🧪 **Intermediate Data Engineering Task: Cleaning, Deduplication & Normalization**

Objective: Clean and prepare a customer dataset by removing duplicates, validating fields, and normalizing inconsistent entries.

📄 **Raw Data Table:**

Customer ID	Full Name	Email	Join Date	Country	Age
301	Ivan Ivanov	ivan@email.com	2021-03-15	Bulgaria	34
302	Maria Georgieva	maria@email.com	2021-03-15	bulgaria	34
303	Ivan Ivanov	ivan@email.com	2021-03-15	BG	34
304	George M	george@email	2022-07-01	Greece	40
305	Elena Georgieva	elena.g@email.com	2022-07-01	GREECE	
306		elena.g@email.com	2022-07-01	Greece	35
307	Petar Georgiev	petar@email.com	2023-01-20	Bulgaria	28

🖌️ **Tasks to Perform:**

1. **Remove duplicate customers**
    - If Full Name, Email, and Join Date are identical → keep only one record
2. **Normalize country names**
    - Convert all country values to consistent format: Bulgaria, Greece
3. **Validate email addresses**
    - Remove rows with invalid emails (missing @ or domain)
4. **Remove rows with missing critical data**
    - If Full Name, Email, or Age is missing → discard the row
5. **Standardize column names**
    - All lowercase, spaces replaced with underscores
6. **Output a cleaned table**
    - Only valid, unique, and normalized records

💡 Bonus Challenge (Optional): Add a new column called customer_segment based on age:

- age < 30 → "Young Adult"
- 30 ≤ age < 45 → "Adult"
- age ≥ 45 → "Senior"

🧹 **Задача по обработка на данни – Ниво „Intermediate“: Почистване, премахване на дубликати и нормализация**

Цел: Да се почисти и подготви таблица с клиентски данни чрез премахване на дублирани записи, валидиране на полета и нормализиране на несъответстващи стойности.

📄 **Сурови данни:**

ID на клиент	Име и фамилия	Имейл	Дата на регистрация	Държава	Възраст
301	Ivan Ivanov	ivan@email.com	15.03.2021	Bulgaria	34
302	Maria Georgieva	maria@email.com	15.03.2021	bulgaria	34
303	Ivan Ivanov	ivan@email.com	15.03.2021	BG	34
304	George M	george@email	01.07.2022	Greece	40
305	Elena Georgieva	elena.g@email.com	01.07.2022	GREECE	
306		elena.g@email.com	01.07.2022	Greece	35
307	Petar Georgiev	petar@email.com	20.01.2023	Bulgaria	28

🖌️ **Задачи за изпълнение:**

1. **Премахване на дублирани клиенти**
    - Ако *Име и фамилия*, *Имейл* и *Дата на регистрация* са идентични → запази само един запис
2. **Нормализиране на имената на държави**
    - Всички стойности да бъдат уеднаквени: Bulgaria, Greece
3. **Валидиране на имейл адреси**
    - Премахни редове с невалидни имейли (липсва @ или домейн)
4. **Премахване на редове с липсващи ключови данни**
    - Ако липсва *Име и фамилия*, *Имейл* или *Възраст* → изтрий реда
5. **Стандартизиране на имената на колоните**
    - Всички с малки букви, интервалите заменени с долни черти
6. **Създай нова почистена таблица**
    - Само валидни, уникални и нормализирани записи

💡 **Бонус предизвикателство (по избор):** Добави нова колона customer_segment според възрастта:

- възраст < 30 → „Млад възрастен“
- 30 ≤ възраст < 45 → „Възрастен“
- възраст ≥ 45 → „Възрастен човек“