Recommender System Test Analysis

By: Mega Septiandara

Technical Description

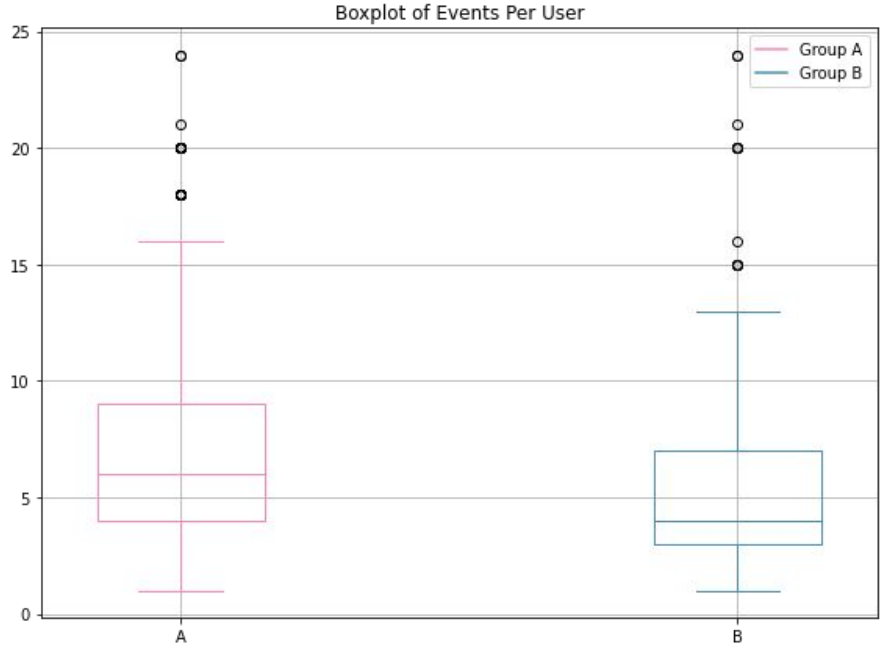
- Test name: recommender_system_test
- Groups: A (control), B (new payment funnel)
- Launch date: 2020-12-07
- The date when they stopped taking up new users: 2020-12-21
- End date: 2021-01-01
- Audience: 15% of the new users from the EU region
- Purpose of the test: testing changes related to the introduction of an improved recommendation system
- Expected result: within 14 days of signing up, users will show better conversion into product page views (the product_page event), product cart views (product_cart) and purchases (purchase). At each of the stage of the funnel product_page → product_card → purchase, there will be at least a 10% increase.
- Expected number of test participants: 6000

Overall Conclusion

- Test running date: 2020-12-07 - 2020-12-29
- Audience: 9% of the new users from the EU region
- Total test participants: 2594 users
- Result:
    a. The number of events per user is distributed equally among all samples. There is a slight difference between group A and B, group A has higher event average while group B has higher maximum.
    b. The number of events is not distributed equally by days. In group A, there is a spike on December 13-25. While for group B it is distributed evenly in the whole experiment period.
    c. Out of the three scenarios we tried, two scenarios (going through each stage once and skipping product cart) result in the purchase conversion of around 30%. One scenario where users skip the product page, the purchase conversion is only around 15%. Users in group B shows lower conversion.
    d. We conducted a comparison of proportions test above, we found no statistically significant difference between the two groups in all events except in the product page.

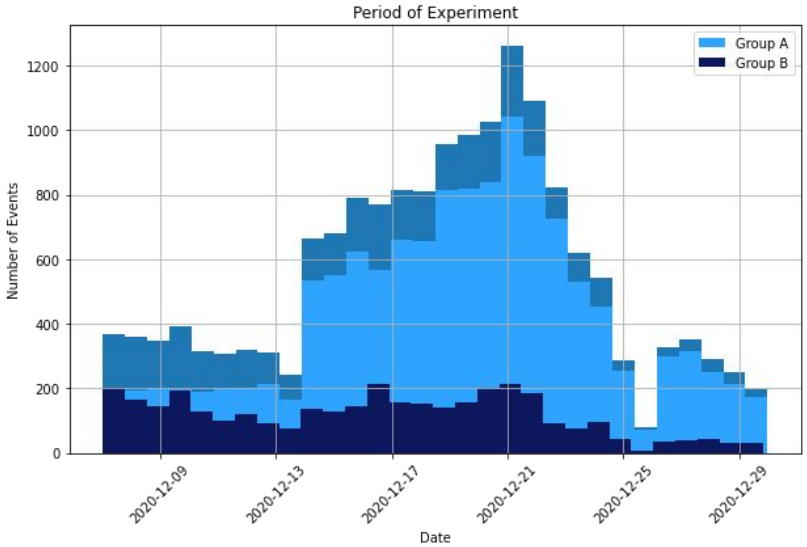
Events Per User

- The average of all groups is around 5.94 events per user, with maximum 24 events. Group A has higher average at 6.2 events per user, while group B has really low average at 5 events per user.
- Based on the Boxplot, we can see that group A has higher quartiles than group B. Some outliers in the group B is around the same level as the 3rd and upper quartiles of the group A at around 10-16 events per user.
- Group A shows the higher average value of events per user, as the distribution of group A's events per user is closer to all groups' distribution, while group B distribution is lower.



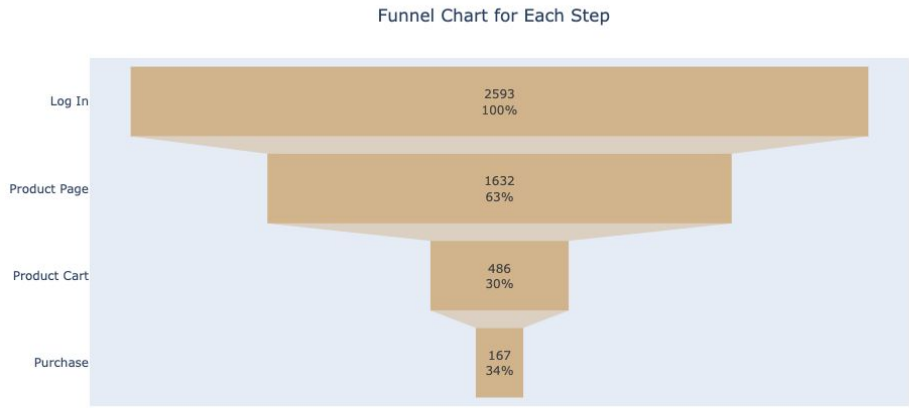
Period of Experiment

- Based on the technical description, the launch date is December 7 and the experiment ends on January 1. The data only shows the final event at December 29.
- It could either be that there is no usage of the online store preceding the new year or the A/B test entry system runs into an error and doesn't capture the final days.
- Based on the histogram on the right, we can see that the distribution of daily events for group A is higher particularly on December 14 until December 25. That would coincide with preparation for the holidays where people would buy gifts for family and friends. Group B on the other hand shows pretty even distribution of events throughout the experiment period.



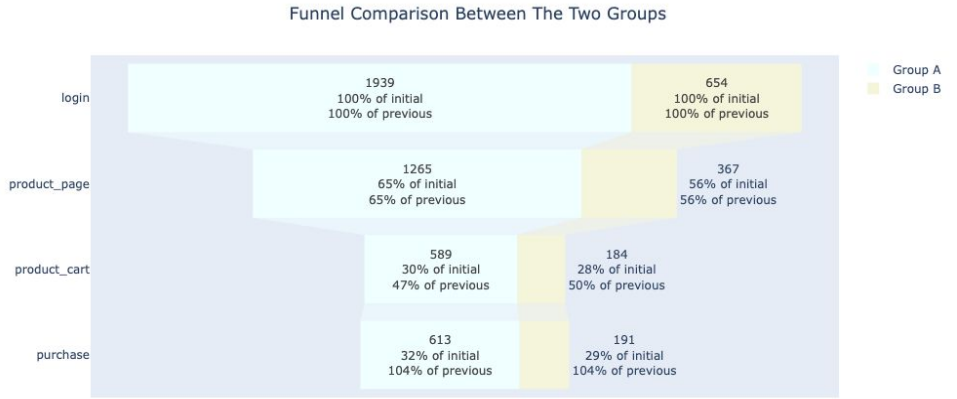
Event Conversion

- In the scenario where users go through each stage, more than 60% of users go to the product page. Less than 30% of users put an item in the cart, then just a bit over 30% of users make the purchase.
- In the second scenario (users skip product page), around 16% of users make the purchase.
- In the third scenario (users skip product cart), the percentage of purchase is higher to the first scenario at around 33%.



Event Conversion Per Group

- Group A shows consistently higher conversion on each stage than group B.
- Group A also shows higher purchase than group B.
- Nevertheless, it looks like both groups show quite similar behaviour on each stage based on the conversion percentage, except in the product page where group B shows much lower conversion (9% less than group A).



Hypotheses Testing

The compare proportions test is used to evaluate if the frequency of occurrence of some event, behavior, intention, etc. differs across groups. We test this hypothesis using sample data from the Control Group and the Test Group.

1. **Null and Alternative Hypotheses**

H0: The true population difference in proportions is equal to 0. This means that the conversion rate of Control Group is the same as the Test Group.

H1: The true population difference in proportions is not equal to 0. This means that the conversion rate of Control Group is not the same as the Test Group.

2. **Criteria for a decision (alpha value)**

We will use the Bonferroni correction as we are performing many independent or dependent statistical tests at the same time. This post hoc test sets the significance cut off at α/n. As we will run the test for each event, we would divide the alpha by 4.

3. **Result**

The only statistically significant difference between the two groups is on product page event. This is reflected in the funnel analysis previously that group B has 9% less conversion in the product page. We also found that when users skip product page, the purchase conversion is only around 16%.