|  |  |
| --- | --- |
| ASSUMPTION | SOLUTION |
| Each “Album”-entry corresponds to only one “Artist”. It turns out that some albums from different artists share the same name | Add a column to the main df to uniquely identify each album  data["album\_id"] = data.groupby(["Album","Artist"]).ngroup() |
| Each artist only has one album with the same name. In reality, some albums by the same artist have been re-edited and published in different years | Rewrite the “album\_id”-column to account for the year  data["album\_id"] = data.groupby(["Album","Artist",”Year”]).ngroup() |
|  |  |