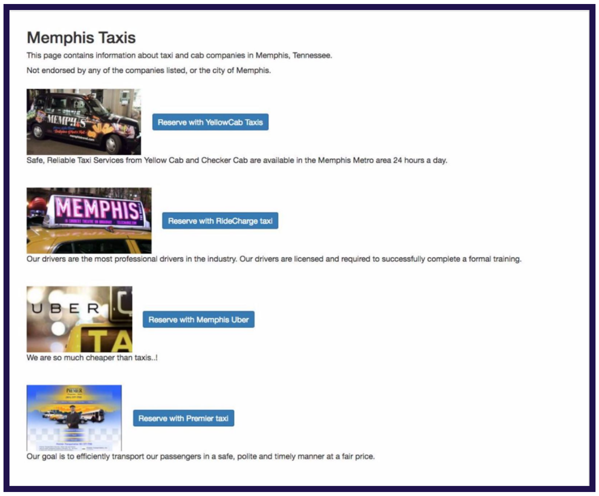
Memphis Taxis: A/B Testing & Eye Tracking

Background

We designed two different versions of Memphis Taxis' website and decided to collect data from A/B and eye tracking tests to determine the version with better user experience.



Memphis Taxis

This page contains information about taxi and cab companies in Memphis, Tennessee.

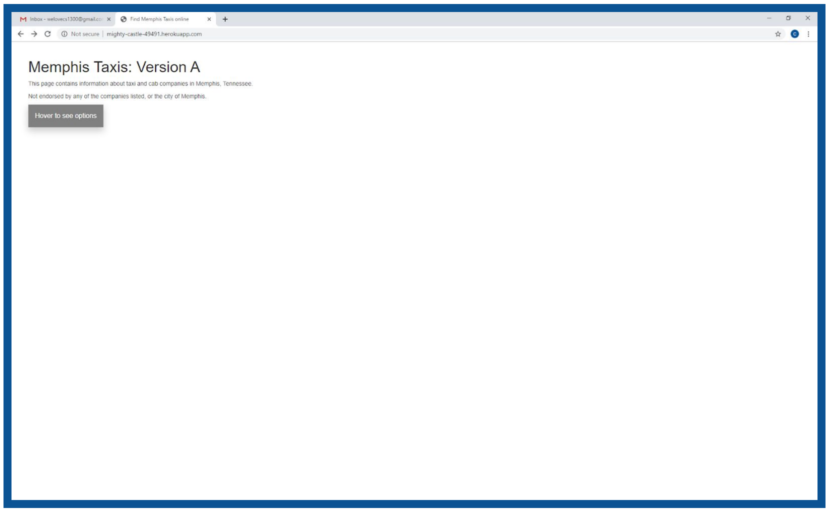
Not endorsed by any of the companies listed, or the city of Memphis.

Safe, Reliable Taxi Services from Yellow Cab and Checker Cab are available in the Memphis Metro area 24 hours a day.

Our drivers are the most professional drivers in the industry. Our drivers are licensed and required to successfully complete a formal training.

We are so much cheaper than taxis..!

Our goal is to efficiently transport our passengers in a safe, polite and timely manner at a fair price.



Inbox - welovecs1300@gmail.co
Find Memphis Taxis online
Not secure
mighty-castle-49491.herokuapp.com

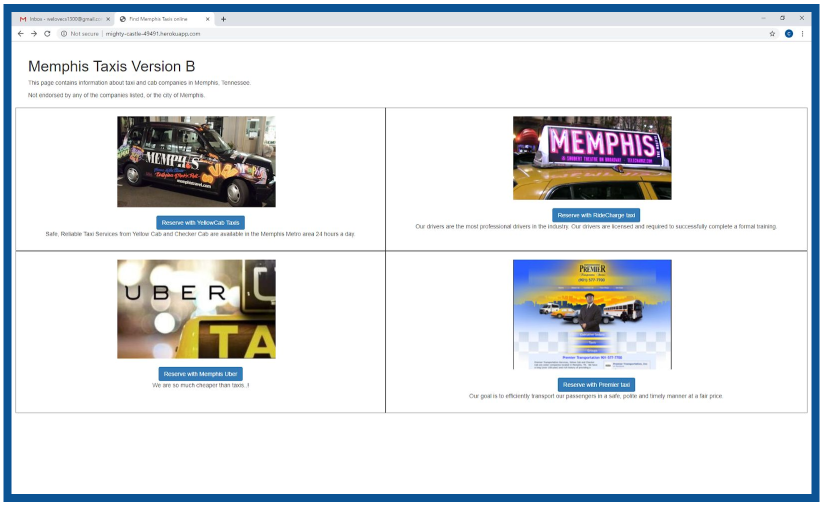
Memphis Taxis: Version A

This page contains information about taxi and cab companies in Memphis, Tennessee.

Not endorsed by any of the companies listed, or the city of Memphis.

Version A

Version B



Inbox - welovecs1300@gmail.co
Find Memphis Taxis online
Not secure
mighty-castle-49491.herokuapp.com

Memphis Taxis Version B

This page contains information about taxi and cab companies in Memphis, Tennessee.

Not endorsed by any of the companies listed, or the city of Memphis.

Reserve with YellowCab Taxis

Safe, Reliable Taxi Services from Yellow Cab and Checker Cab are available in the Memphis Metro area 24 hours a day.

Reserve with RideCharge taxi

Our drivers are the most professional drivers in the industry. Our drivers are licensed and required to successfully complete a formal training.

We are so much cheaper than taxis..!

Our goal is to efficiently transport our passengers in a safe, polite and timely manner at a fair price.

A/B Testing

Hypotheses

	Click through rate	Time to click	Dwell time	Return rate
Null	A and B have the same CTR.	A and B have the same TTC.	A and B have the same DT.	A and B have the same RR.
Alternative	A has higher CTR because there is less information and the user has to make at least one click to see more information	B will have a longer TTC because there is more information presented at the beginning.	A has a longer DT because you enter the second page with less information than B.	A has a higher RR because the user has to go back to the landing page to learn more about other options.

Test Results

We gathered data from 61 unique users who interacted with our site where each was randomly routed to Version A/B (https://mighty-castle-49491.herokuapp.com/).

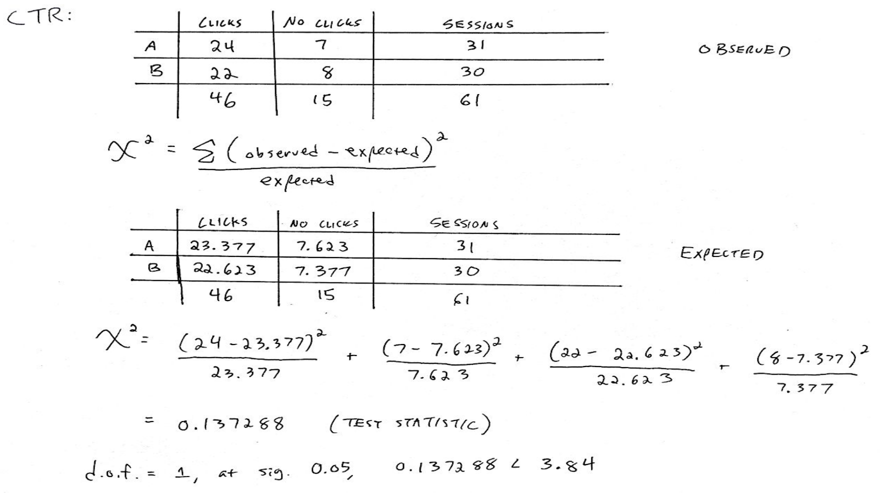
Click Through Rate

We wrote a Python script that first splits sessions A and B, finds # of unique sessions for each variant, and counts # of unique clicks out of unique sessions, ignoring subsequent clicks of session. Then we divide # of unique clicks over # of unique sessions:

CTR of A: $\frac{24}{31} = 0.77$

CTR of B: $\frac{22}{30} = 0.73$

To test for significance, we used a one-sided Chi-Squared test since we have two categorical variables that we are trying to determine a significant association for.



CTR:

	CLICKS	NO CLICKS	SESSIONS
A	24	7	31
B	22	8	30
	46	15	61

OBSERVED

$$\chi^2 = \frac{\Sigma(\text{observed} - \text{expected})^2}{\text{expected}}$$

	CLICKS	NO CLICKS	SESSIONS
A	23.377	7.623	31
B	22.623	7.377	30
	46	15	61

EXPECTED

$$\chi^2 = \frac{(24-23.377)^2}{23.377} + \frac{(7-7.623)^2}{7.623} + \frac{(22-22.623)^2}{22.623} + \frac{(8-7.377)^2}{7.377}$$

$= 0.137288$ (TEST STATISTIC)

d.o.f. = 1, at sig. 0.05, $0.137288 < 3.84$

The test statistic is 0.14, which is lower than the critical value of 3.84 for a significance level of 0.05 at 1 degree of freedom, so the P-value associated with test statistic is greater than 0.05. Therefore, the null hypothesis is not rejected. There is insufficient evidence that A has higher CTR than B.

Time to click

We split sessions A and B and find the average time it takes each session to make their first click on the page:

Average TTC(A): $\frac{296.14}{24} = 12.34\ s$

Average TTC(B): $\frac{338.717}{20} = 16.94\ s$

We then use an independent samples t-test, since we have two distinct sample means (average times for A and B) and we are trying to see whether they represent the same population. Denoting x as the time to first click (in seconds), we know:

	$\sum_{i=1}^{n} x$: sum of differences	$\sum_{i=1}^{n} x^2$: sum of differences squared	n : unique sessions with clicks
A	296.14	13179.5152	24
B	338.717	25704.4305	20

We find our t-statistic is -0.57:

$$\bar{x}_A - \bar{x}_B = -4.597$$

$$(n_A - 1)s_A^2 = \sum_{i=1}^{n_A} x_A^2 - \frac{\left(\sum_{i=1}^{n} x_A\right)^2}{n_A} = 13179.5152 - \frac{(296.14)^2}{24} = 9525.394$$

$$(n_B - 1)s_B^2 = \sum_{i=1}^{n_B} x_B^2 - \frac{\left(\sum_{i=1}^{n} x_B\right)^2}{n_B} = 25704.4305 - \frac{(338.717)^2}{20} = 19967.97$$

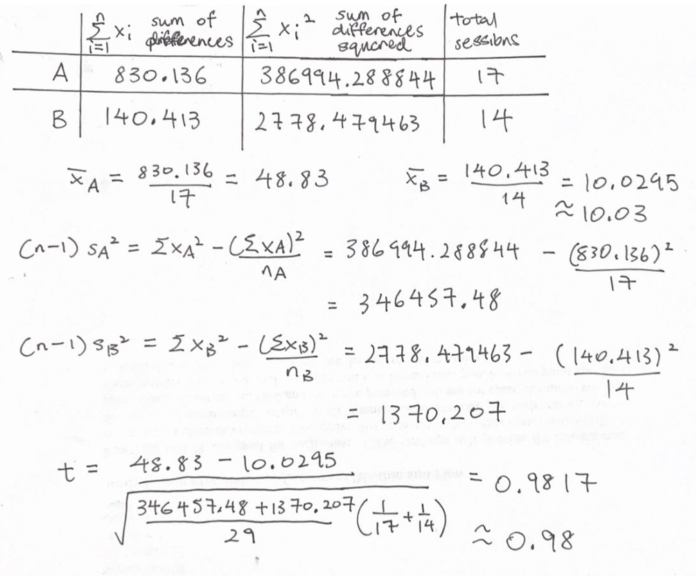
$$t = \frac{\bar{x}_A - \bar{x}_B}{\sqrt{\frac{(n_A-1)s_A^2 + (n_B-1)s_B^2}{n_A + n_B - 2}\left(\frac{1}{n_A} + \frac{1}{n_B}\right)}} = \frac{-4.597}{\sqrt{\frac{9525.394 - 19967.97}{42}\left(\frac{1}{24} + \frac{1}{20}\right)}} = \boxed{-0.573}$$

The critical value for a t-test at a significance level of 0.05 with $24 + 20 - 2 = 42$ degrees of freedom is 1.684. Since $|-0.57| < 1.684$, this means that the P-value associated with t is greater than 0.05. Therefore, the null hypothesis is not rejected. There is insufficient evidence that the time to first click for B is greater than A.

Dwell time

We split sessions A and B and find the average time it takes for each session to return back to the landing page (if at all).

Average DT(A): 48.83 s
Average DT(B): 10.03 s



	$\sum_{i=1}^{n} x_i$ sum of differences	$\sum_{i=1}^{n} x_i^2$ sum of differences squared	total sessions
A	830.136	386994.28844	17
B	140.413	2778.479463	14

$$\bar{x}_A = \frac{830.136}{17} = 48.83 \qquad \bar{x}_B = \frac{140.413}{14} = 10.0295 \approx 10.03$$

$$(n-1)s_A^2 = \Sigma x_A^2 - \frac{(\Sigma x_A)^2}{n_A} = 386994.288844 - \frac{(830.136)^2}{17} = 346457.48$$

$$(n-1)s_B^2 = \Sigma x_B^2 - \frac{(\Sigma x_B)^2}{n_B} = 2778.479463 - \frac{(140.413)^2}{14} = 1370.207$$

$$t = \frac{48.83 - 10.0295}{\sqrt{\frac{346457.48 + 1370.207}{29}\left(\frac{1}{17} + \frac{1}{14}\right)}} = 0.9817 \approx 0.98$$

Again, we chose to use an independent samples t-test, since we have two distinct sample means (dwell times for A and B) and we are trying to see whether they represent the same population.

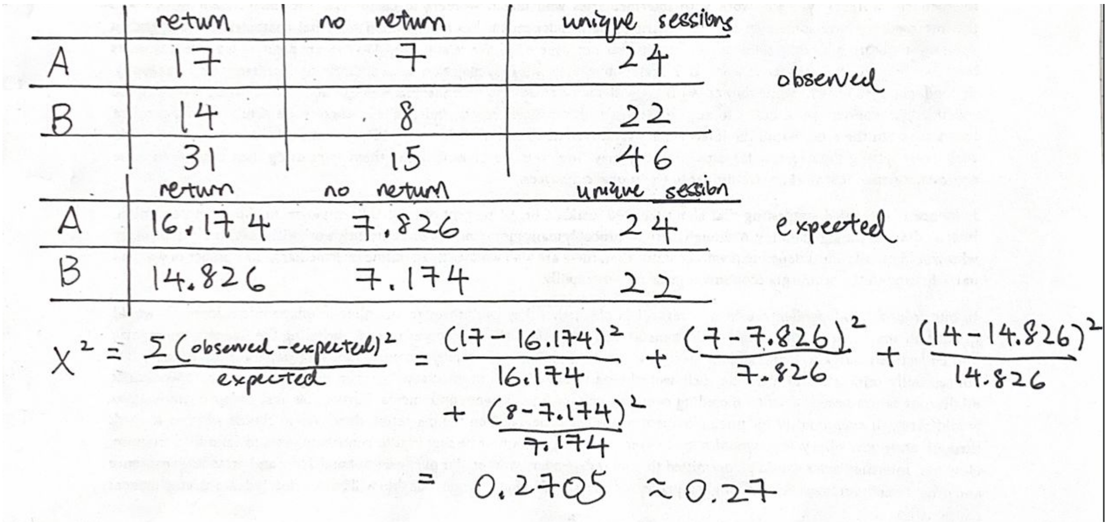
The t-statistic is 0.98, which is lower than the critical value of 1.699 for a t-test at a significance of 0.05 with 29 degrees of freedom. Therefore, the null hypothesis is not rejected. There is insufficient evidence that the dwell-time for A is longer than B.

Return rate

We split sessions A and B and find the proportion of sessions that return to the landing page.

RR(A): $\frac{17}{24}$ = 0.71
RR(B): $\frac{14}{22}$ = 0.64

To test for significance, we use a one-sided Chi-Squared test for the same reason as CTR.



	return	no return	unique sessions
A	17	7	24
B	14	8	22
	31	15	46

observed

	return	no return	unique session
A	16.174	7.826	24
B	14.826	7.174	22

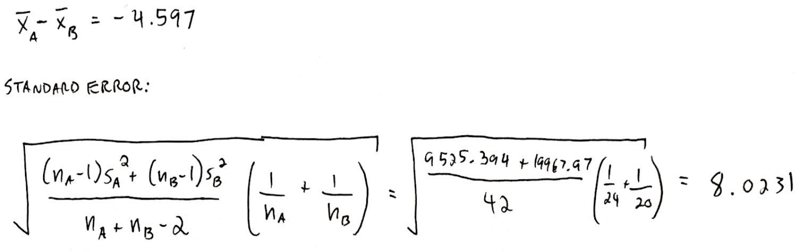
expected

$$X^2 = \frac{\Sigma(\text{observed} - \text{expected})^2}{\text{expected}} = \frac{(17-16.174)^2}{16.174} + \frac{(7-7.826)^2}{7.826} + \frac{(14-14.826)^2}{14.826} + \frac{(8-7.174)^2}{7.174}$$

$$= 0.2705 \approx 0.27$$

The test statistic is 0.27, which is lower than the critical value of 3.84 for a significance level of 0.05 at 1 degree of freedom, so the P-value associated with test statistic is greater than 0.05. Therefore, the null hypothesis is not rejected. There is insufficient evidence that A has a higher return rate than B.

Confidence Interval for Average TTC



$$\bar{X}_A - \bar{X}_B = -4.597$$

STANDARD ERROR:

$$\sqrt{\frac{(n_A-1)s_A^2 + (n_B-1)s_B^2}{n_A+n_B-2}\left(\frac{1}{n_A}+\frac{1}{n_B}\right)} = \sqrt{\frac{9525.394 + 19967.97}{42}\left(\frac{1}{24}+\frac{1}{20}\right)} = 8.0231$$

The critical value we use here is from a 2-tailed test instead of 1-tailed, since we are computing the confidence interval.

CI = - 4.597 ± 2.018 (8.0231) = - 4.60 ± 16.19 = [-20. 79, 11.59]

Eye Tracking

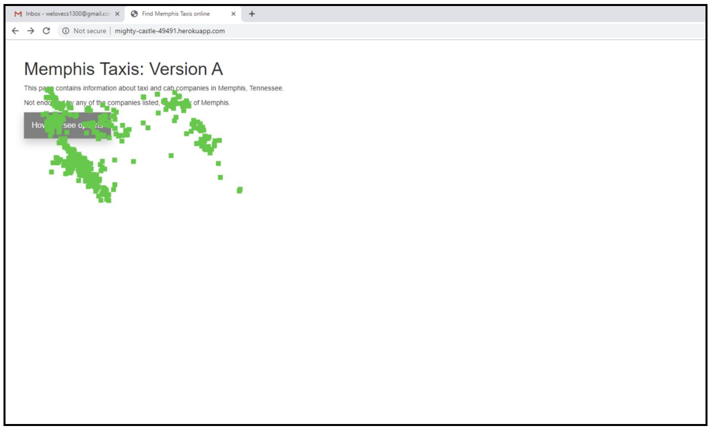
Qualitative Hypothesis

A will have a greater proportion of eye-gazes towards the top-left of the screen than B, since the drop-down menu is located at the top-left and is the only component on A, while B will have a greater proportion of gazes near the images/buttons.

Test Results

We conducted eye tracking tests on two users for the two different versions and generated the following replay screenshots and heatmaps of their eye movements.

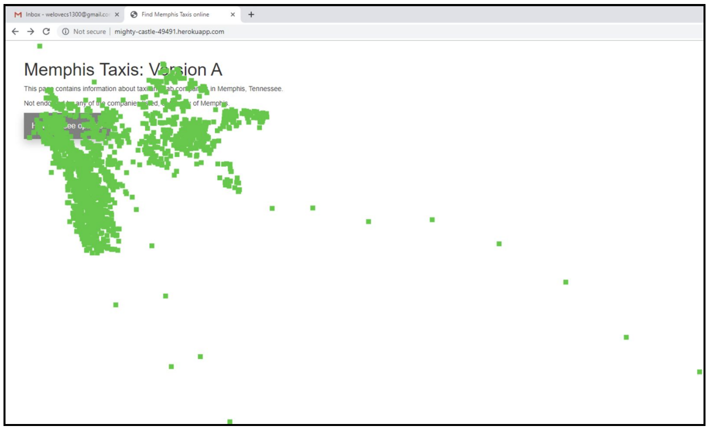
Eye Movement Screenshots


Inbox - welovecs1300@gmail.co
Find Memphis Taxis online
Not secure
mighty-castle-49491.herokuapp.com
How to see options

Memphis Taxis: Version A

This page contains information about taxi and cab companies in Memphis, Tennessee.

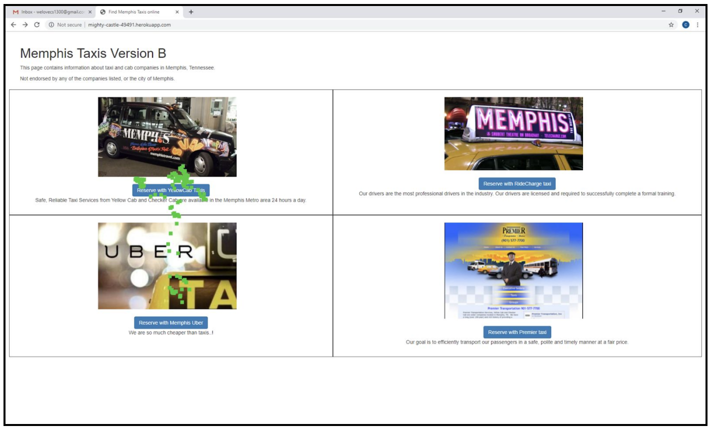
Not endorsed by any of the companies listed, or the city of Memphis.


Inbox - welovecs1300@gmail.co
Find Memphis Taxis online
Not secure
mighty-castle-49491.herokuapp.com
How to see options

Memphis Taxis: Version A

This page contains information about taxi and cab companies in Memphis, Tennessee.

Not endorsed by any of the companies listed, or the city of Memphis.


Inbox - welovecs1300@gmail.co
Find Memphis Taxis online
Not secure
mighty-castle-49491.herokuapp.com
MEMPHIS
UBER
TA
PREMIER
Reserve with YellowCab Taxis
Reserve with RideCharge taxi
Reserve with Memphis Uber
Reserve with Premier taxi

Memphis Taxis Version B

This page contains information about taxi and cab companies in Memphis, Tennessee.

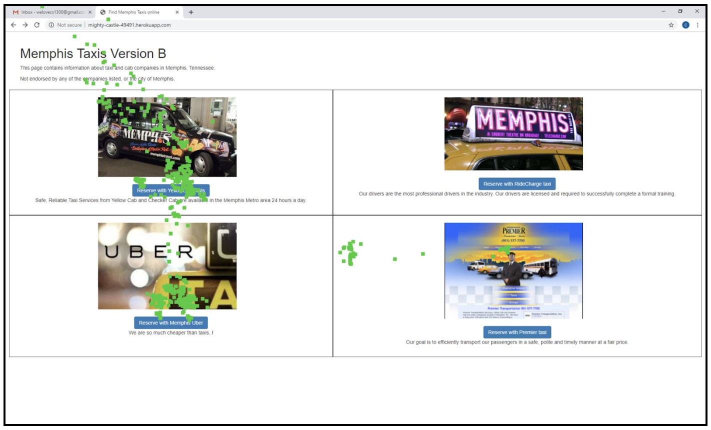
Not endorsed by any of the companies listed, or the city of Memphis.

Safe, Reliable Taxi Services from Yellow Cab and Checker Cab are available in the Memphis Metro area 24 hours a day.

Our drivers are the most professional drivers in the industry. Our drivers are licensed and required to successfully complete a formal training.

We are so much cheaper than taxis..!

Our goal is to efficiently transport our passengers in a safe, polite and timely manner at a fair price.


Inbox - welovecs1300@gmail.co
Find Memphis Taxis online
Not secure
mighty-castle-49491.herokuapp.com
MEMPHIS
UBER
TA
PREMIER
Reserve with YellowCab Taxis
Reserve with RideCharge taxi
Reserve with Memphis Uber
Reserve with Premier taxi

Memphis Taxis Version B

This page contains information about taxi and cab companies in Memphis, Tennessee.

Not endorsed by any of the companies listed, or the city of Memphis.

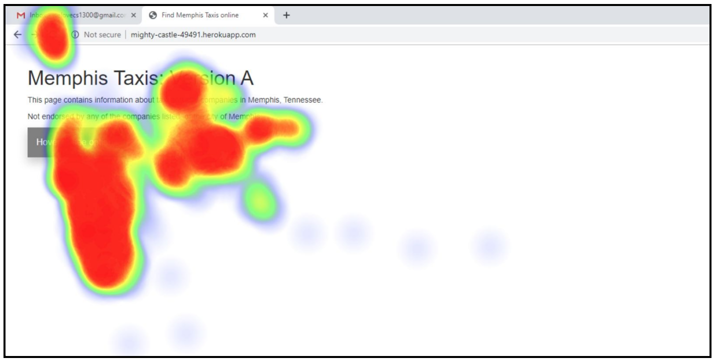
Safe, Reliable Taxi Services from Yellow Cab and Checker Cab are available in the Memphis Metro area 24 hours a day.

Our drivers are the most professional drivers in the industry. Our drivers are licensed and required to successfully complete a formal training.

We are so much cheaper than taxis..!

Our goal is to efficiently transport our passengers in a safe, polite and timely manner at a fair price.

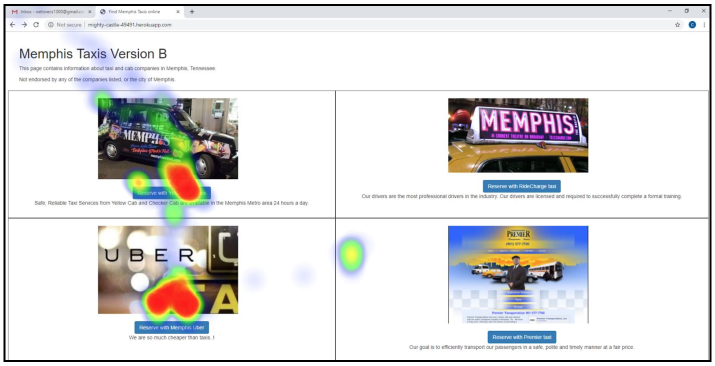
Heatmaps


Find Memphis Taxis online
Not secure | mighty-castle-49491.herokuapp.com
Hov

Memphis Taxis: Version A

This page contains information about t companies in Memphis, Tennessee.

Not endorsed by any of the companies listed city of Memphis.


Inbox - welovecs1300@gmail.co
Find Memphis Taxis online
Not secure | mighty-castle-49491.herokuapp.com
MEMPHIS
UBER
PREMIER
Reserve with RideCharge taxi
Reserve with Memphis Uber
Reserve with Premier taxi

Memphis Taxis Version B

This page contains information about taxi and cab companies in Memphis, Tennessee.

Not endorsed by any of the companies listed, or the city of Memphis.

Safe, Reliable Taxi Services from Yellow Cab and Checker Cab are available in the Memphis Metro area 24 hours a day.

Our drivers are the most professional drivers in the industry. Our drivers are licensed and required to successfully complete a formal training.

We are so much cheaper than taxis..!

Our goal is to efficiently transport our passengers in a safe, polite and timely manner at a fair price.

We can see that our original hypothesis is correct; our heatmap for A was far more concentrated on the left side/top portion of the screen compared to B, since the user had to hover over the bar to view options on A. On B, we see that the gazes are disjointed and concentrated near the pictures.

Comparison

1. As a taxi aggregator we have two sources of revenue: Ad-revenue from our website and commission per redirect to a taxi website. Although the below results are statistically insignificant, they give us a sense of the best way forward.
    - **Time to click:** We see that Version B is better than A because the TTC on B is 37% more than that on A. This makes version B best suited to maximize ad-revenue.
    - **Click through Rate:** We see that version B has a CTR which is just 5.2% less than that of version A. This barely makes A the better candidate to maximize commision based revenue.

   If we look at the eye tracking screenshots we see that version B has two distinct hotspots which gives us more spots to place ads on the screen. Therefore after considering all factors, it seems that the best way forward for us is to maximize ad-revenue by adopting version B.
2. While A/B testing gives us solid quantitative data, eye tracking data can actually help us infer more about the user's interaction with the screen based on their heatmap/replay animations. For example, we saw from the heatmap that the user did not even glance at the RideCharge Taxi option in the top right of the grid, which no data points from our A/B testing could indicate. The main tradeoff between these two methods is that eye tracking data is far more difficult to collect, as it requires expensive equipment and can only be conducted in a lab setting with one user at a time.

3. One metric that can be used unethically is click-through-rate. Websites might try to maximize click-through-rate in order to have more pages that a user must visit which allows the site to host more advertisements, instead of putting all relevant information on one page. Monthly active users is another metric that can be used unethically. Focusing on maximizing monthly active users could mean the website is inclined to use potentially sensitive data to gain insight into your interests and habits and reel you back to the website using that information.