### **Data Engineering Assignment - The Sleep Company**

**Problem Statement 1:**You are given a dataset containing order details for a D2C (Direct-to-Consumer) company. The dataset includes information about the purchase date, order ID, user ID, product category, order value, and quantity.

**Objective:**The goal of this assignment is to assess your ability to perform data transformation tasks that are commonly required in real-world data engineering scenarios. You will be provided with a raw dataset, and your task is to apply specific transformation logic to achieve the desired output.

**Raw Data Sample:**

| Date | OrderId | User Id | Category | Value | Qty |
| --- | --- | --- | --- | --- | --- |
| 13 Aug | 324324 | CXZCS | Mattress | 34999 | 1 |
| 12 Aug | 435345 | FSDSD | Mattress | 28999 | 2 |
| 1 Aug | 453534 | SDFSF | Bed | 68999 | 1 |
| 11 Aug | 322323 | DDFBV | Mattress | 48999 | 1 |
| 21 Aug | 767688 | FFWEW | Bed + Mattress | 106899 | 1 |
| 21 Aug | 424234 | FWRRE | Mattress | 19999 | 3 |
| 9 Aug | 324234 | HRTHR | Bed | 68999 | 1 |
| 8 Aug | 332423 | ERGEE | Mattress | 48999 | 2 |
| 9 Aug | 345435 | GGDBS | Mattress | 28999 | 1 |
| 18 Aug | 243224 | AGSER | Bed + Mattress | 112999 | 1 |
| 17 Aug | 675656 | HTHSD | Bed | 72999 | 2 |
| 23 Aug | 564565 | WEFAA | Mattress | 28999 | 1 |

**Task:**

* **Transform the Data:**
  + **Split the "Bed + Mattress" Category:**
    - For any record where the Category is "Bed + Mattress," split it into two separate records: one for "Bed" and one for "Mattress."
    - Allocate the total "Value" between the two categories, with 80% assigned to "Bed" and 20% to "Mattress."
  + **Retain the Order Information:**
    - Keep the original OrderId, User Id, and Date the same for both records after splitting.
* **Output Data:** The transformed data should look like the example below:

| Date | OrderId | User Id | Category | Value | Qty |
| --- | --- | --- | --- | --- | --- |
| 13 Aug | 324324 | CXZCS | Mattress | 34999 | 1 |
| 12 Aug | 435345 | FSDSD | Mattress | 28999 | 2 |
| 1 Aug | 453534 | SDFSF | Bed | 68999 | 1 |
| 11 Aug | 322323 | DDFBV | Mattress | 48999 | 1 |
| 21 Aug | 424234 | FWRRE | Mattress | 19999 | 3 |
| 9 Aug | 324234 | HRTHR | Bed | 68999 | 1 |
| 8 Aug | 332423 | ERGEE | Mattress | 48999 | 2 |
| 9 Aug | 345435 | GGDBS | Mattress | 28999 | 1 |
| 17 Aug | 675656 | HTHSD | Bed | 72999 | 2 |
| 23 Aug | 564565 | WEFAA | Mattress | 28999 | 1 |
| 21 Aug | 767688 | FFWEW | Bed | 85519 | 1 |
| 18 Aug | 243224 | AGSER | Bed | 90399 | 1 |
| 21 Aug | 767688 | FFWEW | Mattress | 21380 | 1 |
| 18 Aug | 243224 | AGSER | Mattress | 22600 | 1 |

**Submission Requirements:**

Code**:** Provide the SQL code used to perform the transformation.

**Problem Statement 2:**

**Raw Data:** Please find the raw data - [here](https://docs.google.com/spreadsheets/d/1rIHu3Hd1GYIYHIKVjXA7Frr20Zy9GGUkNbYqouE_VL0/edit?gid=2042600181#gid=2042600181)

**Task:** Based on the data shared, write and share SQL queries to answer the following questions:

| 1. Total average monthly revenue by industry. |
| --- |
| 2. Top 2 users by industry based on overall revenue |
| 3. For each user, get the current month, previous month, and previous -1 month revenue and order count. |

**Submission**:

Code: SQL queries for each of the questions

Feel free to reach out (call) if you have any questions

**All the best :)**