

1. رابط الـDataset.

• <https://www.kaggle.com/datasets/spscientist/students-performance-in-exams>

المشاكل التي وجدتھا في البيانات

- لم أجد أي قيم ناقصة (NaN).
- عدد من الصفوف كانت مكررة، وقمت بحذفھا.
- أسماء الأعمدة كانت تحتوي على فراغات وحروف كبيرة، فقامت بتعديلھا لتصبح موحدة (snake_case).

خطوات التنظيف التي قمت بھا:

- استخدمت drop_duplicates() لإزالة التكرارات.
- عدلت أسماء الأعمدة باستخدام قائمة فهمية لتحويلھا إلى lowercase وإزالة الفراغات.
- تأكدت أن أنواع البيانات مناسبة باستخدام df.info().

ملاحظات عامة:

- البيانات نظيفة نسبياً.
- يمكن استخدام هذه البيانات لتدريب موديلات تتنبأ بدرجات الطلاب حسب الخلفية الاجتماعية أو نوع المدرسة أو الجنس.