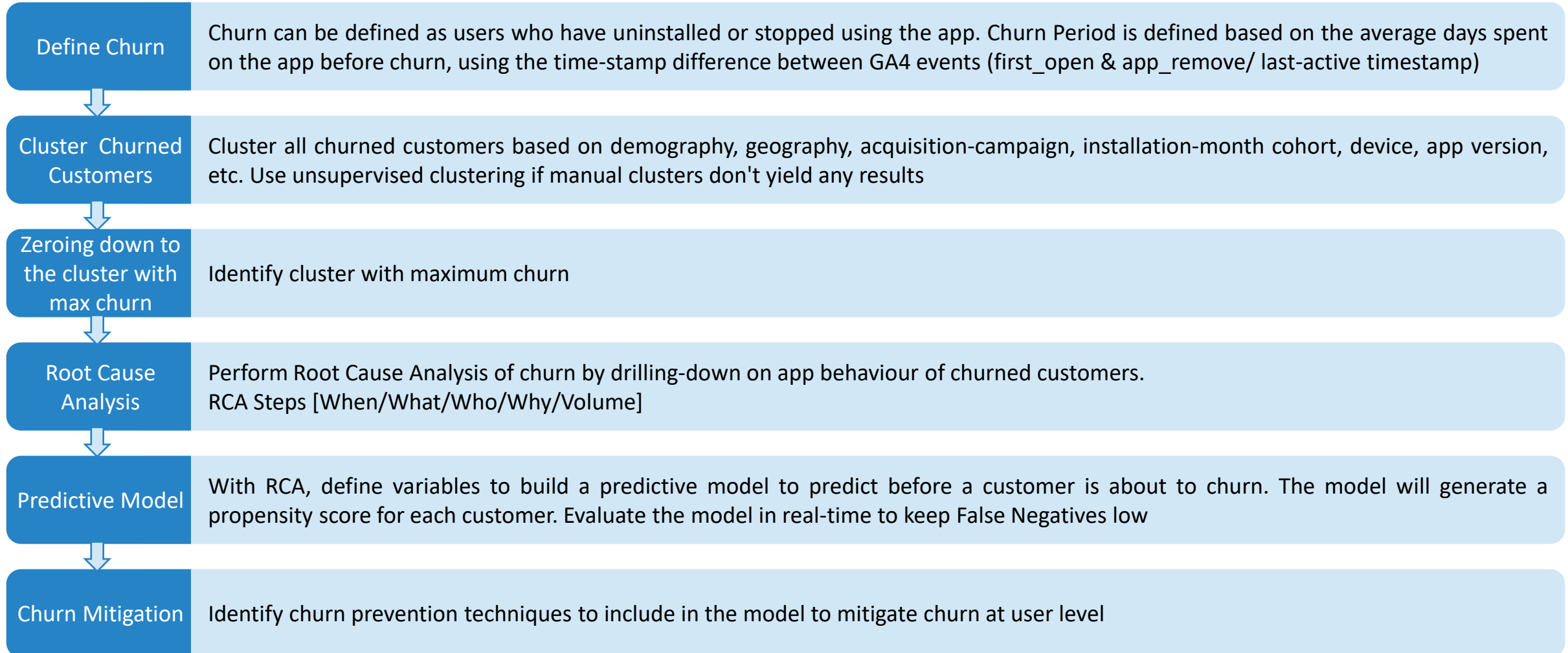# Data Analysis Of
News Application Usage

Gokhul MG

Churn Analysis

Hypothesis: There is likely a minimum number of article impressions/ news cards a new user needs to see to build a habit and thus have a low likelihood of churning.

Dataset : Google Analytics / Firebase Data

Define Churn	Churn can be defined as users who have uninstalled or stopped using the app. Churn Period is defined based on the average days spent on the app before churn, using the time-stamp difference between GA4 events (first_open & app_remove/ last-active timestamp)
Cluster Churned Customers	Cluster all churned customers based on demography, geography, acquisition-campaign, installation-month cohort, device, app version, etc. Use unsupervised clustering if manual clusters don't yield any results
Zeroing down to the cluster with max churn	Identify cluster with maximum churn
Root Cause Analysis	Perform Root Cause Analysis of churn by drilling-down on app behaviour of churned customers. RCA Steps [When/What/Who/Why/Volume]
Predictive Model	With RCA, define variables to build a predictive model to predict before a customer is about to churn. The model will generate a propensity score for each customer. Evaluate the model in real-time to keep False Negatives low
Churn Mitigation	Identify churn prevention techniques to include in the model to mitigate churn at user level

Churn Mitigation using Metrics

CTR

Click Through Rates will let us know if the news or ad users see is in fact relevant to them or not. If CTR is dropping, we need to re-evaluate the curated content.

Total News Impressions

If CTR is low but impressions are high, it possibly implies that users are browsing through the content regularly & are most likely covering the headlines only

Total Session Duration/ Total News Impressions

This would provide the average time spent on each card or news item. A low value suggests that we need to re-evaluate the curated content

Top-News Impressions/ My News Impressions Ratio

If this ratio is high, the user is not navigating to My News beyond Top News. This could mean that users are probably happy with the content curated under Top News.

If this ratio is low, the user skips Top News to navigate to the section of his choice to consume content. We can use section-wise impression metric to curate content for the user.

Evaluating the metrics for churned & non-churned customers will help accept/reject the hypothesis of non-churn customers building a habit.

A/B Testing

Problem Statement

The product team has developed a new UI. The team needs to know if the new UI change will improve the metrics.

UI change:

- Landing screen has changed from Homepage to Top News
- The scrolling in My News has changed from card-swiping to free scrolling

The above UI change should possibly improve Top News impressions & My News impressions, consequentially improving Ad impressions.

Hypothesis

Null Hypothesis (H_0): New UI does not improve metrics.

Alternate Hypothesis (H_A): New UI impacts metrics (positively/negatively).

Note: Since we are testing UI change and not content change, CTR might not be a useful metric. Essentially the UI change should see people spending more time on the app.

Dataset Description

The 25th percentile of all columns are 0.

Upon investigating, found that 26.6% of the population have zero total impressions. This suggests that the sample population that were a part of the experiment were not active users of the app itself.

Since the hypothesis is to identify change in UI experience, including the sample population which have not opened the app would lead to a bias when we are analysing the UI experience.

We will proceed with A/B testing after removing outliers & samples with zero impressions (a combination of zero news & zero ad impressions).
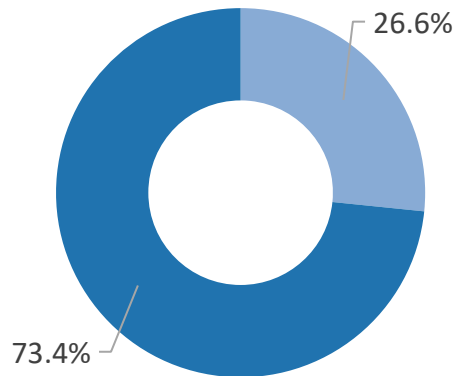
Attribute	Count	Mean	Std	Min	25%	50%	75%	Max
TOPNEWS_IMPRESSION_COUNT	320118	24.78	64.57	0	0	5	21	6307
TOPNEWS_CLICKED_COUNT	320118	2.2	5.99	0	0	0	2	293
ADS_IMPRESSION_TOPNEWS_COUNT	320118	4.09	12.47	0	0	0	3	371
ADS_CLICKED_TOPNEWS_COUNT	320118	0.04	0.36	0	0	0	0	22
MYNEWS_COUNT	320118	25.69	215.44	0	0	0	1	22126
ARTICLE_MYNEWS_CLICKED_COUNT	320118	1.31	7.93	0	0	0	0	535
ADS_IMPRESSION_MYNEWS_COUNT	320118	3.06	25.9	0	0	0	0	2027
ADS_CLICKED_MYNEWS_COUNT	320118	0.02	0.33	0	0	0	0	106
Total_Ads_Imp	320118	7.15	33.83	0	0	0	3	2028
Total_News_Imp	320118	50.47	249.19	0	0	5	26	23003
Total_Ads_Clicked	320118	0.06	0.54	0	0	0	0	108
Total_News_Clicked	320118	3.51	11.68	0	0	0	3	729

Before & After Data Cleaning
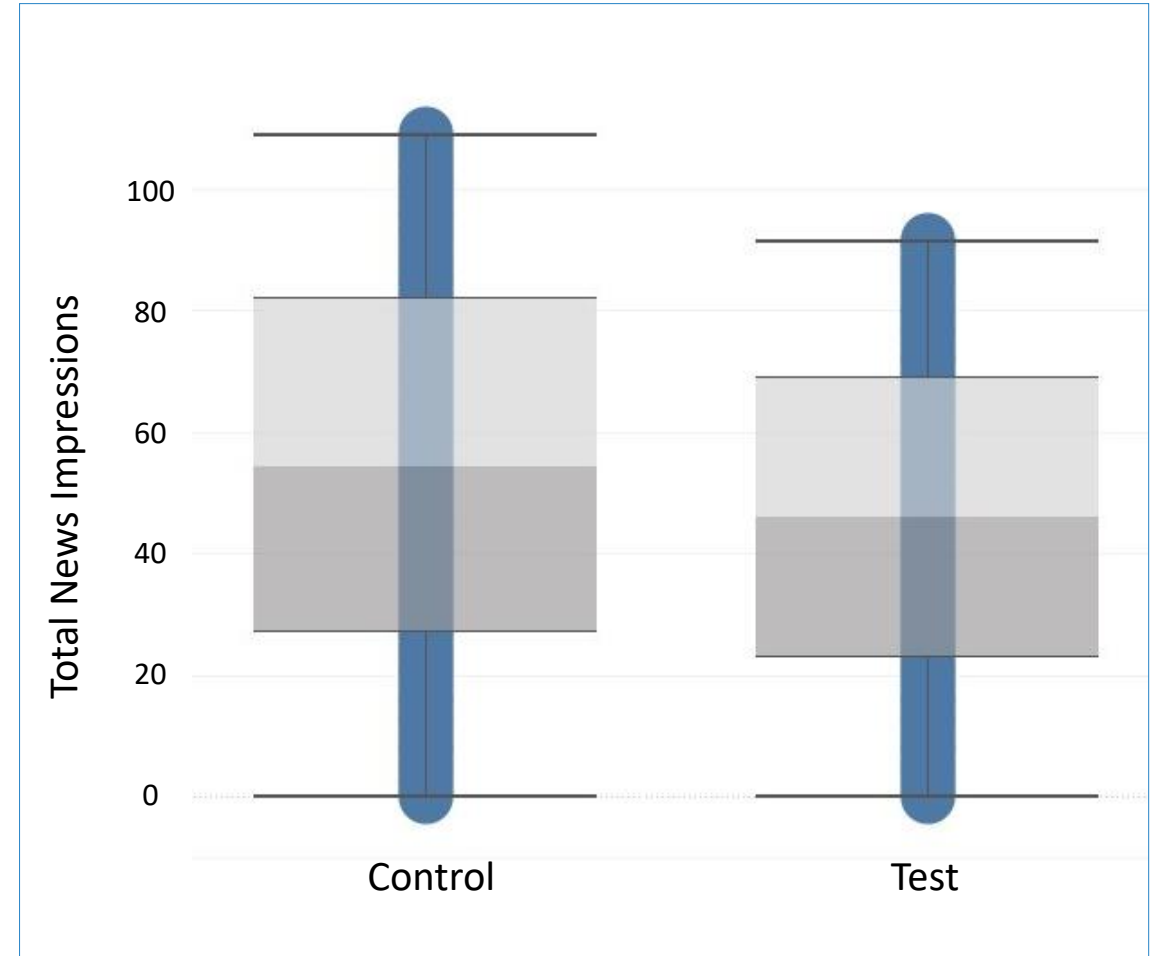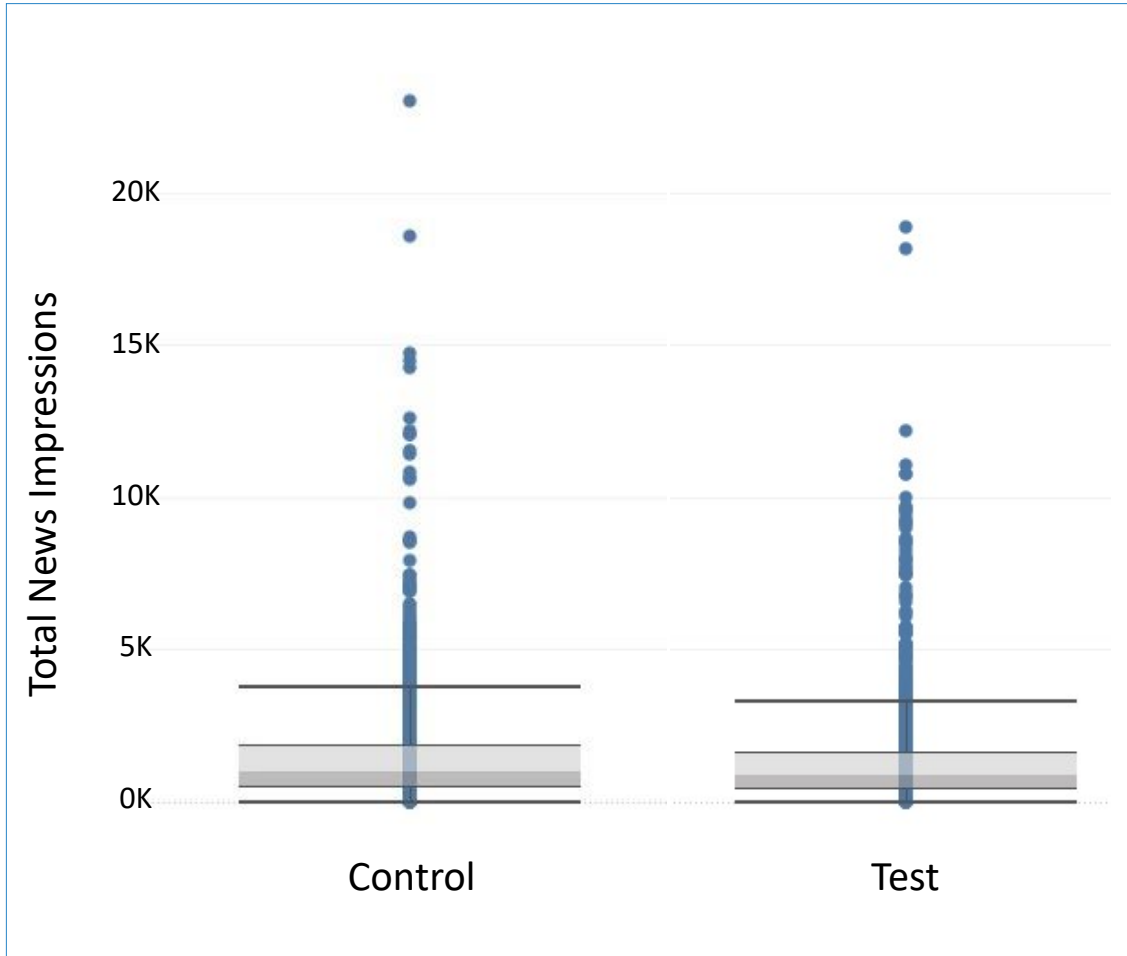
Control vs Test Data Description

Control Group

Attribute	Count	Mean	Std	Min	25%	50%	75%	Max
TOPNEWS_IMPRESSION_COUNT	114194	22.05	25.06	0	3	11	32	76
Total_Ads_Imp	114194	4.64	6.27	0	0	1	7	18
Total_News_Imp	114194	31.22	37.69	0	4	12	46	109
Total_Ads_Clicked	114194	0	0	0	0	0	0	0
Total_News_Clicked	114194	2.79	3.47	0	0	1	4	10

Test Group

| Attribute	Count	Mean	Std	Min	25%	50%	75%	Max
TOPNEWS_IMPRESSION_COUNT	120800	20.87	23.14	0	3	11	30	71
Total_Ads_Imp	120800	3.12	4.32	0	0	1	5	13
Total_News_Imp	120800	26.99	30.8	0	4	13	39	92
Total_Ads_Clicked	120800	0	0	0	0	0	0	0
Total_News_Clicked	120800	2.57	3.3	0	0	1	4	10

The control group (Old UI) has higher mean across all attributes. This possibly implies that the data strongly favours old UI.
We will however proceed with A/B testing to obtain the confidence interval of the data distribution.

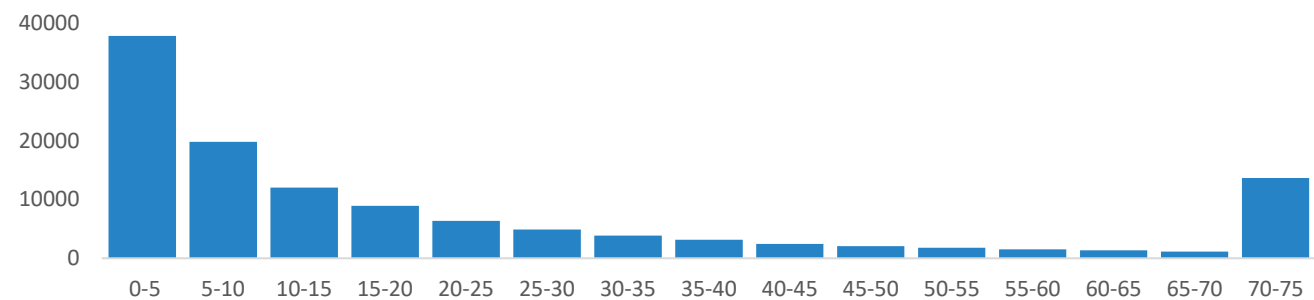# Data Distribution of Control & Test Group

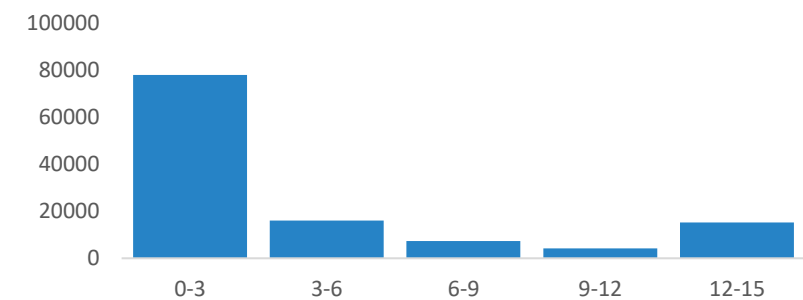Control Group - Top News Impression
Mean: 22.05
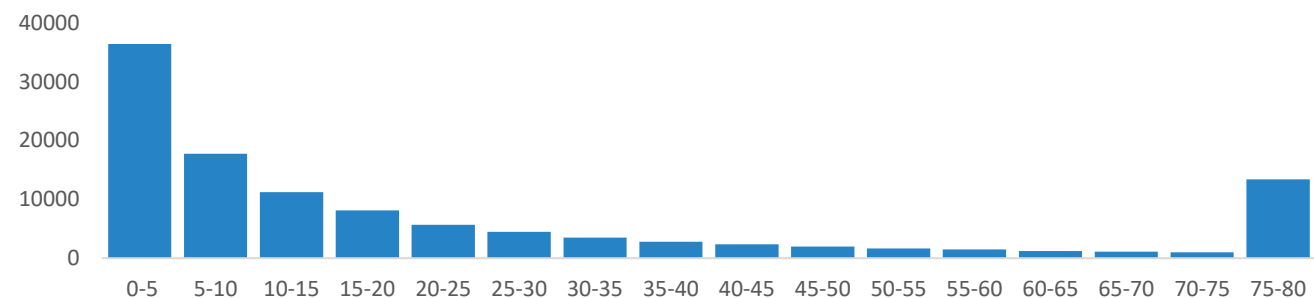
Control Group - Total Ad Impression
Mean: 4.64
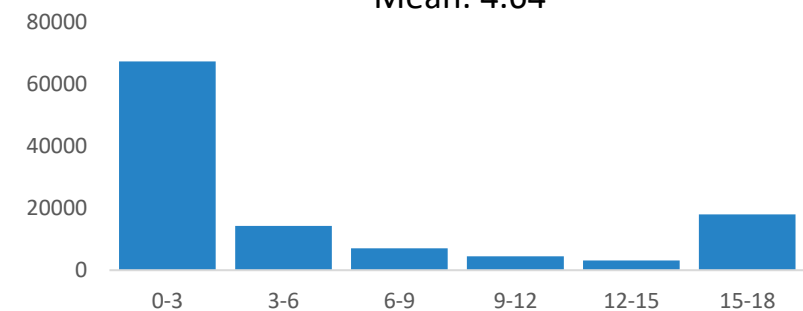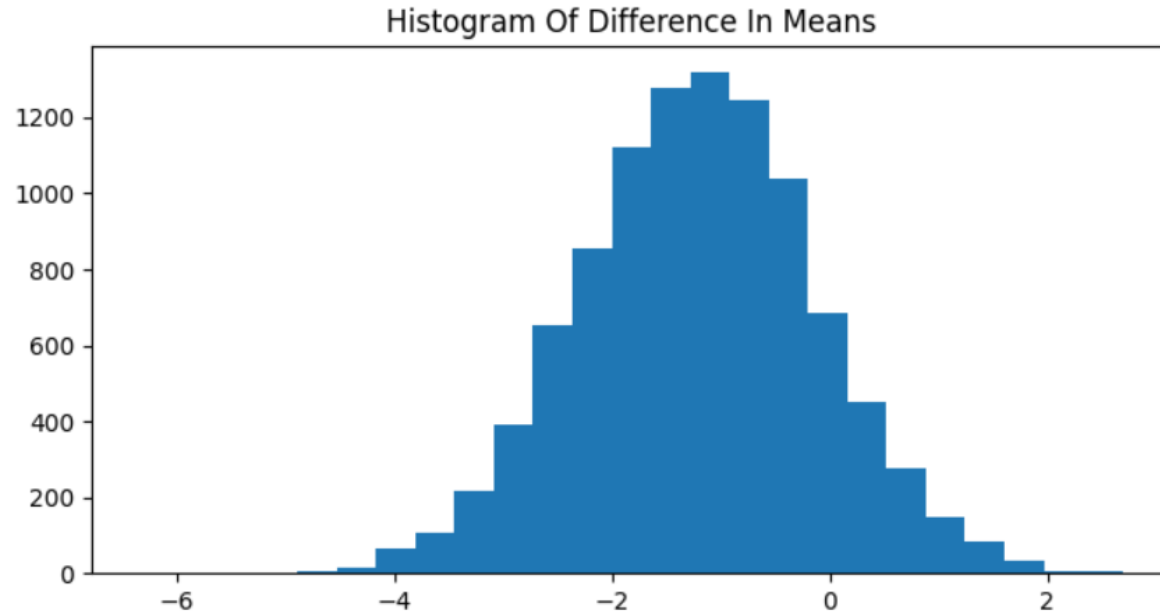
Test Group - Top News Impression
Mean: 20.87

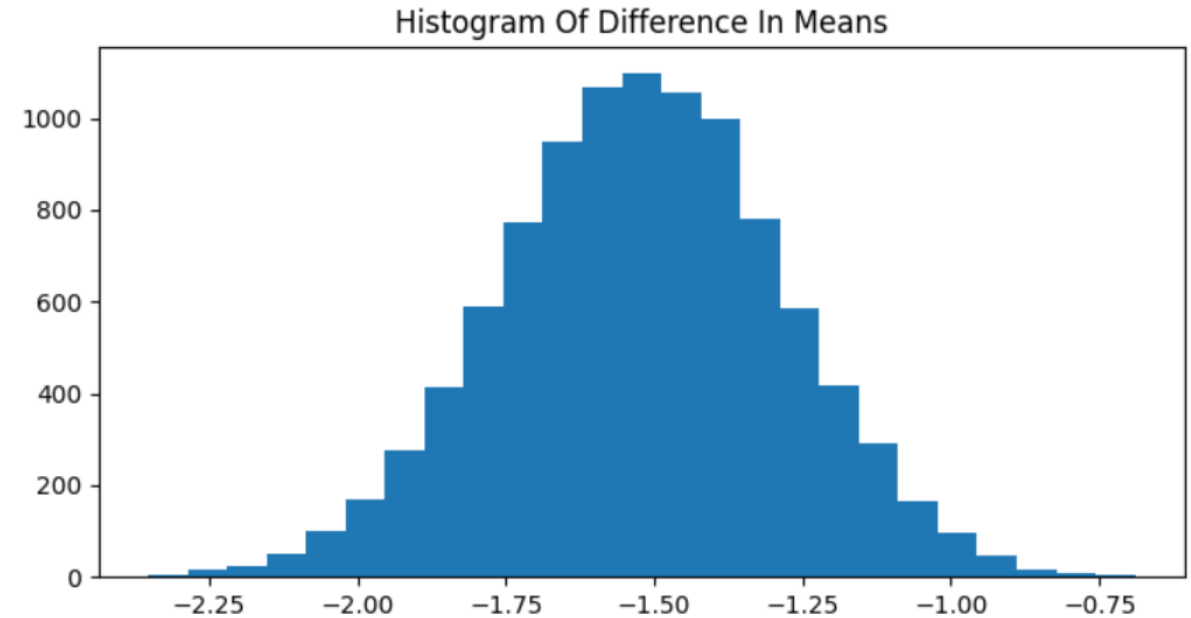Test Group - Total Ad Impression
Mean: 3.12

A/B Testing Results On Impression Metrics

Top News Impression

Total Ads Impression

Top News Impression Results

% probability of Test Mean Being Better Than Control Mean : 12.91 %

Total Ads Impression Results

% probability of Test Mean Being Better Than Control Mean : 0 %

A/B Testing Success Metric

Though the randomization of both control & test group would eliminate bias, the impression metrics considers absolute value of impressions & does not provide a clear picture at a user level.

Defining A Success Metric To Study User Level Success

Since the revenue model is based on Ad impressions & Ad clicks, we will consider Ad Impressions as the success metric.

If a user has had at least 1 Ad impression, it automatically translates to more than 1 news impression.

Success metric = If Total_Ad_Impression>1 Then 1 Else 0

Success of the control & test group would be based on the maximum number of users who have had at least one Ad Impression.

A/B Testing Results

Group	Success Rate	Std Deviation	Std Error
Control	0.615	0.487	0.015
Test	0.546	0.498	0.016

z statistic: 3.13

p-value: 0.002

Confidence-Interval at 95% : [0.585, 0.645]

Conclusion: At p-value of 0.05, the mean of test-group lies outside the confidence interval. The null hypothesis is rejected as there exists a difference between control & test group and at the same time, there is no evidence that the Test (New UI) is better than the control-group. The AB-test shows that the test UI has performed poorly when compared to the existing UI. Multiple iterations of A/B testing on different samples has consistently proved that there is significant difference in the old & new UI, and the new UI change has not positively impacted the metrics. Hence, rejecting the Null Hypothesis and accepting the Alternate Hypothesis.