E-commerce Funnel & Conversion Analysis (500K+ User Interactions)

Ryan Leem

2026-02-02

Data Description

This analysis uses the `2020-Jan.csv` dataset, which contains user interaction data from an online cosmetics e-commerce platform. The dataset captures user behavior during January 2020, including actions such as viewing, purchasing, adding to cart, and removing from cart for individual cosmetic items.

The primary variables used in this analysis are `event_type`, which records the type of user action (e.g., product view or purchase), and `brand`, which in this dataset represents the cosmetic item associated with each interaction. These variables were aggregated to analyze item-level attention and conversion performance.

To ensure meaningful comparisons, records with missing cosmetic item names were removed. Conversion rates were calculated by dividing purchases by view for each cosmetic item, and low-traffic items were filtered out when necessary to avoid misleadingly high conversion rates due to small sample sizes.

Figure 1 below shows the total number of Views, Purchases, Items Removed from Cart.

Shows each event type (Views, Purchase, Remove from Cart, Add to Cart)

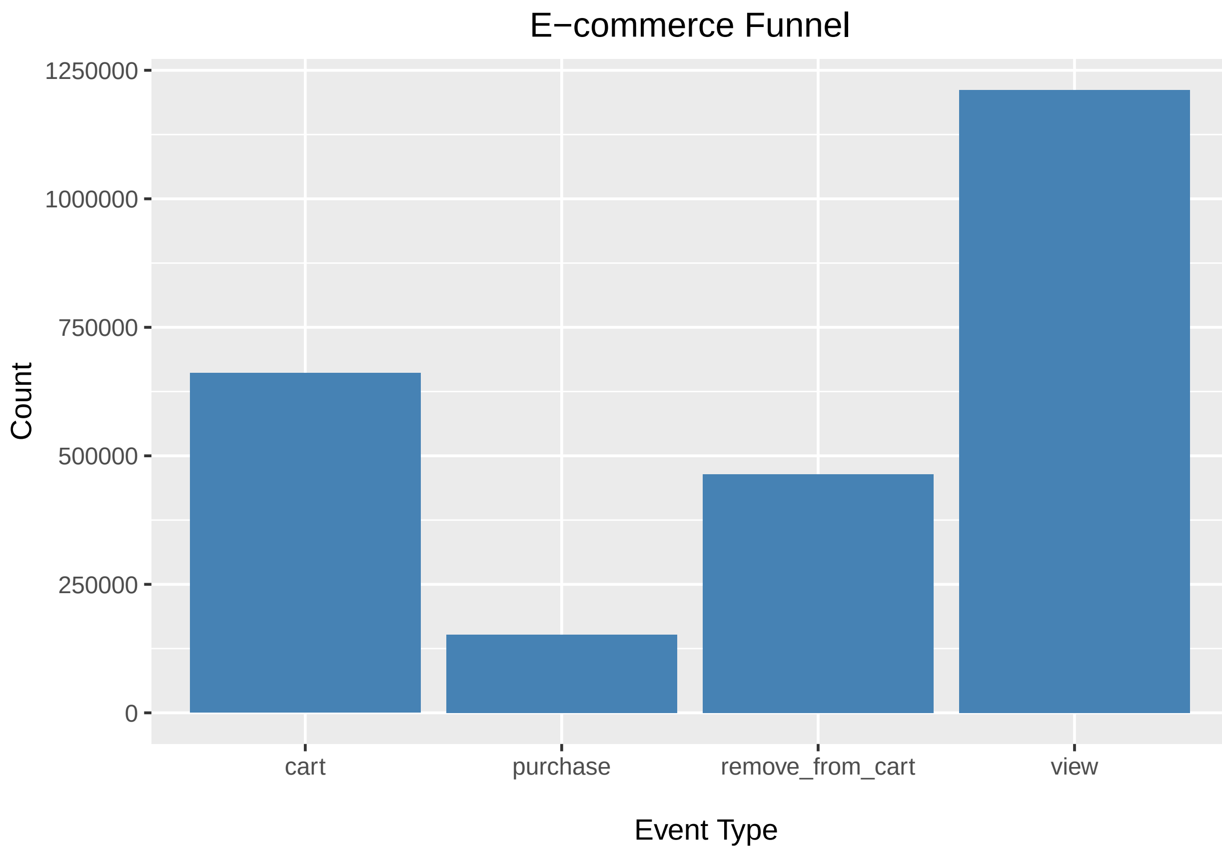


Figure 2 shows the top 10 cosmetic items by conversion rate ordered from the highest to lowest. The conversion rate is calculated as the number of purchases divided by the number of views for each cosmetic item. Only cosmetic items with at least 10,000 views were included to avoid misleadingly high conversion rates driven by very small sample sizes.

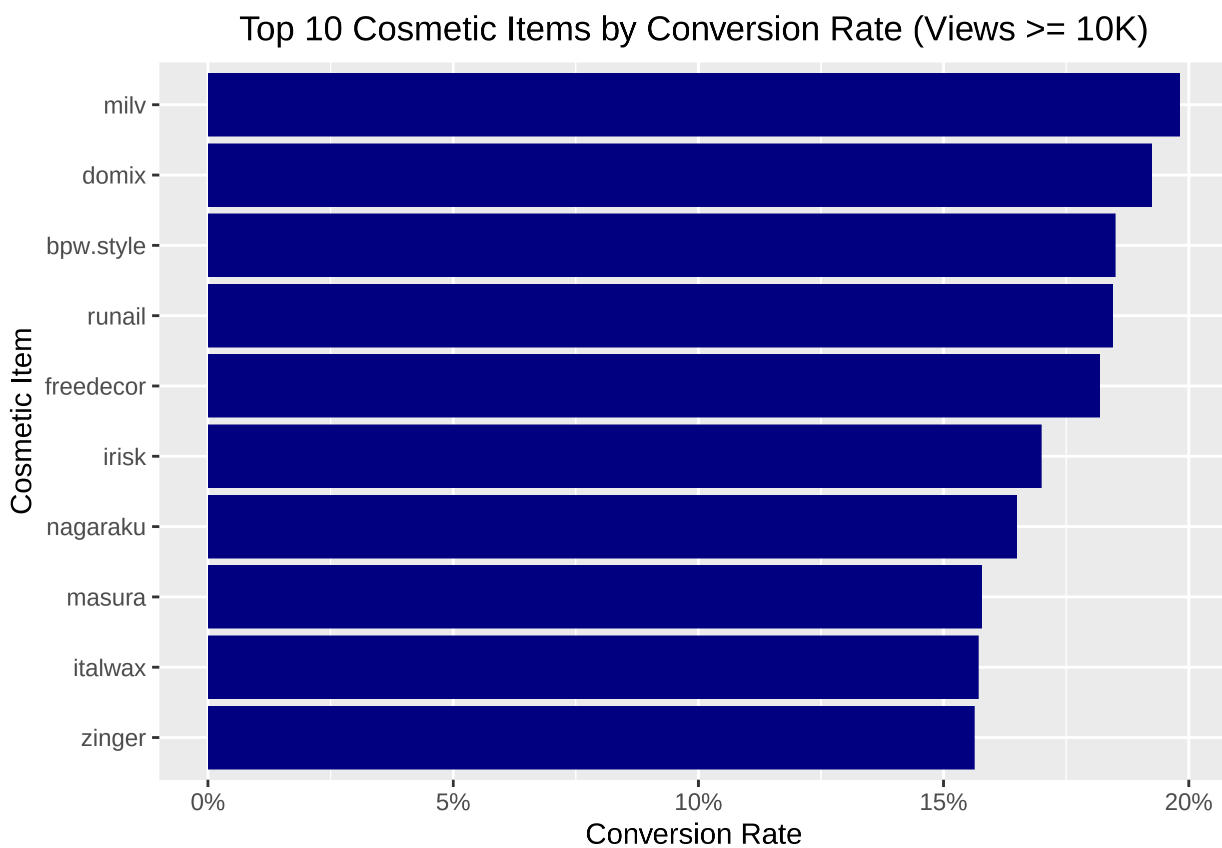


Figure 3 shows the conversion rate of cosmetic items with at least 25,000 views. The top being the highest views and the bottom being the lowest amount of views. Their *conversion rates* are shown as well.

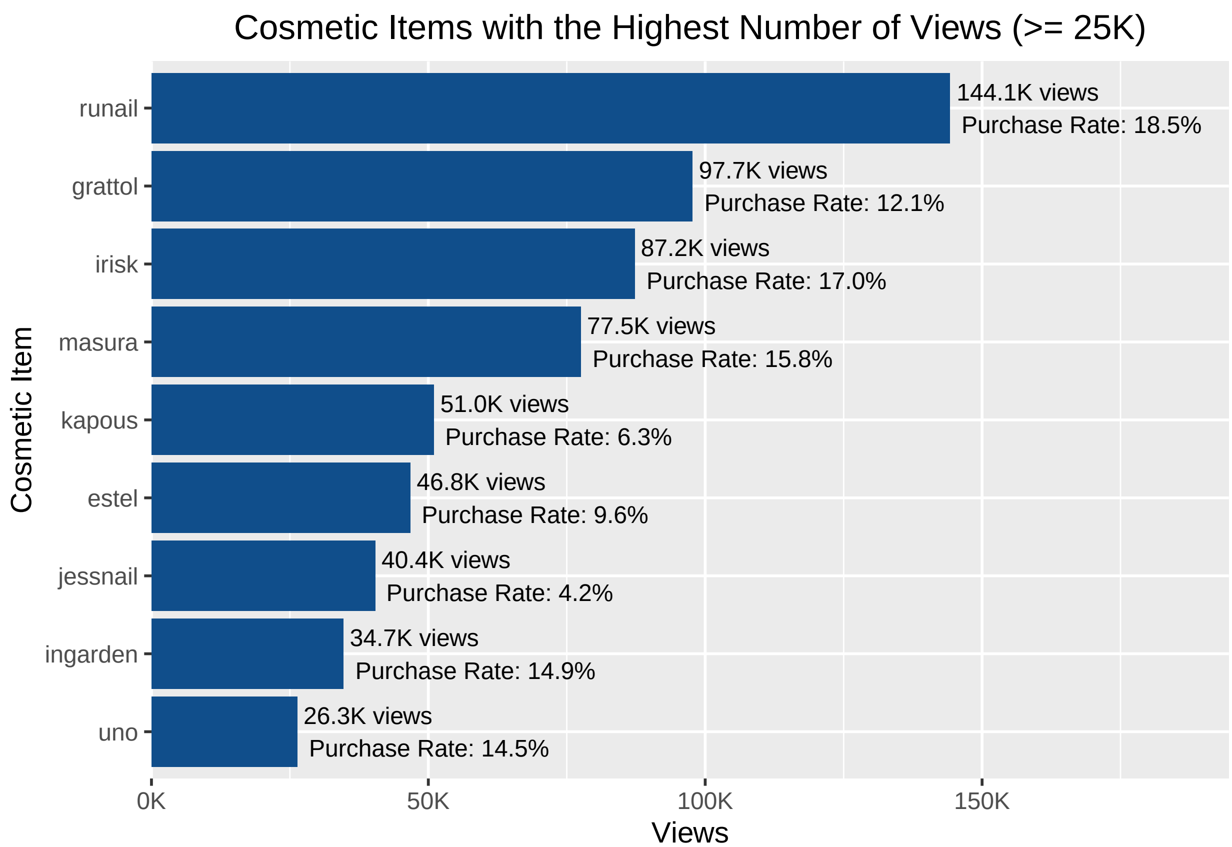
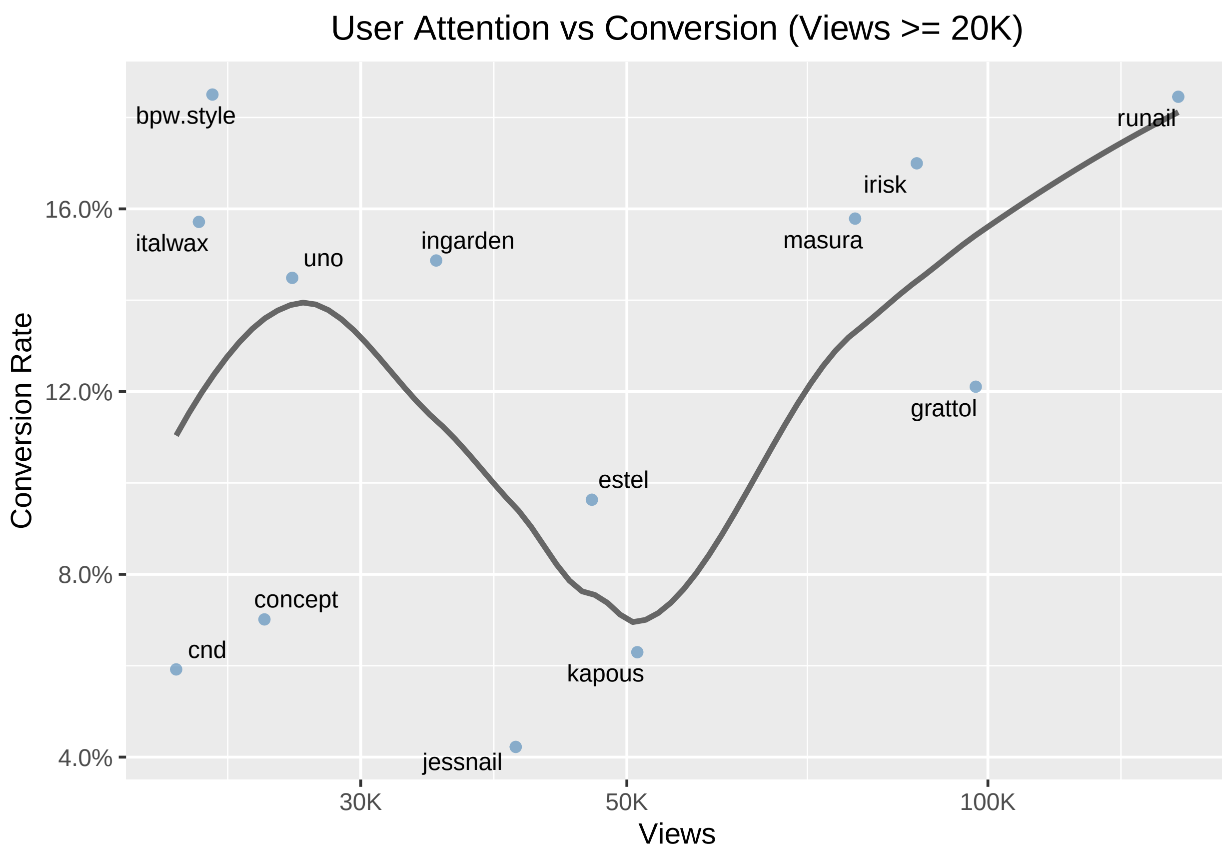


Figure 4 shows cosmetic items with at least 20,000 views and shows their views and percent conversion rate. The smooth trend line summarizes the overall relationship between views and conversion rate.




```r
model <- lm(conversion_rate ~ log(view), data = brand_funnel |> filter(view >= 20000))
summary(model)
```

```
##
## Call:
## lm(formula = conversion_rate ~ log(view), data = filter(brand_funnel,
##     view >= 20000))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.07899 -0.04001  0.02168  0.03352  0.07715
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.12228    0.24391  -0.501    0.626
## log(view)    0.02296    0.02278   1.008    0.335
##
## Residual standard error: 0.05085 on 11 degrees of freedom
## Multiple R-squared:  0.0845, Adjusted R-squared:  0.001276
## F-statistic: 1.015 on 1 and 11 DF,  p-value: 0.3353
```

Statistical Insight To test whether product visibility actually drives purchasing behavior, I ran a simple linear regression using log-transformed views to predict conversion rate for items with at least 20,000 views.

The model produced an R^2 of 0.0845, meaning that only about 8.5% of the variation in conversion rate can be explained by differences in view count. In addition, the relationship was not statistically significant ($p = 0.335$).

In other words, having more views does not reliably translate into a higher conversion rate.

This supports the earlier visual findings: traffic alone isn't enough. Other factors — such as pricing, product appeal, or checkout friction, likely play a much larger role in determining whether a customer completes a purchase.

Findings The results suggest that product visibility and purchasing efficiency are not strongly aligned. While Runail receives the highest number of views, it does not generate the strongest conversion performance. In contrast, brands such as Milv convert a higher percentage of viewers despite attracting less traffic.

The funnel breakdown highlights substantial user drop-off before purchase. Although many users add products to their cart, a noticeable portion remove them rather than completing checkout. This suggests friction in the final stages of the buyer journey.

Taken together, the analysis indicates that increasing exposure alone may not significantly improve sales. Improving product positioning, pricing clarity, and checkout experience may have a greater impact on conversion performance.

Future Work

Future analysis could incorporate pricing data, product category segmentation, or user-level purchase history to better understand the drivers of conversion behavior.