Technical Assessment

As discussed, we will have a face-to-face interview, including a presentation and demonstration of the assignment.

Analyzing the website behaviors of consumers is fundamental in digital marketing. Due to confidentiality, however, we cannot disclose the internal data for this assignment. Instead, we use the publicly available dataset that can infer similar business questions:
https://www.kaggle.com/lipann/prepaired-data-of-customer-revenue-prediction

Please carefully read the following business questions, approach, and presentation method for your preparation.

Assignment: Predicting "**totals_transactionRevenue (at least bigger than 0: >0)**" from Prepaired_data_of_customer_revenue_prediction dataset.

Business questions (provided by the interviewer)

- Can we identify customer generating revenue, that is, "**totals_transactionRevenue (at least bigger than 0: >0)**", from the given dataset regarding the website behaviors?
- What are the critical factors in predicting "**totals_transactionRevenue (at least bigger than 0: >0)**"
- We assume the precisely targeted communication triggers the revenue. What will be critical insight to develop the marketing strategy in terms of customer segmentation and precision targeting in case we want to increase the number of "**totals_transactionRevenue (at least bigger than 0: >0)**"?
  - Whom to communicate – customer demographics or device types, etc.
  - When to communicate – does the time of visit matter?
  - What to communicate – can we identify which type of contents influence?
  - How to communicate – which communication channels?

Presentation

We will use CRISP-DM (Cross-industry standard process for data mining) as a framework for this assignment. Structure your presentation according to each step of CRISP-DM.

- Introduction of the business questions
- Data understanding, EDA, and necessary visualization
- Data preparation (preprocessing including necessary joins of multiple data sets given)
- Connect data using the Kaggle API command
- Data Modeling (classification problem, even though the original competition was designed for regression problem)
- Evaluation (F1, ROC curve, AUC, etc.)
- Can you answer one or more business questions described above?
- Did you find any other interesting insights from the data set provided?

Additional questions that you will answer

- SQL expertise
- Any challenges regarding orchestration/integration of different data sources
- Preferred programming languages and data management platforms when using data sources' API

Method of presentation and demonstration

- Choose any convenient presentation tool for your presentation: MS Powerpoint, PowerBI, or Google Colab, etc.

Tips

- Find column descriptions from Google Analytics Help. (https://support.google.com/analytics/answer/3437719?hl=en)
- If you want to check the original competition, find this link http://kaggle.com/c/ga-customer-revenue-prediction/overview/description.
- In case of unclear data definition or instruction, make your own and best assumption and complete the task.
- If the data size is too large for you to handle, make a slice and complete the task.
- You don't need to spend time on presentation formatting. We are more interested in your motivation of decision in the process of CRISP-DM.
- If you have any further questions, feel free to contact the hiring manager.

Time allocation

- Attendees introduction – 5 min
- Interviewee's presentation, Q&A, discussions - 25 min
- General Interviews – 20 min

Interviewers' information

- Hiring manager
- Future colleague such as a data scientist (To be informed in advance)
- Other marketing management team such as Digital Ads manager, HR, etc. (optional)