CONVERSION PROJECT

Anonymous have hired us to find where the dosh can be found or where the dosh has been wasted

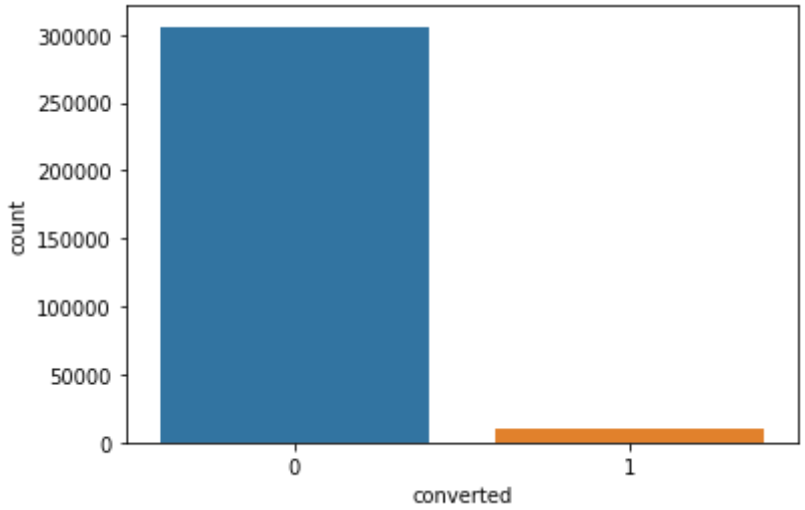
OUR DATASET:

Over 300,000 rows - users who have looked at our site.

Only 6 columns:

- country: what country the user is accessing the site from
- age: the age of the user
- new_user: whether the user has visited the site before
- source: where the user has been directed from
- total_pages_visited: how many pages the user has accessed on our site
- converted: whether the user has spent money on through our website

CONVERSION VISUALIZATION:

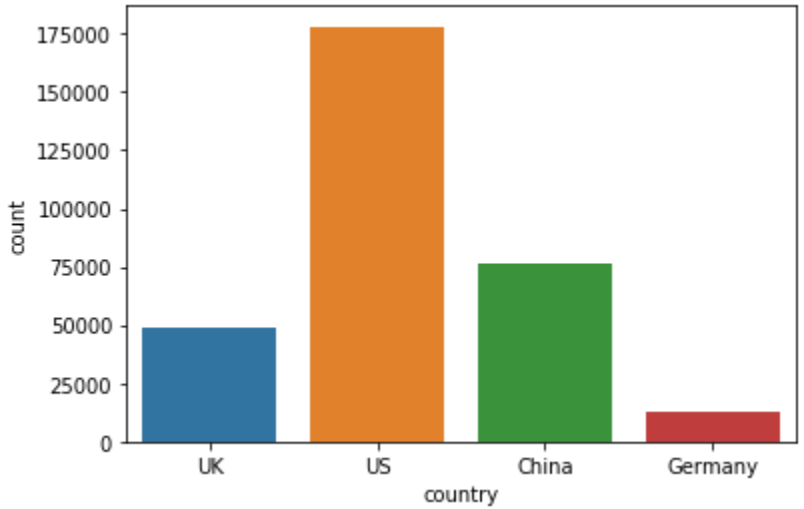


Note: a 3.23% chance of conversion overall =>

Conversions per country:

country	proportion of conversions
US	0.0378
UK	0.05263
China	0.00133
Germany	0.0625

Proportion of our dataset by country:



CONCLUSIONS FROM THE PAST SLIDE

We can see that the UK and Germany have, by far, the highest conversion rates. The UK with many more users, somewhat concretizing their high conversion. The US in third place (just under 4%), representing the majority of the market and finally China whom have little to no chance of conversion (less than 1%) with the second highest user count.

Conclusion:

It would be a good idea to somehow make the content of our website, less china friendly. Maybe running a different style or using different words, that are more likely to put an average user from China from using our site.

Seo – search enginge opmisation	37578
Advertisements	21561
Direct	17463

When looking into how Chinese users have found our site (see table to the left) we see that, the vast majority of them (~78%) have found us through something we are paying for. It would then make sense to cut our advertisements and search engine optimization in china.

CONVERSIONS BY SOURCE

As we can see from the table, Ads gain <0.2% vs Search engine optimization. With Direct venture differing adds by <0.5%. If we remove China from our dataset when it comes to paying for people to join our site, i.e. **leaving ONLY the Direct connection part from china** in our dataset. We get the following results.

source	proportion of conversions
Ads	0.03448
Seo	0.03289
Direct	0.02817

Dataset without china coming from ads or seo

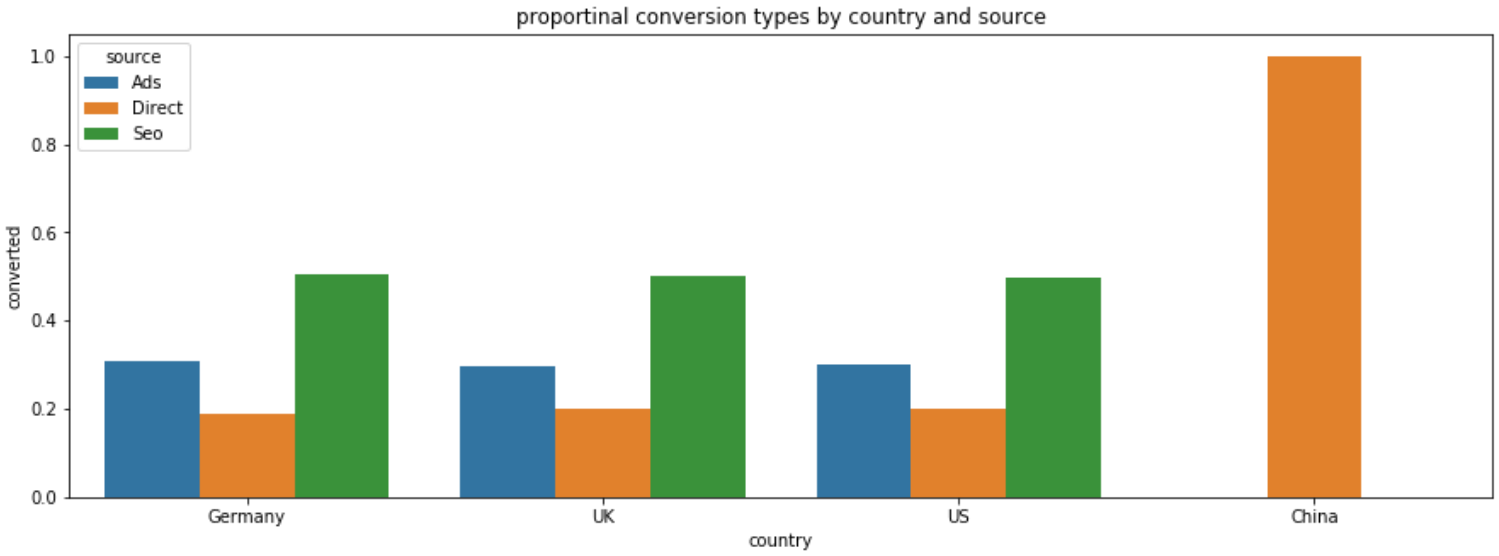
source	proportion of conversions	difference
Ads	0.04507	+1.059%
Seo	0.04303	+1.014%
Direct	0.02817	No change

As we have already noted, this has increased our chance of conversion in general, therefore **note: this will be a permanent change in the dataset from now on to see where we should head from this strong first step.**

CONVERSION BY SOURCE & COUNTRY

source	country	Proportion converted per country
Ads	Germany	0.307598
	UK	0.294902
	US	0.300802
Direct	China	1.000000
	Germany	0.187500
	UK	0.201961
	US	0.200238
Seo	Germany	0.504902
	UK	0.503137
	US	0.498960

We can see that conversions proportional to country and grouped by source show no real difference between country.



OUR 'NEW USER' VARIABLE

New user	Proportion of conversion
No	0.7023
Yes	0.2976

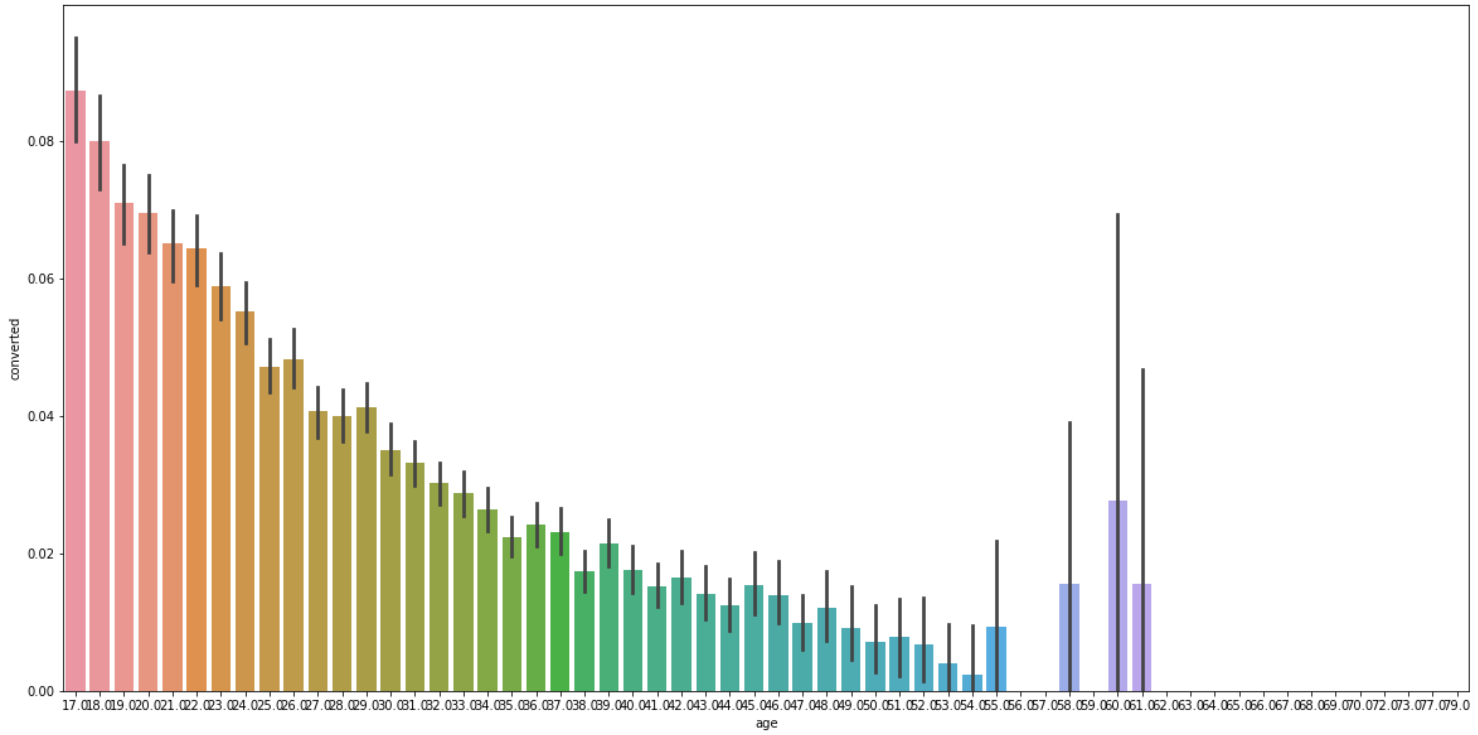
	Proportion of conversion		
Country	**New user**	**NOT new user**	**Proportion NOT**
China	0.000988	0.001383	0.5833
Germany	0.022426	0.058190	0.6957
UK	0.076665	0.175262	0.6957
US	0.197589	0.467497	0.7029

Source	**New user**	**NOT new user**	**Proportion NOT**
Ads	0.085457	0.213693	0.7143
Direct	0.064513	0.137028	0.6799
Seo	0.147698	0.351610	0.7042

We can see that there is a considerable difference in conversion based on a past visit, this may be due to users browsing our service first, then coming back to complete their conversion. Looking at variations according to country and variation according to source we see little change from the above averages over our dataset.

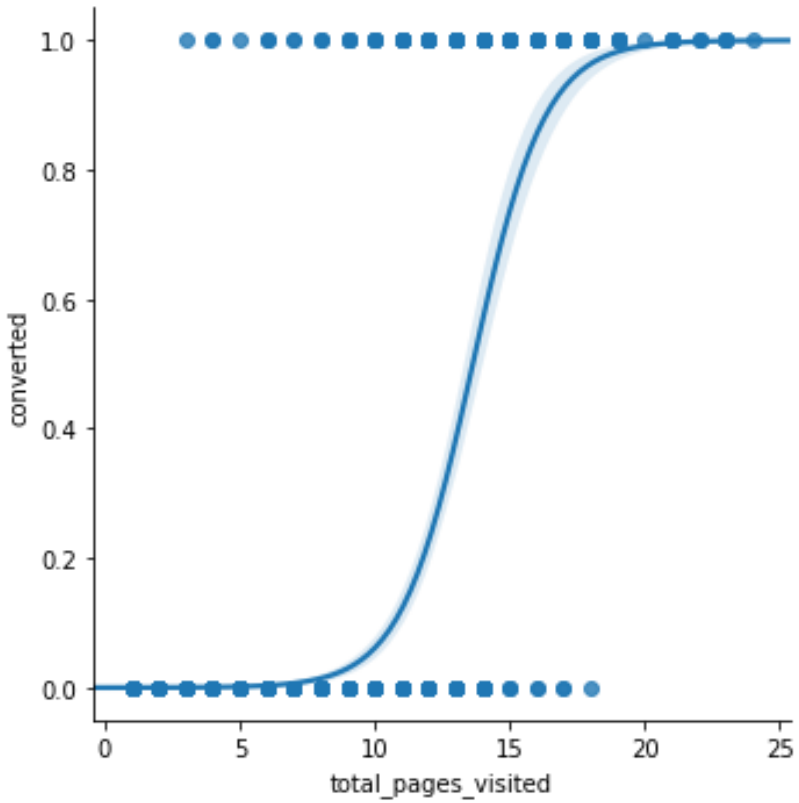
What could be a fruitful query would be looking at how source affects conversion per country.

DETERMINING FACTOR?: AGE



Over our dataset, we see a clear decrease in chance of conversion based on age. With an average user age being around 30, the younger, the better for a chance of conversion, with our youngest user being 17. As we can see the variance increases substantially over 50 years. There was no substantial difference in the graph between countries.

OUR MOST PROMISING PREDICTOR: PAGES VISITED



As we can see from the beautiful inverse sigmoid, there is a clear correlation between conversion and visiting pages. On average, over different countries, the average pages visited was ~5, whilst we see the probability for conversion (>0.5%) occurring around 14/15 pages visited. Slight discrepancies are seen from country to country and source to source but nothing noteworthy, hence adding the plot regarding the totality of our dataset. From the last slide and this one we see some tendencies we should take into account seriously regarding remodeling our business.

CONCLUSIONS AFTER A FIRST LOOK:

- **Age:** (the younger the better)

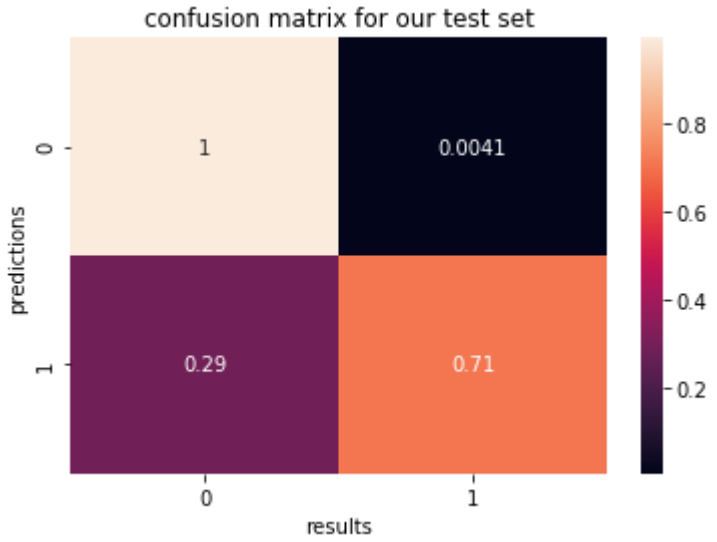
We need to target the younglings! Make the page more youth friendly and correlated with their trends and tendencies. Minimalism or the such, lots of AB testing needed.

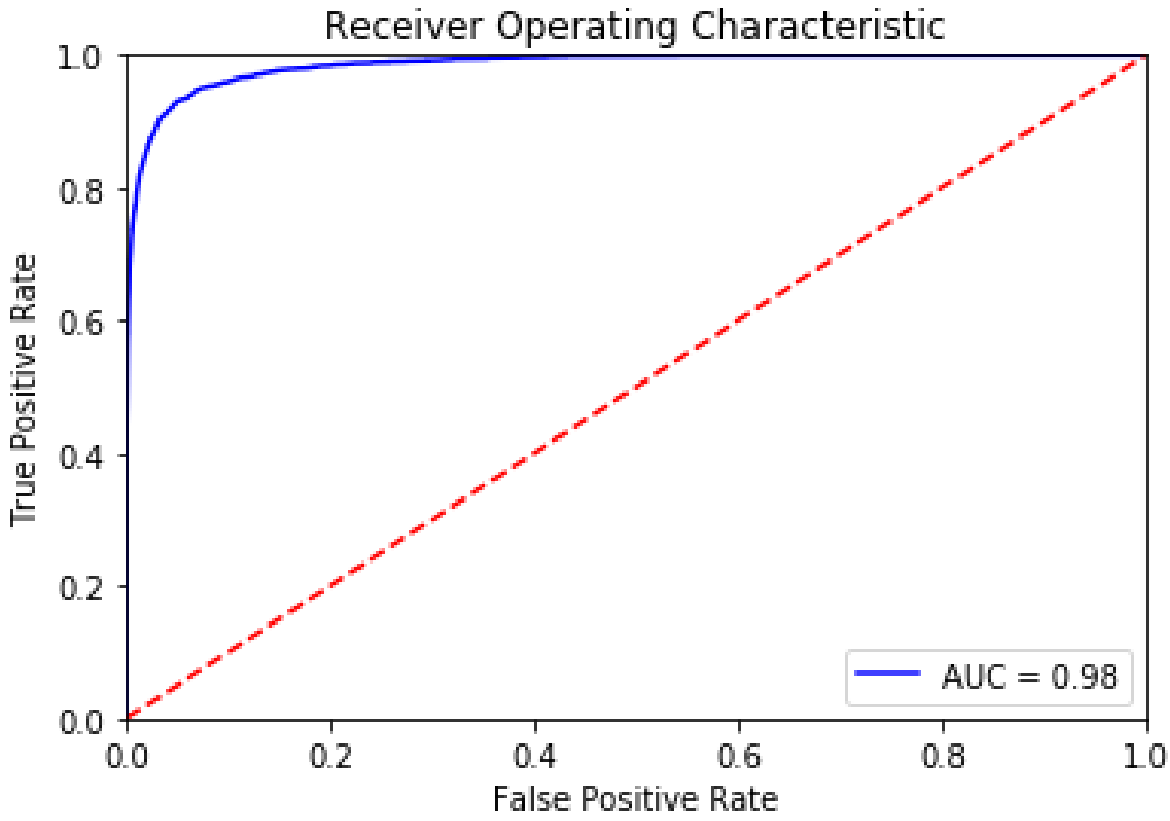
- **Total pages visited:** (the more the merrier)

Keep the people hooked! Increase dopamine doses! Make buttons more satisfying, and pages smaller with more links to new pages. An endless labyrinth of pleasure. Even if this seems counterintuitive, upping the number of pages they visit is our best bet!

LOGISTIC REGRESSION:

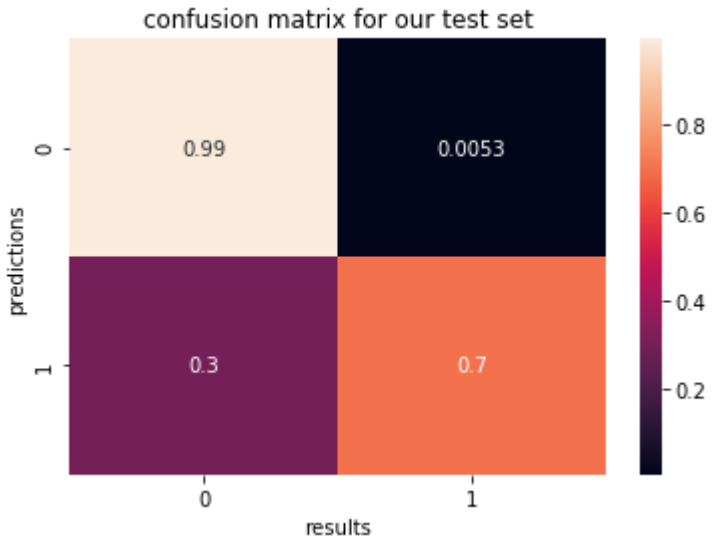
Upon modelling a quick logistic regression, we notice already remarkable results on our test set. Hitting an model score of 1! Notice the following ROC and confusion matrix.

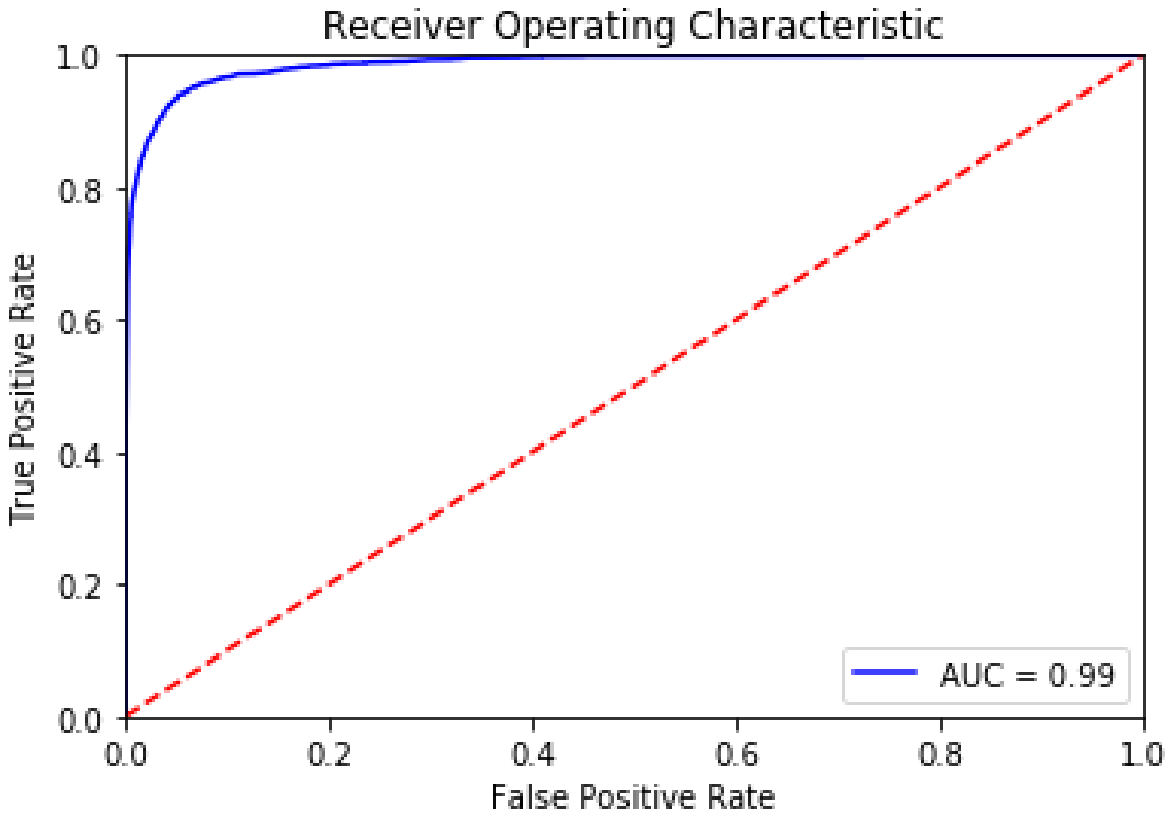




RANDOM FOREST:

After a little grid searching, finding parameters that fit well. We get the following correlation matrix an ROC. Even if the score of the forest model may have been worse, our AUC was more proficient!





MANY THANKS FOR LISTENING!