Data Scientist – Marketplace

Welcome to the Data Scientist – Marketplace challenge! At trivago, we are devoted to the idea that decisions are based on data, not intuition. As Data Scientist in the Marketplace team, you can work on high impact topics and use the power of our petabytes of data to develop creative, data-driven solutions that satisfy the needs of our users and advertisers (online travel agencies, hotel chains, and hoteliers).

Advertisers compete for exposure in trivago's Marketplace by submitting attractive property rates and bidding on a cost-per-click basis. The Marketplace Team is responsible for leveraging this data, so that we can monitor the dynamics of our Marketplace, support decisions on how to continuously improve it, and help advertisers optimize their marketing campaigns on our platform.

We are curious to see your brain in action and have prepared some tasks for you to solve. Our hiring team is looking forward to your results! This challenge is designed to give you a glimpse of the activities you would be tackling in your potential future team. The tasks in this case study are inspired by and very similar to those you need to deal with every day. This means that if you have fun solving this case study you are most likely on the right track! Good luck!

SUBMISSION DEADLINE: 7 days

HOW TO SUBMIT: attach your results via the link provided in the email from your recruiter

THE CHALLENGE

Context:

At trivago one of our core values is “Power of Proof”. This also applies to the development of features in our website. As such, we have run two A/B tests (testing different features, one in February, the other in May) with a classical setup: show a *control group* to the majority of our users (constituting a subset of website visits), and a *test group* (containing the new feature) to the remaining users (and corresponding remaining visits).

You are provided with one csv file for each of these two tests, containing the results collected in terms of our main base metrics. The data is aggregated by ymd (the date), platform (the country-specific trivago platform), and group (control or test). These are the first 3 columns of the files. The remaining columns contain the base metrics:

- visits - how many website visits from our users were logged
- clicks - how many times our users clicked in one of our prices and were re-directed to the website of one of our advertisers. Note that each individual visit may or may not include clicks.

- revenue - how much money (in EUR) was generated for trivago. You can work under the assumption that this revenue is generated under a CPC model.
- bookings - how many accommodation reservations were made in our advertisers' websites.
- booking_amount - the corresponding amount (in EUR) paid by the users for the accommodation reservations.

Questions

Data Exploration

1. What is the most valuable platform for trivago? Why?
2. What are the global and platform average revenue per click?
3. Do you find any trend in the data?
4. Which relation do you think exists between clicks and bookings? And what about visits and bookings? Which one would you consider as more valuable KPI?

A/B test analysis

For each of the 2 tests:

1. Which metrics would you use to evaluate the performance of the *test group*?
2. Can you summarise the performance of the tests?
3. What would be your recommendation regarding the decision for the test? (and why?)
    1. Roll-out the new feature to all users immediately
    2. Develop the feature further before testing again
    3. Run the test for a longer period of time

Preparing your submission

You should treat your case study submission as if you were preparing a report to be read both by your Data Science peers and other business stakeholders.

In particular this means:

- You should include all the code used in all the steps of the analysis.
- In your report you should present results in whatever form you think it is appropriate (free text, charts, tables, etc.)
- You should offer good justification of your choices — we want to understand how you think and how you reached your conclusions

For example a properly laid out notebook (Jupyter or rmarkdown) would be fine, but we leave it up to you to choose an adequate submission format.