## ****Завдання для перевірки навичок Data Analyst/Data Scientist****

### ****Тема:**** Аналіз продажів та прогнозування кількості проданих одиниць товару

### ****Мета:****

Провести аналіз даних про продажі, візуалізувати основні тренди та побудувати модель для прогнозування кількості проданих одиниць товару. Кандидат повинен використовувати надані дані для виконання завдання.

### ****Надані дані:****

1. **train\_data.csv**:
   * **Опис**: Цей файл містить історичні дані про продажі з детальною інформацією про кожну транзакцію, включаючи дату, код товару, опис, кількість проданих одиниць, ціну за одиницю, ідентифікатор покупця та країну.
   * **Мета**: Використовуйте цей набір даних для аналізу та побудови моделі прогнозування.
2. **test\_data\_without\_target.csv**:
   * **Опис**: Цей файл містить аналогічні дані про продажі, але без таргетної змінної Quantity. Це означає, що ви не маєте інформації про кількість проданих одиниць у цьому наборі даних.
   * **Мета**: Після побудови моделі прогнозування, ви повинні використати її для передбачення значень Quantity у цьому тестовому наборі даних.

### ****Завдання:****

#### ****1. Обробка даних та аналіз:****

* Проведіть початковий огляд даних для розуміння структури та вмісту датасету.
* Здійсніть очищення даних: видаліть пропущені або аномальні значення, нормалізуйте дані, якщо це необхідно.
* Виконайте аналіз даних, щоб зрозуміти основні тренди та зв'язки між змінними. Наприклад:
  + Як ціна впливає на кількість проданих одиниць?
  + Чи є сезонні тренди в продажах?
  + Які товари продаються найкраще в різних країнах?

#### ****2. Візуалізація:****

* Створіть мінімум 2 візуалізацій, які наочно демонструють ключові інсайти з даних. Візуалізації можуть включати:
  + Розподіл кількості проданих одиниць за країнами.
  + Тренди продажів за часом (наприклад, по місяцях або днях).
  + Залежність кількості проданих одиниць від ціни.
  + Вплив інших змінних на кількість проданих одиниць.
  + Тощо.

#### ****3. Побудова моделі прогнозування (Опціонально):****

* Використовуючи train\_data.csv, побудуйте модель для прогнозування кількості проданих одиниць (Quantity). Модель може бути простою (лінійна регресія) або більш складною (градієнтний бустинг, випадковий ліс, тощо).
* Після побудови моделі, використовуйте її для передбачення значень Quantity у test\_data\_without\_target.csv.

#### ****4. Документування:****

* В ідеалі якщо код буде написано в форматі ipynb файлу, то добре було б писати як коментарі так і markdowns які будуть пояснювати, вашу логіку думок, висновки і рішення які ви прийняли в ході вирішення задачі.

### ****Необхідні навички та інструменти:****

* **Python** (використання pandas, matplotlib, seaborn, scikit-learn або інші відповідні бібліотеки для аналізу, візуалізації та моделювання).
* Навички роботи з даними, побудови візуалізацій і моделей.

### ****Дедлайн:****

Завдання має бути виконане до 02.09.2024 до кінця дня. Звіти та результати слід надіслати разом із кодом, який ви використовували для виконання завдання.

**На виході результатом виконання ТЗ буде .py або .ipynb файл, з кодом та описом дій, та опціональний файл з результатом предікту в форматі .csv або .xlsx як зручно.**