# Talabat - Analyst Assessment

## Part 1 - Analysis

1. The marketing team wants to understand our vendors a bit better. Their questions are:

a) Is there any order imbalance per vendor in each market?

b) Could we identify some relevant segments of vendors?

c) Are there any other interesting insights you would share with them?

Use the orders data provided. Given the request and the available data:

1. With a tool of your choice (R/Python, SQL etc), use the shared data to build an analysis that could answer their questions.
2. Aside from the available data, what other features would you include for an effective segmentation?
3. How could the marketing team use this output to improve their marketing performance?

*Please share your workings and assumptions made. The goal of this task is to understand how you approach solving a problem with data.*

2. One of our analyses revealed that 30% of the new users ordering for the first time from popular brands (such as McDonalds/BurgerKing) never placed a second order on our platform.

At a first glance, this seems rather high and we want to do a second analysis to better understand these users. To give a direction for the new analysis, we would need to start by compiling a list of assumptions about these users.

a) Could you think about 1-2 assumptions you would like to focus on?

b) What kind of data would you look at in order to test these assumptions?

* Don’t worry too much about which data we have/don’t have about the user, for the moment feel free to think about any possible assumptions.

3. Your colleague ran a linear regression analysis to understand the effect of talabat vouchers on the 1st visit on the total number of sessions (the training data was consolidated on a user level, with voucher = binary variable, with values 0/1 and target variable = total number of orders during the entire lifetime of each user). The results showed an $R^2$ = 0.6 and the estimated coefficient for the voucher was 5.

He needs help interpreting these results and checks in with you before presenting the results to the shareholders. What would you advise?

4. The management team wants to accelerate conversion of customers who placed an order only on food to new verticals recently launched: grocery & tmart (talabat-owned grocery shops).

a) What approach would you suggest considering that there is a small portion of customers who have converted already from food to new verticals?

b) Please suggest a way you could provide a set of insights to our CRM team for setting up new experiments that can validate your approach?

5. One of our brands, OTLOB, was recently merged to be part of talabat. This required a migration for part of the OTLOB existing customer base onto the new platform. Management would like to get a view on how customer behaviour evolved after the migration.

a) What dimensions & metrics would you use to monitor the pre/post migration analysis?

b) How would you present the insights results to management? Feel free to create dummy data to demonstrate how you would tell the story.

## Part 2 - SQL

This task is based on the 'Google Merchandise Store', a real ecommerce store that sells Google-branded merchandise. The questions you'll be answering should provide some insights on the performance of the store and how users interact with it.

You will be using **Google BigQuery** to investigate a data sample of one day (2017-08-01). To be able to view and query the data you'll need to create a new project, and use the following table to answer the questions:

Dataset: bigquery-public-data:google_analytics_sample

Table: ga_sessions_20170801

Before you start, here are some links and references that you might find useful::

Google BigQuery Console

BigQuery Export schema

Some documentation on standard and legacy SQL

Note: Feel free to use either Standard or Legacy SQL dialect to answer the questions.

Please try to follow this format when answering the questions:

> Question {question number}: {The question you are answering}
>
> Query:
> {The query you used to answer the question}
>
> Answer:
> {Your answer}

It's up to you to decide on the best way to answer the question, it can be tabular data, a visualisation, commentary points or whatever you think is the best way to present your insights. We value clarity of response over the fanciness of a solution.

Task:

1. How many unique users visited the website that day?
2. How many sessions did users start on average?
3. What is the website's conversion rate? *(To calculate conversion rate do not consider more than one transaction per session)*
4. What is the conversion rate per traffic medium?
5. How many minutes on average did it take the users to reach the checkout confirmation page (title = 'Checkout Confirmation')?

6. The website tracks add and remove from cart interactions using two events (Action = 'Add to Cart' and Action = 'Remove from Cart' ). Please use that information and investigate the data further to answer the following questions:

   a. How many products in total did users add to cart?

   b. In sessions where users actually added products to their carts, how many products were added per session on average?

   c. What is the most added-to-cart product and what is the most removed-from-cart product (mention the product name)?

Bonus questions (optional):

1. Did you find any insights / observations while you were investigating the data that you think are interesting to share?
2. Do you have any thoughts, notes, doubts or suggestions on the data structure?