AB Testing Data Analysis Exercise

Executive Summary

Two web pages (simply called ‘new_page’ and ‘old_page’) were distributed to users to investigate if the landing page that the users visited impacted their conversion rate. The users of the old page were selected to be in the control group, while its variant, the new page, was provided to users in a contrast group. The number of users that were converted after visiting each page was recorded.

Data

The data can be downloaded from the link in a description below.

Hypothesis

Investigate if the landing page that a user visits has any effect on their conversion rate. The null hypothesis is that there is no difference in effectiveness of the new_page in converting users compared to the old_page. Check to see if the null hypothesis could be rejected.

Task

1. Explore the data using pandas
    - Check the shape of the data: There was an excess of around 400,000 users that visited the web-pages more than once (possibly visited both web-pages).
    - Check the number of unique entries from the “user_id” column

Having some users access both pages would make the results of the analysis invalid. Therefore, the users in the control group were paired only with the ‘old’ landing page and those in the treatment group were paired only with the ‘new’ landing page

2. With the information above, determine the number of users that landed on both pages (old and new) and drop the duplicate entries.
3. Plot a bar chart to visualise the number of users in the control group and treatment groups.
4. Plot a bar chart to show the proportion of users that were converted by either web-pages against those that were not converted to investigate the overall effectiveness of either pages in converting users.

5. Calculate the mean for the number of users that were converted for the control group and that for the treatment group.
6. Plot a bar chart showing the proportion of users that were converted in the control condition vs the treatment condition for comparison.
7. Determine if the null hypothesis could be rejected.