**Globox A/B Testing analysis report**

**Summary:**

I recommend further prolonging of the A/B test for at least 4 weeks period to rule out the novelty effect and the specific customer behavior keen to dates accruing at the end of January (very high traffic looking for exotic goods).

We need to make sure that the banner is making enough revenues for the front-page space he is getting, there was a total of 1270$ increase in revenue for the twelve days period.

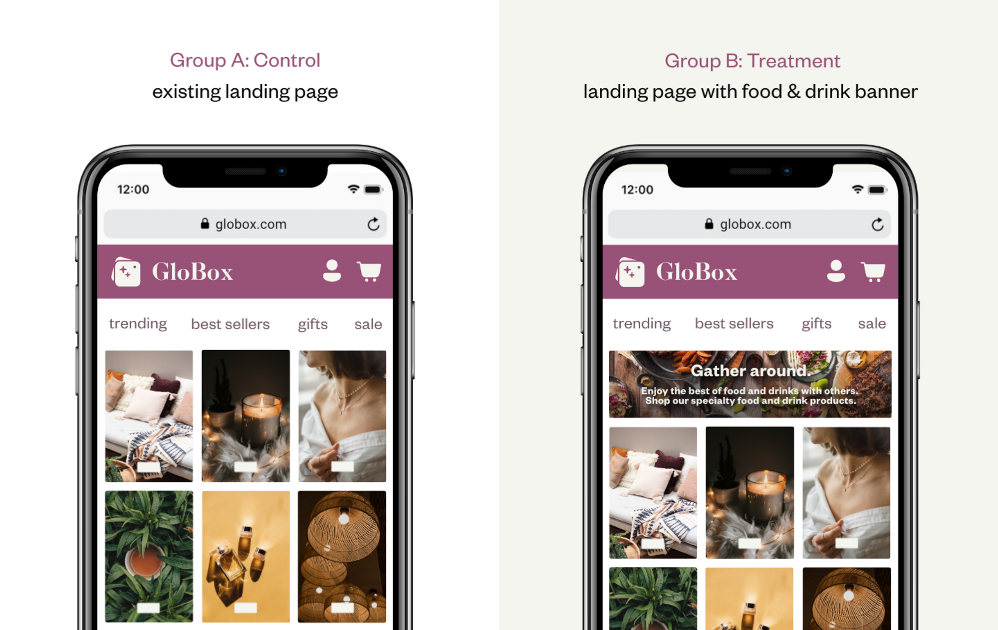
I would further go by dividing the spent column to the different categories appearing on the site to see if the banner directly effects the food and drinks category sales.

**Context:**

GloBox is primarily known amongst its customer base for boutique fashion items and high-end decor products. However, their food and drink offerings have grown tremendously in the last few months, and the company wants to bring awareness to this product category to increase revenue.

**A/B Test Setup**

The Growth team decides to run an A/B test that highlights key products in the food and drink category as a banner at the top of the website. The control group does not see the banner, and the test group sees it as shown below:



The test ran for 12 days from 25th of January to the 6th of February.

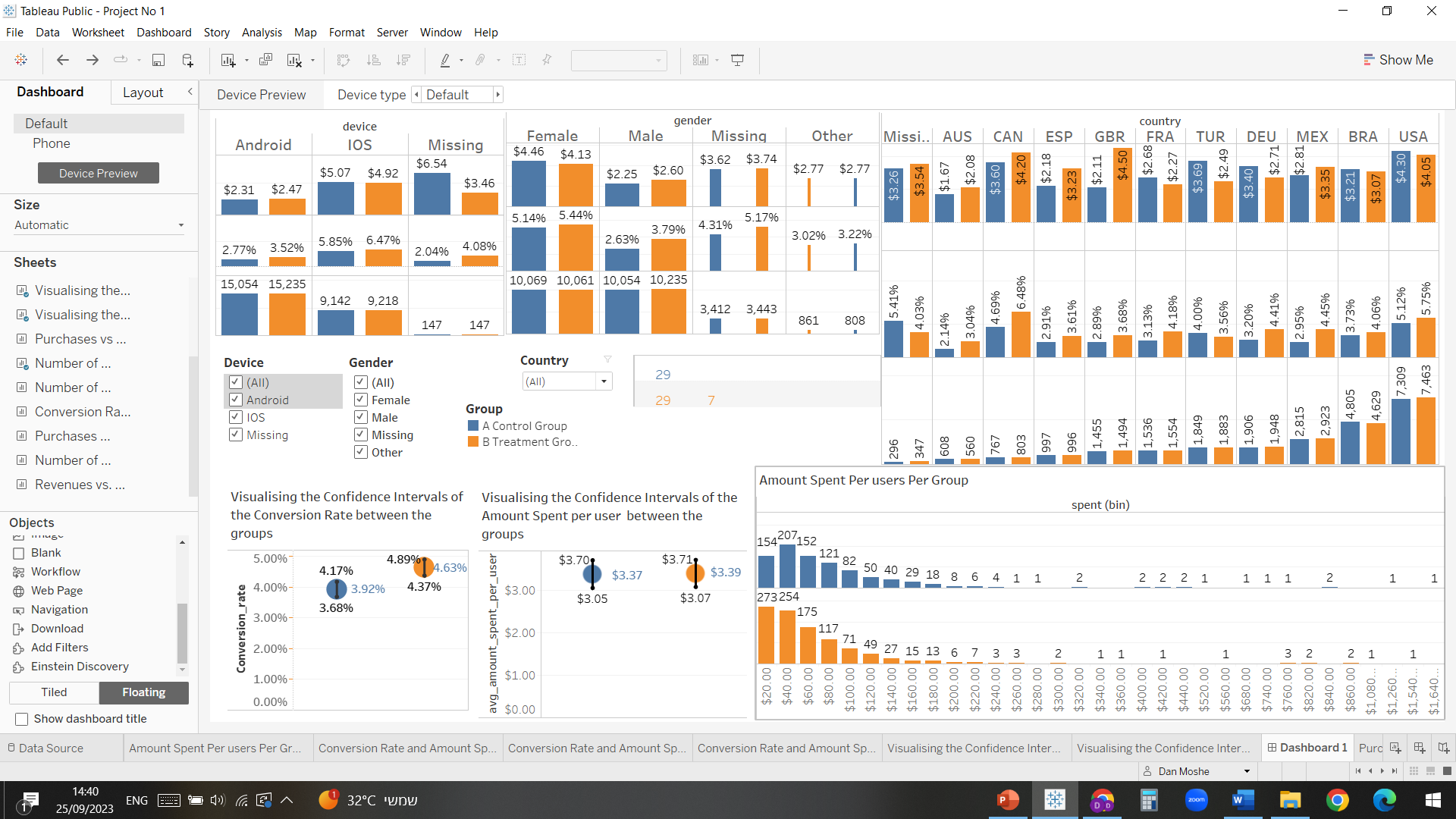
There were 24343 users in the control group, 24600 users in the treatment group and 48943 in total.

**Results:**

In order to determine whether there was a difference in revenue per user between the two groups, we ran two hypothesis test, one to check the conversion rate and one to check the average amount spent per user.

We saw a statistically significant difference between the 2 groups at the 5% significance level (p=0.01), the 95% CI for the difference in conversion rate between the 2 groups is (0.0034, 0.0106).

We didn't saw a statistically significant difference between the 2 groups at the 5% significance level (p=0.94), the 95% CI for the difference in average amount spent between the 2 groups is (-0.439,0.471).



תמונה שמכילה טקסט, צילום מסך, תוכנה, סמל מחשב

התיאור נוצר באופן אוטומטי

We took care of all the blank values with the following function in sheets:

=IF(isBlank(cell\_no),"Missing",cell\_no)

And copied the outcome on the previous blank values column.

There were 6855 missing values in the gender column, 294 in the device column and 643 in the country column.

We used box plot to see the statistical deviation of the purchase's amounts between the treatment and control group for relevant dates:

תמונה שמכילה טקסט, צילום מסך, תוכנה, סמל מחשב

התיאור נוצר באופן אוטומטי

As can be seen on almost all dates the box is smaller for the treatment group which indicates us there where smaller size purchases then in the control group, which in turn correlates to the banner goal.

We computed conversion rate and average amount spent for different countries:



And saw that the bigger the user base from each country the smaller the CI for each parameter.

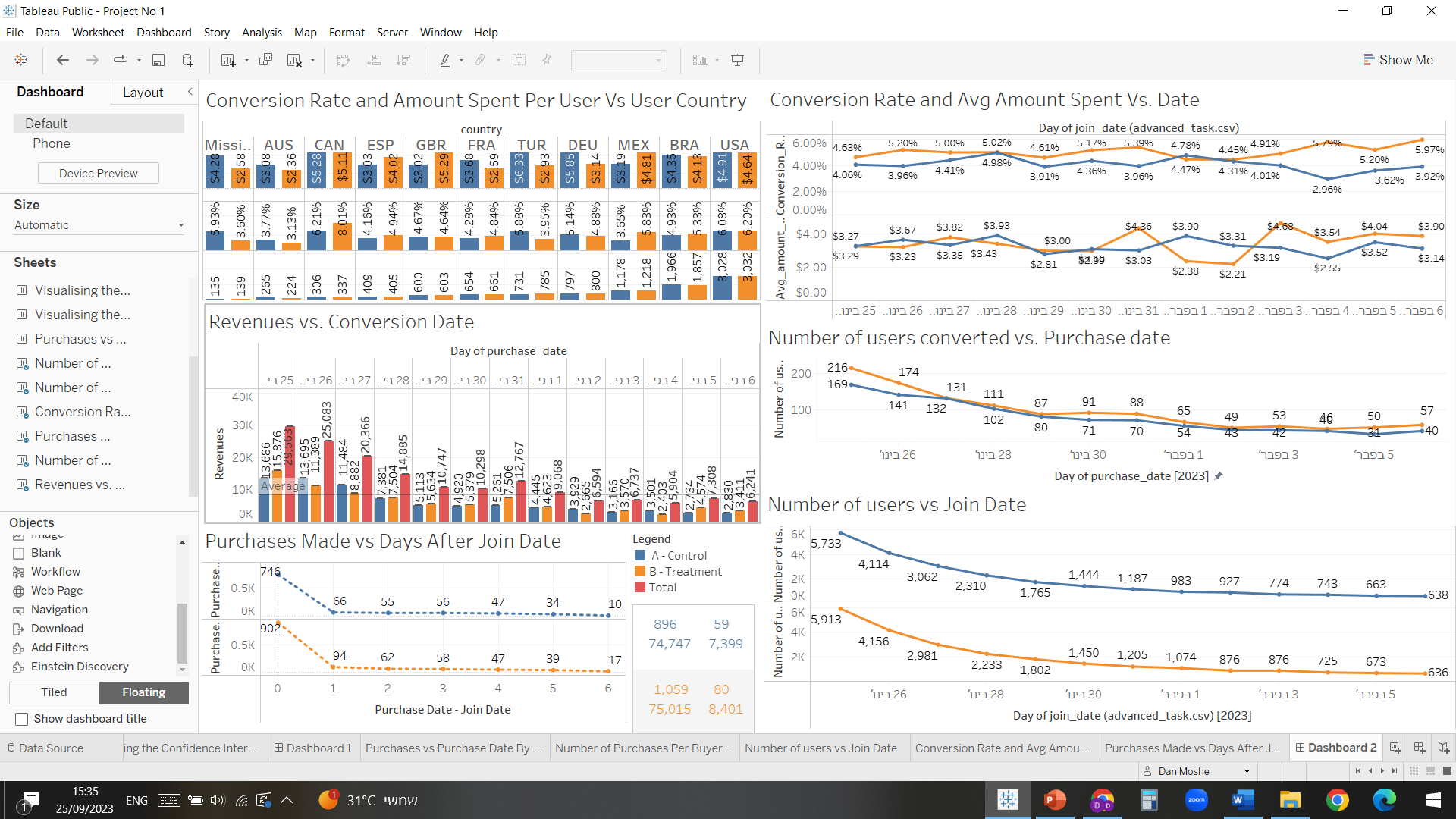
We noticed that male users have higher conversion rate in addition to higher average amount spent in relations to the control group, so does android users.

IOS users spend more money on Globox on average, female users aswell.

|  |  |
| --- | --- |
|  |  |

**Advanced tasks:**

We checked for novelty effect using a new table we extracted from the database using **SQL code for advanced tasks** this time with dates.

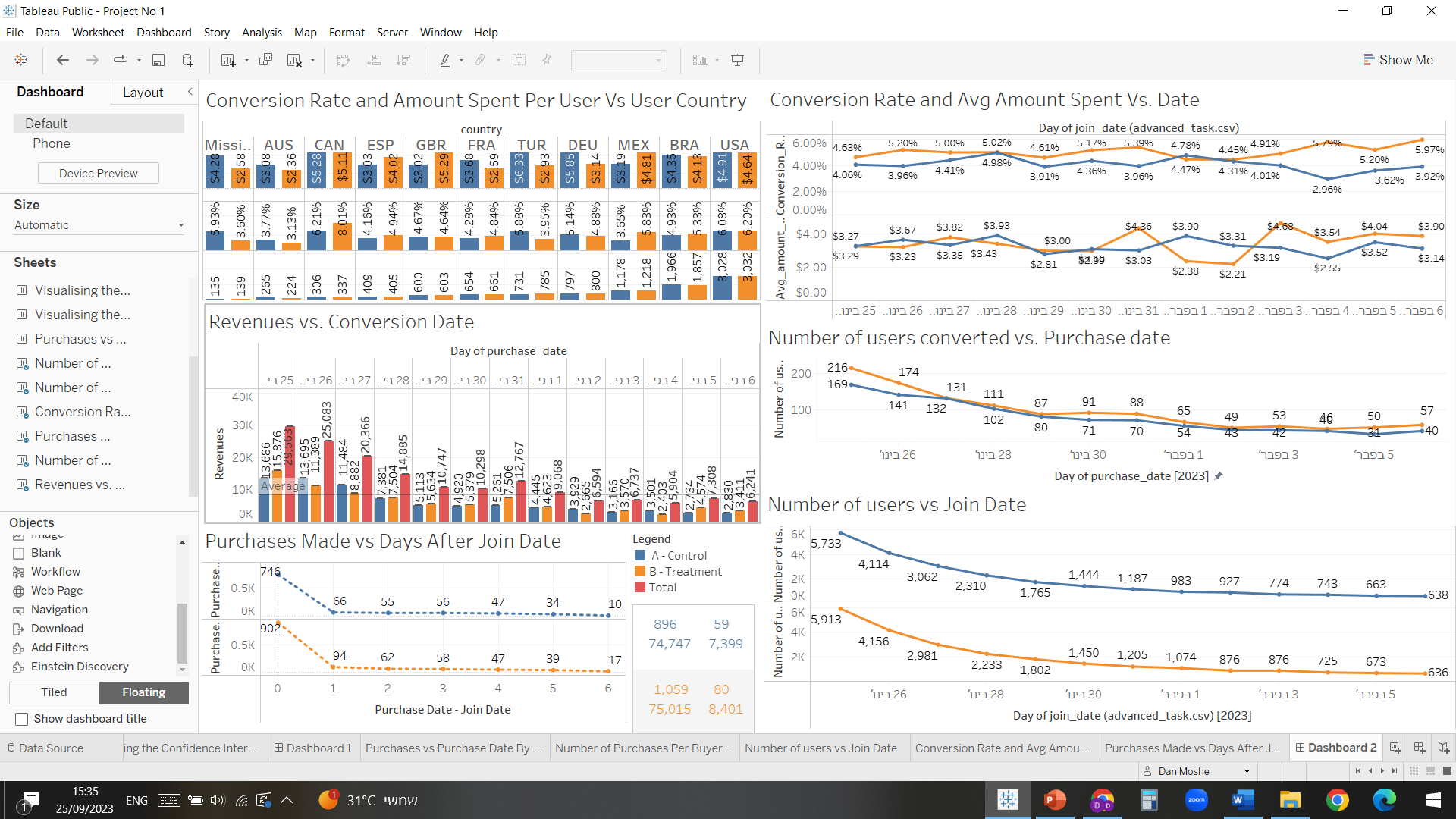


Conversion rate and avg amount spent vs. join date allowed us to see daily changes in the data and to evaluate when a trend is starting to form (February 3rd ).

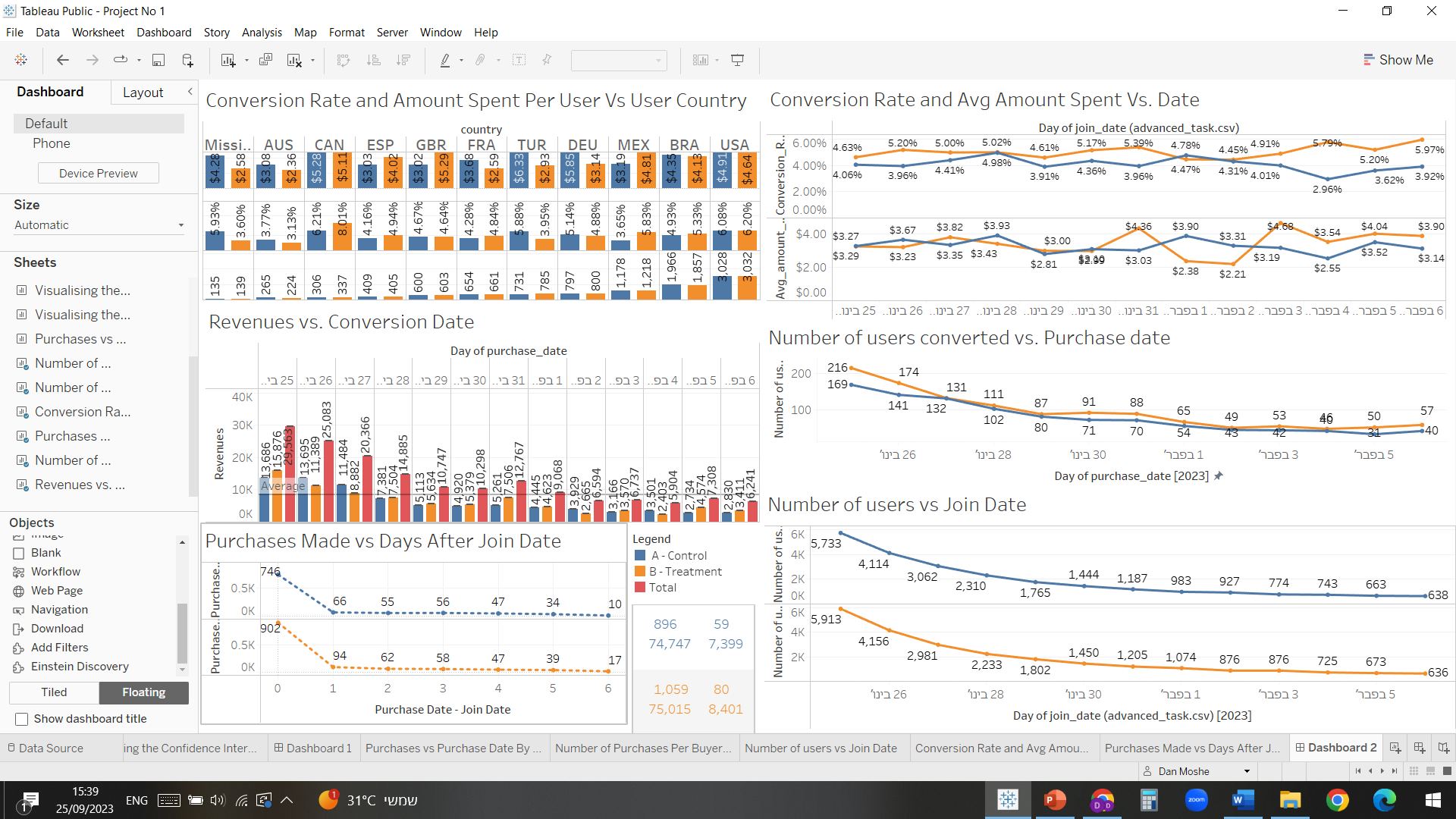
תמונה שמכילה טקסט, צילום מסך, קו, עלילה

התיאור נוצר באופן אוטומטי

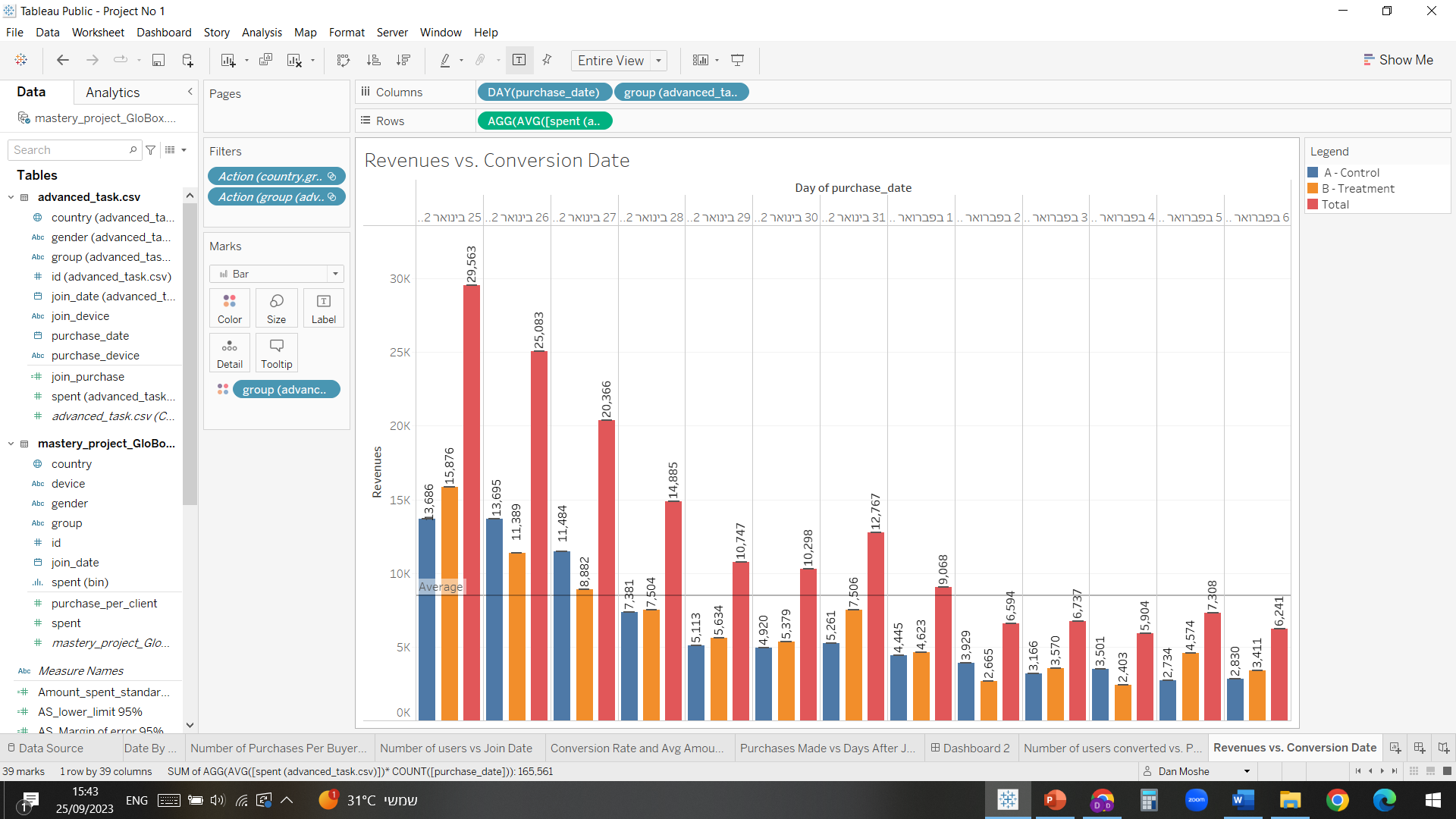
Converted users vs. purchase date allowed us to see when people where actually making purchases and to look and identify further trends.



Number of user vs join date allowed us to evaluate the traffic the site was getting and to be aware of outside influences over our A/B test.



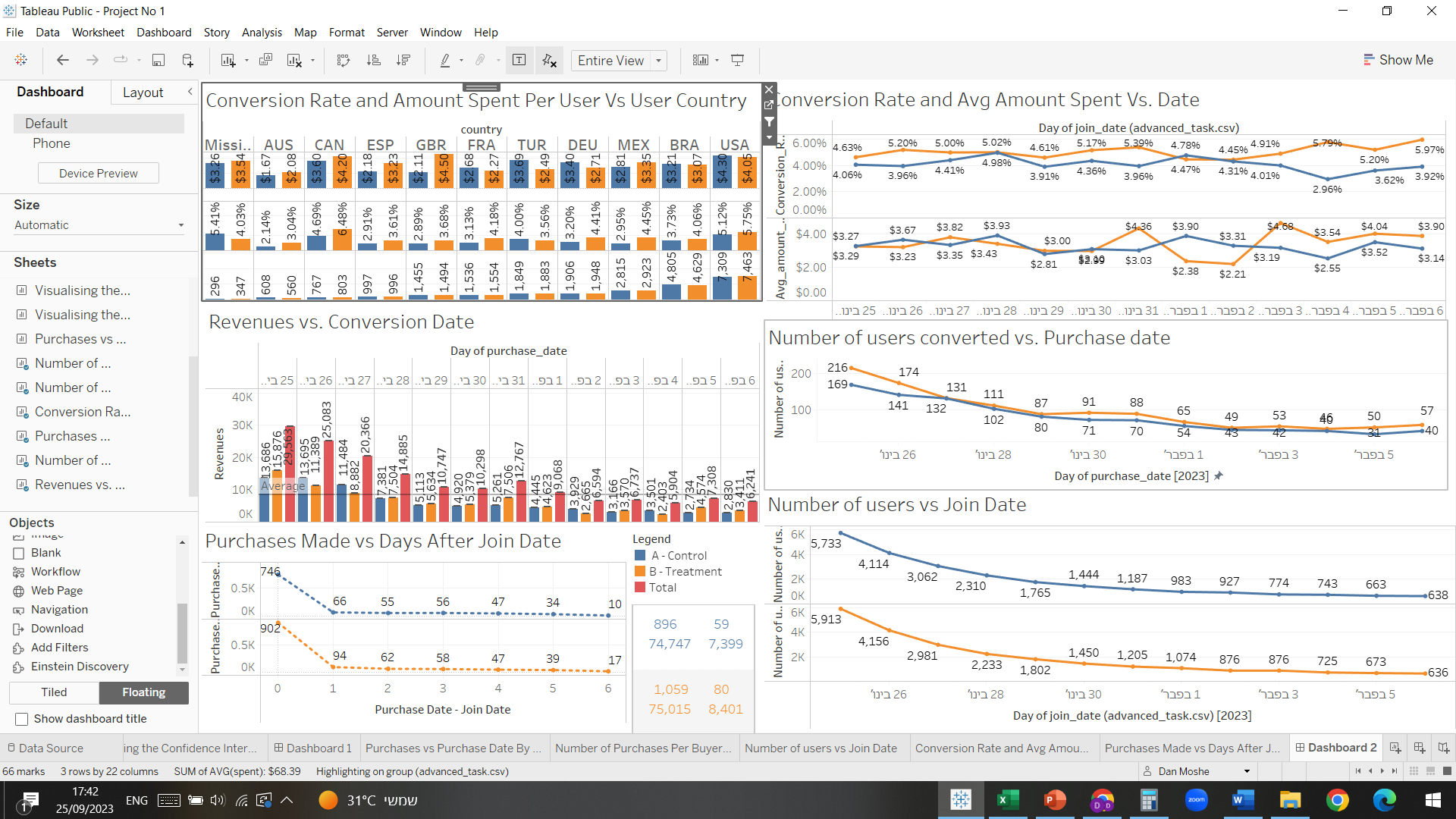
The purchase date – join date vs. purchases date made allowed us to evaluate how many days after users joined the experiment they made a purchase and help us evaluate the novelty effect, as we can see most people made the purchase the day they joined the experiment, and the data doesn't show any real differences between the 2 groups except for the one day difference and the 6 day difference.



תמונה שמכילה טקסט, צילום מסך, תוכנה, סמל מחשב

התיאור נוצר באופן אוטומטי

This chart shows us daily revenues per conversion date and allowing us to see the exact daily differences between the revenues of both test group, we can conclude that the treatment group is making more revenues but only by a marginal amount.



This table summarizes the number of users converted once and twice and total revenues, we had a higher percentage of 2nd time buyers for the treatment group in comparison to the control one.

The Novelty Effect happens when the engagement and interaction with one of our variations is substantially higher than previously, but only temporarily – giving us a false positive.

In our case because the banner is new and the test is only being ran for 12 days, we can't rule out this effect entirely and we need a higher duration A/B test to check and see if the banner actually increases the conversion rate in returning customers and not just new ones, we should recommend at least a 4-week test to check for the novelty effect further.

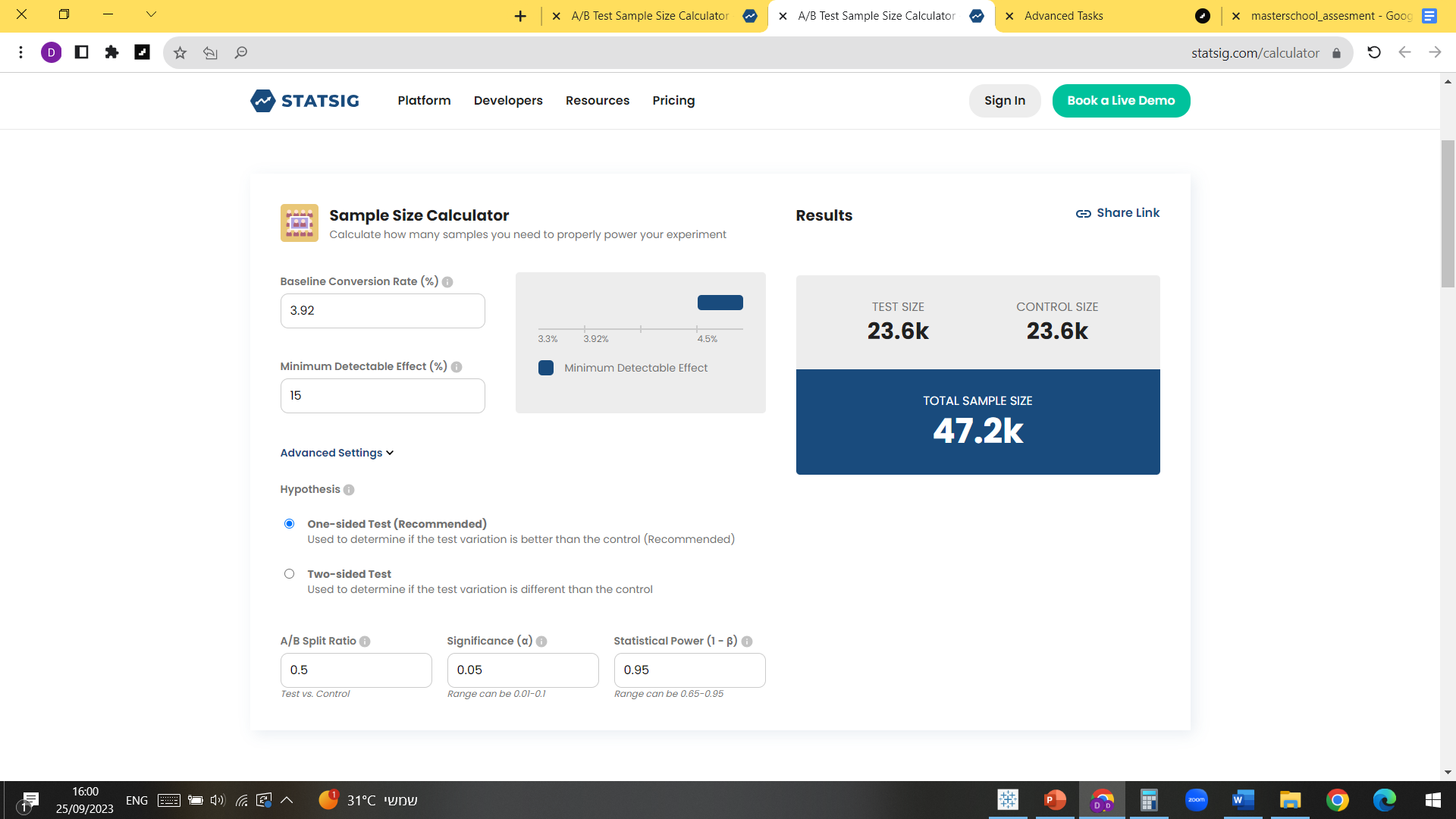
Possible reason the site saw a lot more traffic at the last days of January in relation to the start of February.

According to [Capstone Logistics](https://www.capstonelogistics.com/blog/post-holiday-spending-what-do-consumers-buy-in-january/): **Self-Indulgent Shoppers.**eBay advertising data shows that from December 24 through January 24, searches for “designer” items rose by 190 percent, and searches for “luxury” products rose 148 percent ([eBay](http://advertising.ebay.co.uk/news/2016/self-indulgent-spending-set-soar-january)). This suggests that spending on big-ticket items—like jewelry, designer handbags, and cars—increases in January. - **GloBox is primarily known amongst** its customer base for boutique fashion items and high-end decor products.

**Recommendation**

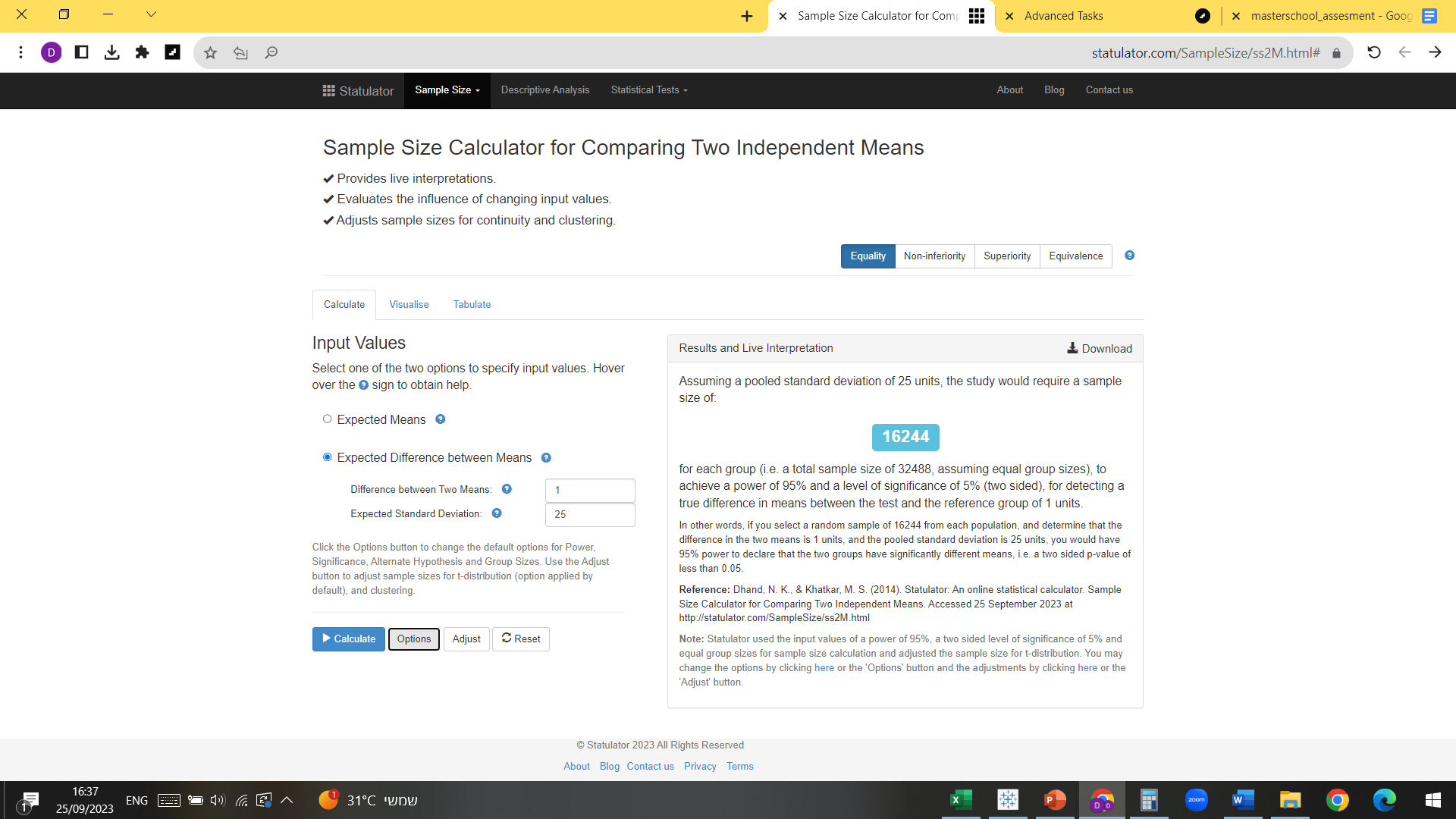
Based on the results above, we need a longer A/B test in order to truly understand if we should launch the banner, if there is only favorable results on the conversion rate side and not on the average amount spent we shouldn't launch it after the following test which should be conducted for at least 4 weeks and with a sample size of 47.2K.

According to Statsig, by choosing the baseline conversion rate – the control group conversion rate as 3.92% (according to data) for a MDE of 15% which gives us a 4.5% conversion rate (data from the treatment group) and filling the other parameters to match our hypothesis test we can conclude that we need a sample size of 47.2K to get significantly accurate results.



We have a sufficient sample size to determine.

According to Statulator, by choosing a difference between the means of at least 1$ we assure that the banner will make enough revenues to justify the front-page space he is getting, the standard deviation was calculated in the excel sheet and was taken from there, and the rest of the variables is according to our hypothesis T test.



We need a sample size of 32,488 to determine, our sample size is enough to determine.