Applying Iterative Design Principles to a Live Product

Name: Sérgio C.
Date: 23 July 2021

Step 1

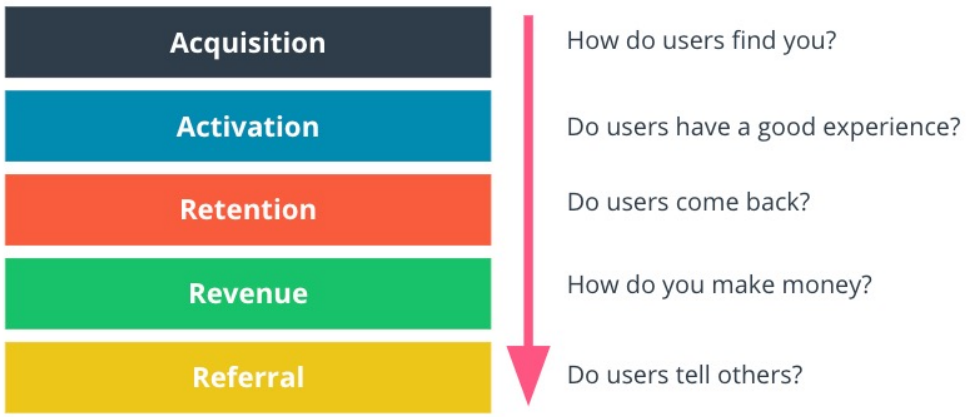
Select KPIs

&

Evaluate Previous Multivariate Experiment Results

Select KPIs for Flyber Analyses

- The **Flyber** customer journey can also be “broken down” into categories of ***Acquisition, Activation, Retention, Revenue, and Referral***. Consequently KPIs can fall under these categories.



- Given the dataset I have available, there are KPIs that fall into the above categories that truly match Flyber's business goals:
    - **Acquisition**
        - KPI 1 - number of new users per day
            - Calculation: Nr Users / Event Time Day
    - **Activation**
        - KPI 2 - number of rides / number of users making rides per day
            - Calculation: Nr Users with Event Type
        - KPI 3: number of users per day, week, month (DAU, WAU e MAU)
            - Calculation: Nr Users

Select KPIs for Flyber Analyses

- **Retention**
  - KPI 4 - ratio of users that search and make a ride
    - Calculation: overall rides count per ride event / overall rides count per search event in the same period of time
  - KPI 5 - ride frequency by neighbourhood
    - Calculation: overall rides count per neighbourhood (considering the ride event) in a period of time
- **Revenue:**
  - N/A
- **Referral**
  - N/A

Select KPIs for Flyber Analyses

There are other important KPIs for Flyber that fall into the above categories that cannot be calculated as there is no data available in the data set:

- **Acquisition**
  - number of new users per day, per week, per month, ...
- **Activation**
  - The time users spend searching on the first visit, ...
- **Retention**
  - The average time length a user remains a customer, The number of users who renew their subscriptions, The number of repeat users, The number of users support tickets/complaints, ratio of Day/week/month/quarterly ride cancellation, ...
- **Revenue:**
  - Customer Acquisition Cost (CAC), Customer Lifetime Value (LTV), Average Revenue per User (ARPU), Revenue per Month, ...
- **Referral**
  - NPS, Customer Satisfaction, Customer Effort Score, Trust Index, ...

Describe the First Multivariate Experiment

In the multivariate experiments different elements will be tested. Let's consider the images below:

Control

Experiment 1

Experiment 2

Experiment 3

Describe the First Multivariate Experiment

In the **multivariate experiments** different elements will be tested.

Based on the image presented above, it is possible to identify the **two variables that are considered** in the multivariate experiment:

1. the text field "***Tip Included***"
2. the text on the button (Value 1: "***Fly Now***" | Value 2: "***Book Flight***").

Experiments group:

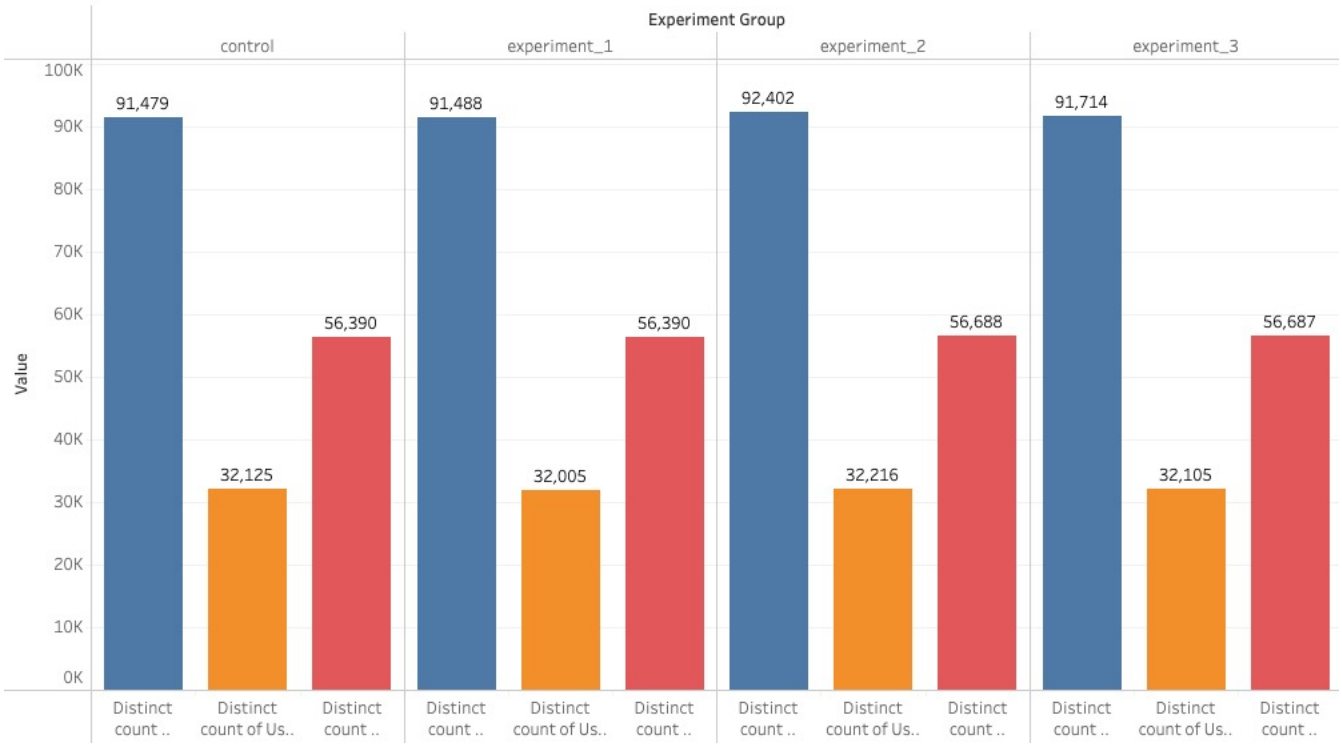
- **Control**: Button with text “Book Flight” and text field “Tip included”
- **Experiment 1**: Button with text “Fly Now” and text field “Tip included”
- **Experiment 2**: Button with text “Book Flight” without text field
- **Experiment 3**: Button with text “Fly Now” without text field

The visualisations presented in this document can be obtained from the public viz of the tableau here.

Review Multivariate Test Results: Visualization

Using the data provided, it will be possible to **analyse the impact of the experiments in terms of the conversion ratio of users to booking a flying ride**.

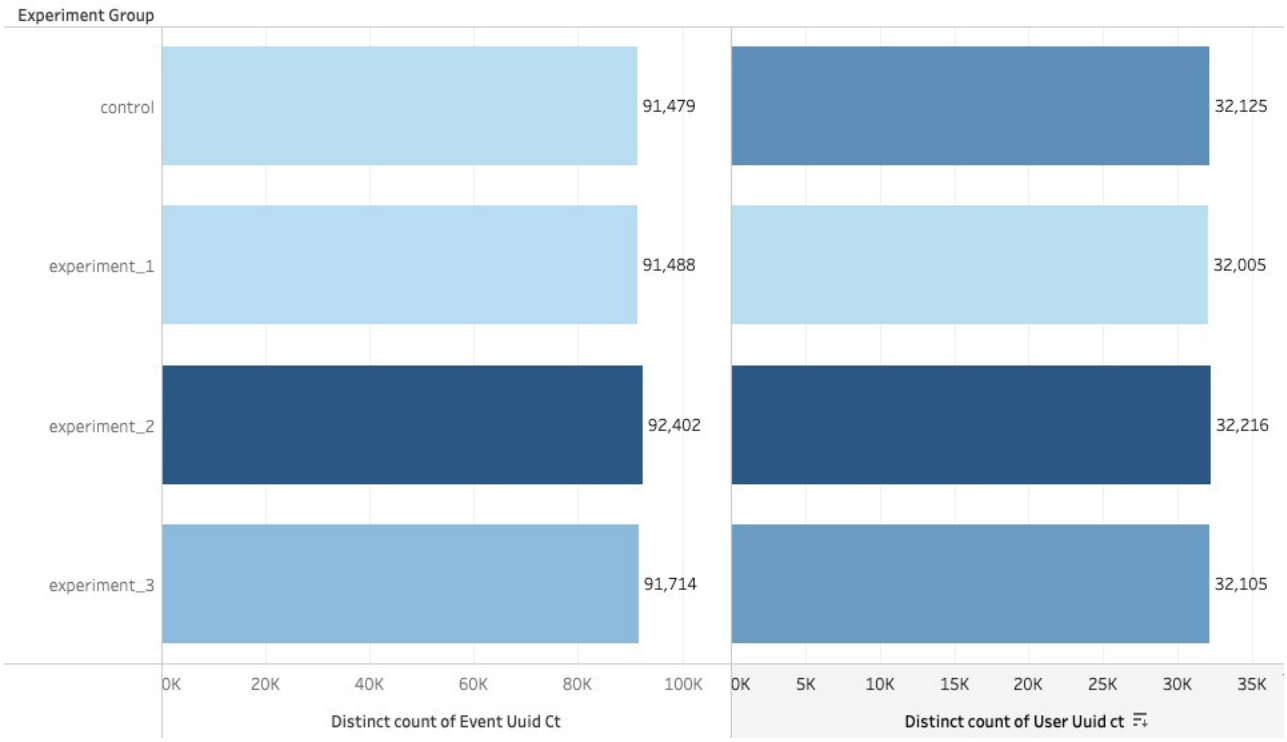
Analysis 1: Count the number of distinct events and number of users.



Review Multivariate Test Results: Visualization

Using the data provided, it will be possible to **analyse the impact of the experiments in terms of the conversion ratio of users to booking a flying ride**.

Analysis 2: Multivariate Test - Experimentations



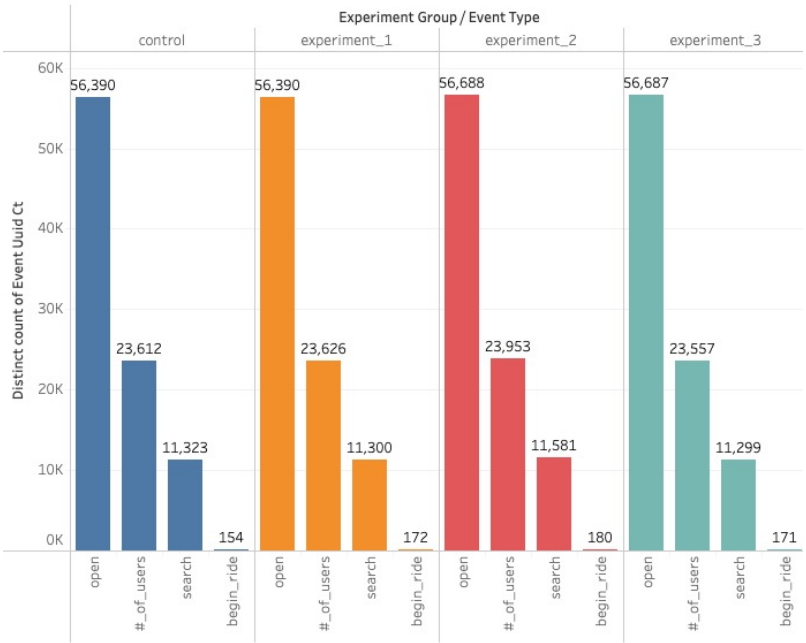
Review Multivariate Test Results: Visualization

Using the data provided, it will be possible to **analyse the impact of the experiments in terms of the conversion ratio of users to booking a flying ride**.

Analysis 3: Multivariate Test – Conversion Rate & User Drop-Off

	Experiment Group			
Event Type	control	experiment_1	experiment_2	experiment_3
open	56,390	56,390	56,688	56,687
#_of_users	23,612	23,626	23,953	23,557
search	11,323	11,300	11,581	11,299
begin_ride	154	172	180	171

	Experiment Group			
Event Type	control	experiment_1	experiment_2	experiment_3
open	0.00%	0.00%	0.00%	0.00%
#_of_users	-58.13%	-58.10%	-57.75%	-58.44%
search	-79.92%	-79.96%	-79.57%	-80.07%
begin_ride	-99.73%	-99.69%	-99.68%	-99.70%



Review Multivariate Test Results: Significance Test

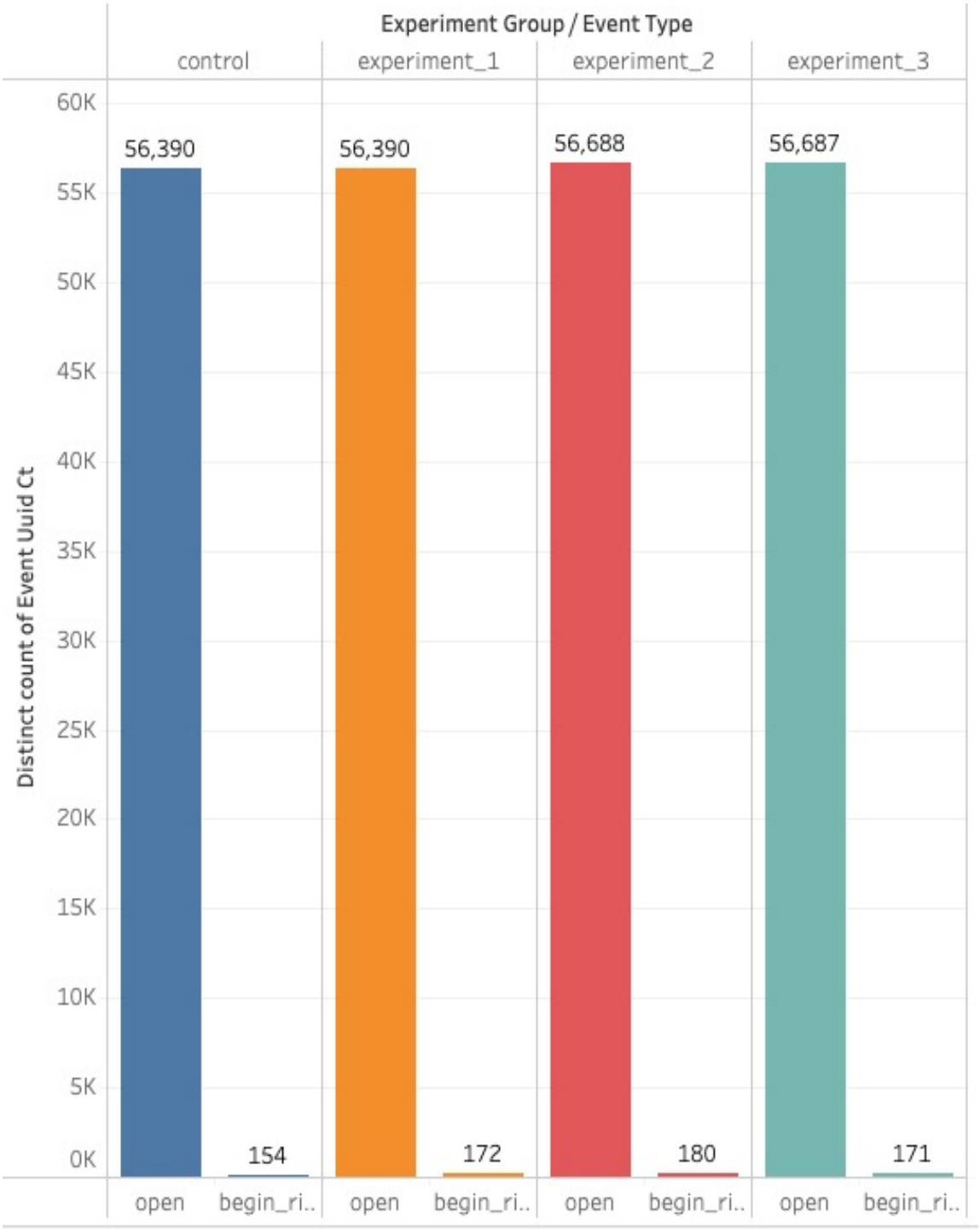
Considering the experiments carried out, I try to understand if there are *big differences between the experimental and control*.

Thus, in order to perform the t-test, I will follow the following steps:

1. Identify the test variable that best fits the scenario - Open event type
2. As the main objective of the test I define the conversion of searches in confirmed rides
3. I assume as hypotheses the possibility of non-conversion of control vs test users or the opposite (state the null hypothesis and alternative hypothesis)
4. Compare the control with the other experiments on an individual basis
5. Consider a 95% confidence threshold interval with significant results
6. Run the multivariate experiment
7. Perform a T-test using this tool - I need to perform a T-test to see if the differences you see in the experiment are significant

Review Multivariate Test Results: Significance Test

Data available for event type and control/experiment groups:



Review Multivariate Test Results: Significance Test

Calculators for Experiments – Control vs Experiment 1

AB Testguide

Is your test result significant? Does it have enough power?
Play with the controls and get a better feel for how a lower confidence level will boost the power or how an increase in test size can make a small CR-difference significant!

Pre-test calculation or post-test evaluation?

- ○ Pre-test analysis
- ◉ Test evaluation

Test data

Visitors A	Conversions A
56390	154
Visitors B	**Conversions B**
56390	172

Apply changes

Settings

Hypothesis (?)

- ○ One-sided
- ◉ Two-sided

Confidence (?)

- ○ 90%
- ◉ 95%
- ○ 99%

save & shar

Test result

The test result is *not* significant.
The observed difference in conversion rate (11.69%) isn't big enough to declare a significant winner. There is no real difference in performance between A and B or you need to collect more data.

A 0.27%
B 0.31%

The expected distributions of variation A and B.

Conversion Rate Control Conversions A / Visitors A	Conversion Rate B Conversions B / Visitors B	Relative uplift in Conversion Rate $CR_B - CR_A / CR_A$
0.27%	0.31%	11.69%
Observed Power	**p value**	**Z-score** $(CR_B - CR_A) / SE_{difference}$
31.55%	0.1590	0.9984
Standard error A $(CR_A * (1-CR_A) / Visitors_A)^{1/2}$	**Standard error B** $(CR_B * (1-CR_B) / Visitors_B)^{1/2}$	**Std. Error of difference** $SE_{difference} = (SE_A^2 + SE_B^2)^{1/2}$
0.00022	0.000232	0.00032

Review Multivariate Test Results: Significance Test

Calculators for Experiments – Control vs Experiment 2

AB Testguide

Is your test result significant? Does it have enough power?
Play with the controls and get a better feel for how a lower confidence level will boost the power or how an increase in test size can make a small CR-difference significant!

Pre-test calculation or post-test evaluation?

- ○ Pre-test analysis
- ◉ Test evaluation

Test data

Visitors A	Conversions A
56390	154
Visitors B	**Conversions B**
56688	180

Apply changes

Settings

Hypothesis (?)

- ○ One-sided
- ◉ Two-sided

Confidence (?)

- ○ 90%
- ◉ 95%
- ○ 99%

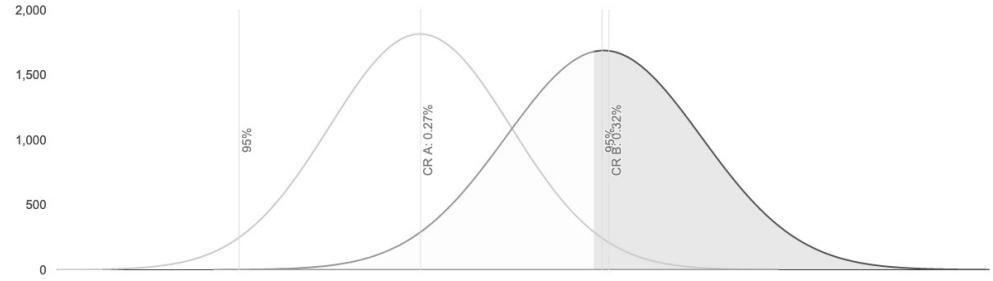
Save & share

Test result

The test result is *not* significant.
The observed difference in conversion rate (16.27%) isn't big enough to declare a significant winner. There is no real difference in performance between A and B or you need to collect more data.

A 0.27%
B 0.32%

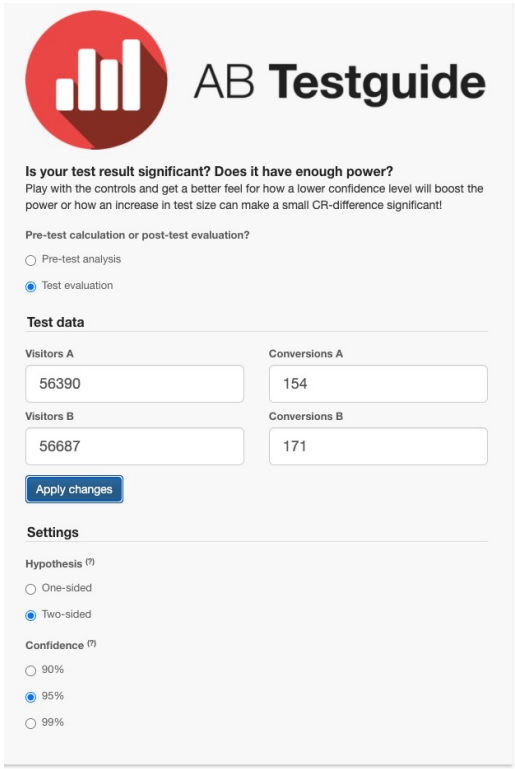
The expected distributions of variation A and B.

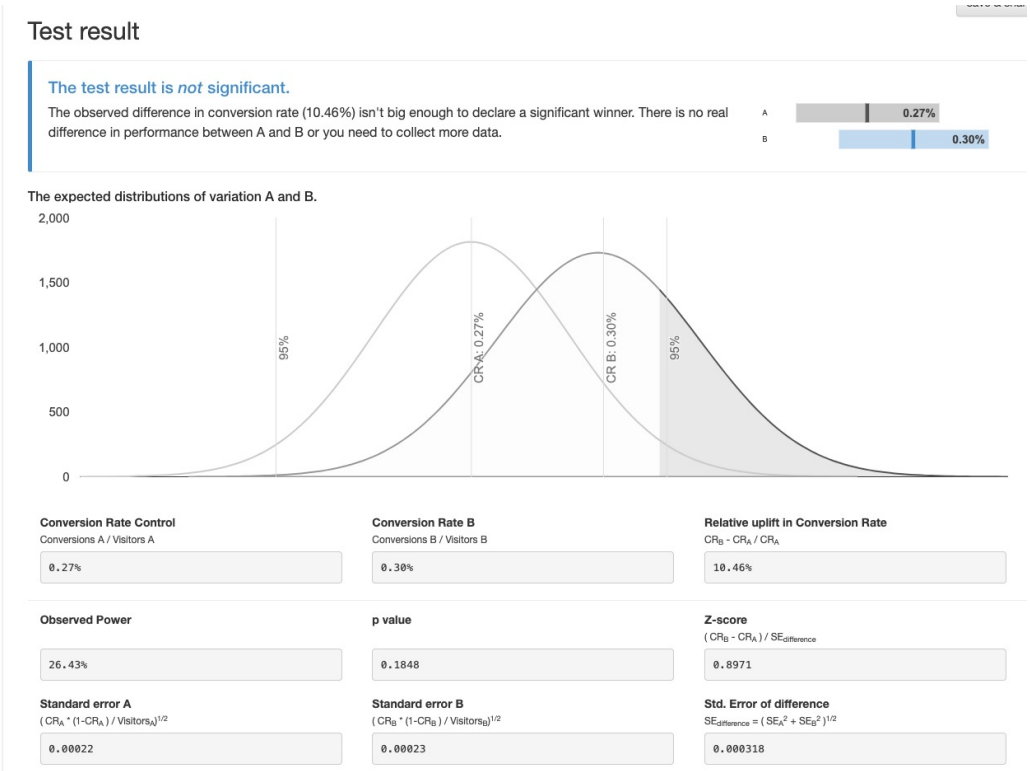


Conversion Rate Control Conversions A / Visitors A	Conversion Rate B Conversions B / Visitors B	Relative uplift in Conversion Rate $CR_B - CR_A / CR_A$
0.27%	0.32%	16.27%
Observed Power	**p value**	**Z-score** $(CR_B - CR_A) / SE_{difference}$
52.29%	0.0843	1.3768
Standard error A $(CR_A * (1-CR_A) / Visitors_A)^{1/2}$	**Standard error B** $(CR_B * (1-CR_B) / Visitors_B)^{1/2}$	**Std. Error of difference** $SE_{difference} = (SE_A^2 + SE_B^2)^{1/2}$
0.00022	0.000236	0.000323

Review Multivariate Test Results: Significance Test

Calculators for Experiments – Control vs Experiment 3





Review Multivariate Test Results: Significance Test

For a multivariate test, I compared the “Control” to experiments “Exp. 1” and then “Control” to “Exp. 2” and then “Control” to “Exp. 3” to see if any test have a significant effect.

Considering the above results, comparison between control and each of the experiments, it is concluded that the results, based on the p-value, are insignificant for the 3 cases at 95% confidence level.

Step 2

Funnel & Cohort Analyses

User Funnel

Identifying the different stages the user funnel

Considering the data available I can conclude that there are four (events) steps in the customer journey to book a flying ride:

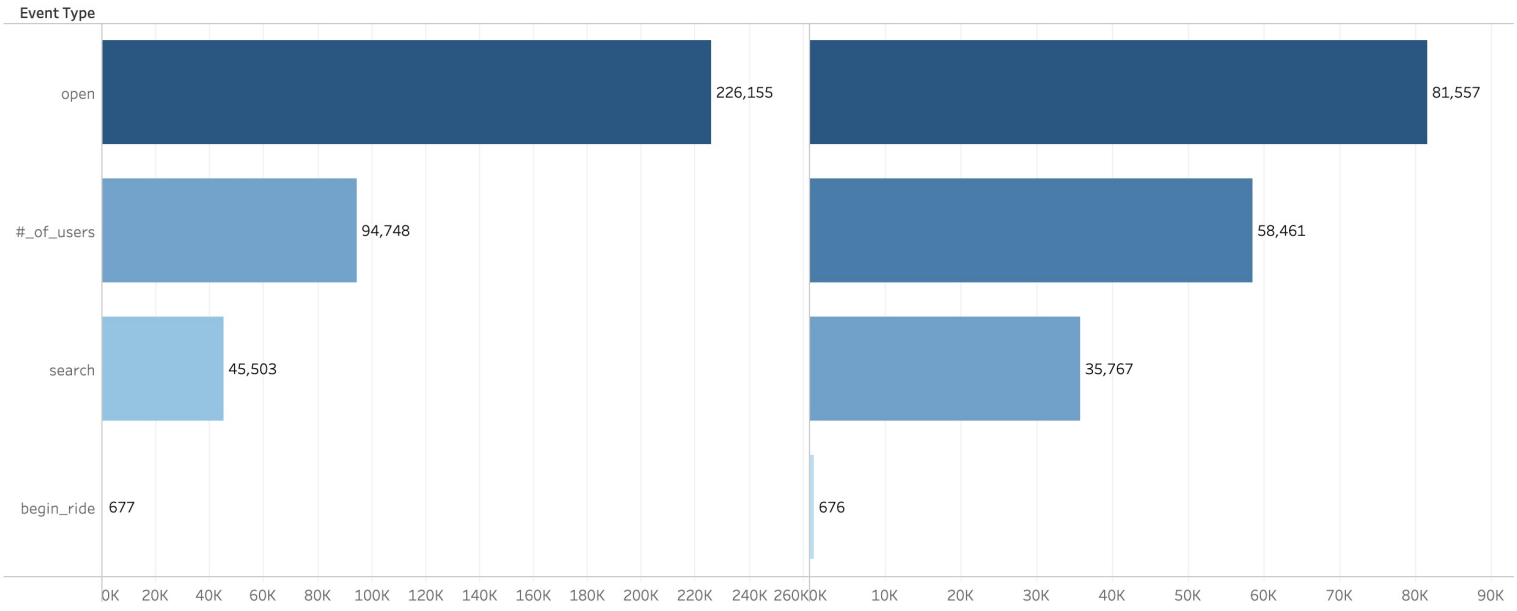
1. open - assuming that all incoming costumers open the mobile application;
2. #_of_users - complete the login and enter/navigate the application. Refers to the number of passengers to book the ride. After opening the app and search for a ride, then you specify how many passengers to take that ride.
3. search - search by cab, location, time period and time, etc ...;
4. begin_ride – the users complete the search step and proceed to booking the ride and begin riding.

User Funnel

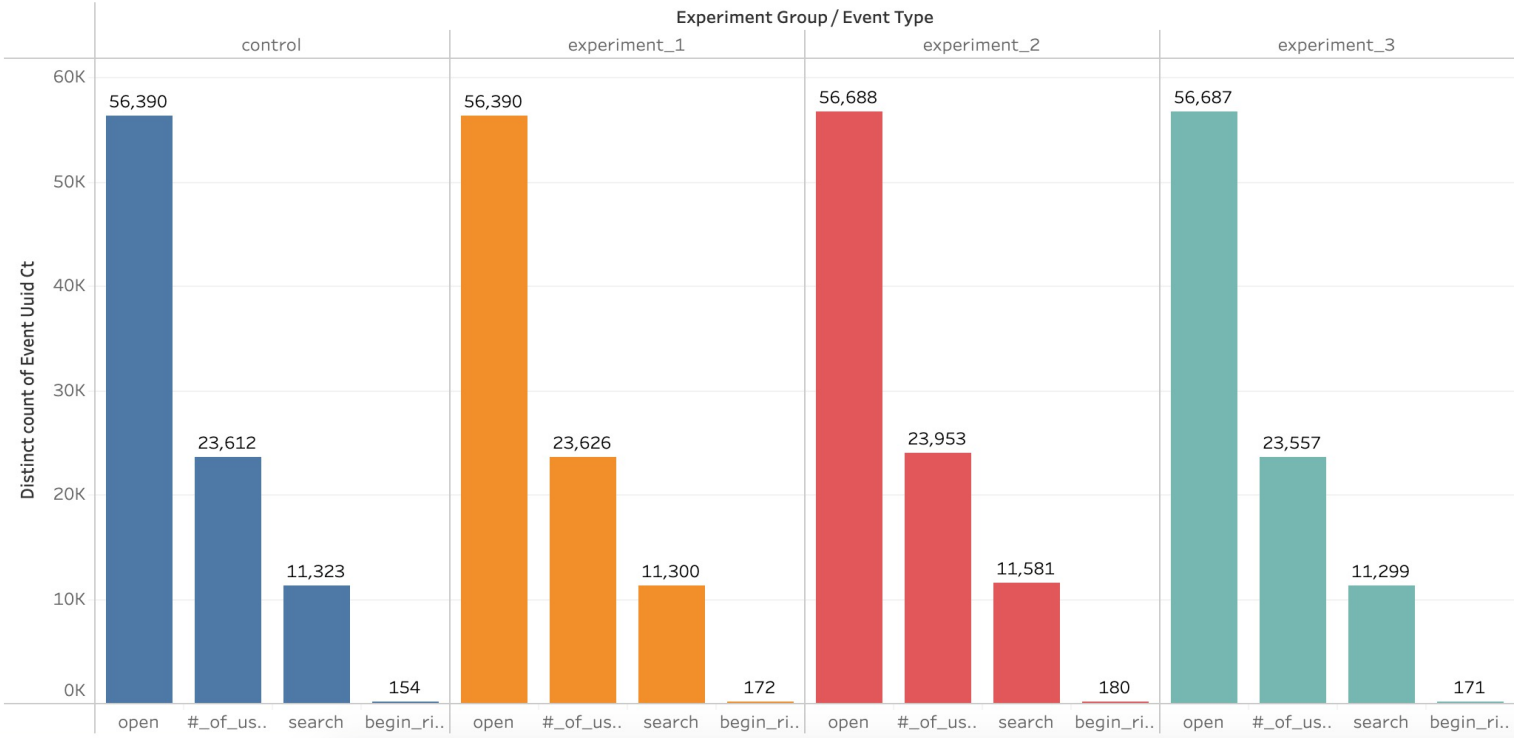
Identifying the different stages the user funnel

Funnel from step to step, including drop off rates

Events Funnel Analysis



User Drop

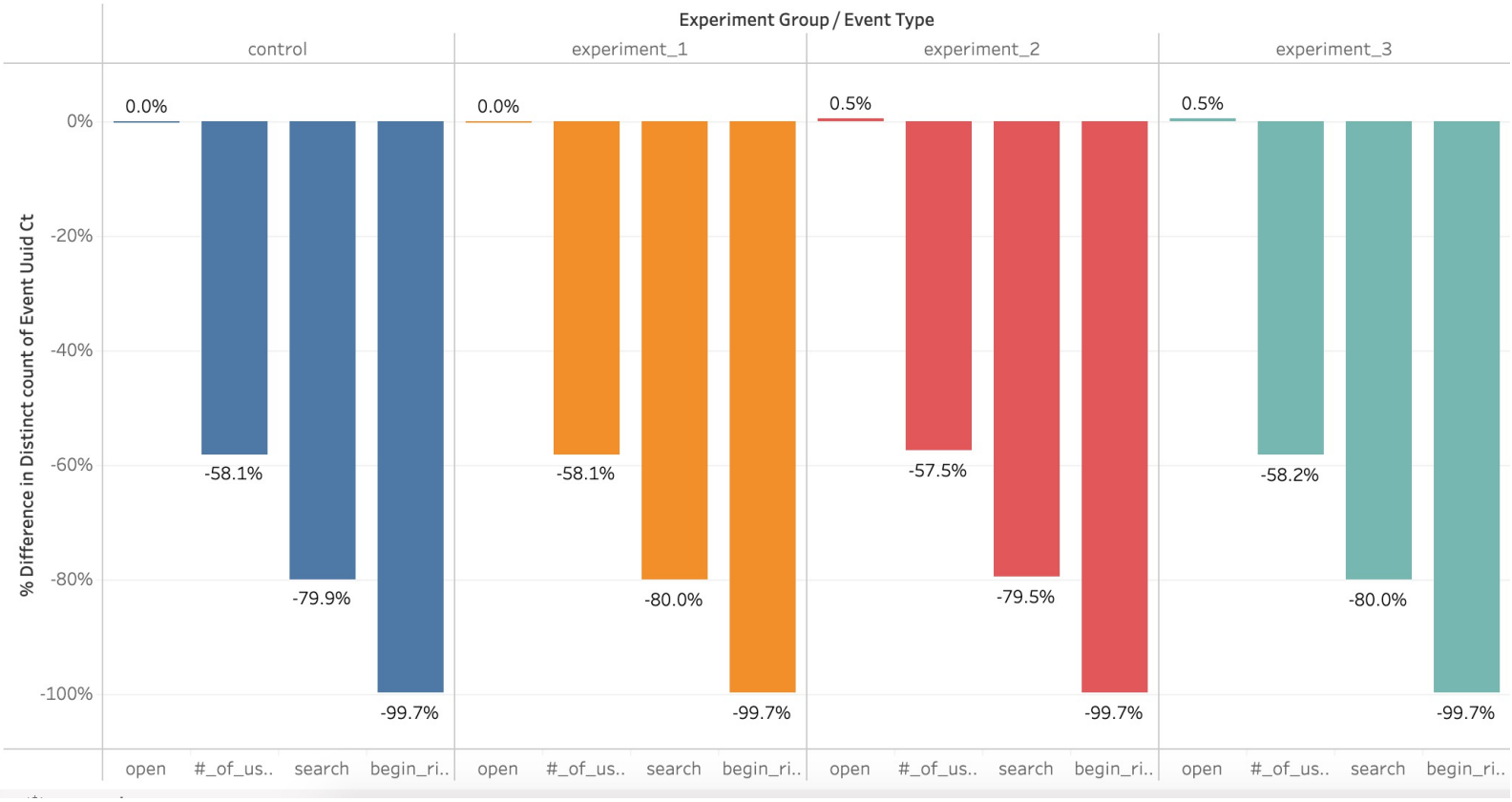


User Funnel

Identifying the different stages the user funnel

Funnel from step to step, including drop off rates

User Drop conversion



Conversion Rate

	Experiment Group			
Event Type	control	experiment_1	experiment_2	experiment_3
open	0.00%	0.00%	0.00%	0.00%
#_of_users	-58.13%	-58.10%	-57.75%	-58.44%
search	-79.92%	-79.96%	-79.57%	-80.07%
begin_ride	-99.73%	-99.69%	-99.68%	-99.70%

User Segments

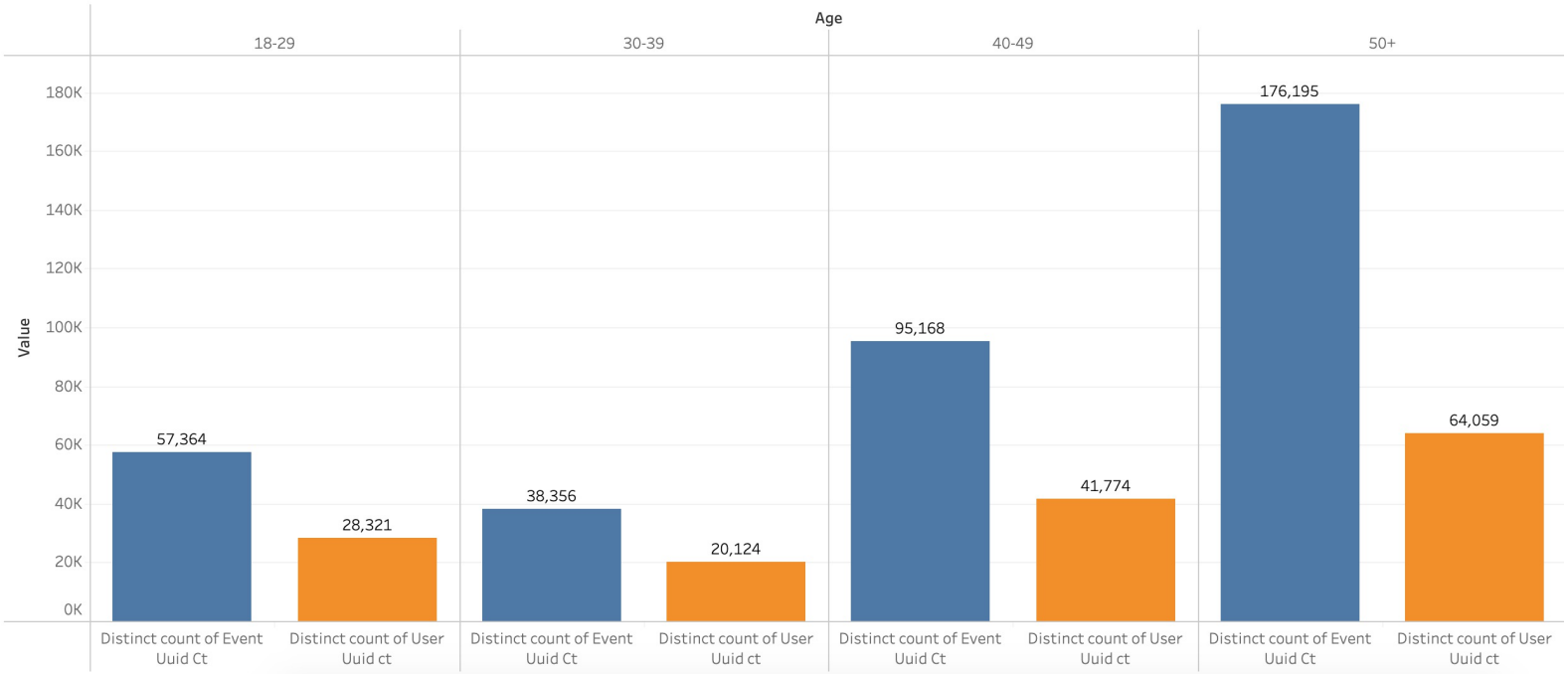
Given the data we have available, it is possible to identify **two demographic attributes** that allow for segment analysis:

1. Age group
2. Neighbourhood

User Neighborhoo	
Bronx	10,802
Brooklyn	73,880
Manhattan	257,259
Queens	18,088
Staten Island	7,054

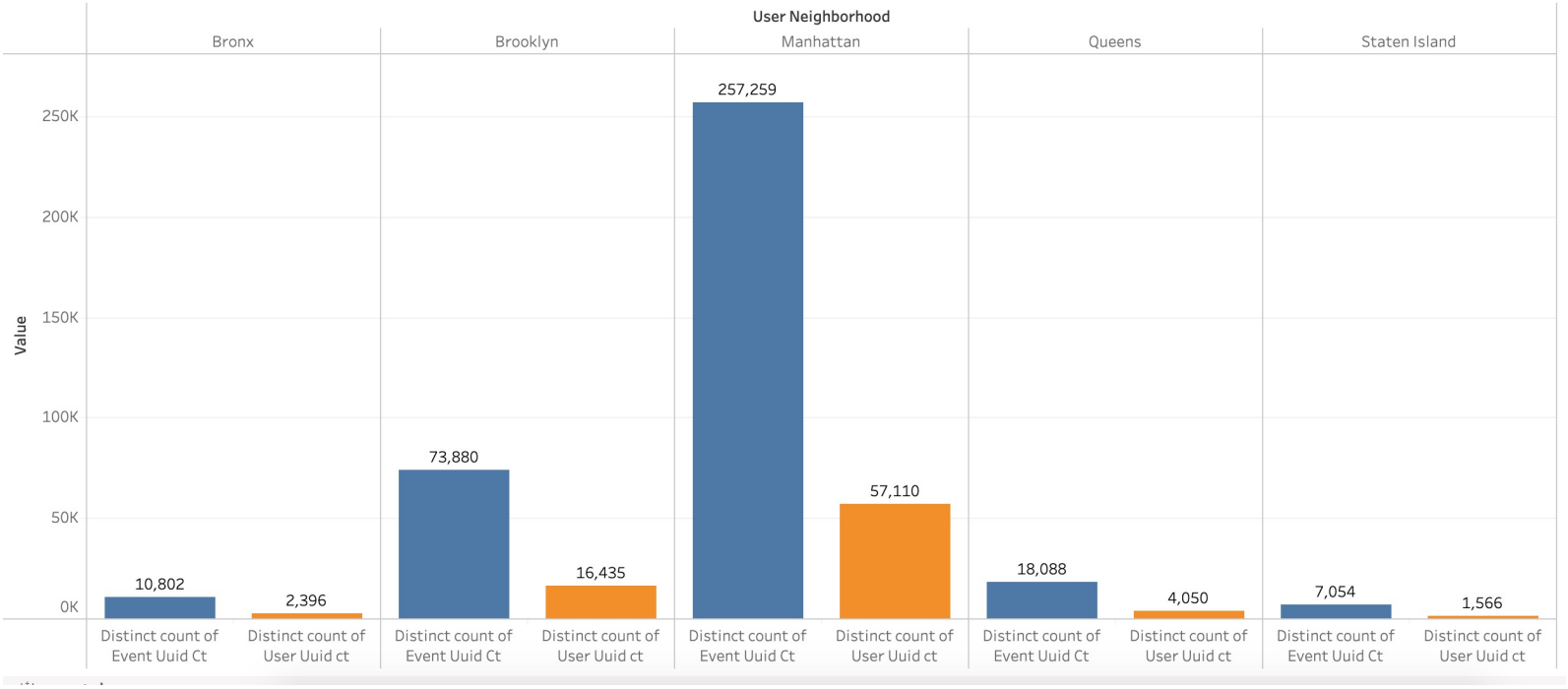
Age	
18-29	57,364
30-39	38,356
40-49	95,168
50+	176,195

Age Distribution



User Segments

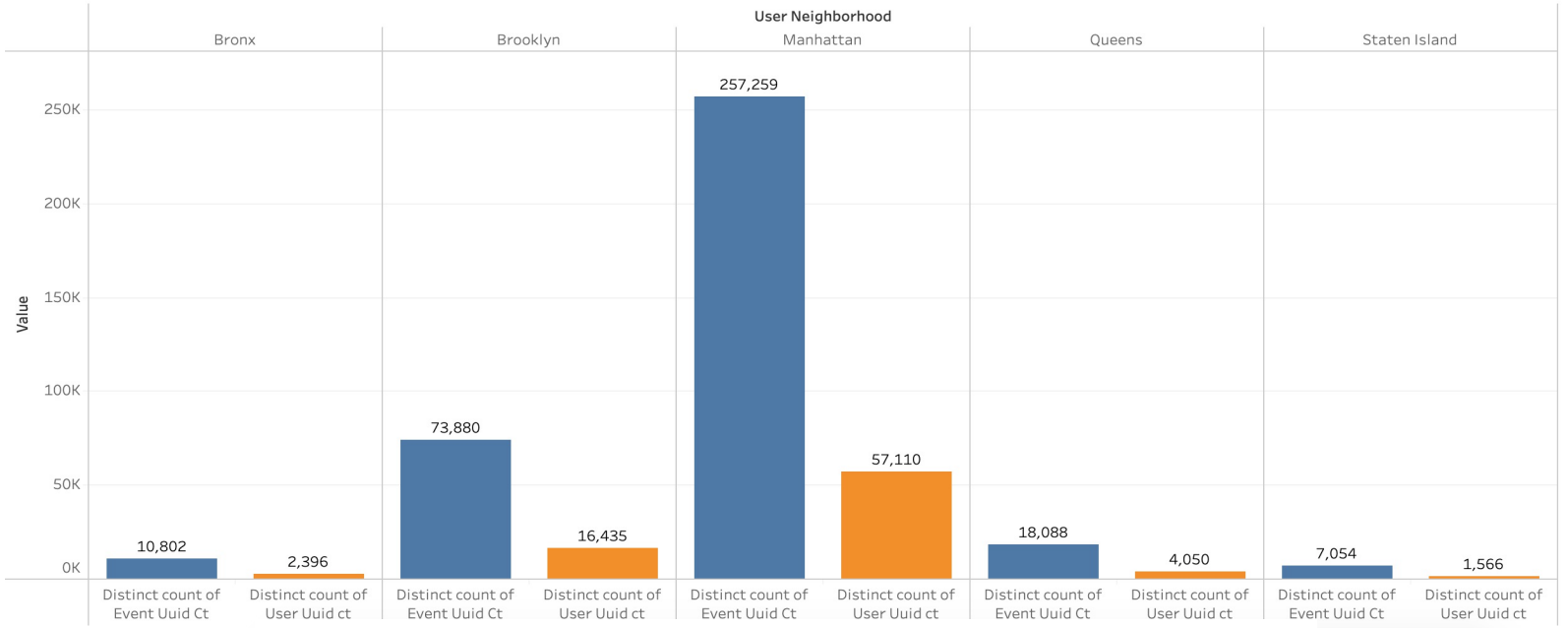
Neighboard Distribution



	Age			
User Neighborhood	18-29	30-39	40-49	50+
Manhattan	40,165	26,850	66,599	123,645
Brooklyn	11,641	7,677	19,183	35,379
Queens	2,835	1,930	4,720	8,603
Bronx	1,686	1,186	2,740	5,190
Staten Island	1,037	713	1,926	3,378

User Segments

Neighboard Distribution



	Age			
User Neighborhood	18-29	30-39	40-49	50+
Manhattan	40,165	26,850	66,599	123,645
Brooklyn	11,641	7,677	19,183	35,379
Queens	2,835	1,930	4,720	8,603
Bronx	1,686	1,186	2,740	5,190
Staten Island	1,037	713	1,926	3,378

	Experiment Group			
Age	control	experiment_1	experiment_2	experiment_3
18-29	14,350	14,083	14,613	14,318
30-39	9,570	9,635	9,641	9,510
40-49	23,363	24,036	23,768	24,001
50+	44,196	43,734	44,380	43,885

	Experiment Group			
User Neighborhood	control	experiment_1	experiment_2	experiment_3
Manhattan	64,068	64,048	64,636	64,507
Brooklyn	18,362	18,574	18,687	18,257
Queens	4,547	4,312	4,653	4,576
Bronx	2,683	2,748	2,723	2,648
Staten Island	1,819	1,806	1,703	1,726

User Segments

Using the visualisations presented above:

- Manhattan had the **highest number of users**, and age group over 50+ was the most active, followed by 40-49 group.

- Whereas in Manhattan the **lowest amount of users was in the 30-39 age group** and **Staten Island had the lowest amount of users**.

- The 30-39 age group had the least number of users and the 50+ age group had the most users.

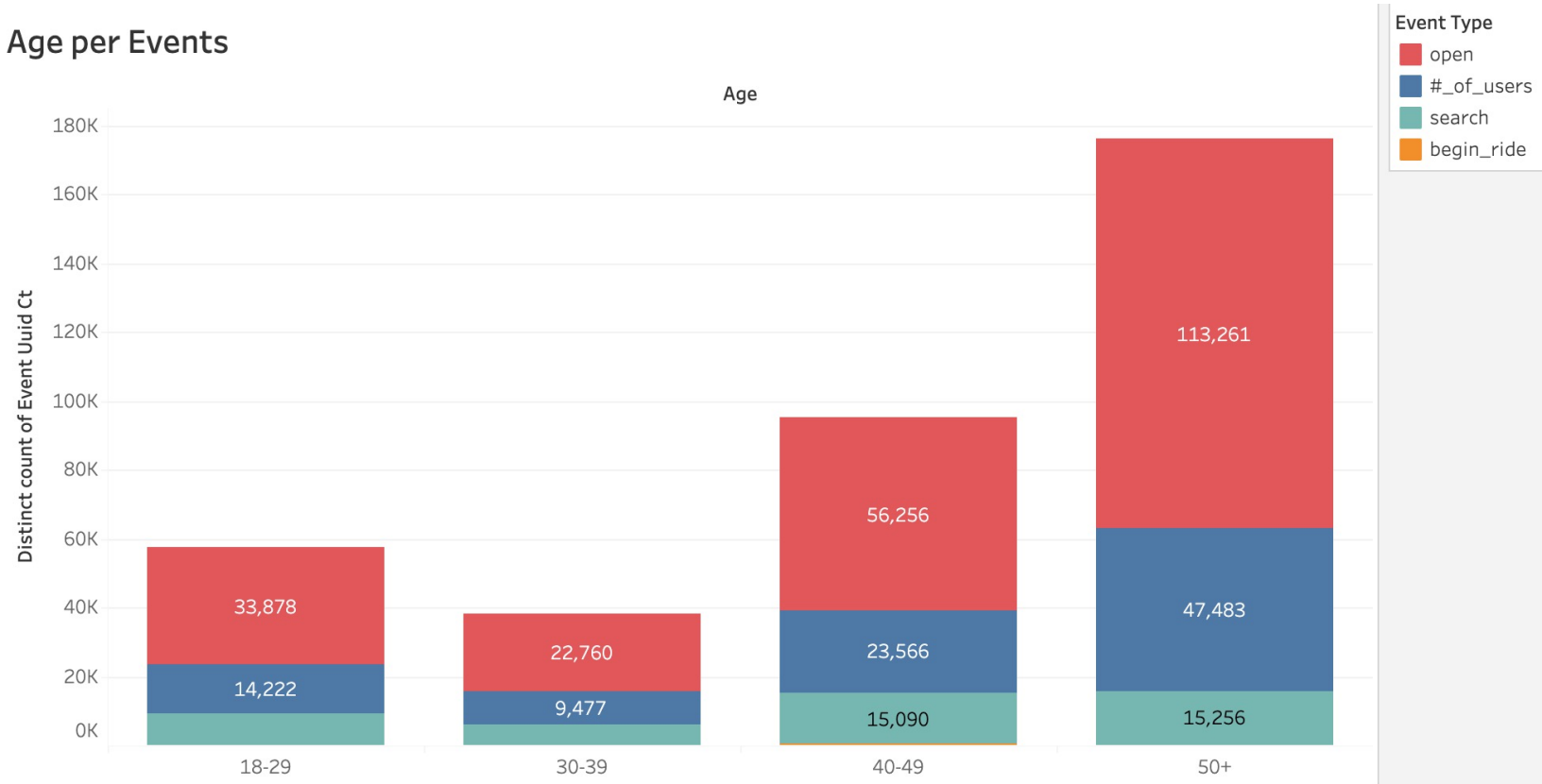
Segment Analysis of Funnel

Identify Opportunities for Improvement

Perform a funnel analysis by segment for Age Group

Age Group Funnel Drop Off

	Age			
Event Type	18-29	30-39	40-49	50+
open	33,878	22,760	56,256	113,261
#_of_users	14,222	9,477	23,566	47,483
search	9,138	6,019	15,090	15,256
begin_ride	126	100	256	195



Segment Analysis of Funnel

Identify Opportunities for Improvement

Perform a funnel analysis by segment for Age Group:

After analysis it was found that the **50+ age group** shows a great interest in using the service compared to the other age groups, but after analysis it was found that the actual use of the service is very low. Contrary to the age groups 18-29 age group and 30-39 age group which show a more in line behaviour, that is, they have a relatively low service demand rate, and a slightly higher service utilization rate.

On the other hand, **the 30-39 age group shows a lower performance**, which is strange because it is a group with very specific needs and they are somehow aligned with the 40-49 age group, as they should present similar needs and behaviour both in social and professional terms. *I'll need to analyse the qualitative data to try to understand why.*

No doubt that the **50+ and 30-39 age groups** (and probably 18-29 age group too) **need a more specific and deep analysis.**

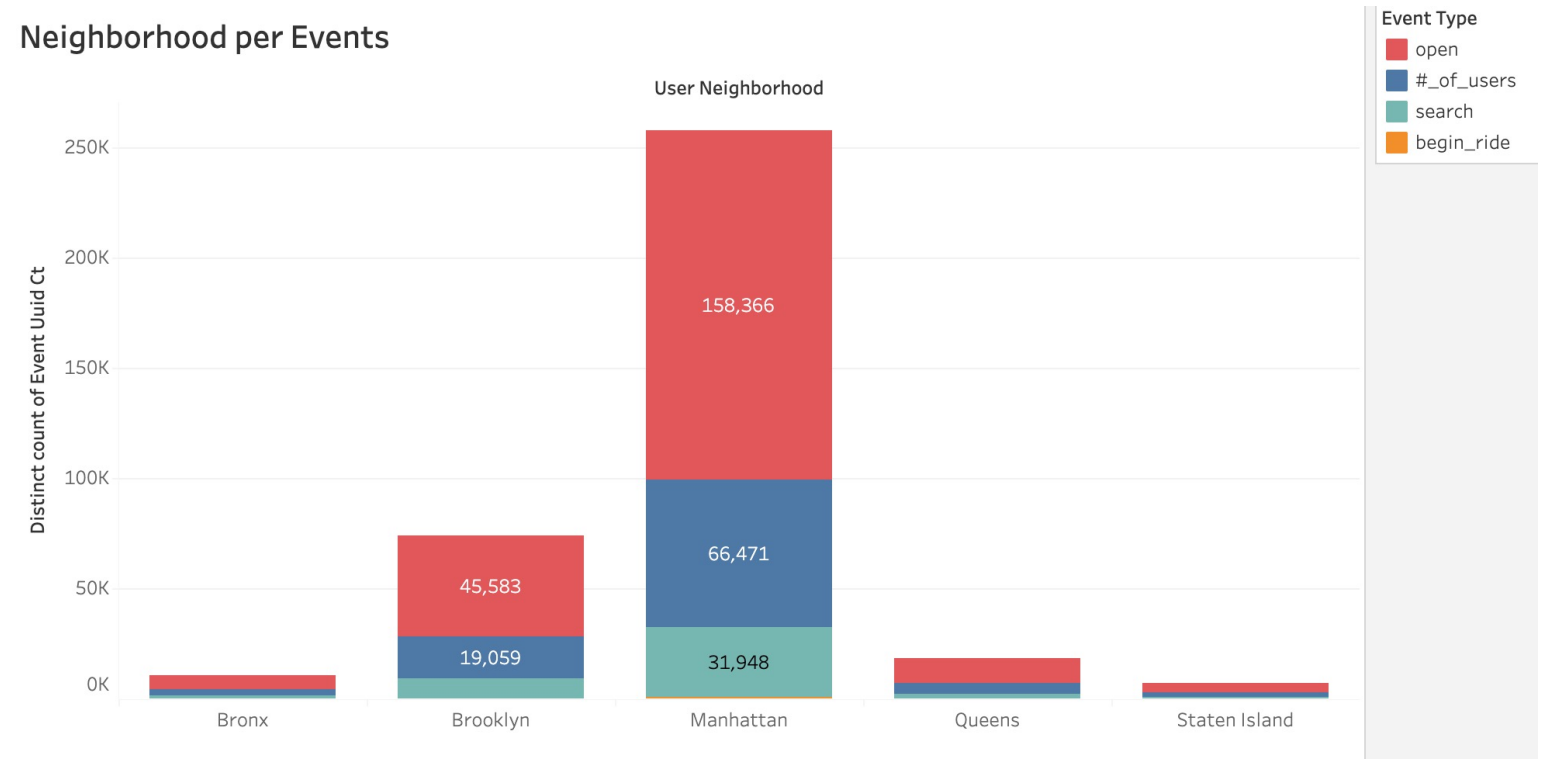
Segment Analysis of Funnel

Identify Opportunities for Improvement

Perform a funnel analysis by segment for Neighborhood

Neighborhood Funnel Drop Off

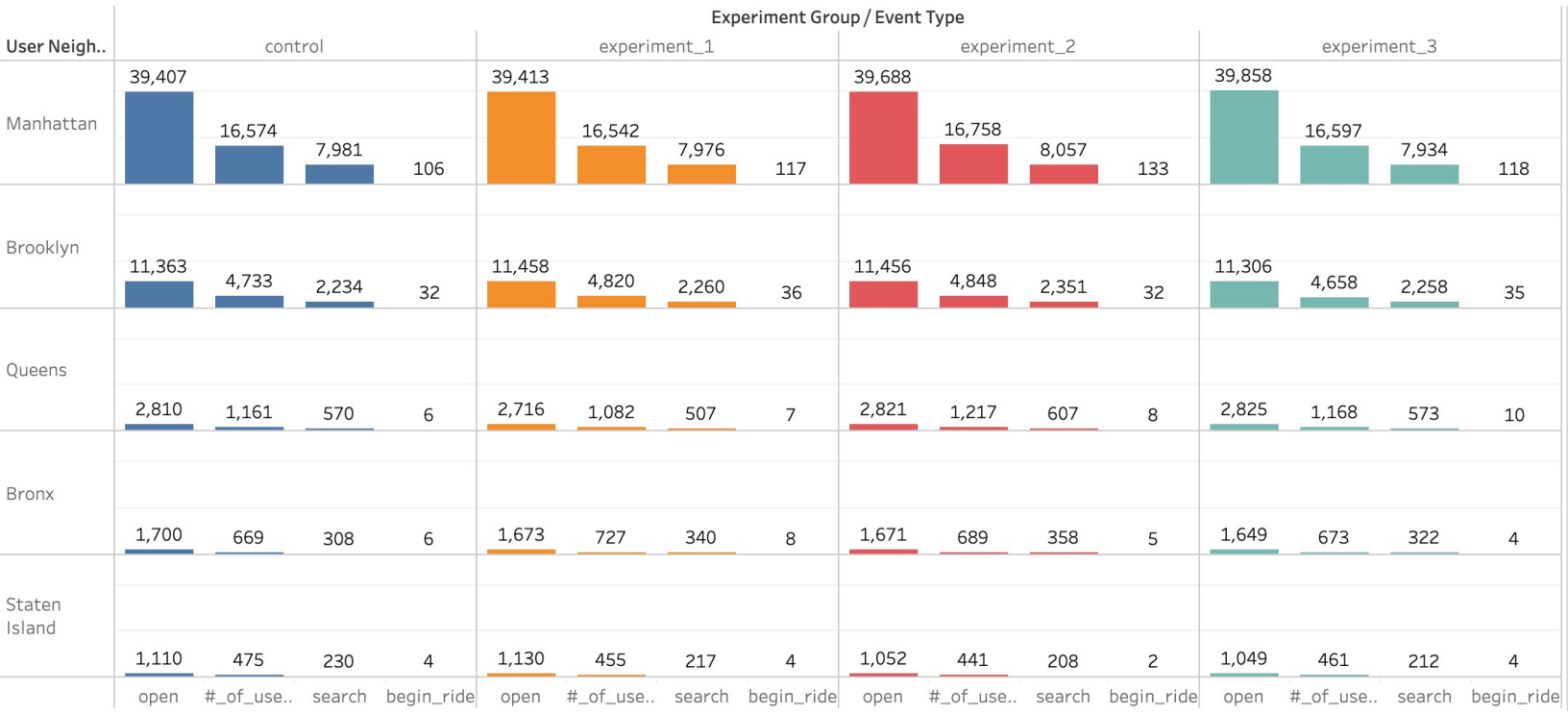
	User Neighborhood				
Event Type	Manhattan	Brooklyn	Queens	Bronx	Staten Island
open	158,366	45,583	11,172	6,693	4,341
#_of_users	66,471	19,059	4,628	2,758	1,832
search	31,948	9,103	2,257	1,328	867
begin_ride	474	135	31	23	14



Segment Analysis of Funnel

Identify Opportunities for Improvement

Perform a funnel analysis by segment for Neighborhood



There is no **doubt that Manhattan has the highest number of service usage**, followed by Brooklyn.

The **drop off is more pronounced in Manhattan in the event of a search**, and the island state has the lowest number of users.

In terms of **best success rate of rides, I find Bronx**.

The **best conversion rate** is found in **experiment 2**.

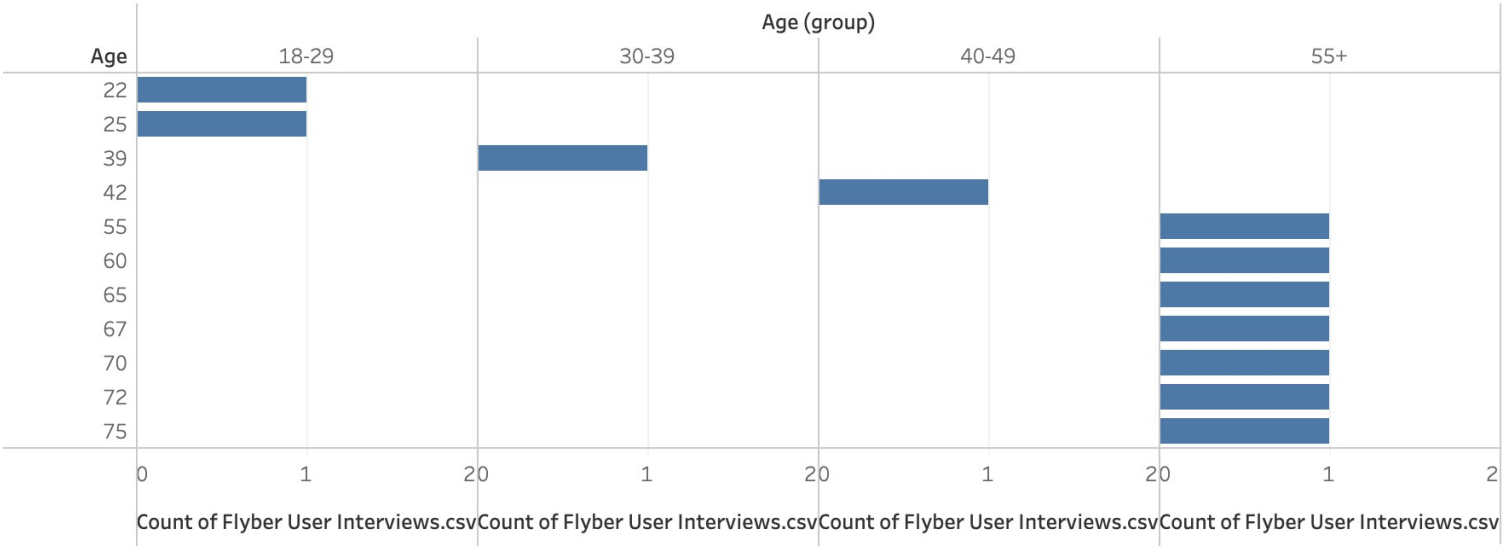
Step 3

Hypothesis & Next Steps

Review Qualitative Data

After analysis it can be seen that the user interview log file (**qualitative data**) is in line with the quantitative data because I can see that the **age group 50+** is in majority.

Age Distributions



How Often Do You Use Flyber?

	Age (group)			
How Often Do You Use Fly..	18-29	30-39	40-49	55+
1X a month			1	1
1X a week	1			
2X a month				1
3+ times a week		1		1
5+ times a month, during ..				1
Every couple months	1			1
I've used in a few times.				1
Tried it once				1

Review Qualitative Data

At the “level of usage” I can see that there is a **great variation between all age groups**.

At the level of **behavioural patterns** I can identify at the 50+ level that: **(1)** prefer to call a taxi, **(2)** use a chauffeur to drive the car, **(3)** lack of smartphone use and **(4)** old ingrained habits.

The age groups 18-29 and 30-39, showed **low levels of effective use (booking rides)**, so it’s important to understand what is happening at the qualitative level in these groups. In the qualitative data I found three users who offer different views between them.

In theses there are those who use the service three times a week, but say they *look for another alternative like Uber to save money*, and the rest use the service less than once a week.

Review Qualitative Data

I can see the statements of the interviewees "broken down" by age group (*quotes from users*).

What Do You Use When You Don't Use Flyber?

What Do You Use When You Don't Use Flyber?	Age (group) 18-29	30-39	40-49	55+
Before Flyber, I'd call a taxi service on the phone.				1
Before I started using Flyber, I would use a normal uber.			1	
Drive my car or call a taxi service.				1
Honestly, I thought about using Flyber to suprise my grandson or granddaughter with a visit to one of their sporting games. Luckily my daughter was around to h..				1
I call up our local pilot, Bob. He's not always available but I don't need to fiddle around with an app and hitting tiny buttons. He knows where I tend to be and w..				1
I have a personal car service on call. My assistant books Flyber whenever I'd be travelling during peak NYC traffic hours. Time is money and Flyber saves me tim..				1
I just hail a taxi or tell my phone to call a cab to go to a certain address (I'm always on the phone, so I just use voice commands with my phone most of the ti..				1
I usually use Flyber when I need to get from the office to a client meeting. Before Flyber, we used a car service.				1
If the timing isn't different, I'll take a taxi or uber to save money.		1		
My family has a private jet, so I only use Flyber when I need to get around in the city. Sometimes my friends and I take limosines ..or I have 3 personal cars I can ..	1			
Sponsors often pay for me to use their vehicles or set up transportation for me when visiting their restaurants or events.	1			

I like that Flyber saves me time over traffic. — Esmae Wagner

Drive my car or call a taxi service. — Tanner Cherry

If the timing isn't different, I'll take a taxi or uber to save money. — Keelan Flores

Sponsors often pay for me to use their vehicles or set up transportation for me when visiting their restaurants or events. — Pharrell Campos

Honestly, I thought about using Flyber to suprise my grandson or granddaughter with a visit to one of their sporting games. Luckily my daughter was around to help me book the ride. I usually just use Uber because it remembers my addresses and has all my favorite places saved, so I guess I always just open that up since it is so convenient and saves me time. Though now that I say that, I really should use Flyber again since it would save more time when it comes to fighting traffic! — Robbie Gates

Suggested Features & Experimentation Plan

Three hypothesis using the following format:

"We believe [observed quantitative effect] Because [hypothesized user "why"] And that by [general change/opportunity for Flyber to improve] for [targeted cohort] we will see [expected effect]"

Hypothesis 1 - Drop off in 18-19 & 30-39 age groups

We believe that the two age groups between 18 and 39 do not have a reasonable level of usage against the needs of this age group because they are probably not aware of the brand and service provided by Flyber due to lack of marketing and advertising campaigns, targeted at these age groups. And that by having new marketing and advertising campaigns aimed at these age groups in the appropriate channels (these age groups are on social networks, youtube, ...) we will see a growth in demand and a higher using and effectiveness rate in 3 months.

Features for Hypothesis 1

- Create a monetary incentive that allows to add and increase $ after using and effectively booking rides for 4 times. After 4 rides you are entitled to 20% of the total value of the 4 ride made previously

- Allow the user to send a link to a friend to refer them to register and use the app with the guarantee of having a monetary incentive to travel (both the referrer and the installer and app user)

Suggested Features & Experimentation Plan

Hypothesis 2 - Drop off in 18-19 & 30-39 age groups

We believe that the two age groups between 18 and 39 do not have a reasonable level of use in face of the needs of this age group because they are probably used to using other transport services like Uber or Fly. And that by having a incentive in the application after the first booking, as a monetary incentive to make rides, we will see a growth in demand and a higher rate of usage and effectiveness at the end of 1 month, converting them after a memorable ride and experience.

Features for Hypothesis 2

- Allow the user to send a link to a friend to refer them to register and use the app with the guarantee of having a monetary incentive to travel (both the referrer and the installer and app user)

- Create a monetary incentive that allows to add and increase $ after using and effectively booking rides for 4 times. After 4 rides you are entitled to 20% of the total value of the 4 ride made previously

Suggested Features & Experimentation Plan

Hypothesis 3 - Drop off in the 50+ age group

We believe that the age group of 50+ despite having a very high rate of use, shows signs of abandonment of the application at the search step because they do not book the ride due to the fact that they are very used to call a taxi by phone or because they can not use / handle the application correctly, and that by adding features that allow booking by call for this age group it is likely that we will see increases in the level of booking rate and decrease in the level of bounced rate.

Features for Hypothesis 3

- Make the booking through the app in a call form, which allows users to simulate the behaviour of the manual booking like call a taxi-cab

- Simplify the manual booking request that allows the GPS to auto-complete the locations and other configurations

- Allow to record in history the rides, allow to give a specific name to the ride, and through voice request the same ride again with the same characteristics without having to do all the configurations

Suggested Features & Experimentation Plan

To ensure that the differences shown in the different groups of experiments are guaranteed to be caused by changes in Flyber's service and not caused by different user characteristics, when conducting the various experiments it's important to ensure that the different groups have similar characteristics.

Who should be exposed to the experimental changes for Hypothesis 1 and 2

Although there is a possibility of slightly different behaviour within this age group, based on the qualitative data survey, the groups that should be exposed would be. It is necessary to ensure that these users are placed as a target in the control and in the experiments.

- 18-29 age group - strong focus on incentives + marketing campaigns
- 30-39 age group - strong focus on incentives + marketing campaigns

Who should be exposed to the experimental changes for Hypothesis 3

It is necessary to ensure that these users are placed as a target in the control and in the experimente, and also ensuring that other characteristics, such as region and income, are taken into account and are properly randomly balanced between the experimental groups and control.

- 50+ age group - focus on the use of the new feature

Suggested Features & Experimentation Plan

When launching an experiment, it is important to track user activity and user characteristics.

Additional metrics that can be gathered from the suggested characteristics:

- User information, such as mobile device information, version, operating system, ...
- The number of unique visitors to the app where they will test the new design, ...
- Identify which users used a certain feature considered in the hypotheses (referencing, call, voice, ...)
- Understand which was the user's experimentation group
- Identify the usage rate vs. achievement rate for those who used the features
- Understand what kind of users in terms of characteristics used these features
- Qualitative data - Possibility to launch a short survey in the satisfaction app after a successful ride, to identify the level of user satisfaction and what they would like to see as new features

Suggested Features & Experimentation Plan

In order to perform the t-test, I'll follow next steps:

1. Identify the test variable that best fits the scenario - **Open** event type
2. As the main objective of the test I define the **conversion of booking rides**
3. I assume as hypotheses the possibility of **non-conversion of control vs test users** or the opposite:
    - **null hypothesis** - there is no difference in the booking ride rate between users in the control and test groups. This is what I want to reject
    - **alternative hypothesis** - there is a difference in the booking ride rate between the test and control groups. This is what I want to accept
    - I must make sure that users in the control and the test/experiment group share the same characteristics.
4. Compare the control with the other experiments on an individual basis
5. Consider a 95% confidence threshold interval with significant results
6. Run the multivariate experiment (A/B test could be followed up with multivariate experiment to further optimize headlines)
7. Perform a **T-test** using this tool
8. *"P value: The value that indicates the probability of obtaining results as extreme as the observed results of a statistical hypothesis test, assuming that the null hypothesis is correct. A smaller p-value means the evidence in favor of the alternative hypothesis is stronger"*

Appendix

Additional Info

Tableau Viz: *https://public.tableau.com/app/profile/s.rgio.da.costa*

T-Test tool: *https://abtestguide.com/calc*