

# Take Home Test

## Data Analyst

As a Data Analyst, you will be responsible for analyzing data, and working with various teams to help understand business and customer problems and solutions to solve them.

This test is for us to gauge your analytical ability in a setting close to how you would be working at GYG. The data has been anonymised, but please do not share, post this to any forum, or put it out in the public domain; this hinders our ability to hire new analysts.

Your time is your own, however we would not recommend spending more than 4-5 hours on this task. Our expectations of what you will provide are based on this time frame. For example, we are not looking for you to train a deep learning model.

Choice of tools used is entirely at your discretion, insights matter more than tooling. However we value ability in SQL, Python, Scala, or R.

We have provided definitions for most fields in the .csv files provided. If a definition does not exist for a particular field we want you to interpret the meaning of this field.

### **When should I deliver?**

Please deliver this within 72 hours of receipt.

### **What format?**

Please provide your thoughts/proposal in English in whatever format you choose (Google Slides, PPT, Word, Google Docs, PDF, etc.)

## Task 1

### [Data](#)

You have been given a data set which contains bookings. You can find the definitions of fields in the appendix below.

1. Clean the dataset. Point out anything you find, that required cleaning.
2. Plot a count of bookings over time. Point out anything you notice about the timeseries.
3. Analyse our activity categories. Tell us anything interesting you find about how they differ in net revenue (NR) and gross merchandise value (GMV). Are there other differences between categories that you notice?
4. Some devices have unique booking behaviour. What is unique about device\_ids 4 and 5?

## Task 2

### [Data](#)

This data set is based on our shopping cart data. We would like you to discover an insight into our business using this data set. You can take your analysis in any direction you like. We are looking to understand how you think about the business and how you would analyse a data set with no specific question being asked.

Tip: A good place to start is to think of a business question based on the data provided.

## Task 3

Based on the insight from Tasks 1 or 2, write a short description of an experiment that you would like to run. Please include the details of the experiment, including populations that you are testing, the metrics you would use to evaluate the experiment, and any other relevant information about your experiment design.

## Task 4

Our Customer Relationship Team occasionally asks us to generate lists from our booking data to send emails to our customers.

They ask you to generate a list of customer\_ids and first checkout date for customers who have made their second checkout within 30 days of their first check out. Using the fact\_checkout table below, write the SQL needed to generate the list.

Using the fact\_checkout table below, write the SQL needed to generate the list.

## Appendix

### Question 1 Data Definitions

#### Table Name: fact\_orders

order_id	Uniquely identifies a shopping cart.
booking_id	Uniquely identifies a booking.
customer_id	Uniquely identifies a customer
net_revenue	This is a calculated metric that equals GMV less our basic costs
gross_merchandise_value	The Gross Merchandise Value (gmv)
date_of_checkout	Indicates the date and time of checkout. Format is: YYYY-MM-DD HH:MM:SS
date_of_travel	Indicates the date when activity will take place. Format is: YYYY-MM-DD HH:MM:SS
activity_category	Our groupings of activities.
device_id	Uniquely identifies a device. Device refers to technology devices, such as desktop, mobile and tablet

### Question 2 Data Definitions

#### Table Name: fact\_checkout

date_of_creation	Indicates the creation date. Format is: YYYY-MM-DD HH:MM:SS.
date_of_checkout	Indicates the date and time of checkout. Format is: YYYY-MM-DD HH:MM:SS
bookings	Indicates the total number of bookings associated with a shopping cart. A shopping cart can have one or more bookings.

<b>purchase_type_id</b>	Uniquely identifies the type of purchase. Typical purchase types include 'Repeat' which indicates a repeat customer, and 'Same Trip' which indicates the shopping cart is for the same trip.
<b>is_fraud</b>	Indicates if fraud is involved
<b>week</b>	Indicates the week of creation.
<b>tour_category_id</b>	Uniquely identifies a tour category.
<b>date_of_travel</b>	Indicates the date when activity will take place. Format is: YYYY-MM-DD HH:MM:SS
<b>commission_rate</b>	A calculated metric that equals: NR divided by GMV. This is a crucial metric for GYG because it defines what portion of the GMV we retain.
<b>shopping_cart_id_a</b>	Uniquely identifies a shopping cart. Shopping carts are also commonly referred to as baskets or orders.
<b>customer_id_a</b>	Uniquely identifies a customer
<b>customer_address_id_a</b>	Uniquely identifies a customer address
<b>billing_id_a</b>	Uniquely identifies a billing. You can think of the billing as the 'transaction' associated with payment for bookings in the shopping cart. That is, a billing is essentially a shopping cart that has been paid for.
<b>coupon_id_a</b>	Unique identifier of a coupon, if applied.
<b>reseller_id_a</b>	Uniquely identifies a reseller. Resellers are also commonly referred to as 'partners'
<b>reseller_campaign_id_a</b>	Uniquely identifies a campaign associated with a reseller
<b>gyg_campaign_id_a</b>	Uniquely identifies a GetYourGuide (gyg) campaign.
<b>currency_id_a</b>	Uniquely identifies the currency
<b>shopping_cart_status_id_a</b>	Uniquely identifies a status for a shopping cart
<b>coupon_segment_id_a</b>	Uniquely identifies a coupon segment. A coupon segment is a grouping of a set of coupons
<b>attribution_id_a</b>	Uniquely identifies an attribution. Attribution indicates the marketing channel from which the booking originated
<b>device_id_a</b>	Uniquely identifies a device. Device refers to technology devices, such as desktop, mobile

	and tablet
<b>exchange_rate_a</b>	Indicates the multiplication factor for converting to euros. If no conversion is required then this column has a 1.
<b>gmv_a</b>	The Gross Merchandise Value (gmv)
<b>gmv_after_coupons_a</b>	The Gross Merchandise Value (gmv) expressed in the local currency
<b>visitor_id_a</b>	Uniquely identifies a visitor to a site or app. A potential customer is considered a visitor until they either register with us or purchase a booking.
<b>session_id_a</b>	Uniquely identifies a user's session. A session is the period between a user opening a GYG web page or mobile app and ending with a logoff or close of the browser/app