

Data Science Interview Project

Congratulations on making it to the Zocdoc data science interview project! This part of the interview will give you an opportunity to get a better sense of the kind of work the team does and to showcase your problem solving and analytical skills for the team.

We will be evaluating a number of factors:

1. Analytical problem solving and ability to structure an ambiguous question
2. Technical skills and accuracy of analysis
3. Communication skills and ability to create a cohesive narrative from your results

To help gauge how much depth to go into, the entire exercise typically takes 6 to 10 hours. If you are spending more time than that, you're likely spinning your wheels on something you don't need to be. Questions are encouraged - please reach out with any questions as you are working through this!

Background

We've recently experienced a number of Google algorithm updates that have lowered our average page rankings for our URLs on Google. We need to analyze our URL data to determine which URL qualities drive our ranking. We will use this information to develop a roadmap to improve our rankings.

Part 1. Which features impact rankings?

We are able to pull a sample dataset from one of our search engine optimization tools with data at the URL level. We want to use this dataset to understand which features of a URL impact how the URL ranks on the Google search result page (SERP).

The dataset includes the following fields:

- Google Rank: the average ranking in the Google search results for that URL, across all the keywords that the URL ranks for
- % Cross Listed Results: % of Zocdoc search results that do not match the specialty that was searched for, i.e. a user may search for depression and a primary care provider could be shown in the results if they treat depression, even though a psychiatrist would be the default specialty for that type of care
- % Templated Words: % of words on the page that appear to the Google crawler to be templated; we expect that the Google crawler rewards unique words vs templated
- Distance: the average distance between the providers shown on the search result page and the searcher
- # Reviews: the number of providers with reviews in the search results
- Number of Preview Profiles: the number of Zocdoc search results that are not active Zocdoc customers; they are profiles with information we have gathered about the provider, but they do not provide the option to book.

All the features in the dataset are an average of the last 30 days. Use the data to answer the following questions:

1. Which of these features are drivers of URL rankings?
2. Which of these features drive the largest impact on URL rankings?

You may use any tool to run this analysis – python, R, excel, etc. Please treat the analysis as a deliverable for the project, regardless of which tool you choose. Make sure the file (python notebook, excel workbook, etc.) is readable, well formatted, and has comments so that it can be easily evaluated.

Part 2. Propose and size a new feature

Based on the findings from your analysis above, propose and size a change to the SEO landing pages. The initial dataset and analysis above should allow you to translate the feature into a rankings impact. Use the assumptions below to translate the rankings impact into a session and bookings impact:

- Each URL in the dataset received 1,000 impressions per month on a Google SERP

- A rankings improvement of 1, i.e. going from the 5th slot to the 4th slot, improves the click through rate by 1% on an absolute basis
- The conversion rate from an SEO landing page to booking is 20%

You will need to make assumptions about the impact the feature will have on rankings as well. Take your best guess at the impact – try to get to a reasonable estimate, but most importantly, make sure your assumptions and logic are clear and well documented!

Part 3. Designing and testing new features

You've sized your new proposed feature and presented it to the team. They are excited about the bookings opportunity but they want to test the feature before rolling it out. The test will not be to validate the rankings improvements as this can take several months to be observed. They want to test if the feature has any impact on the conversion rate (i.e. the % of sessions that convert into a booking). Describe how you would design a test to measure the conversion rate impact of the new feature.

Part 4. Summarizing your findings and recommendations

We've finished running the test and found that the conversion rate was flat, which was enough for the team to feel comfortable rolling out the feature. Over the next few months, we've seen a commensurate lift in rankings, sessions, and bookings to what you predicted in your sizing. Now, we need to present the full analysis and findings to our leadership team. Consolidate your findings and recommendations into a single deck or document with a cohesive narrative. Here are some areas to focus on in the deck:

1. Summarize the key findings from your analysis for the product / business.
 - a. Be sure to include context on the analysis (as the leadership team will not have been as close to the project), the results that were observed, and what the takeaways are.

- b. Since you don't have actuals for the test or post roll out period, describe this conceptually and use the data provided in the sizing section as needed.
2. Provide recommended next steps in terms of further analysis, additional features, and testing.

Tips for your deck: try to keep the main narrative concise and streamlined; include all additional data and supporting material in the appendix (it is important to include this but it doesn't need to be in the main flow of the deck – remember, this is a presentation to the leadership team!).