

# تقرير تنظيف البيانات

## • رابط الـ Dataset:

<https://data.montgomerycountymd.gov/api/views/v76h-r7br/rows.csv?accessType=DOWNLOAD>

## • المشاكل التي تم اكتشافها:

- وجود قيم ناقصة (Missing Values) في بعض الأعمدة.
- وجود صفوف مكررة (Duplicate Rows) تكررت فيها نفس البيانات.

## • خطوات تنظيف البيانات:

- حذف الصفوف المكررة باستخدام الدالة `drop_duplicates()` لتقليل التكرار وتحسين جودة البيانات.
- حذف الصفوف التي تحتوي على قيم ناقصة باستخدام `dropna()` لتجنب مشاكل في التحليل لاحقاً.
- يمكن أيضاً (اختيارياً) تعويض القيم الناقصة بقيمة مناسبة، مثل المتوسط أو القيمة الأكثر تكراراً، حسب طبيعة العمود.

## • ملاحظات إضافية:

- تم فحص عدد الصفوف والأعمدة، وأنواع البيانات، قبل وبعد التنظيف باستخدام `shape` و `dtypes`.
- يمكن لاحقاً تحسين التنظيف أكثر عن طريق تغيير نوع البيانات لبعض الأعمدة إلى أنواع أكثر دقة (مثل `int`, `datetime`, إلخ).
- من الأفضل تحليل الأعمدة التي تحتوي على الكثير من القيم الفارغة قبل اتخاذ قرار بالحذف أو التعويض.