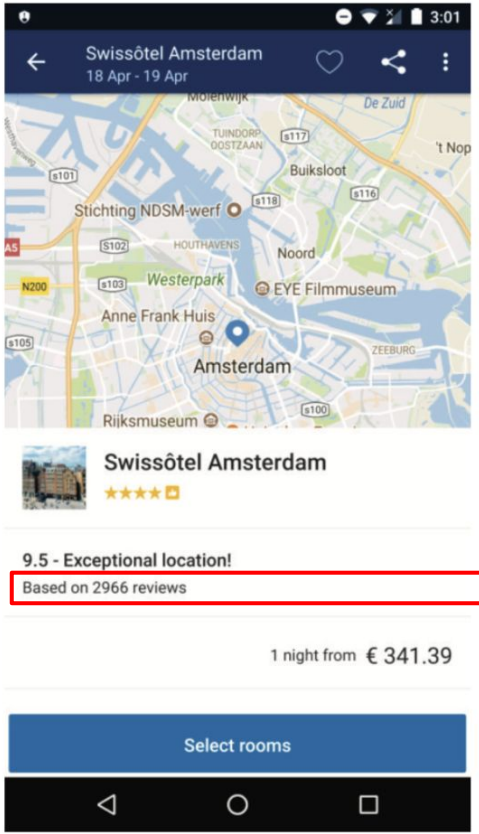
Booking.com

GitHub Repository: https://github.com/CinnyLin/dma

Cinny, Ellie, Jackie

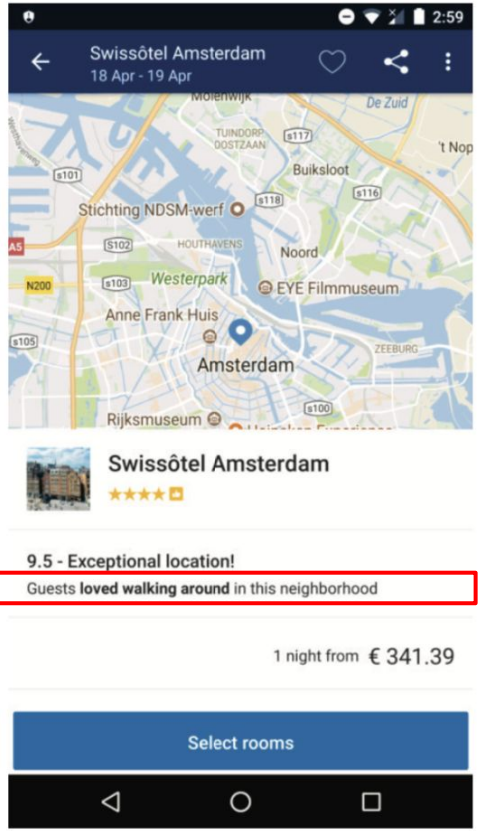
"A", The Control (Champion)

Shows current practice



"B", The Treatment (Challenger)

Adds walkability assessment (user reviews)



Booking.com

Executive Summary.

```
# set parameters
alpha = 0.05     # significance level
num_sides = 2    # one-sided=1 or two-sided=2 test
```

95%: z=1.96

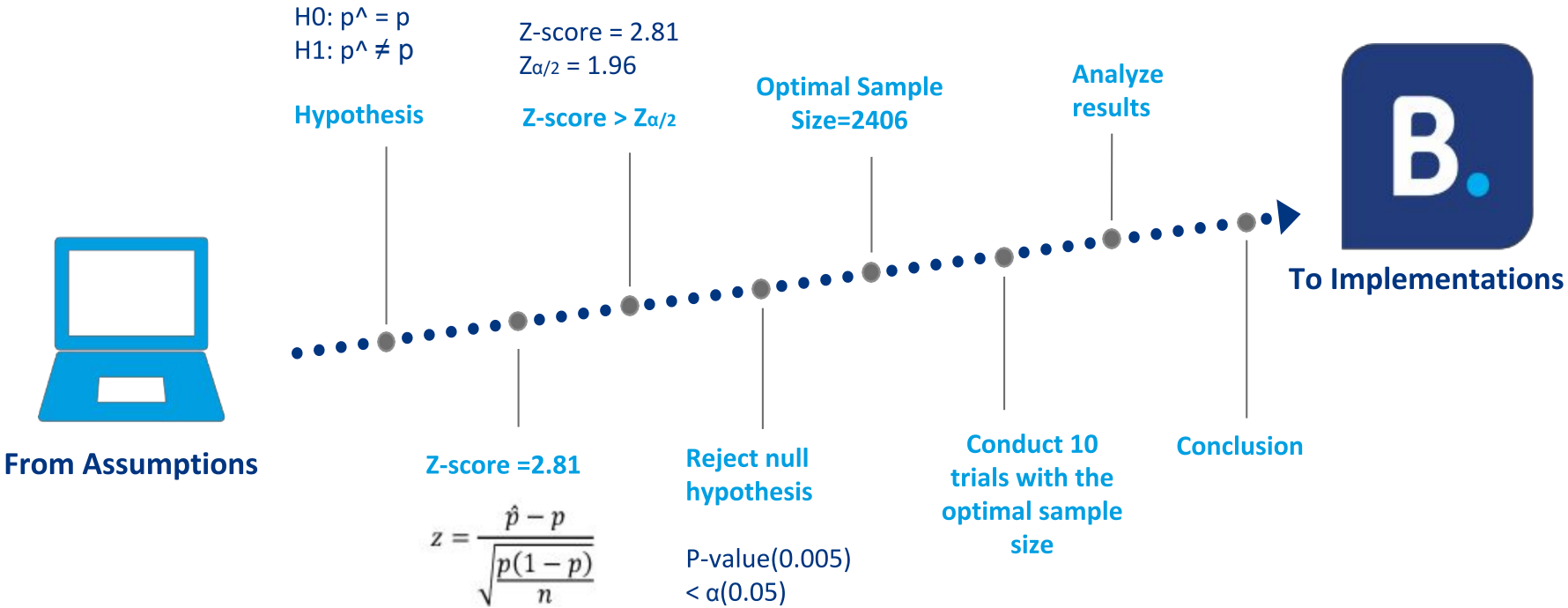
```
For month beginning 2020-08-01, Variant A had 5000 exposures (15.1%) and Variant B had 5000 exposures (16.5%)
Conversion rate for Variant A: 15.1%
Conversion rate for Variant B: 16.5%
Using all Variant B, reject null T/F?: True
z-score = 2.81 and p-value = 0.5%
```

The walkability assessment feature increase the conversion rate from 15.1% to 16.5% at 95% significance level.

Booking.com

Method Explanation.

p is the conversion rate for control A
p^ is the conversion rate for treatment B



Method Explanation.

Because the **variable is proportion** and we assume there is **large enough data** to use **normal approximation to binomial distribution**, we calculate the optimal sample size using the formula:

$$n^* = \left(t_{\alpha/2}\sqrt{2\bar{p}(1-\bar{p})} + t_{\beta}\sqrt{p_0(1-p_0)+p_1(1-p_1)}\right)^2 \frac{1}{\delta^2}$$

-Alpha is significance level of 0.05
-Power of the test is 0.80
-Desired minimum detectable effect(mde=P1-P0) is 0.03
-p0 is the conversion rate for control A
-p1 is the conversion rate for treatment B
-p_bar is the average of the two conversion rates
-$t_{\alpha/2}$ = 1.96
-t_{β} = 0.84

Our resulting optimal sample size is 2406.

Booking.com

Result Discussion.

For 10 samples of optimal sample size 2406, 40.00% rejected the null.

Trial	1	2	3	4	5	6	7	8	9	10
Z-score	1.2998	2.8292	2.2175	2.6253	2.3194	1.7077	0.5861	0.4842	0.4842	0.8920
P-value	0.1937	0.0050	0.0266	0.0087	0.0204	0.0877	0.5578	0.6283	0.6283	0.3724
Reject Null	False	True	True	True	True	False	False	False	False	False

Out of 130,000 data, we randomly pulled 2406 samples for both variant A and B and compare them for ten times. For 4 out of 10 times, we were able to obtain a p-value smaller than 0.05. Thus, we are able to reject the null hypothesis that there is no difference between the conversion rate of Test A and B.

Result Discussion.

What are potential barriers during real life implementation?

- Are there other potential external factors behind the scene? Such as the weather and the season at the time when the A/B Testing was conducted?
- In addition, though in short-term Variant B proves to produce a higher conversion rate, will such an effect remain in the long term?
- Is the sample from the data we collected representative of the whole population? We need more information on the demographics, psychographics, and perhaps even the purpose of why the users search for this particular location.
- Though we do obtaining an significant result, we also need to consider the time and effort it will take in applying the changes. What is the cost? How much time will it take?
- Variant B only shows a certain perspective of the location. There may be cases when people would rather scroll through comments and reach a conclusion on their own.