product level (N= 26810) in order to find out the relation among events and revenue earned from particular product (see table 7).

Table 7. Correlation among web events and revenue gained by product

	Revenue	*N of views*	*N of cart*	*N of removed from a cart*
N of views	**0.70**			
N of cart	0.45	**0.66**		
N of removed from a cart	0.49	**0.67**	**0.91**	
N of purchase	**0.52**	**0.70**	**0.94**	**0.92**

All variables had some medium to high degree of positive correlation with each other meaning the bigger number of one variable the bigger number of the other variable and vice versa.

Revenue gained from particular product correlated with all events but mostly with the views, the more product was viewed the more revenue that product generated and vice versa. Some correlations might seem counterintuitive due to positive relation meaning that i.e., the same product could be removed from the cart many times, but many items of this product could still be sold. So, the removal from a cart not necessarily indicated low purchases of particular products. The product could be frequently added or removed from a cart but still have decent number of purchases.

Knowing that for particular product events being medium to high correlated, it was not surprise that some part of top 10 products removed from a cart appeared in the lists of top 10 most viewed, top 10 most put in the cart, top 10 purchased (see table 8).

Table 8. Products by revenue gained and number of events

Top 10 by Revenue	Top 10 by Views	Top 10 by Cart	*Top 10 by Removals*	**Top 10 by Purchase**
5850281	***5809910***	***5809910***	***5809910***	***5809910***
5560754	5909810	***5802432***	***5809912***	***5854897***
5809910	5877454	***5700037***	***5854897***	***5802432***
5751422	*5809912*	***5854897***	***5815662***	***5700037***
5751383	5886282	***5815662***	***5802432***	***5809912***
5877454	5877456	***5809912***	***5700037***	***5833330***
5909810	5649236	5836522	*5809911*	**5304**
5846437	*5809911*	5843836	***5751422***	***5751422***
5849033	5769877	**5304**	***5833330***	***5815662***
5792800	5856186	5700046	***5751383***	***5751383***

3 out of 10 mostly viewed products appeared among those products which generated the most revenue. 6 out of 10 mostly removed products were among the mostly put in the cart products and 9 out of 10 among those the mostly purchased. 4 out of 10 mostly viewed products were among top 10 those which were mostly removed, and 5 appeared among those which were the mostly purchased. Most removed products appeared in 9 out of 10 pairs of mostly bought together products (see table 9).

Table 9. Most bought together product pairs

Product pairs	N times bought together
5809910, 5809912	325
5809910, 5809911	272
5751383, 5751422	173
5809911, 5809912	152
5809910, 5816170	150
5751422, 5849033	114
5833325, *5833330*	111
5809910, 5816166	105
5700037, 5802432	102
5833325, 5833326	94

Out of 28968 orders during December of 2019 first pair of products was bought together only 325 times, from a first sight it does not seem a huge number. What if in general orders are formed with the same product in high volume instead of multiple products, and if so, consumers might remove from their shopping carts not the multiple products but they just lower the quantity of the same product – further product level analysis on consumer behavior needed to address this question.

Conclusions and future recommendations

This analysis was initiated after concerning number of "remove from cart" was noticed but before naming it as if it is an issue to be urgently solved it was important to understand whether product item removals had negative effect for this online business. It was argued that "remove from cart" actions not necessarily mean that customers won't purchase anything – on the contrary, consumers' tendency to change the mind while making purchase decisions could lead business to even greater revenue.

Firstly, we analyzed how cart abandoners differed from purchasers. Analysis revealed that on average cart abandoners viewed, put in the cart, removed from cart less items than purchasers. Their average initial cart value was smaller compare to purchasers. Their engagement with the web product was lower than purchasers' engagement and expected value from these customers were also lower. The biggest drop of users in a funnel was between "view" and "cart" events, the bounce rate for this online store page for this month was 66,69%, this signaled to us that the business pain point was at this stage of consumers' journey and in order to solve this issue we should further explore this stage. The recommendation for further analysis would be to explore what kind of monetary value products customers are viewing and put in the cart the most and compare between the group of purchasers and cart abandoners.

Analysis of purchasers with and without removals supported assumption that "remove from cart" event could lead business to the greater revenue for this particular online store since purchasers with removals were more engaged with web product and average revenue gained per consumer was greater than from purchasers without any removals. This tendency to remove items from a shopping cart might look differently taking into account additional demographical variables, such as age, gender. In case this tendency is more prevalent in one or another age or gender group, business could get advantage of it by adjusting their selling strategy accordingly.

Postponing the act of purchase might led to the greater revenue gained, meaning that business could take advantage of some stumbling consumer behavior. Our analysis supported this assumption since on average the other day purchasers bought more items and greater revenue per consumer was gained from their purchases. Removals from a shopping cart was more abundant in the group of the other day purchasers compare to the same day purchasers but it was more related to the overall engagement with a web product than a concerning issue. This online shop during reviewed period had more the same day purchasers than the other day purchasers, but value per consumer was greater for the other day purchasers. Business management could ask the question: which segment of consumers we would like to grow and the answer indicate the choice between “less money but now” or “more money but later”. At least regular monitoring on these two types of consumers might be beneficial for further selling strategies.

While analyzing purchasers by number of bought items we found out that the group of low volume purchasers who bought from 2 to 10 items were most abundant, but the value gained per customer was the greatest in the group of high volume purchasers. High volume purchasers were in general the most engaged with the web product by viewing, adding to the cart and finally buying product items. “Remove from a cart” events for these groups seemed to be rather a form of engagement with a web product than an issue. It would be beneficial for this online store to investigate what kind of products (cheap/medium/high price) were purchased by these four groups and make up strategies to encourage other than high volume groups to buy in higher quantity.

From a product level analysis for this online store we can conclude that the same product could be frequently added or removed from a cart but still have decent number of purchases: 9 out of 10 mostly removed products were among those the mostly purchased. “Remove from cart” event indicated rather the popularity of certain product than i.e.: price, functionality issues related to the product. The cheaper products were dropped from a cart more often, but in general cheap price range of goods was the most abundant, so the cheapest products had the greater possibility to be dropped from a cart. From a business perspective, it would be beneficial to do constant monitoring of the products by number of removals together with their prices to detect negative tendencies if any.

Limitations

The dataset consisted of information from one month period but online store activity is ongoing process, so certain events for particular consumers appeared in a dataset without having prior actions, i.e.: some users had “cart” event for a product without “view” event for the same product. This could happen due to strict cut on certain date and time.

Segmentation based on certain consumer behavior: purchasers/cart abandoners, purchasers with removals/purchasers without removals, purchasers by volume could be even more beneficial having additional information such as consumer’s demographics, geographics, lifetime relation with this store (recency/ monetary).

Having only product code instead of true product title didn’t let us fully understand which kind of goods, or categories were accountable for the most revenue, could be sold together – it is hard to give recommendations for selling strategies or marketing improvements not knowing what actually was hidden under product code.

Finally, we have analyzed activity from one month period and even if we identified statistically significant differences between customer groups, certain conclusions about business decisions should be based on longer period analysis.

Overview of data used

Dataset consisted of more than 3,5mln rows and was taken from the source: https://www.kaggle.com/datasets/nowingkim/ecommerce-data-cosmetics-shop. Time frame for this dataset was one month from 2019-12-01 to 2019-12-31 of a multi-category online Cosmetics store in the form of customers' web events. The initial raw file variables:

time	Time when event happened at (in UTC).
event_name	4 kinds of value: purchase, cart, view, remove*from*cart
product_id	ID of a product
category_id	Product's category ID
category_name	Product's category taxonomy (code name) if it was possible to make it.
brand	Downcased string of brand name.
price	Float price of a product.
user_id	Permanent user ID.
session	Temporary user's session ID. Same for each user's session. Is changed every time user come back to online store from a long pause.
category_1	Largest class of product included
category_2	Bigger class of product included
category_3	Smallest class of product included

Due to high number of missing values in brand and category related variables (>90% of missing values), only variables: time, event_name, product_id, price, user_id, session were used for further analysis. Raw file in CSV format consisted of more than 3,5 million records was divided by rows to 4 files with Python pandas and loaded to BigQuery as 4 events tables for further analysis with SQL.

Snippet of Python code used to cut dataset into parts for fluent uploading into Big Query:

```
chunk_size = 900000
for i in range(len(min_events) // chunk_size + 1):
    min_events[i*chunk_size:(i+1)*chunk_size].to_csv(f"events_{i:02d}.csv",
                                                    sep=";", index=False)
```