Data Analyst Intern - Assignment **Part 1 : SQL**

**Instructions:**

You’ll be working with BigQuery in ***Sandbox*** mode to test your SQL skills. Steps:

1. Go to BigQuery Sandbox page and click Go to BigQuery , and wait for the project to be created. (You might need to sign up with Google prior to this).

(If you are a first-time BigQuery user, you’d be greeted with a Welcome dialog. Tick the checkbox and click “Agree and Continue”. In the Create Project page, make no changes and click “Create”.)

2. On the left, type “**bigquery**” in the search bar, and click “Broaden search to all projects”. Hover over ‘**bigquery-public-data**’ and click the icon to pin it, and click to go back. Now expand ‘***bigquery-public-data****’* and look for *‘****thelook\_ecommerce****’.*

3. **Thelook\_ecommerce** is the database which is to be utilised for this session, and it consists of 7 tables. Go through the schema to understand the DB.

4. On the right, is the Editor which is the SQL workspace area.

5. You can query the table by replacing <table\_name> with the required table, in the FROM clause of the query. **`bigquery-public-data.thelook\_ecommerce.<table\_name>`**

This dataset describes an eCommerce clothing site’s database. The dataset contains information about customers, products, orders, logistics.

**Mandatory: Define SQL Queries which are entirely automated *(i.e. the query should be framed in such a way that it prints out the required answer directly)***

1. What is the number of orders that got ‘Cancelled’ last year?

2. What are the top 5 best selling ‘Jeans’ brands?

3. We are planning on giving out gift vouchers to the top 10 spending customers. Identify them.

4. What is the Total Sale Price of ‘Fashion Hoodies and Sweatshirts’ purchased by Chinese female customers?

5. The brand ‘Allegra K’ is running their Global Ad campaign on Facebook. What’s the impact of their ad campaign in the US, talking in terms of the number of total products sold?

6. Which category of products has the second most unsold items? *(make use of the ‘sold at’ column from inventory\_items)*

7. Which distribution centre location houses the highest number of products from the ^above obtained product category?