Data Scientist Tech Interview Candidate Email

Hi Amanda De Silva,

Congratulations on securing an interview for the Associate Data Scientist-Marketing position at Canva! This email confirms the content of the interview in conjunction with a calendar invite (coming through shortly!).

The interview will be broken up into two sections.

Part 1: Discussing your career history and achievement
You will be meeting Data Scientist team members to discuss your previous experience, achievements, and educational background. They may ask you to discuss some projects or initiatives you've worked on as well as pose some scenario-based questions with respect to how you would tackle certain problems.

The best thing you can do to prepare for this portion of the interview is to have some examples of various projects you've worked on fresh in your memory to discuss as well as be familiar with what you've shared on your CV.

Part 2: Technical Challenge
In this session, you will be given an experimentation scenario and will be asked how to solve it in a real world situation. You can choose either Python or R for this section. The aim of this section is to assess your technical ability in the following areas: reporting, data wrangling, experimental setup, statistics, ad-hoc investigations, and visualisation. Please make sure to review the pre-interview resources and follow all recommended steps.

Final tips to approach the challenge!

- Keep an open mind
- Don't be afraid to think laterally
- Ask clarifying questions
- Ask for more information when required

You will also have time to ask questions at the end of your interview.

Congratulations and best of luck in your interview - don't hesitate to reach out if you have any other questions.

We like to focus on really getting to know you, without having to worry about just taking notes, so this interview ***may*** be recorded and transcribed using brighthire.ai. If you would like to opt-out of this recording, you can do so HERE. We respect your privacy, so be assured, opting out will not impact your application. Please note that we cannot provide recordings of the interview.

Regards
Emma

Data Scientist Tech Interview Candidate Resources

The Data Scientist tech interview will test your ability to analyse an experiment, using Python or R.

This guide will provide some useful context around the problems that you will look at, and will help you prepare for the interview. This guide does not contain any of the actual questions and is best used to help you familiarise yourself with anything mentioned in the context that you may not be familiar with.

There is no need to have written any R/Python prior to the interview.

Prior to the interview, please ensure that you have your preferred notebook for R/Python. You will be asked to share your screen, so please test to make sure everything works as expected. Please be aware that many workplaces block screen sharing outside of their company tooling or accounts.

Interview Structure

The technical interview will follow this format:

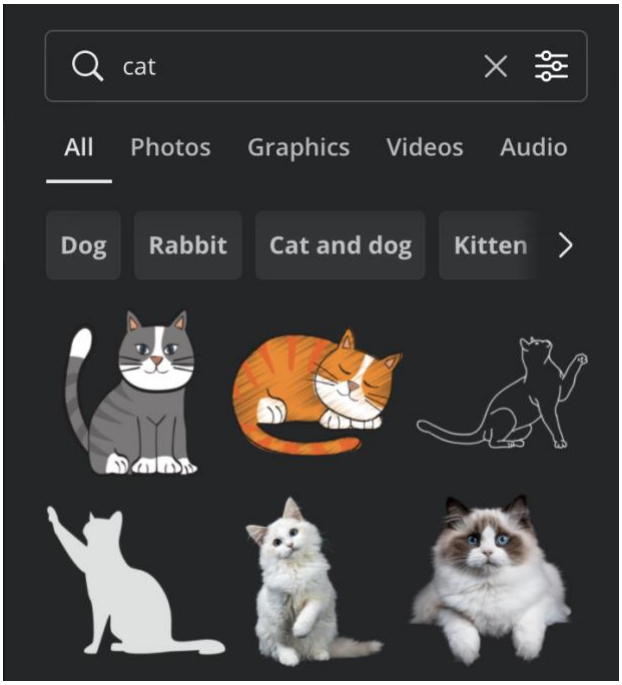
- Your Career Chronology
- Working through the experimentation challenge covered below
- Questions for the interviewers

Experimentation

The data file can be found here.
Please make sure you download the data file prior to the interview.

Scenario

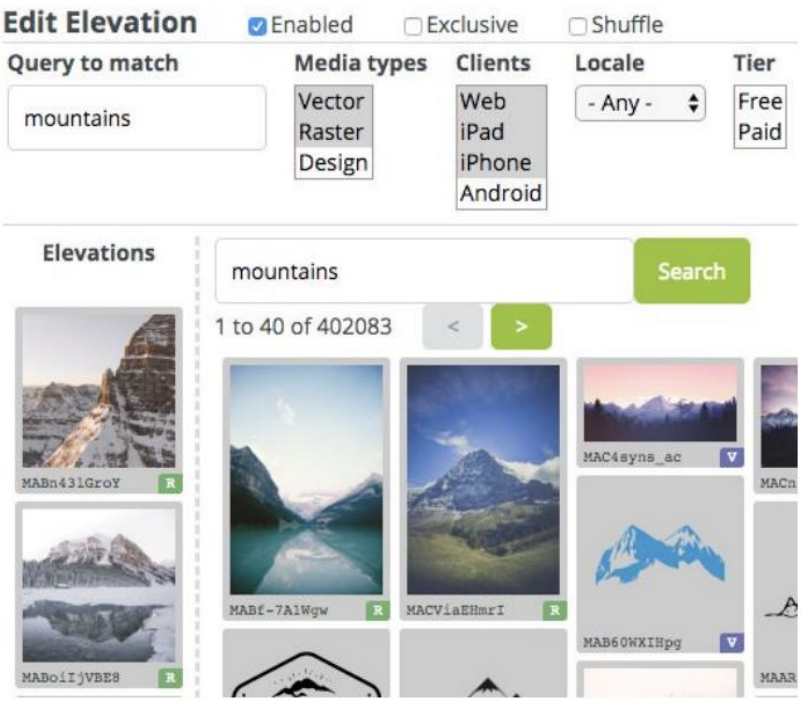
Canva's design editors have built-in image search functionality, allowing users to find a wide range of content to include in their designs.



At the time of this scenario (many years ago), our search functionality was powered by an open source search engine and provided results purely based on the similarity of image metadata with the user query string. Relying on this similarity works reasonably well, but it's usually easy to find examples of queries with poor results.

To work around this, our engineers developed a tool allowing Canva administrators to administer "manual elevations". These are manually curated collections of search results which will appear first for a given query.

In particular, for any given combination of query string and media type (raster and/or vector), a privileged user can find existing search results and drag them into the 'elevations' section on the left. Those results would then be shown first when users searched for the relevant query string. The tool’s interface is shown in the screenshot below.



In this scenario, a new mechanism has been developed that calculates elevations automatically, by processing a large quantity of search data. The automated mechanism recalculates elevation sets for tens of thousands of queries daily and has demonstrated a strong overall positive impact on search results. The automated elevations mechanism calculates at most 50 elevations per query.

We have performed an experiment to assess the ongoing value of manual elevations. A subset of users were randomly assigned to one of the following experiment groups:

- **Group A:** Both automated and manual elevations were applied, with manual elevations taking priority
- **Group B:** Only automated elevations were applied to search results

Each user query was recorded along with each result click. In our data warehouse ELT process this has been aggregated and joined with publish and billing data to produce a single record per user query. This data is filtered so that only user queries that correspond to manual elevations are included.

The fields in this file are:

- `search_id`
  - A unique identifier for each search made by a user
- `user_id`
  - A unique identifier for each user
- `experiment_day`
  - 1 for the first day of the experiment, 2 for the second day, and so on
- `experiment_group`

    - The experiment group of the user at the time of the query
    - possible values: 'A' or 'B'
- `user_tier`
    - Indicates whether or not someone has a paid Canva subscription
    - possible values: 'free' or 'paid'
- `query`
    - The user query string
- `media_types`
    - The media type filters applied to the query (raster and/or vector filter)
    - These filters can be set in the UI with the All/Photos/Graphics selector visible in the example 'cat' query
    - possible values: 'R', 'V', 'RV'
- `num_elevations`
    - The number of search results with manual elevations (these are shown to users before any other result)
- `num_clicks`
    - The number of distinct images that were clicked on in the search results
- `num_exported_results`
    - The number of distinct images that were clicked on in the search results and then downloaded in a design
- `num_licenses`
    - The number of distinct images that were clicked on in the search results and then paid for while downloading a design

Tooling to prepare

- Python/R notebook environment with data exploration and visualisation packages installed
- Ability to share your screen through Zoom (many workplaces block screen sharing outside of their preferred tools and accounts, so please double check everything works)