

Data links:

1. The following links are the original 3 datasets and to be used for the first notebook which is in *Notebooks folder* -> **Data cleaning and processing**:

Transactions Dataset:

https://drive.google.com/file/d/1So5F6WkQvdZza3zaT1YeLBnAZZprNCqO/view?usp=share_link

Customers Dataset:

https://drive.google.com/file/d/1Maf3J1WGfRyHJDa6jzQ9XBZXVH47qyuz/view?usp=share_link

Articles Dataset:

https://drive.google.com/file/d/10j5jPppv8miBl_xvD1ZXz9EgaPmjW5ms/view?usp=share_link

Merged_df1:

https://drive.google.com/file/d/1Rd8Zotr6Nw6qDvO6yugRC0Elyb4yya4K/view?usp=share_link

The updated clean versions are able to be downloaded via csv in jupyter notebook at the end of the data cleaning and processing notebook.

2. The clean versions of the above files are used for **EDA** notebook which is in *Notebooks* folder. Please download the cleanfiles from the Datacleaning and processing notebook. The clean files are then merged into one csv table called merged_df1.
3. Merged_df1 is used in Baseline Model 1 which is in *Models folder*
4. Merged_df1 is used in Decision Tree Modeling which is in *Models folder*
5. Cleancustomers, cleantransaction and cleanarticles are used for the final notebook MarketBasketAnalysis which is in *Models folder*