

تعد أهم المراحل في دورة حياة علم البيانات يجب على عالم البيانات تصحيح وتجهيز البيانات التي تم جمعها في مرحلة التنقيب للتأكد من مناسبةها لمرحلة التحليل اللاحقة وعند دمج البيانات من مصادر متعددة تزيد احتمالية تكرار البيانات أو تداخلها الأمر الذي يتطلب عملية تصحيح وتصويب لتلك البيانات وكذلك هو الحال إذا وجدت بيانات تالفة أو منسقة بشكل غير صحيح أو مكررة أو خاطئة أو حتى غير مكتملة تكمن أهمية تصحيح تلك البيانات فإن الرؤى أو الإستنتاجات المستمدة في مرحلة التحليل تلك البيانات ستكون خاطئة وسيصعب للغاية استنتاج ما إذا كانت المشكلة ناشئة من أخطاء في خطوات التحليل أو أن البيانات نفسها لم يتم تصحيحها ولهذا السبب فإن عملية تنظيف البيانات والتحقق من صحتها جيدا قبل تحليلها يعد أمرا مهما للغاية للعملية بأكملها