Data Science Case Study

Werkspot/Instapro/Travaux

Introduction

One of the company's key value propositions is that we make it easier for homeowners to hire quality tradespeople.

Let's assume that the user journey can be defined as this:

1. A Homeowner posts a Job by filling the Job form. The form asks specific questions about the size of the job, a general description and attachments (optional).
2. Tradespeople can see all Jobs on the 'Pick & Choose' page when logged in. They can view and propose to these Jobs.
3. The Homeowner receives these proposals, reviews them and can "shortlist" the Tradespeople they want to contact. When a Tradesperson is shortlisted, Werkspot charges them a lead fee.
4. The shortlisted Tradespeople then receive the contact details of the consumer and contact them off the platform (via phone/email/etc.)

The very first step of the process is when a Homeowner publishes a Job on our platform. Not all Homeowners go directly to werkspot.nl or travaux.com to post a Job, actually, the majority of Jobs come from internet searches. One of our Marketing Team's focus is SEM (Search Engine Marketing) when it comes to acquiring Jobs. They define and manage campaigns to get enough Jobs published in our platform. They set up campaigns on different web search engines, Google being the main one.

Problem

As in any investment, you want to get the highest return, so the optimization of the performance of SEM campaigns is key to maximize the margin. This is one of the Marketing Team's main goals, and you are here to help them.

The underlying objective is to generate insights from campaigns data that will help the Marketing Team reach their goals. You will be given a dataset that contains data from SEM campaigns and published Jobs in our platform.

Dataset description

The dataset contains 54.729 records of unique jobs that were published on our platform between February 1st, 2020 and February 1st, 2021. Those published jobs have associated a revenue (sum of lead fees of all shortlists that happened for that job) and a cost (if the job was acquired through a paid marketing campaign in Google). The jobs that were acquired through a google campaign, have associated an identifier of the campaign. More that one job can be acquired from the same campaign. Each job is also connected to a specific service and profession.

date_id [datetime]: date and time when the job was published.
source_medium [string]: there are three categories ‘google / cpc’ (acquired through a google ad), ‘ google / organic’ (acquired through a google non-paid link), and ‘(direct) / (none)’ (acquired through a direct landing to our site).
campaign_id [string]: unique identifier for a marketing campaign.
channel [string]: source from which the job request was acquired. PAID refers to acquisition through SEM strategy (BRANDED means that the marketing campaign contain the keyword “werkspot” and NONBRANDED if brand name is not in the keywords), ORGANIC refers to non-paid traffic coming through a Google search, DIRECT refers to direct traffic landing on the website, and OTHER couldn’t be allocated to any channel.
acquisition_id [string]: unique identifier of a Job request.
city [string]: name of the location where the job is related to.
service_name [string]: name of the professional service related to the job.
profession_name [string]: name of the profession related to the service. One profession can have many services, but one service is associated with only one profession.
campaign_type [string]: can be “SERVICE” (when the campaign only contains keywords that describe the service, for example “painting”) or “GEO” (when the campaign contains location search words, for example “painting in Amsterdam”).
revenue [numeric]: amount of money in Euros that the company earned from all shortlists on that particular job.
cost [numeric]: amount of money in Euros that the company was charged from paid campaigns for that particular job.
nr_clicks [numeric]: number of necessary clicks that lead to the job publication.

Questions to address

Part 1: Generate insights from the Marketing dataset using advanced analytics. Use Python as the programming language to develop your solution and feel free to report any exploratory analysis that was relevant to you during the exercise.

Job volume forecast. Being able to anticipate the demand is something that can help marketing in defining their strategy. We ask you to forecast the demand for each of the services that you will find in the dataset for the first 2 weeks of February 2021. You can build a predictive model of your choice. Marketing is interested in having the demand volume in terms of **number of published Jobs** at a **weekly and country level**. They also would like to know how accurate your forecast is.

Part 2: Open-ended questions. We strive for pragmatism and product iteration. Our approach is to get to a “good enough” first version in production, learn from the experience/data and then iterate.

a. Marketing's main success metric is total margin (revenue - cost) and they want to maximize it. Given the information you have in the dataset, how would you use it to make recommendations in order to optimize the performance of their campaigns? There is no specific answer expected here and you can make any assumptions you think are necessary.

b. What other information/data would you like to have in order to improve SEM campaigns' performance? Could be internal to the company, external or both.

c. If you are to productionize the model that forecasts the demand:

   i. How often would you choose to train the model?

   ii. Briefly define the technology stack you would choose for the end-to-end automated solution.

   iii. What information do you need from Marketing to ensure the adoption of your solution?

   iv. If Marketing uses your predictions to set up campaigns, what business metrics would you use to evaluate the online performance of your model?

d. Assuming that this is an end-to-end machine learning project: How does the project timeline of this first version look like?

e. Extra credit: Which of those do you think is better to maximise total margin, tROAS, tCPA, or another strategy?