

**LEFT LANE Data Engineer Case Study**

**Summer 2022**

Below you will find the case study outlines and materials for the Technical Investment Associate. The purpose of this work is to better understand your skillset, how you approach work, your communication style, and ultimate quality of work output. The case studies are meant to test the three areas of focus as outlined in the initial interview: company sourcing, company analysis, and automation / tooling within Left Lane.

As stated in the initial interview, this role is quite open-ended and entrepreneurial. The case study is meant to bring out this entrepreneurial spirit and most questions are open-ended. This is on purpose as we want to see what interesting ideas and approaches you come up with.

All work done within this case study is highly confidential (as outlined in the NDA) and will not be shared with any other groups or individuals outside of Left Lane.

Do not hesitate to reach out with clarifying questions as you work through the different sections. Nothing is intended to be confusing or trick you and, like all workstreams, this can be a collaborative process to better understand your working style.

We look forward to seeing your work.

**PART 1: COMPANY SOURCING**

Please select at least **ONE** of the following prompts. Feel free to work on more but we expect one complete body of work for at least one of the prompts.

**PROMPT 1: SIMILAR WEB API SOURCING**

Using the [SW API](https://docs.api.similarweb.com/) come up with a tool that will surface promising companies for Left Lane to reach out to as potential investment targets. You can select which attributes (visits, unique visitors, page views, category rank, etc.) are the characteristics you want to use as the model input.

Using whatever tool you wish, design and create an API script that will pull down company websites that could make interesting investment targets for Left Lane.

Left Lane can run the code / program on our end once the code is shared as we have the auth/secret on our end but are not allowed to share outside of our organization.

**PROMPT 2: APPTOPIA DATA API SOURCING**

Like Prompt 1, this prompt leverages the Apptopia Data API to surface interesting investment opportunities within the mobile app ecosystem.

Using the [Apptopia API](https://dev.apptopia.com/) come up with a tool that will surface promising companies for Left Lane to reach out to as potential investment targets. You can select which attributes (installs, rank, revenue, MAUs, etc.) are the characteristics you want to use as the model input.

Using whatever tool you wish, design and create an API script that will pull down apps that could make interesting investment targets for Left Lane.

Left Lane can run the code / program on our end once the code is shared as we have the auth/secret on our end but are not allowed to share outside of our organization.

**PROMPT 3: OPEN ENDED “OTHER” DATA SOURCE**

If you have another data source in mind that you think would uncover interesting investment opportunities within the consumer internet industry, please outline and design a sourcing tool to leverage this dataset.

Assuming Left Lane does not currently have access to this dataset, please share the way you would go about interfacing with the dataset and if you have access, share the output of the model you build.

**PART 2: COMPANY DATA ANALYSIS**

Using the below raw transaction file, please analyze and draw out insights that you think are important to consider in an investment process. You are welcome to use any analysis tool (Python, R, SQL, Alteryx, Tableau, PowerBI, etc.) you wish but please share the underlying workflow along with the output.

The aim of this prompt is to understand how you attack a dataset and gain insights from it to better understand a company and its customer’s behavior. Please draw out insights around growth, retention, user behavior, concentration, predictive behavior, etc. along with any other things you think are important to point out.

Please share a set of visualizations of the outputs along with qualitative context around the visualizations providing your opinion and takeaways.

Raw file available for download [here](https://drive.google.com/file/d/1GXfFHWTtdw1RwJNgdeiCFrnPRCVKSLFB/view?usp=sharing).

The transaction file is from a real money gaming company that runs games for people to “buy into” on their phones and win the prize pool if they are successful in the game.

**Raw file column overview:**

User\_id = the unique identifier of the end user

Registration\_timestamp = the DateTime of when the User\_id first signed up for the platform

Transaction\_timestamp = the DateTime of when that specific transaction record occurred

Transaction\_type = the string identifier of the record type.

Order = order within a specific game

Deposit = dollar deposit into the platform to play a game

Prize claimed = dollar amount won from winning a game

Transaction\_value = dollar amount of that record’s specific order, deposit, or prize won

Transaction\_revenue = dollar amount of end revenue to the Company for that specific transaction

State = state identifier of where the transaction took place (where the UserID is registered geographically)

**PART 3: INTERNAL AUTOMATION & TOOLING**

Please select at least **ONE** of the following prompts **AND** come up with one other idea of your own that you think could help the Left Lane team automate and drive further efficiency within the sourcing, analysis, and post-investment workflows.

**PROMPT 1: LEFT LANE COMPETITIVE BENCHMARKING**

Left Lane ingests data from 1,000-1,500 companies every year and has one of the most complete private company datasets in the world. But this data sits across disparate folders on our shared Drive which makes quick, automated, and accurate benchmarking a manual task.

Detail how you would go about building an internal procedure, aggregation / automation tool, hosting platform, and overall process to pull together this data across companies, industry, stage, etc. Outline where this data would sit and how Left Lane could interact with it on an ongoing basis to pull competitive sets on growth, financials, etc.

**PROMPT 2: AUTOMATED RAW DATA ANALYSIS AND VISUALIZATION TOOLING**

For almost every company Left Lane looks at we pull a raw transaction file (much like the one in Part 2 of this case study) and use Alteryx to summarize the data. These summaries / outputs must then be shared in visual format to the entire team so it can be easily digested and discussed during our investment committee. This basic process is repeated dozens of times each week by each team member. This process takes 1-3 hours depending on the complexity of the company’s data and the depth of analysis we are looking to do.

We are looking to build automated tooling across various steps of this process. From the data transfer from the company to the raw data cleaning / cleansing to the common summaries (growth, retention, etc.) that we want to see for every company.

Please outline and design ways you would automate these steps to drive efficiency and save time across the Left Lane investment process.

Examples of the steps that could be automated:

1. Ingestion / transference of the data from the company – what is the most efficient way for the company to share data? Where should this data sit?
2. Cleaning / standardizing of the data – all transaction files are slightly different (column order, field names, data structure, etc.). How can we automate this standardization?
3. Analysis – build the top 5-10 analysis items that we look for in every company (growth, retention, etc.)
4. Visualization – automate the visualization of this analysis