HW3: Experiment Analysis Report

Student Name: Bini Chandra

Date: 02/07/2025

Part 1: Getting to know your data

1. What data is in the file "t1_users_active_mins.csv"?

This file contains data about the users' active minutes after the experiment started. It has three columns: uid, dt, and active_mins. Each row represents the total active minutes(active_mins) spent by a user(uid) on a specific date (dt) after the experiment began. If a user does not visit the site on a given date, they will not have an entry for that date.

2. What data is in file "t2_users_variant.csv"?

This file contains data about the users' treatment assignments. It has four columns: uid, variant_number, dt, and signup_date. Every user belongs to either the control group(variant_number=0) or the treatment group(variant_number=1). Each row represents the experiment assignment of a unique user. The column 'dt' is 2019-02-06 for all users, while 'signup_date' varies per user.

3. What data is in file "t3_users_active_mins_pre.csv"?

This file contains data about the users' active minutes before the experiment started. It has the same structure as "t1_users_active_mins.csv," with three columns: uid, dt, and active_mins. However, the dates in this file (dt) are from before the experiment started. This will help us compare user activity before and after the update.

4. What data is in file "t4_users_attributes.csv"?

This file contains data about some user attributes. It has three columns: uid, gender, and user_type. Each row represents a unique user's attributes, such as gender and activity level ('new_user', 'non_reader', 'reader', or 'contributor').

5. What data is in file "table_schema.txt"?

This file explains the structure of all the dataset files and helps us understand what each column represents.

Part 2: Organizing the Data

The overall objective of this study is to see if the new layout increases user engagement. For this, we need data on how long users stay active and which group they are in (control or treatment). The file *t1_user_active_min.csv* has user activity but doesn't say which group a user belongs to. That information is in *t2_user_variant.csv*, so I have merged both files using *uid*. The final dataset is now structured with userID, date, active minutes, and experiment group and ready for statistical analysis.

```
Merged Dataset:

   uid          dt  active_mins  variant_number
0    0  2019-02-22          5.0               0
1    0  2019-03-11          5.0               0
2    0  2019-03-18          3.0               0
3    0  2019-03-22          4.0               0
4    0  2019-04-03          9.0               0
```

Part 3: Statistical Analysis

Control Group (A) = Users who did not receive the new platform update.
Treatment Group (B) = Users who received the new platform update.
Metric = Total active minutes spent on the platform.

Let's now perform hypothesis testing to check if the update significantly increased engagement.

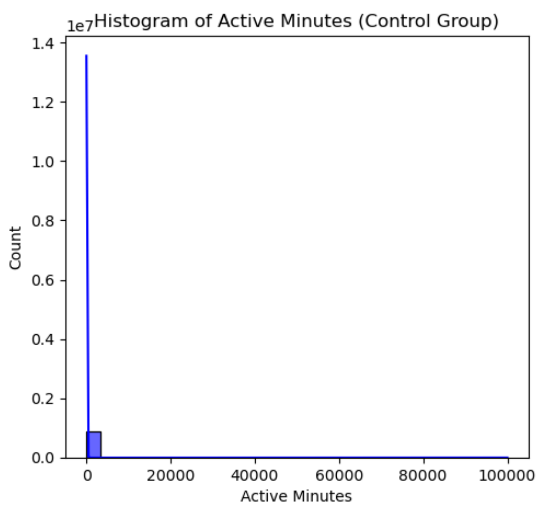
Null Hypothesis: The new platform update does not increase active minutes.

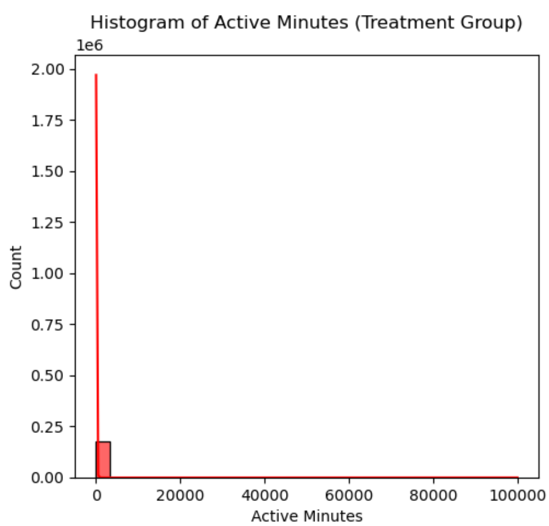
Alternative Hypothesis: The new platform update does increase active minutes.

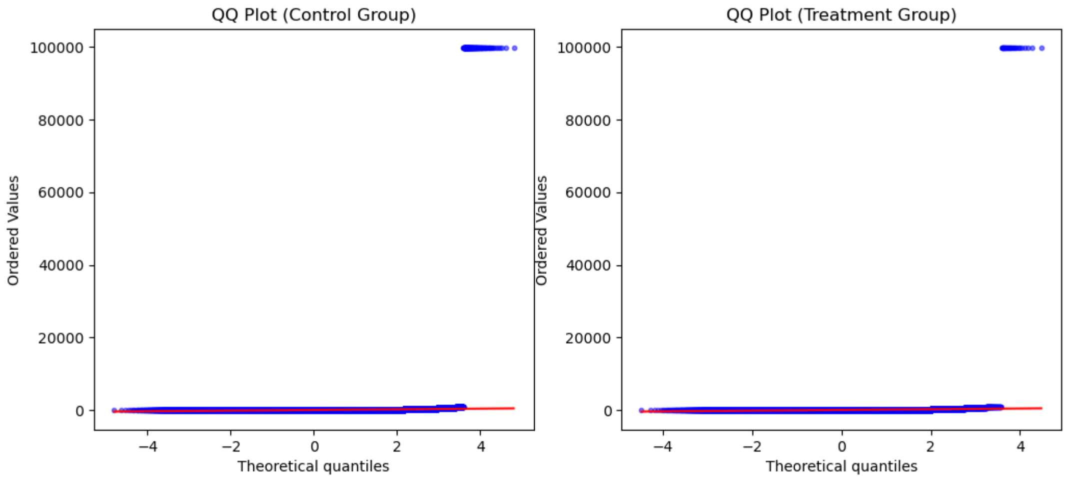
Let's set $\alpha = 0.05$. If p-value < 0.05, we will reject the Null Hypothesis and conclude that the update improved engagement.

Now, let's check whether our data is distributed normally by plotting Histogram and QQ plots.

If it is normal, we will use a T-test; otherwise, we will use a Mann-Whitney U-Test.







The histograms show a right-skewed distribution and also the QQ plots indicate that the data is not normally distributed. Therefore, we will use the Mann-Whitney U-Test instead of T-test.

```
Mann-Whitney U-Test Results:
MannwhitneyuResult(statistic=np.float64(70467369895.5), pvalue=np.float64(0.0))
```

Th results of the Mann-Whitney U-Test shows that p-value is less than 0.05, so we reject the Null Hypothesis. This means the new platform significantly increased user engagement.

1. Is there a statically difference between group 1 and group 2?

Yes, there is a significant difference. The p-value = 0.0 which is less than 0.05, so we reject the Null Hypothesis.

This means the new platform significantly increased user engagement.

2. What is the mean and median for group 1 and group 2?

The treatment group (new platform) shows a higher average and median active time compared to the control group. However, the high variance and standard deviation suggest that the data has large fluctuations. This indicates the presence of outliers that might be skewing the results.

```
                    mean  median          std           var
variant_number
0              35.344199     5.0  1265.733184  1.602080e+06
1              40.240408     7.0  1293.703072  1.673668e+06
```

3. What can you conclude based on that data?

The treatment group has higher active minutes than the control group, suggesting that the new platform is encouraging users to stay longer. The p-value is statistically significant (0.0), so we reject the Null Hypothesis.

However, the high variance and standard deviation as well graphs indicate the presence of outliers, which may be affecting the results.

Part 4: Digging a Little Deeper

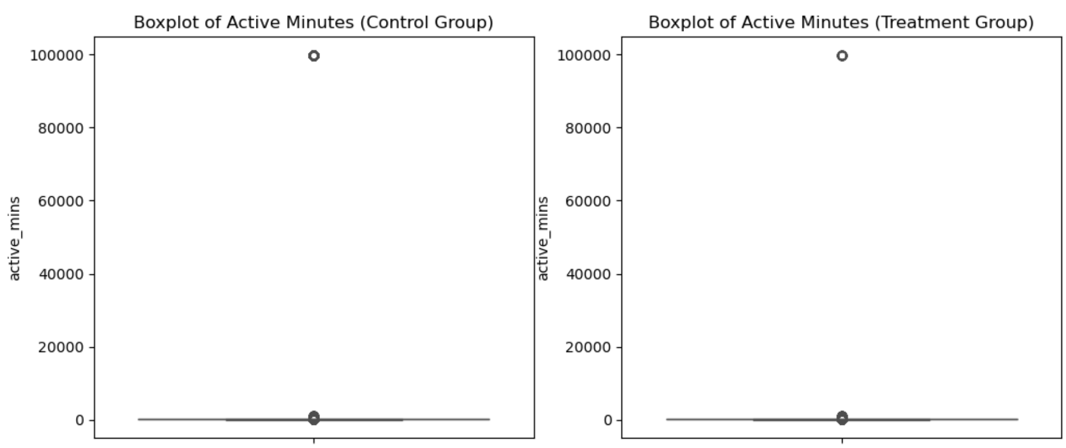
1. Can you trust that the results? Why or why not?

No, we cannot trust our results in Part3 because we noticed that some users have extremely high active minutes, which means outliers. Outliers can distort the mean and give misleading results. Hence, before making a final conclusion, we need to check if removing these outliers changes our conclusion.

2. Is the data normally distributed?

No, the data is not normally distributed. We already checked this in Part 3 by plotting histograms and QQ plots. The histograms show right-skewed distributions with long tails. The QQ plots also confirmed that data is not normally distributed.

3. Plot a box plot of group 1 and group 2.



4. Are there any outliers?

Yes! The boxplots confirm that there are extreme outliers in both the control and treatment groups. The dataset shows that some users logged thousands of minutes of activity per day.

5. What might be causing those outliers?

These outliers might be caused by data entry errors or by some automated systems like bots.
Some users logged thousands of active minutes per day, which is impossible (since there are only 1, 440 minutes in a day).

6. Remove any data point that might be causing outliers.

To remove outliers, I have applied the IQR method using formula: IQR = Q3 - Q1
Lower Bound: Q1 - 1.5 * IQR
Upper Bound: Q3 + 1.5 * IQR

```
Outliers removed: 134449
Max Active Minutes After Outlier Removal: 39.0
```

A total of 1134449 outliers are removed from the data.

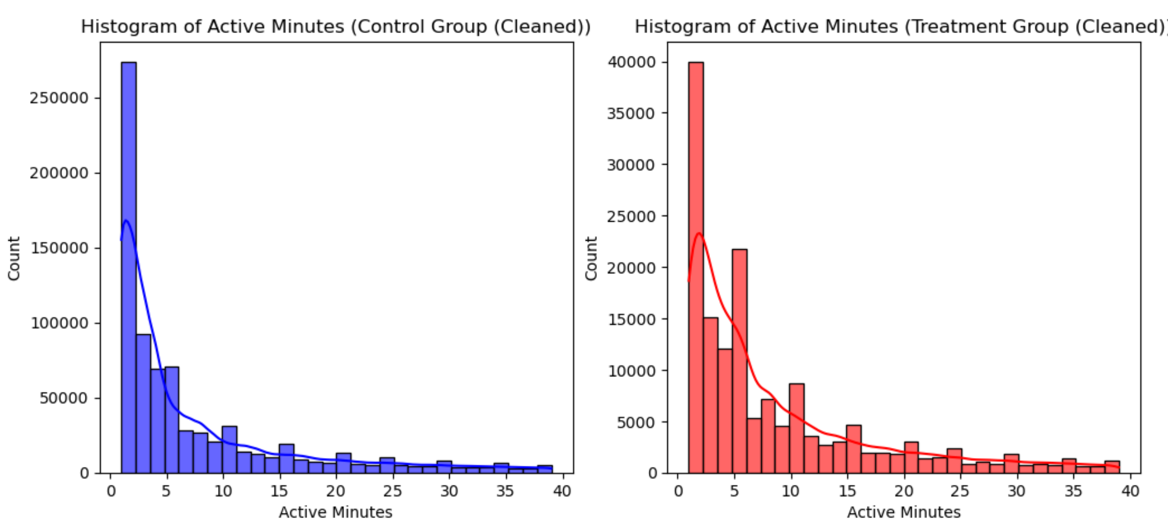
7. Redo part 2 and 3 with the new data without those data points.

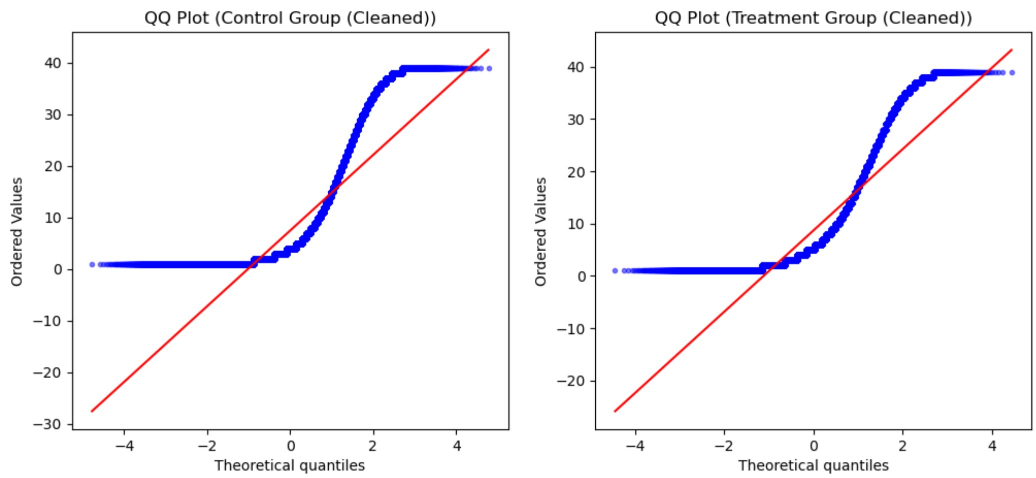
Now after removing the outliers, let's recompute the mean and median.

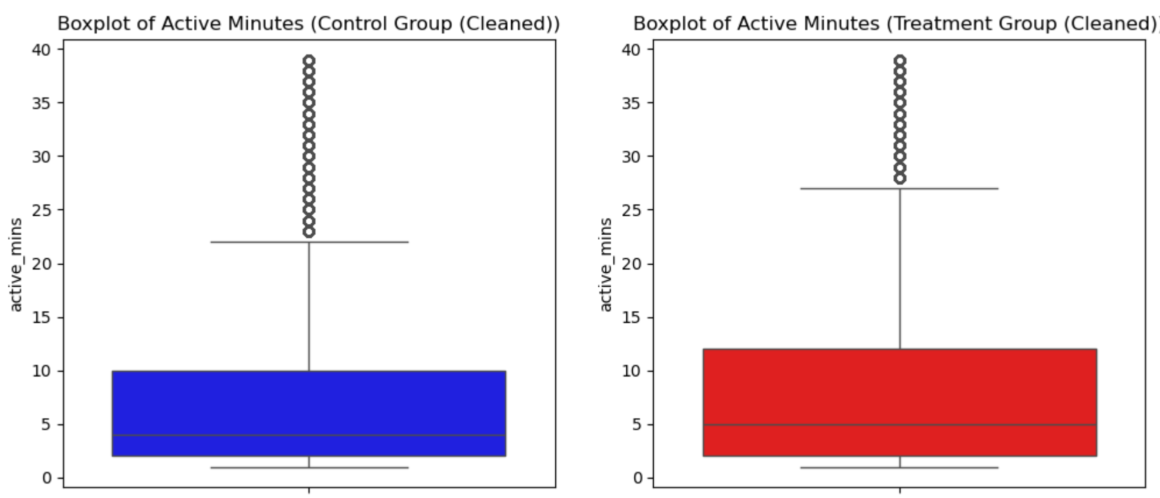
	Control	Treatment
Mean	7.480607	8.676429
Median	4.000000	5.000000
Std Dev	8.491041	8.688977
Variance	72.097780	75.498315

The treatment group still has a higher mean and median than the control group.

Let's again check normality of data by plotting histogram, and boxplots.







The histograms , QQ plots and box plots indicate that the data is still not normally distributed. Therefore, we will use the Mann-Whitney U-Test.

```
Mann-Whitney U-Test Results:
MannwhitneyuResult(statistic=np.float64(52327781622.0), pvalue=np.float64(0.0))
```

The Mann-Whitney U-Test still resulted in p-value = 0.0 between the groups.

8. What is the new conclusion based on the new data?

In Part 3, before removing outliers, the Treatment group had significantly higher engagement (p-value = 0.0).
However, even after removing outliers, the p-value remains 0.0, meaning the difference is still statistically significant.

This means the new platform update has actually improved engagement.

Part 5: Digging Even Deeper

1. Why do we care about the data from t3?

We care about this file because it contains user engagement before the experiment started. We need to check if the treatment and control groups were equal before the update. If one group was already using the platform less, then their increase in engagement might not be fully caused by the update.

2. Accounting for the data from t3 rerun part 2 and 3.

- Merged t3_user_active_min_pre.csv (pre-experiment data) with t2_user_variant.csv (user assignments)
  This allowed us to see which users were in the control and treatment groups before the update.

```
Pre-Experiment Data:

   uid          dt  active_mins  variant_number
0    0  2018-09-24          3.0               0
1    0  2018-11-08          4.0               0
2    0  2018-11-24          3.0               0
3    0  2018-11-28          6.0               0
4    0  2018-12-02          6.0               0
```

- Removed outliers from the pre-experiment data.

```
Outliers removed: 161328
Max Active Minutes After Outlier Removal: 39.0
```

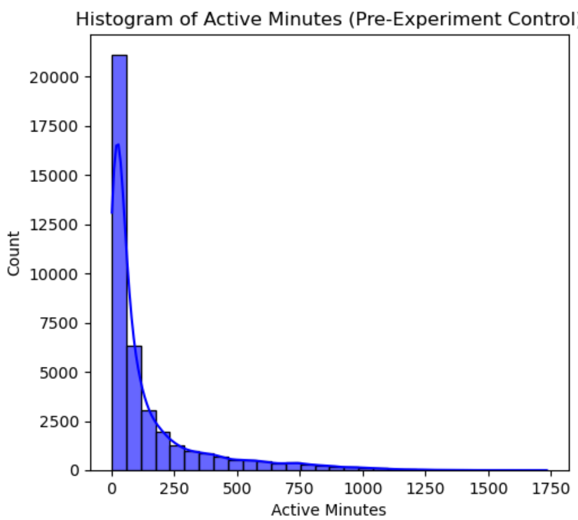
- Added up total pre-experiment active minutes per user. So, instead of looking at daily data, we calculated total minutes spent by each user before the experiment.
- After this, we re-evaluated and :
  - calculated pre-experiment engagement levels across both groups.
  - calculated statistical tests (mean and median).
  - performed a Mann-Whitney U-Test to check if the groups were different before the update.
  - calculated engagement gain (post-experiment minutes - pre-experiment minutes).

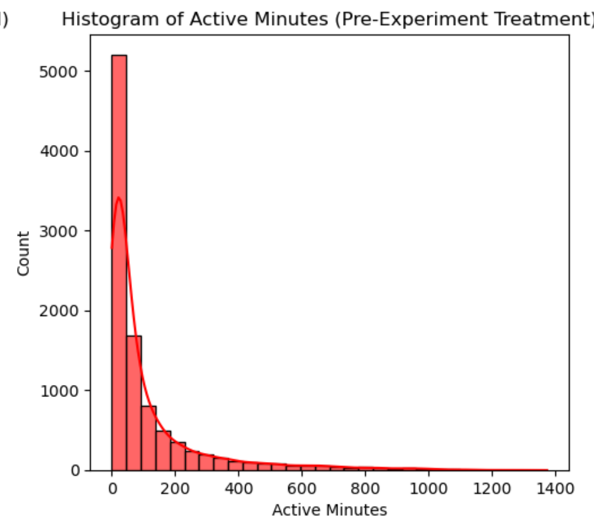
```
Summary Statistics for Pre-Experiment Engagement:
                  count        mean         std  min   25%   50%    75%     max
variant_number
0               39734.0  140.323149  210.727292  1.0  18.0  52.0  158.0  1733.0
1                9910.0  111.041171  174.618763  1.0  15.0  42.0  121.0  1373.0
```

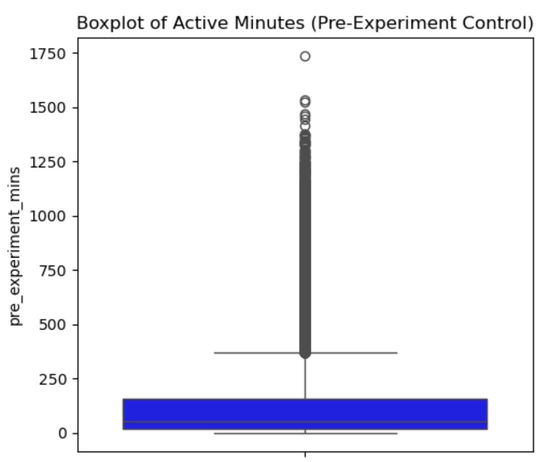
```
Mann-Whitney U-Test Result (Pre-Experiment Engagement): MannwhitneyuResult(statistic=np.float64(212410365.5), pvalue=np.float64(4.665735409092042e-34))
```

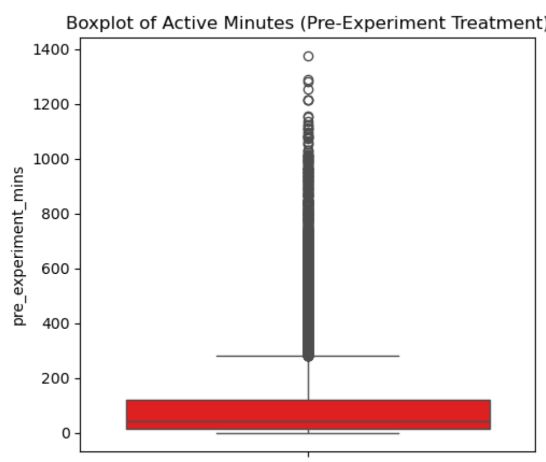
```
Engagement Gain Data:
   uid  pre_experiment_mins  post_experiment_mins  variant_number  engagement_gain
0    0                 70.0                  43.0               0            -27.0
1    1                245.0                  99.0               0           -146.0
2    2                 37.0                  17.0               0            -20.0
3    3                 69.0                  77.0               0              8.0
4    4                 66.0                  39.0               0            -27.0

Summary Statistics for Engagement Gain:
                  count       mean         std    min   25%  50%   75%     max
variant_number
0               37269.0   7.681049  144.818105 -968.0 -33.0 -3.0  27.0  1236.0
1                9148.0  27.097398  133.534396 -884.0 -16.0  9.0  56.0  1326.0

Mann-Whitney U-Test Result (Engagement Gain): MannwhitneyuResult(statistic=np.float64(142198500.5), pvalue=np.float64(8.183504155046488e-134))
```

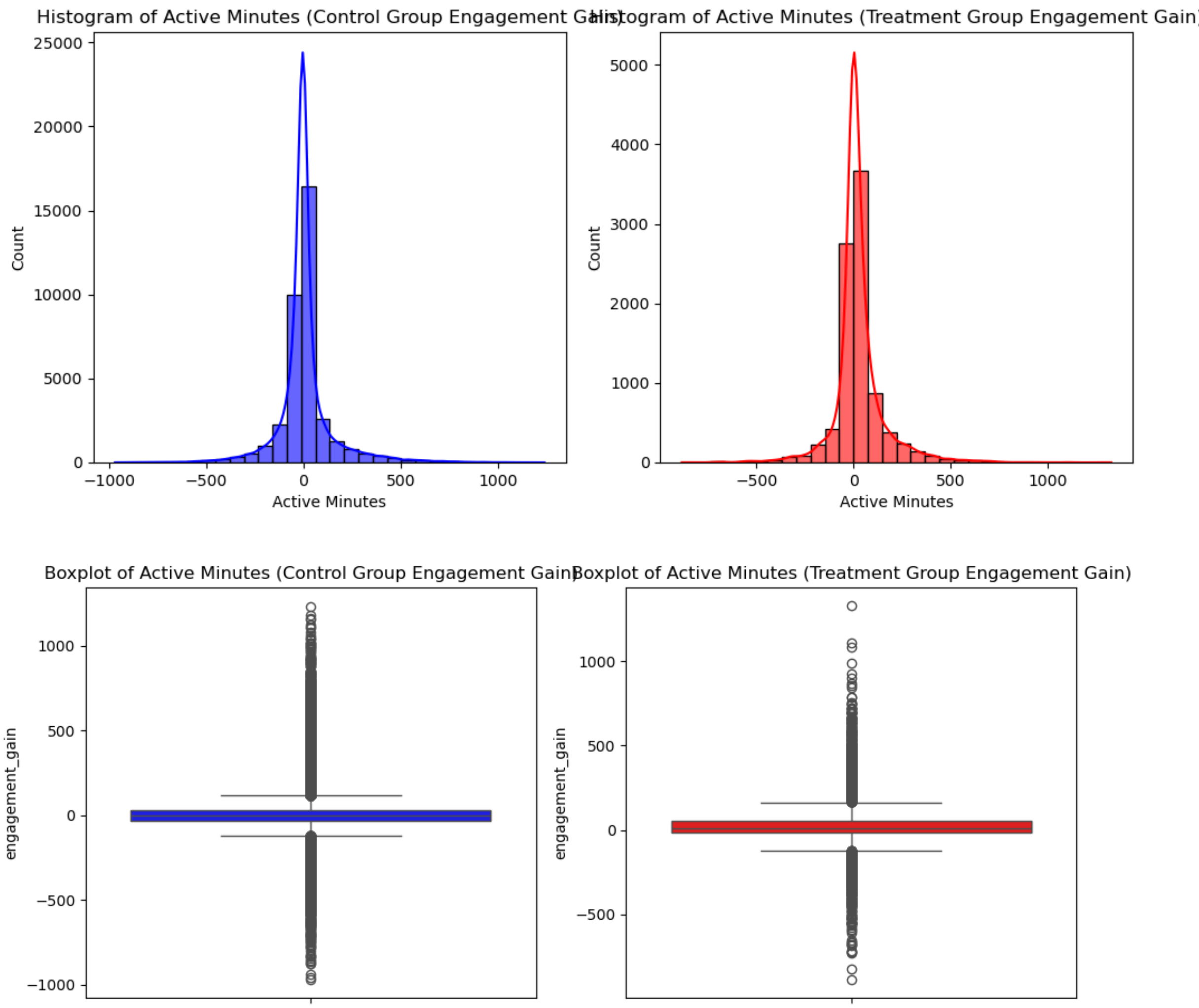
- Plotted histograms and boxplots to see if one group was using the platform more than the other before the update and also created plots for engagement gain











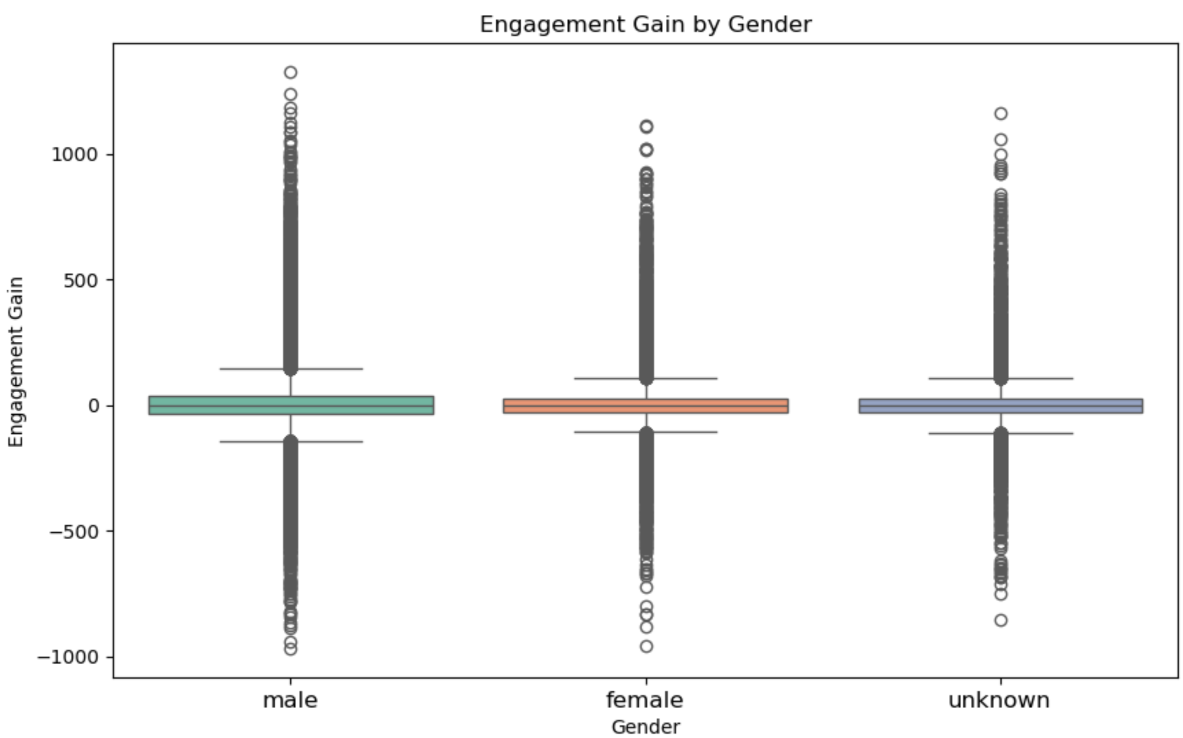
3. Are there any new conclusion?

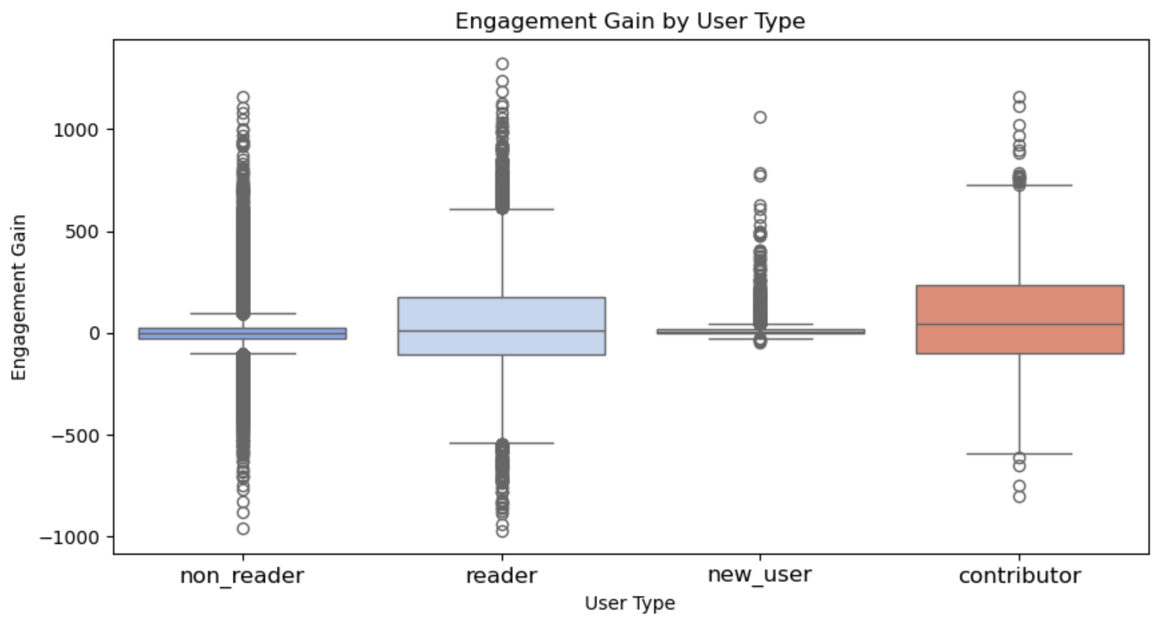
- Before the update, the treatment group was using the platform less than the control group. But, after the update, the treatment group showed an increase in engagement, while the control group had smaller changes. This means the two groups were not equal from the start.
- The p-value(0.0) confirms that the update had statistically significant impact on engagemnet. However, since the treatment group initially had lower engagement, some of the increase comay be due to natural variation rather than the update alone.

Part 6: Exploring other conclusions

In this part, we merged user attributes (t4_user_attributes.csv) with engagement data to check differences by gender and user type and then calculated mean, median, and standard deviation for engagement gain in each group. We have created boxplots for visualization and performed a Mann-Whitney U-Test to check if male and female engagement changes were statistically different.

```
Engagement statistics by gender:
               mean  median         std
gender
female    11.502486    -1.0  132.529854
male      12.442161    -1.0  151.002844
unknown    8.091792    -2.0  130.261084

Engagement statistics by user type:
                   mean  median         std
user_type
contributor   69.236791    47.0  275.152293
new_user      19.505100     5.0   57.301231
non_reader     2.663239    -4.0   97.377543
reader        39.407921    14.0  256.437156

Mann-Whitney U-Test Result (Male vs. Female Engagement Gain): 0.6829576078906725
```





Gender Analysis:

- Mean engagement gain for males (12.44 min) and females (11.50) is similar, but it is slightly lower for unknown gender users (8.09 min).
- The p-value (0.68) and box plot also suggest that there is no significant difference between male and female engagement.

User Type Analysis:

- Contributors gained the most engagement (mean = 69.23 min, median = 47 min), while non-readers had minimal changes (mean = 2.66 min, median = -4 min). Readers and new users also had notable increase.
- Boxplot reveals larger engagement variance in contributors and readers, while non-readers had lower engagement overall.

Conclusion:
The update's effect was not uniform across all users. Instead, it benefited contributor and readers the most, while non-readers showed minimal response. Gender did not play a major role in engagement changes.

Part 7: Summarize Your Results

Part 1: Understanding the Data
We explored the datasets to understand what each file contains. The files track user engagement before and after the experiment, assign users to groups (control or treatment), and provide user attributes. This helped us prepare for analysis.

Part 2: Merging the Data

We merged post-experiment user activity (t1_user_active_min.csv) with user assignments (t2_user_variant.csv) so that we could compare engagement between the control and treatment groups.

Part 3: Initial Analysis

We tested if the platform update increased engagement. The treatment group had significantly higher engagement, with a p-value of 0.0. However, we found outliers with unrealistic active minutes, which could affect the results.

Part 4: Removing Outliers

We removed outliers using IQR method. After cleaning the data, the difference between the groups decreased, but the p-value remained 0.0, confirming the update had a real effect.

Part 5: Checking Pre-Experiment Engagement

We analyzed t3_user_active_min_pre.csv (pre-experiment engagement) to check if the two groups were equal before the update. Before the update, the treatment group had lower engagement than the control group. But post-update, the treatment group showed an increase in engagement. This means part of the increase might be due to natural shift, rather than entirely due to the update.

Part 6: Exploring Other Insights

Here we analyzed how engagement gain differs by gender and user type with the help of user attributes (t4_user_attributes.csv). We found that the update was most beneficial for contributors and readers. Gender didn't play a major role in engagement changes.

Final Conclusion:

The new platform update significantly increased user engagement, as seen from our tests. But we also noticed that the treatment group was already using the platform less before the update. This means some of the increase in engagement may be influenced by their initially lower usage, rather than solely by the update. We also found that different types of users reacted differently. Gender didn't seem to make a difference in engagement changes.