Causal Project

A/B Testing on App's Booking Page

Group 10: Yu Chun Peng, Chien-Chu Hsu, Chia-Yen Ho, Ng Ka Lok Carol

Roadmap

Problem & Data

1

Threats to Causal Inference

3

Experiment

5

2

Experiment Hypothesis

4

Visualization & Check

6

Recommendations & Limitations
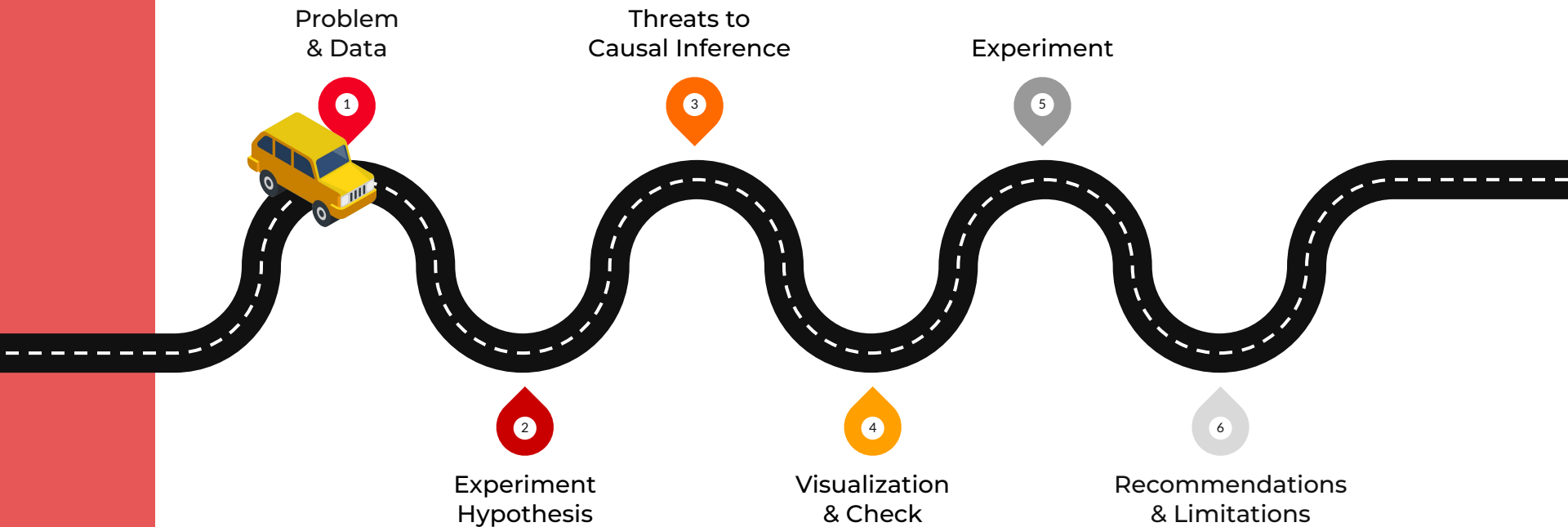
Background
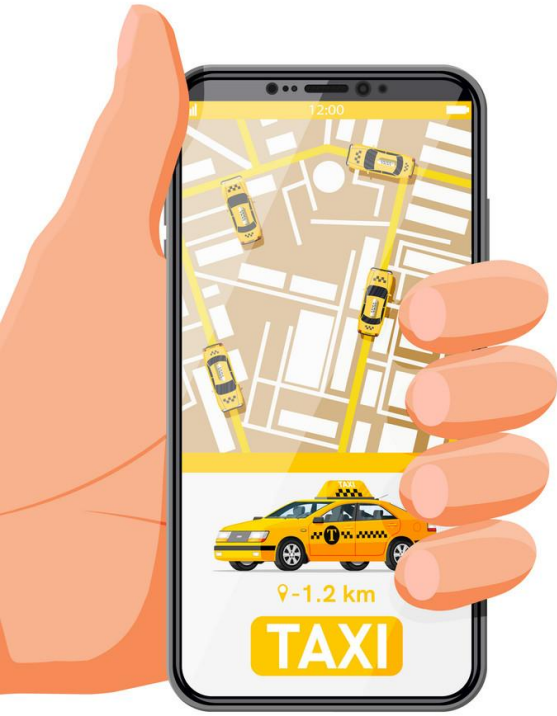
Flyber app provides a flying-taxi service in one of the most congested cities in America - New York City.

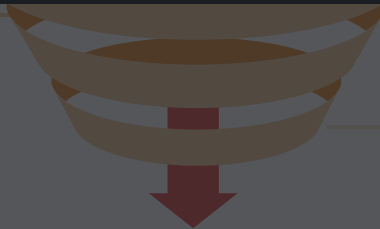See More →

User Journey of Booking Page

Open Flyber — 1

2 — Enter Number of Riders

Search Vehicle — 3

4 — Begin Ride

User Journey of Booking Page

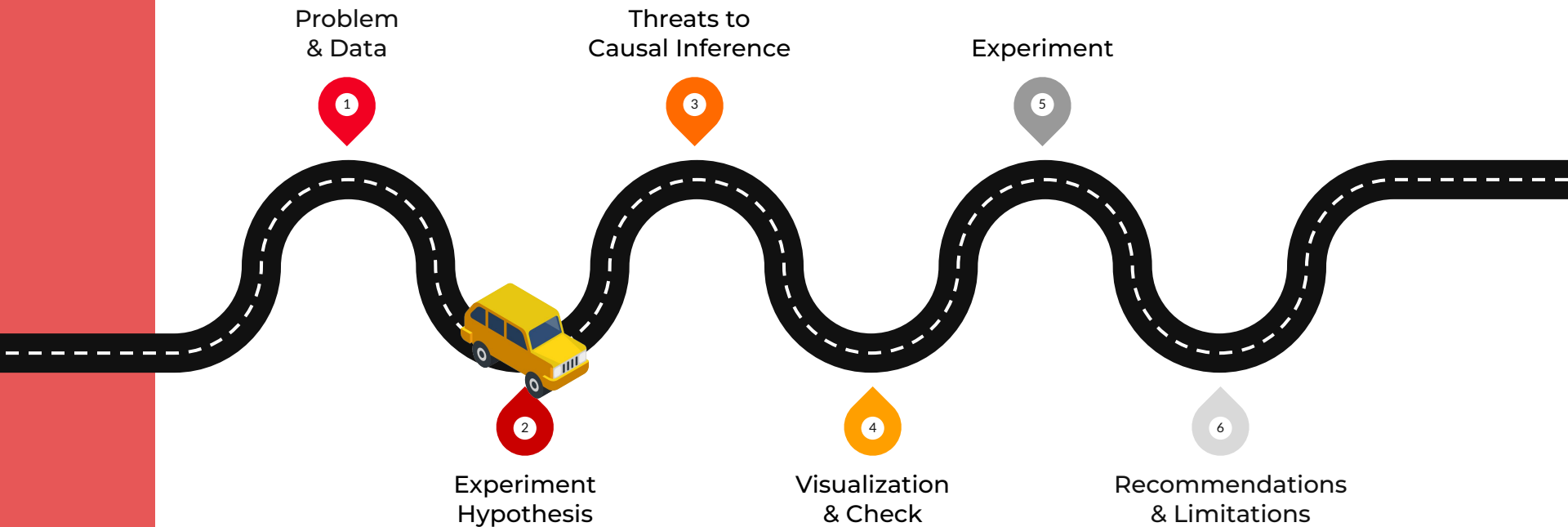Search Vehicle

3

4 Begin Ride

Low Conversion Rate

Problems for Flyber

▶ How people engage with the booking page?

▶ How can we improve UI/UX to generate more bookings?

Dataset Description

- **user_uuid** - an unique id for each user

- **experiment** - group - to identify 1 original and 2 experiment versions

- **event_uuid** - trip booking event id

- **event_time** - time when user triggered a trip

- **age** - four age groups including 18-29, 30-39, 40-49 and 50+

- **session_uuid** - an unique id for each session

- **user_neighborhood** - neighborhoods in New York City

- **event_type** - indicates the stage of user journey

Roadmap

Problem & Data

1

Threats to Causal Inference

3

Experiment

5

2

Experiment Hypothesis

4

Visualization & Check

6

Recommendations & Limitations

Experiment hypothesis

Hypothesis 1: Redesign the text on the booking button would drive more conversions
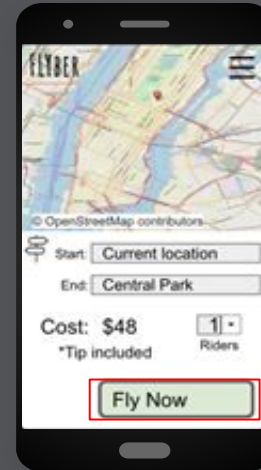

Control

Treatment 1

Experiment hypothesis

Hypothesis 1: Redesign the text on the booking button would drive more conversions

Control   Treatment 1

Experiment hypothesis

Hypothesis 1: Redesign the text on the booking button would drive more conversions
Hypothesis 2: Remove "Tip included" text would drive more conversions
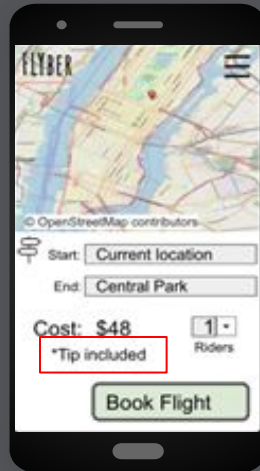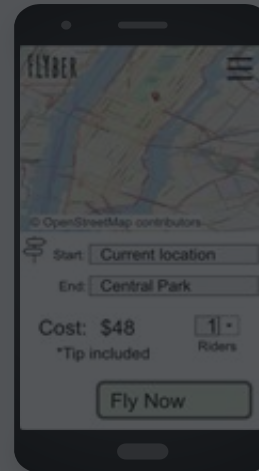

Control

Treatment 1

Treatment 2

Experiment hypothesis

Hypothesis 1: Redesign the text on the booking button would drive more conversions

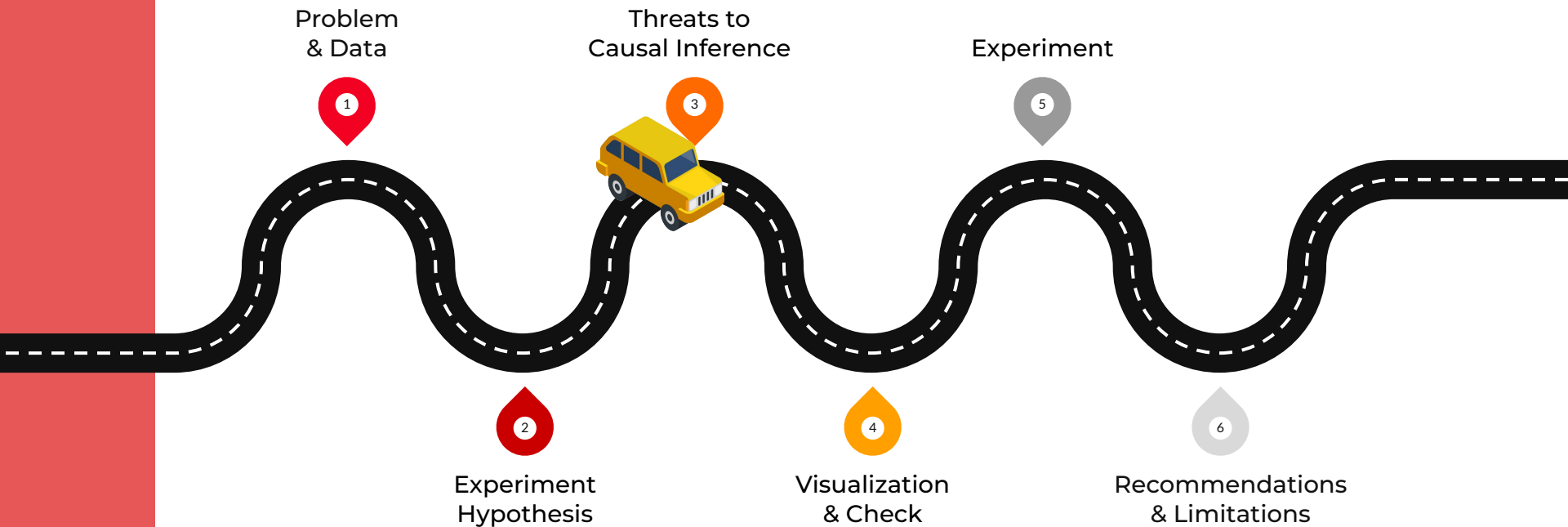**Hypothesis 2:** Remove "Tip included" text would drive more conversions

Control   Treatment 1   Treatment 2

Roadmap

Problem
& Data

1

Threats to
Causal Inference

3

Experiment

5

Experiment
Hypothesis

2

Visualization
& Check

4

Recommendations
& Limitations

6

Threats to Causal Inference

Omitted Variable Bias

User's decisions might be affected by other variables e.g. salary

1

2

Selection Bias

Users could share the same preference or habit of riding

Users might not be aware of design change

Measurement Error

3

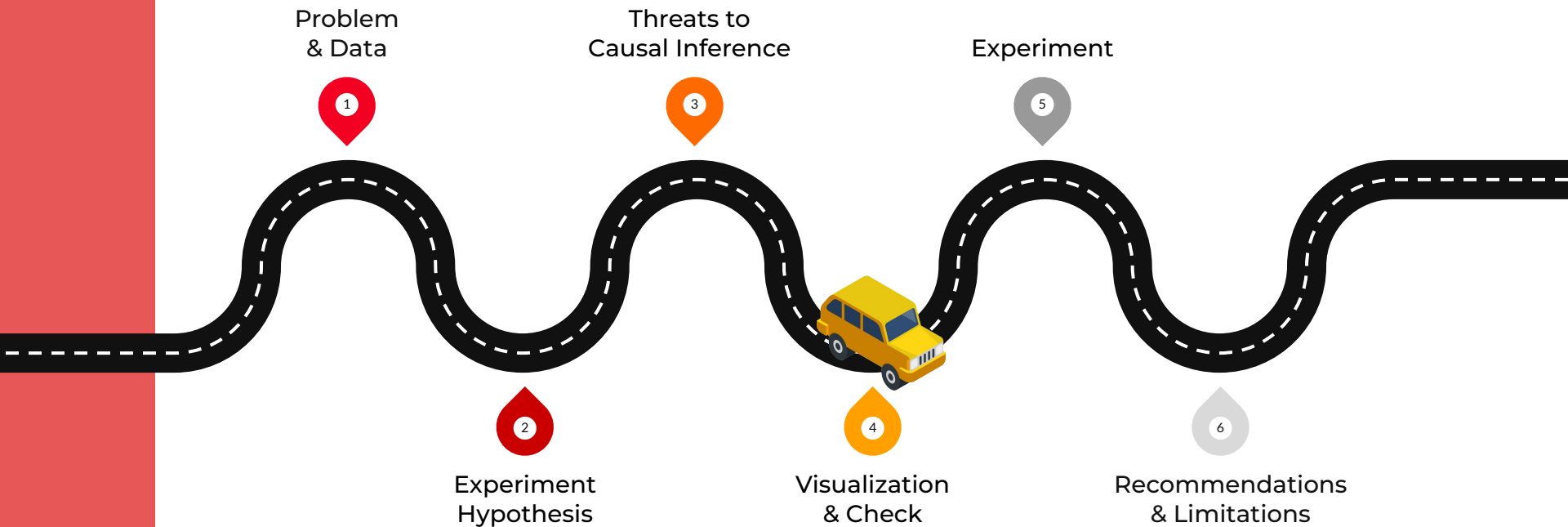**4**

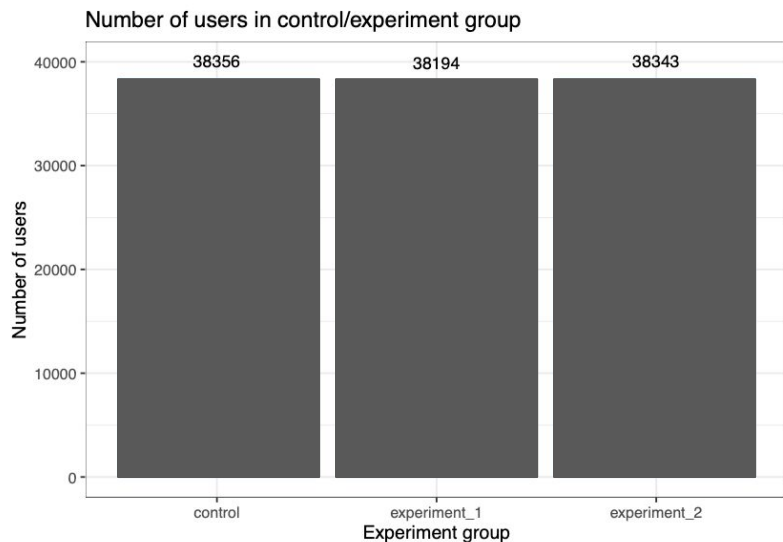No obvious simultaneity bias in this case

Simultaneity Bias

Roadmap

Problem
& Data

1

Threats to
Causal Inference

3

Experiment

5

2

Experiment
Hypothesis

4

Visualization
& Check

6

Recommendations
& Limitations

Visualization
Number of users in each group are equal

Number of users in control/experiment group

- ➔ Control group 38356 users
- ➔ Experiment group 1: 38194 users
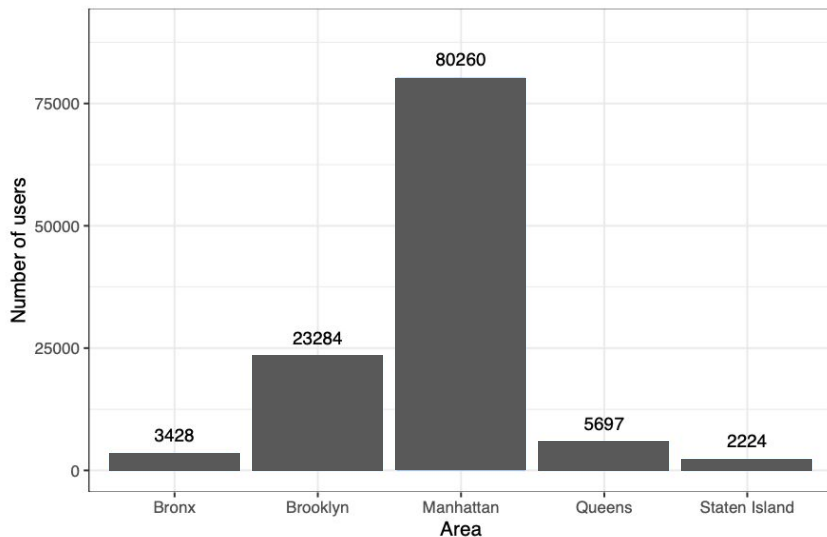- ➔ Experiment group 2: 38343 users
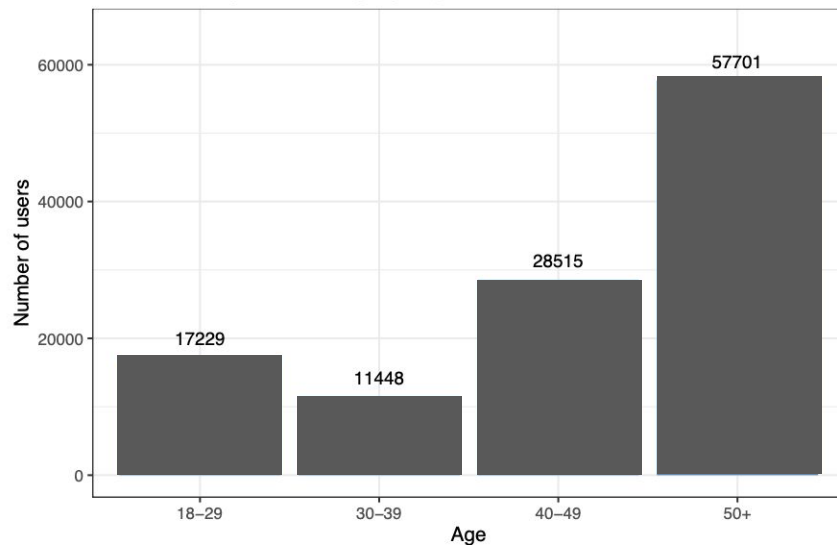
Visualization

Distribution of Variables

➜   Area across Bronx, Brooklyn, Manhattan, Queen, and Station Island
➜   Age group contains 18-29, 30-39, 40-49 and 50+

Number of users in each area

Number of logs in each age group

Visualization
Conversion Rate

Conversion Rate (search/open app)
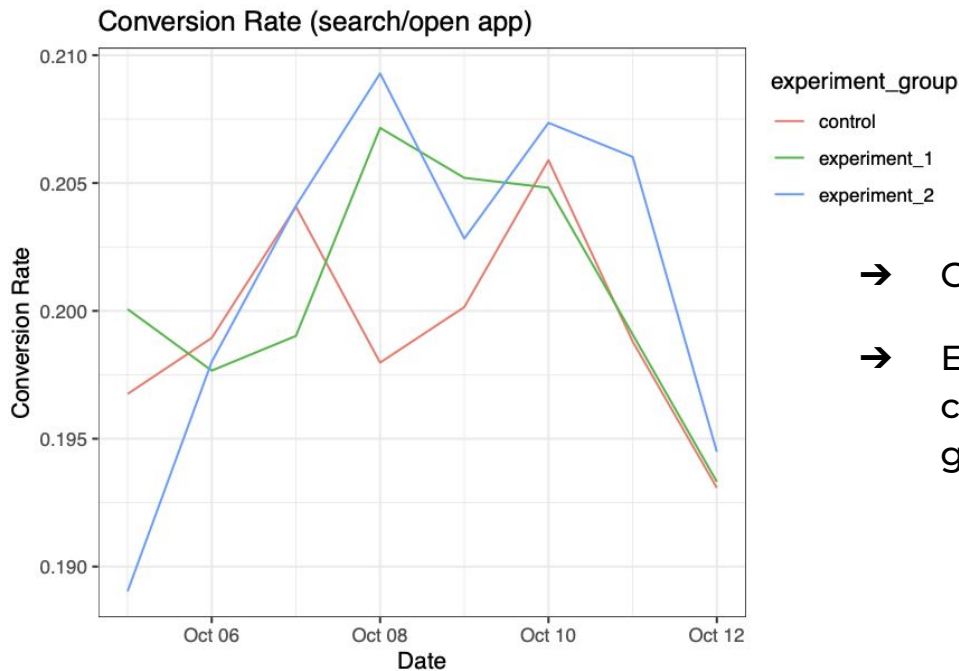
- ➔ Conversion rates are between 19% to 21%

- ➔ Experiment group 2 has a roughly higher conversion rate compared to other groups

* Conversion rate: the percentage of users open and search on the app divided by users open the app

Sanity Check

Use T test to check if there is an difference in age/neighbor between control and treatment group

➔   If p-value is greater than 0.05, two groups are similar.

➔   16/18 of test pass, **users in the control and treatment groups are randomized**

```
# Check randomization: control vs experiment 1
# User neighborhood
t.test(control$Bronx, experiment1$Bronx)


##
##   Welch Two Sample t-test
##
## data:  control$Bronx and experiment1$Bronx
## t = 0.49245, df = 76546, p-value = 0.6224
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.001804100  0.003014853
## sample estimates:
##  mean of x  mean of y
## 0.03011263 0.02950725
```
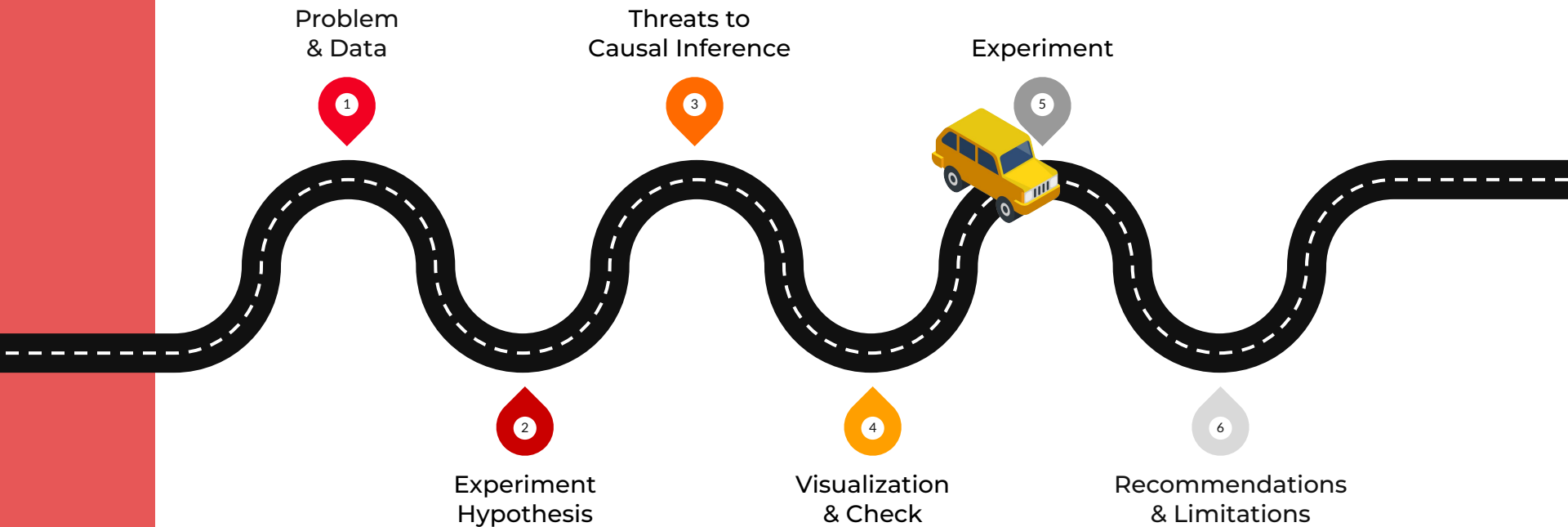
Sample Size Check

Expect to increase the conversion rate by 1% in the experiment

➔   The current sample size can only detect a difference of 0.8%

➔   Our experiment appears to be **underpowered to detect the effect**

```
power.t.test(n=38194, power=.8, sig.level=0.05, sd=0.4)
```

```
##
##      Two-sample t test power calculation
##
##               n = 38194
##           delta = 0.008115087
##              sd = 0.4
##       sig.level = 0.05
##           power = 0.8
##     alternative = two.sided
##
## NOTE: n is number in *each* group
```

Roadmap

Problem
& Data

1

Threats to
Causal Inference

3

Experiment

5

2

Experiment
Hypothesis

4

Visualization
& Check

6

Recommendations
& Limitations

Results of the Experiment
Hypothesis 1

Redesign the text on the booking button would lead to more conversions

```
t.test(control$search, experiment1$search, conf.level = 0.95)

##
##  Welch Two Sample t-test
##
## data:  control$search and experiment1$search
## t = -0.4468, df = 76545, p-value = 0.655
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.006968298  0.004381106
## sample estimates:
## mean of x mean of y
## 0.2000469 0.2013405
```

p-value > 0.05, failed to reject the null hypothesis

➔ The conversion has **no difference** between the control and experiment 1.

Results of the Experiment
Hypothesis 2

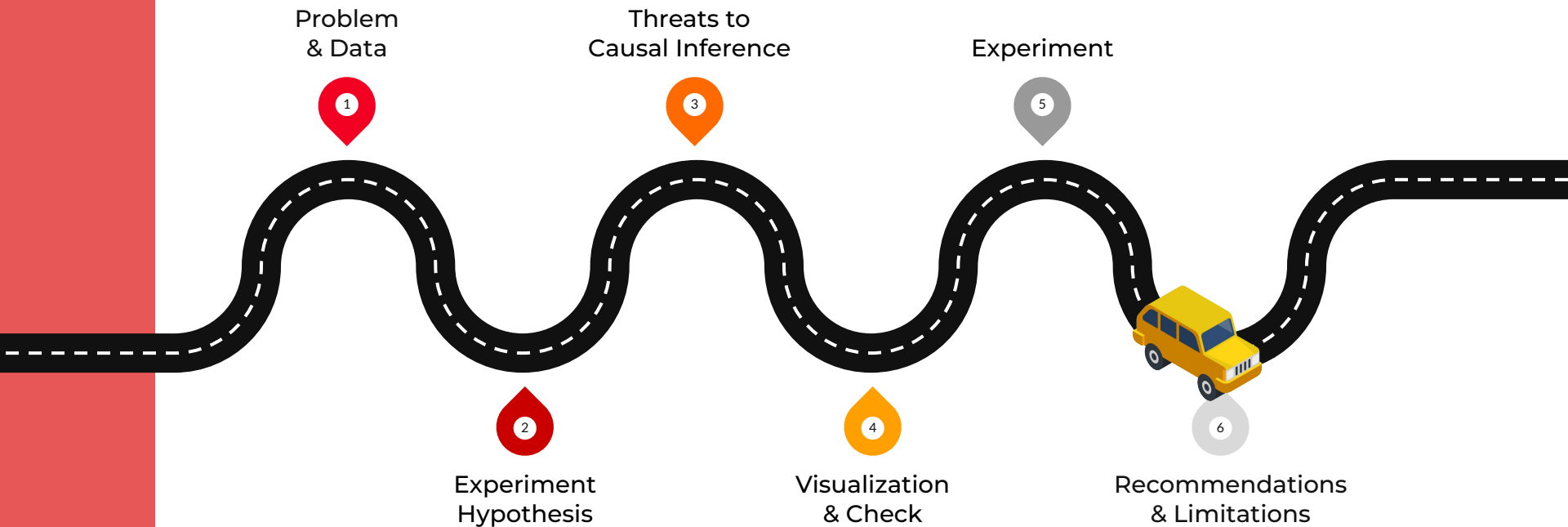Remove the "tip included" text would lead to more conversions

```
t.test(control$search, experiment2$search, conf.level = 0.95)

##
##  Welch Two Sample t-test
##
## data:  control$search and experiment2$search
## t = -0.90597, df = 76695, p-value = 0.365
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.008299864  0.003052460
## sample estimates:
## mean of x mean of y
## 0.2000469 0.2026706
```

p-value > 0.05, failed to reject the null hypothesis

➔ <u>The conversion has **no difference** between the control and experiment 2.</u>

Roadmap

Problem & Data

1

Threats to Causal Inference

3

Experiment

5

Experiment Hypothesis

2

Visualization & Check

4

Recommendations & Limitations

6

Recommendations

▶ **Don't change these two features on Flyber's UI**

▶ **Conduct Usability Testing to find out the root cause of poor conversion rate**

Limitations

⚠️   Funnel impact may take a long term to get into effect

⚠️   KPI requires long time periods and a very large sample to test for reliability

⚠️   Experiments are underpowered to detect the effect