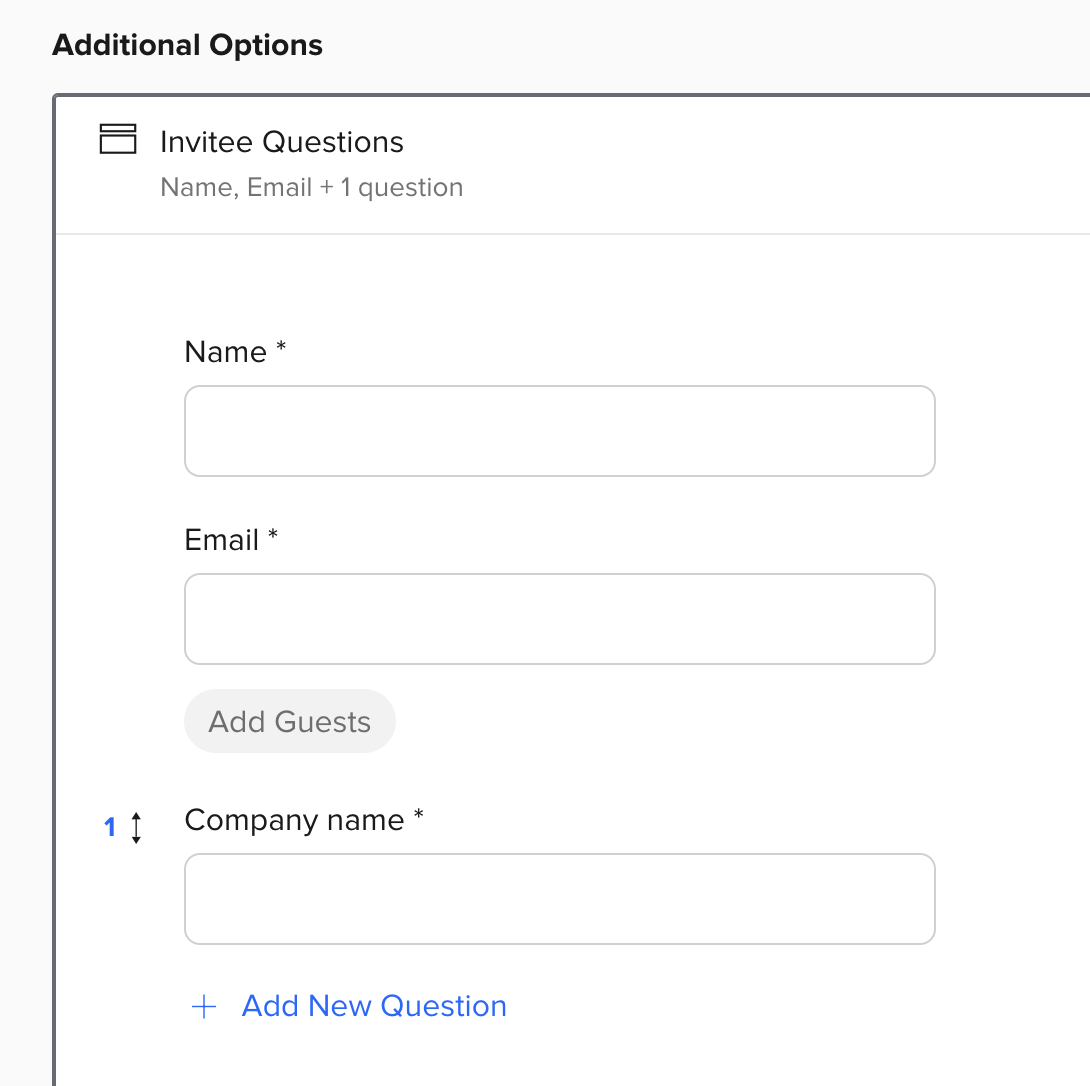
## **Technical Exercise**

This is an exercise to help us understand your ability to work with ETL/ELT pipelines, SQL, Python, Google Products, and use API. You are free to choose the technologies, however, please use Python at least for one of the solutions or a part of a solution.

### Task 1

SBTi uses Calendly as a way for companies to book slots for the validation service.

* Calendly is not used for its traditional purpose here, rather a timeslot represents team capacity and not a time booking. 30 minutes represents a slot for the validation of near-term targets, and 60 minutes is for the validation of net-zero targets. Suppose SBTi has the ability to do 100 30-minute slots a month.
* The date of the booking is the date when a decision on the validation should be delivered by the SBTi team. And the contract start date is 30 days before the delivery date for near-term validation and 60 days for net-zero validation.
* The booking page should contain the name, email, and company name of the person booking the slot.



Using Calendly’s API, create a dashboard in Google Sheets that would contain all booked slots for the month of December. The dashboard should contain:

* The company name
* The name of the company contact
* Email
* Type of the validation (near-term or net-zero)
* Delivery date
* Date of the start of the contract

Please provide a link to the Google sheets and the booking page of Calendly to test the dashboard and the code for review.

### Task 2

On the SBTi website, we have [a public dataset, “Companies taking action,” available for downloading](https://sciencebasedtargets.org/companies-taking-action). This dataset is a result of years of manual entry and updates and has many quality issues.

1. Create SQL-database based on that table making the data structure more appropriate for a relational database. You are free to propose the data structure you see most appropriate. For example, instead of 3 columns “Near term - Target Status”, “Near term - Target Classification” and “Near term - Target Year” it could be a table “near-term\_targets” with company\_id, status, classification and target\_year.
2. Populate the database using web-scrapping from the website.
3. Examine the need for data cleaning. Propose 2-3 data transformation strategies to combat data quality issues present in the dataset.

Please provide the database, the code and detailed description of how you approached the task. Keep in mind that there is a lot of SBTi-specific information about this dataset that we do not expect you to know. We are interested in the approach that you have taken, and are not expecting a perfect result.

**Additional context about the “Companies taking action” dataset:**

The company A.G.Barr has set 3 types of targets, and they were combined together in one big target language. That how it could be broken down:

1. **Overall Net-Zero Target** A.G. Barr plc commits to reach net-zero greenhouse gas emissions across the value chain by FY2050 from a FY2020 base year.
2. **Near-Term Targets** A.G. Barr commits to reduce absolute scope 1 and 2 GHG emissions 60% by FY2030 from a FY2020 base year. A.G. Barr also commits to reduce absolute scope 3 GHG emissions from purchased goods and services, upstream transport and distribution and downstream transport and distribution 25% within the same timeframe.
3. **Long-Term Targets** A.G. Barr plc commits to reduce absolute scope 1 and 2 GHG emissions 90% by FY2035 from a FY2020 base year. A.G. Barr plc also commits to reduce scope 3 GHG emissions from purchased goods and services, upstream transport and distribution and downstream transport and distribution 90% by FY2050 from a FY2020 base year.

There are 3 type of commitments: standard, Business Ambition 1.5 and net-zero. Examples in the dataset:

* Company A&L Goodbody has made a standard commitment.
* 3B-Fibreglass has made Business Ambition 1.5 commitment, and
* A.G. Barr plc has made net-zero commitment.

Column date is a date of the last update of the company status.