Technical Interview – Analytics Engineer at Treasure Data

General Overview:

1. **Candidate Presentation** (take-home) - 20-25 mins
2. **Q&A after the presentation** - 5-10 mins
3. **SQL hands-on exercise** (live coding) - 10-15 mins
4. **Conceptual Data Questions** (focusing on data pre-processing / ETL pipeline type of methods and data-transformations) - 10-15 mins

Guidelines for Each Section:

SECTION 1:   Candidate Presentation (take-home) - 20-25 mins

For this exercise, please choose **ONE** of the two options below:

OPTION 1

Present your own Dashboard Project:

Pick a project you have worked on (could be for school or work - if you are comfortable sharing actual metrics and data-context or have a way to anonymize or generalize the data without sharing sensitive customer information). Project should have involved building an ETL pipeline that powers a Dashboard that can be presented to business users / decision makers. Prepare a presentation deck as though you would be presenting this in front of a client (***assume audience will be less technical and more business oriented, so try to keep it simple and boil down complex tech processes in to easy-to-understand to a non-technical person format***), which should cover the points below:

Presentation Structure Guidelines

- Briefly summarize the purpose of the dashboard and the Insights that you plan to summarize in the final production-ready widgets ***(2-3 mins, 1 slide)***

- Briefly summarize the context of the input tables used to build the data model that powers the final dashboard and the important columns that the end user needs to be aware of in order to understand dashboard output better *(2-3 mins, 1 slide)*
- Mention any key data-cleaning, JOINs, Aggregations, custom-field calculations, filters etc. that you had to perform during the data pre-processing steps before you send the clean, production-ready tables to final data-model for the dashboard *(2-3 mins, 1 slide)*
- The rest of the deck should include screenshots of the important widgets of the dashboard and a walkthrough of the important metrics and insights, displaying any interactive widgets or ability to apply global filters (if such exist in the dashboard). Please, share your own observations on important business metrics and patterns that you observed in the data and provide recommendations on how the end-user / client can benefit from the dashboard. *(10-15 minutes, can include as many slides as needed, depending on how many widgets your dashboard contains and how you want to group those widgets. Just make sure presenting the slides will fit into the suggested 10-15 min window. If entire dashboard is very large, just pick some of the most important metrics to present)*

OPTION 2

___Build and present Dashboard using sample data provided by TD-team:___

*If you do not have a project from school / work you can share, please use the dataset attached and the instructions below to build a dashboard using a tool of your choice **(free trial version of Tableau, Power BI or any other popular BI tool or using simple PowerPoint graphs, Python visualization libraries etc.)***

Please follow the same **Presentation Structure Guidelines as in OPTION 1 described above.

The Task

Treasure Bikes is Treasure Data's new product line, capitalizing on the growing popularity of bicycles. To boost sales in the new year, Treasure Bikes' marketing team would like to send promo codes for a free helmet with each purchase. However, there is a cost to send out codes to each user. Thus, the marketing team has to be strategic on **which segment of the market to send this promotion to**.

With the provided data, design and put together a dashboard that you think will give marketers the most valuable information to pick which segment of their users to push this promotion to. You may use any dashboarding tool you like, and if you do not have access to one, you may put together images of graphs onto a single PowerPoint slide.

Do keep in mind that we're aware you might not be a marketing expert. What we're looking for is your intuition when it comes to making presentations and how you communicate your findings and your work. Feel free to come up with your own metrics and explanations for why you think a certain segment of the market is best served by this promotion. In addition, do show any

additional customer or business insights extracted from the data you think will be useful to the Treasure Bikes marketing team.

The Data

Provided is a `.csv` file with 20,000 rows of pageview data from the Treasure Bikes website with 12 columns. You may choose to do your preprocessing however you feel most comfortable.

1. **td_client_id** - this is the 3rd party cookie from the web-browser (associated to the specific device/browser user is accessing the site with, so the same person can have multiple td_client_id associated with them if they accessed the site from multiple devices/browsers)
2. **td_path** - the path to the page this row of data refers to (comes from URL)
3. **td_url** - the **full** URL, including UTM parameters (more on UTM below)
4. **td_referrer** - referral-page url = page visited right before current URL
5. **td_host** - the host site (these will all have the same value)
6. **td_browser** - the browser the user accessed the site with
7. **td_ip** - the user's IP address (a way to geolocate users)
8. **td_os** - the user's operating system
9. **td_screen** - the user's screen dimensions
10. **td_language** - the main language detected in browser / web page
11. **td_canonical_id** - the user's Treasure Data Unique ID (assigned after an ID-unification algorithm is applied to the original data)
12. **time** - the timestamp, in Unix time

UTM Parameters

Within the values in td_url, you'll notice there are parameters that extend past the webpage URL that you'll want to use as features in your dashboard. These are [UTM parameters](#) (follow the hyperlink for more information). For example:

https://www.treasurebikes.com/cart/?utm_source=google&utm_campaign=retargeting&utm_medium=search

---

SECTION 2:   Q&A After Presentation - 5-10 mins

Be ready to anticipate and answer questions from the TD team during and after your presentation about any of the slides you presented. You can expect both technical and business insights related questions, as though they are coming from the actual client team you prepared the presentation for.

SECTION 3:   SQL Hands-on Coding Exercise (live coding) - 10-15 mins

---

You will be presented with some simple tables that might require certain types of JOINs and other transformations and aggregations and asked to write a few SQL Queries that output the data in a certain format. You can just do this in any text editor of your choice. So, please have one open and ready to be used and screen-share during the interview.

SECTION 4:   Conceptual Data Questions (live coding) - 10-15 mins

---

This section is meant to demonstrate your ability to think strategically and problem-solve creatively, so no code-writing will be required. You might be presented with some example tables or some messy data and asked how you would approach certain data pre-processing / ETL pipeline type of tasks and accomplish an end goal. Some questions could also be in the form of brainteasers or just a brainstorming, project-planning type conversation.

For any clarifying questions on the tasks above feel free to email the people below directly:

Dilyan Kovachev: dilyan.kovachev@treasure-data.com

Yish Lim: yish.lim@treasure-data.com

Thank you for your time and we're looking forward to seeing your presentation and chatting to you soon!