

شرح خطوات ال Data Cleaning :

1 - Include Libraries :

استدعاء مكتبة Pandas لقراءة ملفات CSV و JSON و EXCEL من اجل عملية تنظيف البيانات

2 - Read Data From files :

قراءة البيانات من داخل الملفات لمصفوفات للبدء بتنفيذ العمليات

3 - Know my Dataset :

التعرف على طبيعة البيانات التي سيتم العمل عليها من خلال طباعة بعض القيم بداية ال Data ومعرفة أسماء الاعمدة لكل Dataset

4 - Datasets Cleaning :

البدء بعملية تنظيف البيانات ، الخطوة الأولى كانت تهدف الى توحيد اسماء الاعمدة فقمنا باعادة تسمية الاعمدة لل Dataset الثانية والثالثة حسب ال Dataset الأولى ، بعد ذلك قمنا بحذف الاعمدة Unamed و Random Feature من ال Dataset الأولى والثالثة ، بعد ذلك تحققنا من عدم وجود قيم فارغة NaN في الصفوف للمصفوفات كاملة ، بعد ذلك انتقلنا لتنظيف كل dataset منفردة

5 - Dataset2 Cleaning :

طاعة صفين من محتوى Dataset2 و Dataset1 للتعرف على طبيعة البيانات ، بعد ذلك قمنا بمعالجة الاعمدة لاجل توحيد قيم البيانات لكليهما .

- Fuel Type Column :

تعديل محتوى العمود بحيث يتوافق مع ال Dataset الأخرى وتغيير نوع البيانات

- Paint Type Column :

تعديل محتوى العمود بحيث يتوافق مع ال Dataset الأخرى

- Doors Column :

تعديل محتوى الاعمدة لل Dataset1 و Dataset2 وتعديل نوع البيانات لها

- Weight Column :

تعديل نوع البيانات للعمود بحيث يتوافق مع الاعمدة الأخرى لل Dataset

5 - Dataset3 Cleaning :

طاعة صفين من محتوى Dataset3 و Dataset1 للتعرف على طبيعة البيانات ، بعد ذلك قمنا بمعالجة الاعمدة لاجل توحيد قيم البيانات لكليهما .

- Paint Type Columns :

تعديل محتوى العمود بحيث يتوافق مع ال Dataset الأخرى

6 - Get Shape of Datasets :

بعد الانتهاء من عملية التنظيف قمنا بالتأكد من ابعاد ال Datasets

7 - Drop Columns from Datasets :

حذفنا بعض الاعمدة من ال Dataset بعضها يمكن ان نستنتجها من أعمدة أخرى و العمود Age Group تم حذفها بسبب اختلاف قيم ال metadata لكلاهما وكذلك عدم وجودها في ال Dataset الثالثة

8 - Get Head From Datasets :

طاعة صفين من محتويات ال Datasets للتعرف على طبيعة البيانات

9 - Know Information After Cleaning Datasets :

عرض المعلومات عن ال Datasets الجديدة

10 - Integrate My Datasets :

دمج كل ال Datasets بعد عملية المعالجة

11 - Check my New Dataset :

التحقق من ال Dataset الجديدة بعد الربط حيث ظهرت قيم فارغة في عمود ال Vehicle Age

12 - Clean my New Dataset :

معالجة ال Dataset الجديدة حيث قمنا بعمل الوسط بدلا من القيم الفارغة وحولنا نوع البيانات للعمود

13 - Dataset After Cleaning :

إعادة أسماء الأعمدة الأصلية وعرض ال Dataset بعد عملية المعالجة

14 - Check My New Dataset :

التحقق بعد عملية المعالجة لل Dataset الجديدة

15 - Write Dataset to CSV File :

حفظ ال Dataset المعالجة في ملف CSV