Akanksha Arun- Christian Ingul - Juhil Ahir - Lavanya Deshmukh
Paul Nguyen - Sagar Patel

Agenda

01 What are we seeking to answer?

02 What is our Hypothesis?

03 What is the ideal experiment to examine the issue?

04 Data cleaning and data manipulation

05 Describe the data and highlight hot topic

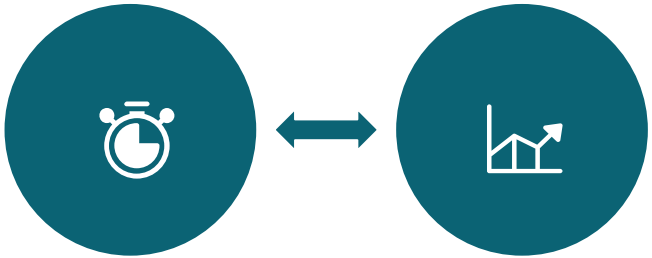
06 Provide details and results of our main analysis method.

Why would a data-driven approach be useful to obtain insights regarding ad click-through rates?

Marketing is often one of the biggest expenses a business can incur. Businesses can use a data-driven approach to optimize their advertising strategies and expenditures. It can help them single out high-performing ads and curb spending on ads with low ROI. If businesses can understand what causes a customer to convert, they can optimize their ad plan accordingly.

The Hypothesis

Time Spent on Ad vs Click Through Rate



Null Hypothesis
Time spent on website does NOT lead to better clicked on ad rates.

Alternate Hypothesis
Time spent on website does lead to better clicked on ad rates.

The Ideal Experiment

Impact of Ad Content

Analyze how the average time spent on ads differs depending on the ad's content

Impact of Buzz-Words

Analyze whether specific words lead to higher click-through rates.

A/B Test for Time Spent

An ideal experiment for this scenario would be A/B testing on time spent on the website.

The Goal

To determine the ideal amount of time users need to spend on advertisements to lead to a conversion

Data Cleaning and Data Manipulation

01 Blended datasets for ‘Continents’ based on existing ‘Country’ column

02 Creating Dummy Variables to Handle Categorical Data

03 Dividing Data about ‘Age’ , ‘Time of Day’ into Buckets

04 Adding a column for Ad Length based on existing ‘Ad Topic’ column

CTR Prediction Data Breakdown

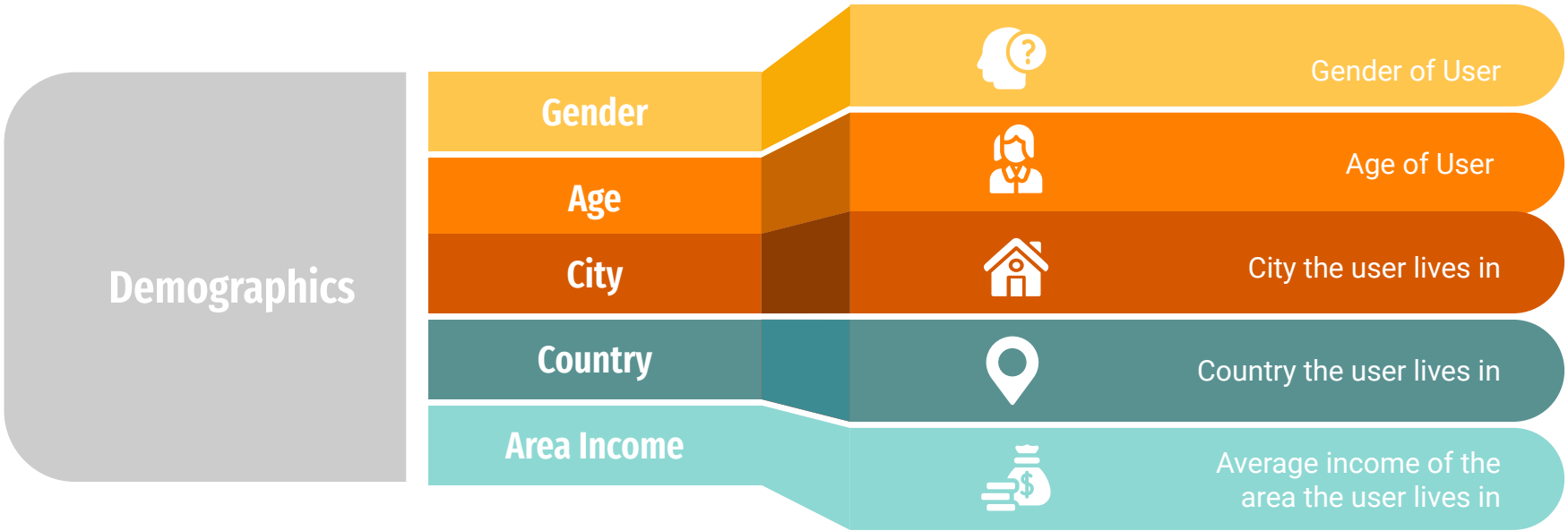
Behaviour on Website

- **Daily Time Spent on Site** — The amount of time spent on site (minutes)
- **Daily Internet Usage** — The amount of time spent on the internet (minutes)
- **Ad Topic Line** — The topic line ad on the website
- **Time Stamp** — The time that the person entered the website

Outcome

- **Clicked on Ad** — A binary value for whether or not the person clicked on the ad

CTR Prediction Data Breakdown (cont.)



Descriptive Statistics at a glance..

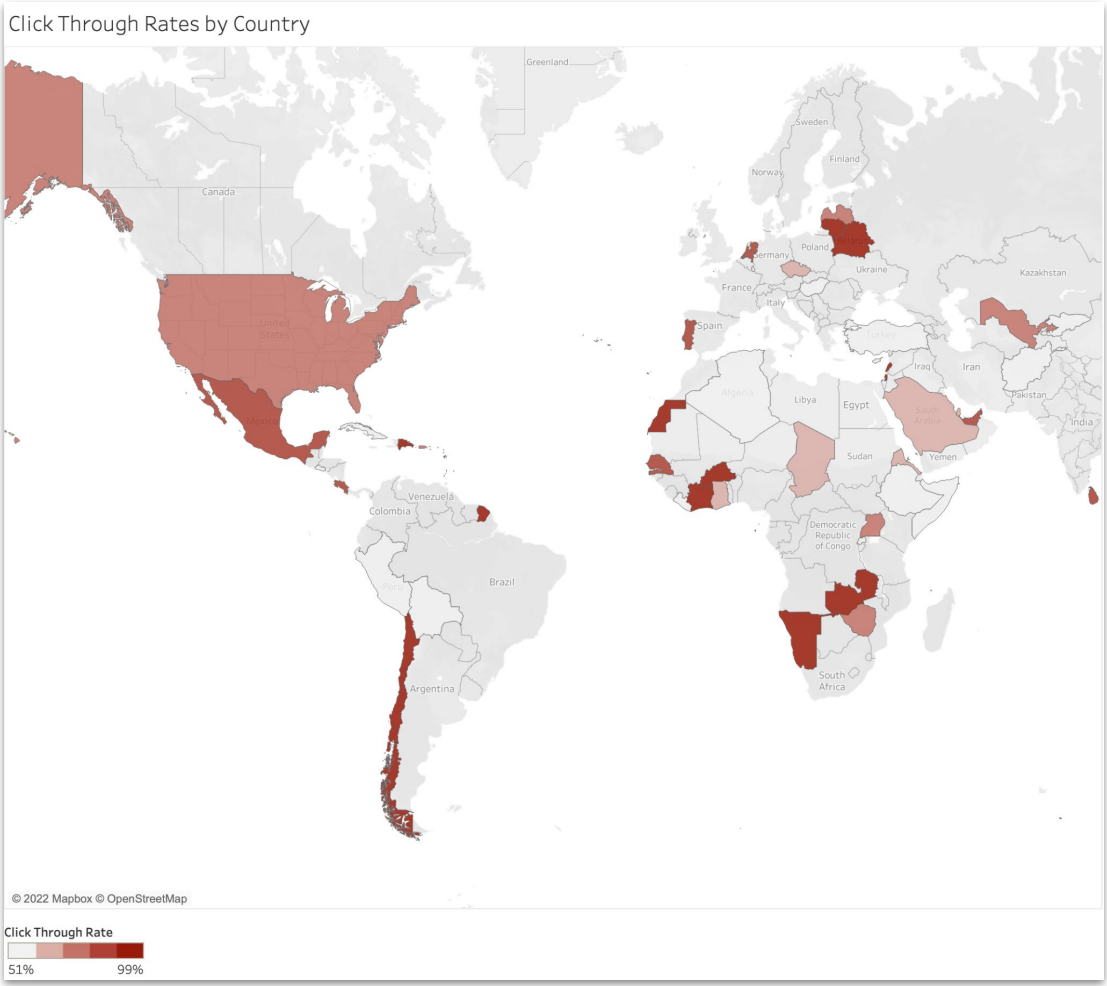
	Daily Time Spent on Site	Age	Area Income	Daily Internet Usage	Clicked on Ad
count	50000.000000	50000.000000	50000.000000	50000.000000	50000.000000
mean	64.133188	35.835400	53927.726446	174.776637	0.497640
std	14.842587	8.865057	11413.630035	42.023941	0.499999
min	32.600000	19.000000	13996.500000	104.780000	0.000000
25%	49.840000	29.000000	47575.440000	136.180000	0.000000
50%	66.630000	34.000000	55993.680000	167.860000	0.000000
75%	76.440000	41.000000	63100.130000	213.750000	1.000000
max	91.370000	60.000000	79332.330000	269.960000	1.000000

Avg Daily Time Spent

Avg Age

Clicked on Ad

Investigating the Data through Visualizations



Questions we aimed to answer:

1. **Which countries are generating the highest click through rate?**

Key Insights:

Highest clicked on ad rate:

North-America: The U.S

South-America: Chile & French Guiana

Africa: Burkina Faso & Zimbabwe

Europe: Lithuania & Portugal

Asia: UAE

Clicker Demographics

Gender	Continent	Avg. Age	Avg. Area Income	Avg. Daily Time Spent on Site
Female	Africa	38	$51,916.96	64
	Americas	37	$54,778.64	64
	Asia	37	$54,297.27	63
	Australia	33	$54,506.44	65
	Europe	38	$54,133.92	64
Male	Africa	38	$53,222.71	63
	Americas	36	$53,642.27	64
	Asia	38	$53,113.35	63
	Australia	31	$53,859.93	65
	Europe	39	$55,673.78	62

Questions we aimed to answer:

1. **What does an average clicker look like for these countries?**

Key Insights:

For Male and Female across the different continents **our clicker demographics looks fairly similar.**

Time of Day Analysis

Time of Day..	Click Through Rate	Total Ads
Morning	46%	15,145
Afternoon	38%	10,105
Evening	61%	12,280
Night	53%	12,470

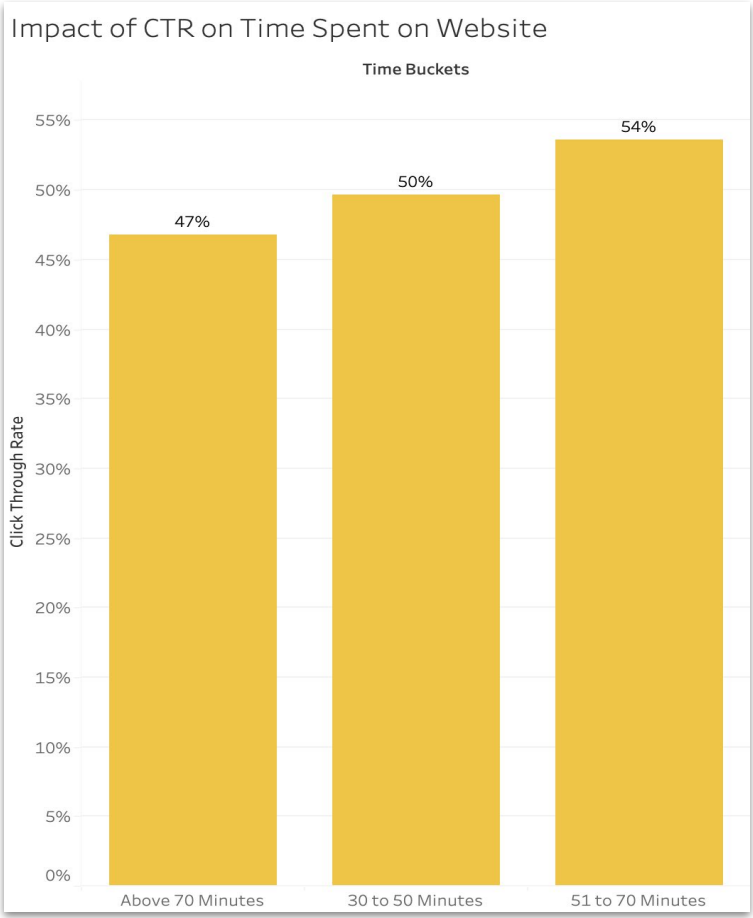
Morning : 6am -11:59 am
Afternoon: 12pm - 4:59 pm
Evening: 5pm - 8:59 pm
Night: 9pm onwards

Questions we aimed to answer:

1. **What time of the day are clickers clicking?**

Key Insights:

Evening time saw the the highest click through rate and Afternoon the lowest.



Questions we aimed to answer:

1. How long does someone spend on the website before they click on an ad?

Key Insights:

- The CTR declines, or in general does not seem to vary significantly based on the time spent on the website.

- The marketing department can use a different strategy to increase the clicked on rate.

Regression Model 1

OLS Regression Results

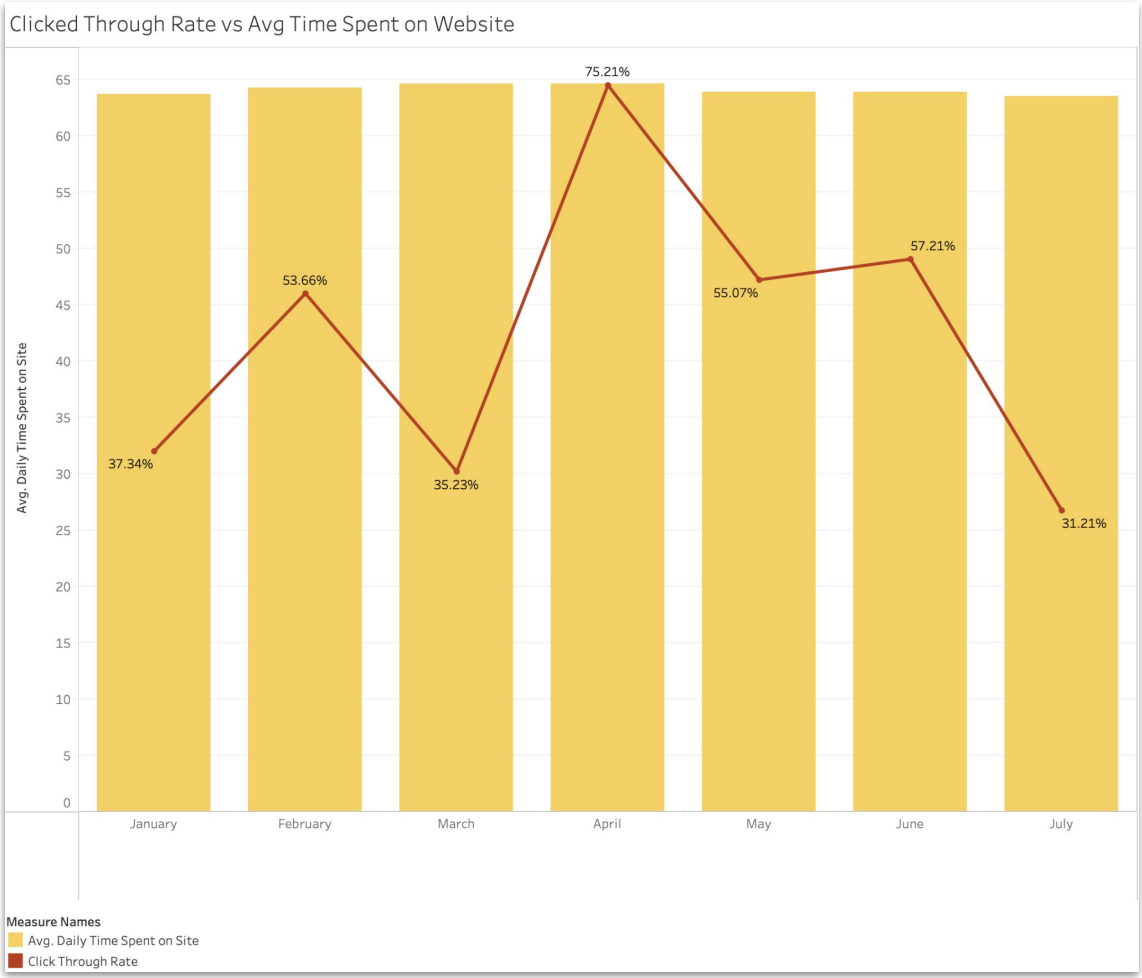
Dep. Variable:	Clicked on Ad New	R-squared:	0.256
Model:	OLS	Adj. R-squared:	0.256
Method:	Least Squares	F-statistic:	1432.
Date:	Mon, 28 Nov 2022	Prob (F-statistic):	0.00
Time:	00:40:48	Log-Likelihood:	-28904.
No. Observations:	50000	AIC:	5.783e+04
Df Residuals:	49987	BIC:	5.795e+04
Df Model:	12		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-0.1714	0.019	-8.801	0.000	-0.210	-0.133
Daily Time Spent on Site	-0.0012	0.000	-9.439	0.000	-0.001	-0.001
Age	0.0244	0.000	108.152	0.000	0.024	0.025
Area Income	-4.676e-07	1.7e-07	-2.753	0.006	-8.01e-07	-1.35e-07
Daily Internet Usage	-0.0015	4.69e-05	-32.804	0.000	-0.002	-0.001
Ad classification	0.0158	0.004	3.883	0.000	0.008	0.024
Africa	0.1632	0.007	22.312	0.000	0.149	0.178
Americas	0.0728	0.007	10.064	0.000	0.059	0.087
Antartica	-0.2317	0.080	-2.883	0.004	-0.389	-0.074
Asia	0.1179	0.008	15.536	0.000	0.103	0.133
Australia	0.2280	0.034	6.665	0.000	0.161	0.295
Europe	0.1059	0.008	13.903	0.000	0.091	0.121
Quarter Year	0.0377	0.003	12.392	0.000	0.032	0.044

Omnibus:	9399.754	Durbin-Watson:	1.995
Prob(Omnibus):	0.000	Jarque-Bera (JB):	2063.985
Skew:	0.133	Prob(JB):	0.00
Kurtosis:	2.041	Cond. No.	2.30e+06

Limitations with this model

- Why is the “Time Spent on Site” coefficient negative? Because it is a linear model it fails to capture non-linearity.
- We care about daily time spent on site and controlling for things that might be correlated with time spent on site and the outcome variable.
- Other factors that could be bringing the R-Squared down: We think that there is something we are not including in the model.

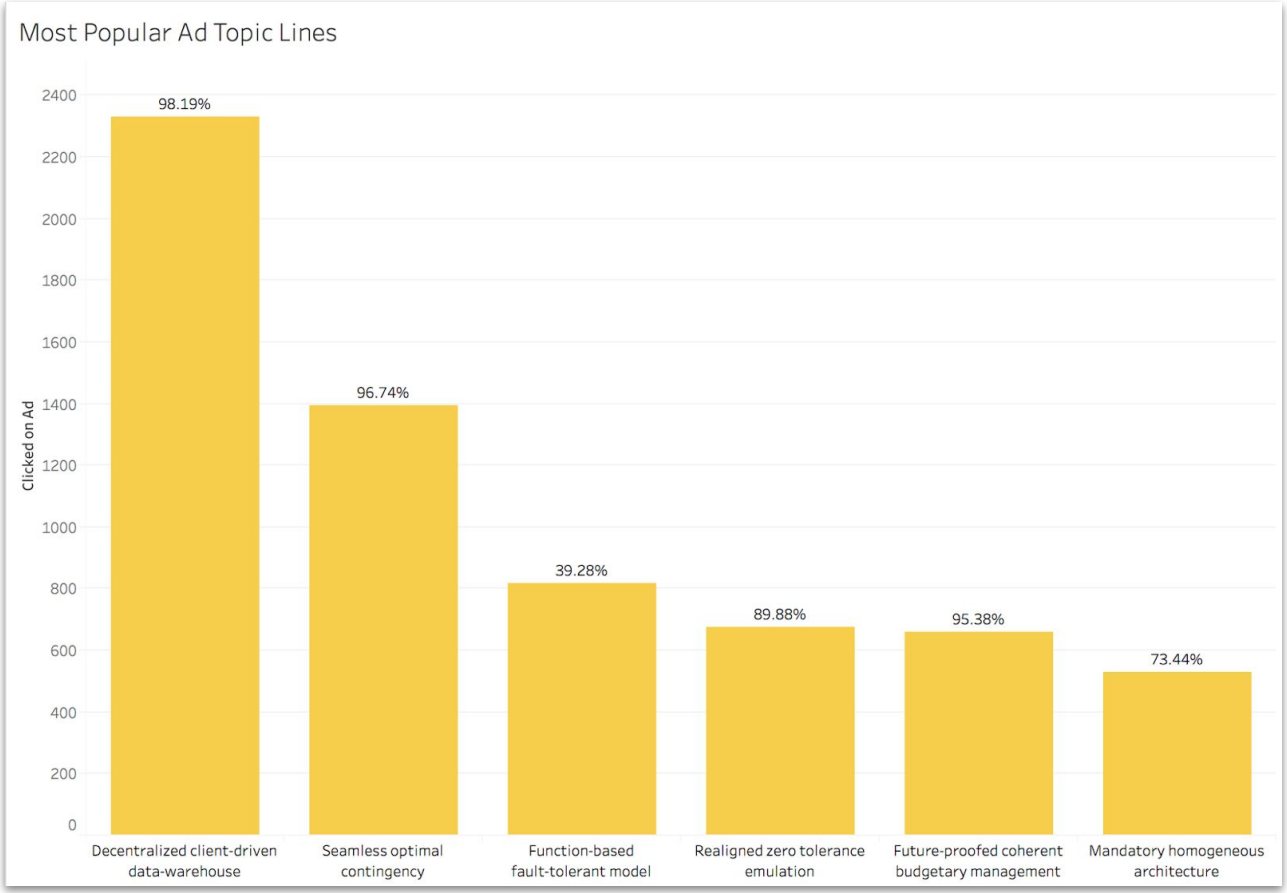


Observations:

- The average time spent on the website per month barely varies or changes
- However, the click through rate has significant variation or increase on a monthly basis.

Take-away:

There could be a different factor influencing CTR



Observation:

The top 6 ad_titles makes up 26% of all ads that were clicked on

Conclusion:

Run another regression with the 6 most popular ad topic lines included

Regression Model 2

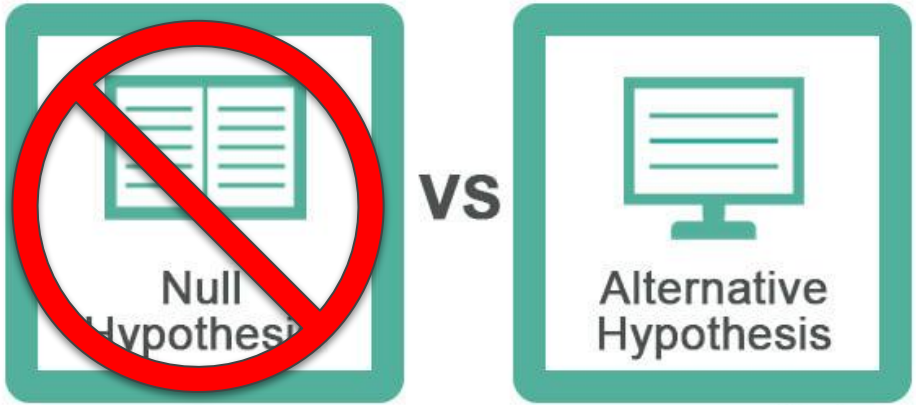
OLS Regression Results

Dep. Variable:	Clicked on Ad New	R-squared:	0.282
Model:	OLS	Adj. R-squared:	0.282
Method:	Least Squares	F-statistic:	1636.
Date:	Tue, 29 Nov 2022	Prob (F-statistic):	0.00
Time:	18:06:37	Log-Likelihood:	-28006.
No. Observations:	50000	AIC:	5.604e+04
Df Residuals:	49987	BIC:	5.615e+04
Df Model:	12		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025
const	-0.1436	0.019	-7.510	0.000	-0.181
most_popular_titles	0.2312	0.005	42.930	0.000	0.221
Daily Time Spent on Site	-0.0014	0.000	-10.636	0.000	-0.002
Age	0.0224	0.000	99.397	0.000	0.022
Area Income	6.204e-07	1.69e-07	3.677	0.000	2.9e-07
Daily Internet Usage	-0.0017	4.61e-05	-36.202	0.000	-0.002
Africa	0.1500	0.007	20.861	0.000	0.136
Americas	0.0714	0.007	10.058	0.000	0.058
Antartica	-0.2018	0.079	-2.557	0.011	-0.357
Asia	0.0991	0.007	13.281	0.000	0.084
Australia	0.1697	0.034	5.048	0.000	0.104
Europe	0.0906	0.007	12.102	0.000	0.076
Quarter Year	0.0341	0.003	11.399	0.000	0.028

Omnibus:	10590.041	Durbin-Watson:	1.995
Prob(Omnibus):	0.000	Jarque-Bera (JB):	2300.760
Skew:	0.190	Prob(JB):	0.00
Kurtosis:	2.021	Cond. No.	2.30e+06

Final Conclusions



- Statistically, at a 95% confidence interval (0.05 alpha), we **REJECT** our null hypothesis.
- Business perspective: Ad topic Line is more impactful than time spent on the site. (0.2312 vs -0.0014, respectfully)
- Variation in click rate is hard to predict; many factors outside of our model might affect a user's decision to click on an ad.
- Other factors that could be bring the R-Squared down: Content, graphics, sound effects, how much space is the ad taking up on the website and how frequently the ad is shown.