

Users table contains duplicate record for the primary key column ['_id']

```
users['_id'].value_counts(dropna = False)
```

```
54943462e4b07e684157a532    20
5fc961c3b8cfca11a077dd33    20
5ff5d15aeb7c7d12096d91a2    18
5fa41775898c7a11a6bcef3e    18
59c124bae4b0299e55b0f330    18
..
6004a965e257124ec6b9a39f     1
600746fd6e64691717e8cfb5     1
60074b996e64691717e8f11a     1
60074246325c8a12289e22a0     1
60088e5d633aab121bb8e5cf     1
Name: _id, Length: 212, dtype: int64
```

brands table contains lot of missing values.

```
brands.isnull().sum()
```

```
_id          0
barcode      0
category    155
categoryCode 650
name         0
topBrand     612
brandCode    234
cpg_ref      0
cpg_id       0
dtype: int64
```

receipt_item table contains lot of missing values.

```
receipt_items.isnull().sum()
```

```
_id          0
barcode     3851
description  381
finalPrice   174
itemPrice    174
partnerItemId 0
quantityPurchased 174
discountedItemPrice 1172
originalReceiptItemText 1181
dtype: int64
```