**Data Engineering Intern Assignment Bureau.id**

**Problem Statements:**

SQL: Database Schema on Page 2 (Just mention the logic and SQL)

Consider dummy data

1. List of booking id that was shipped by Air with departure date in Jan 2021
2. List the number of bookings by each mode of shipment id with departure date in Jan 2021
3. List the bookings (by booking\_id) invoiced to FurniturePlus (customer id: 214598) that departed in Jan 2021
4. List the highest valued 10 invoices (by invoice\_id) invoiced in GBP that were sent in Jan 2021
5. Return the list of customer names that had more than 10 unique invoices sent in Jan 2021.
6. Return the the list of legs and the amount invoiced to each leg in invoice currency of FurniturePlus’ (customer id: 214598) first ever created shipment. Hint: the integer id in the booking table is unique and auto-increment.
7. Return the list of the top 10 customers, in descending order, by sales (in invoice currency) invoiced in GBP in Jan 2021. In addition, for each of these customers, return the next customer (by sales in invoice currency) from in the same country, the amount of sales and the difference in sales between the current and the next customer from the same country. Hint: columns needed:
   * customer\_name
   * total\_sales
   * lag\_customer\_name\_same\_country
   * lag\_total\_sales\_same\_country
   * difference\_in\_sales

**Note: Please find attached .sql file for solution.**

***Python:***

1. Write a python function which reads json data from public API ( [api.coindesk.com/v1/bpi/currentprice.json](https://api.coindesk.com/v1/bpi/currentprice.json)) and do below
   1. Flatten the data in meaningful manner
   2. insert the data into table
      1. Keep a CSV file instead for the table.
2. Create a Analysis report on the above data basis your inferences.

**Note: Please find attached .ipynb and .pdf files for solution.**

