

Pandas و Python باستخدام حقيقية بيانات وتنظيف جمع: مشروع

### Dataset:

<https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction>

### البيانات على عامة نظرة:

- الصفوف عدد: 918
- الأعمدة عدد: 12
- (object) ونصية (int, float) عددية: البيانات أنواع

### عليها العثور تم التي المشاكل:

- يوجد لا: ناقصة قيم وجود
- يوجد لا: مكررة بيانات وجود

### التنظيف خطوات:

- `df.drop_duplicates()` باستخدام المكررة الصفوف حذف
- `df.astype()` باستخدام يلزم ما وتعديل البيانات أنواع من التأكد
- (توجد لا مبدئية بصورة) المتطرفة أو المنطقية غير القيم من التحقق

### ملاحظات:

- آلة تعلم نموذج تدريب أو لتحليل ومناسبة عام بشكل نظيفة البيانات
- حقيقية بيانات مع التعامل في مفيدة تجربة