

ransaction_ID	Customer_Name	Age	Email	Join_Date	Total_Purchase
1001	Ahmed Ali	28	<a href="mailto:ahmed@mail.com">ahmed@mail.com</a>	2025-01-10	250
1002	Sara Omar	NaN	<a href="mailto:sara@mail.com">sara@mail.com</a>	2025-02-15	300
1003	Ali Saleh	35	NaN	2025-03-20	150
1004	Nada Hassan	42	<a href="mailto:nada@mail.com">nada@mail.com</a>	NaN	400
1005	Omar Khalid	NaN	<a href="mailto:omar@mail.com">omar@mail.com</a>	2025-05-05	NaN
1006	Ahmed Ali	28	<a href="mailto:ahmed@mail.com">ahmed@mail.com</a>	2025-01-10	250

استنادًا إلى جدول البيانات أعلاه، استخدم مكتبة Pandas لكتابة سطر الكود المطلوب !:

Df['Join_Date']= Pd.to_datetime(df['Join_Date'])	تحويل عمود Join_Date إلى تاريخ.
Df[df.isnull().sum(axis=1) > 1]	تحديد الصفوف التي فيها أكثر من قيمة فارغة.
Df.info()	معرفة نوع البيانات وعدد الصفوف والأعمدة.
Df.isnull().sum()	تتحقق من عدد القيم الفارغة في كل عمود.
Df[df['Age'] > 30]	تحديد الصفوف التي العمر < من 30 سنة.
Remaining_rows = df.dropna().shape[0]	معرفة كم صف يتبقى بعد حذف الصفوف التي تحتوي على أي قيم فارغة
Df['Age']= Df['Age'].fillna( df['Age'].mean() )	استبدال القيم الفارغة في عمود Age بالمتوسط.
Df['Total_Purchase']= Df['Total_Purchase'].fillna(0)	استبدال القيم الفارغة في عمود Total_Purchase بالرقم 0.
Df.drop_duplicates(inplace=True)	إزالة الصفوف المتكررة.
Df[df.duplicated()]	معرفة الصفوف المتكررة قبل حذفها.

## أجب عن الأسئلة التالية:

- في جدول البيانات السابق أي الأعمدة في الجدول تحتاج تنظيف قبل التحليل؟ ولماذا؟

- (NaN) يحتوي على قيم فارغة **Age: عمود**
- (NaN) يحتوي على قيم فارغة **Total\_Purchase: عمود**
- لتسهيل عمليات (datetime) يجب تحويله إلى صيغة تاريخ **Join\_Date: عمود**
- الفرز والتحليل المتعلقة بالوقت

- ما أثر الصفوف المكررة على التحليل؟

- 1- تشويه الاحصائيات والوصف العام.
- 2- تظليل في مؤشرات الأداء.
- 3- انحياز في التقسيم.
- 4- التأثير على النماذج التنبؤية.
- 5- إعاقة فهم الحجم الحقيقي للبيانات.

- إذا كنت تريد تحليل متوسط العمر للعملاء، ما الحل الأمثل للتعامل مع القيم الفارغة في عمود Age؟

أفضل خيارين الاحلال بالوسيط

`fillna(df['Age'].mean())` للمحافظة على حجم العينة.

الوسيط أفضل لو التوزيع ملتوي.

أو إسقاط الصفوف التي تفتقد Age إن كانت قليلة جداً ولا تؤثر على حجم العينة(لكن ستقلل البيانات).

- لماذا من المهم تحويل عمود Join\_Date من نص إلى نوع تاريخ قبل أي تحليل؟

لتمكين عمليات زمنية صحيحة:

استخراج السنة/الشهر/اليوم ، الفرز الزمني ، حساب الفترات (Retention , Cohorts) ، تجميع شهري/ فصلي ، والتعامل مع القيم غير القابلة للتحويل ك NaT بدل نص.

- بناءً على الجدول، ما الملاحظات التي يمكنك استخراجها حول العملاء ومشترياتهم؟

- تاريخ انضمام العميل: يمكن تحديد أقدم عميل وأحدث عميل بناءً على عمود Join\_Date.
- Total\_Purchase سلوك الشراء: يمكن معرفة أكثر العملاء شراءً بناءً على عمود
- خصائص العمر: يمكن تحديد الفئة العمرية التي تنفق أكثر، أو معرفة متوسط عمر العملاء.
- النمو بمرور الوقت: يمكن تحليل عدد العملاء الجدد الذين انضموا في فترة زمنية محددة

























