### README

### Ensure that the input data files required for the code in the EDA\_Stage1.ipynb notebook are downloaded from the appropriate source ([**https://www.kaggle.com/datasets/nadyinky/sephora-products-and-skincare-reviews**](https://www.kaggle.com/datasets/nadyinky/sephora-products-and-skincare-reviews)). Verify that the correct local file paths are specified in the code section where the data is imported. This step is critical for the successful execution of the notebook.

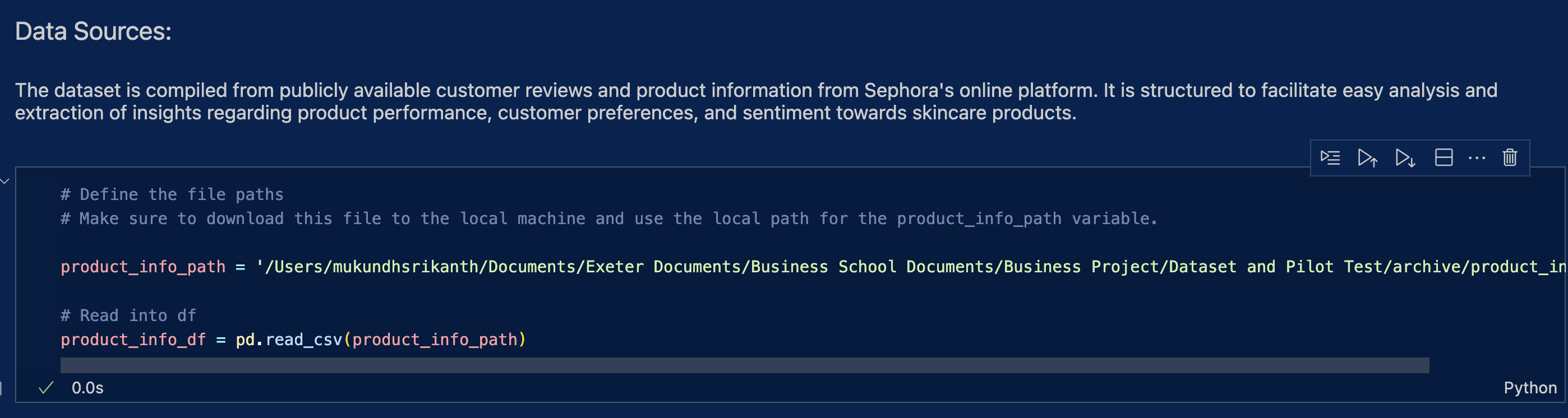


Figure 1 - Make sure product\_info\_path points to the correct local path. Source: Author's own work.

### The primary output of the **EDA\_Stage1.ipynb** notebook is a dataframe named *filtered\_product\_cluster\_df*. This dataframe should be saved to a local path of your choice. When running the second notebook, **Stage1\_Stage2.ipynb**, make sure that this output file is correctly referenced by specifying the appropriate path in your local system.

A computer screen shot of a computer program

Description automatically generated

Figure 2 - Please ensure product\_info\_clusterd\_path points to the exact location of the filtered\_product\_cluster\_df file. Source: Author's own work

### Ensure that all necessary **review data files** have been downloaded. It's important that the local paths for these files are correctly specified in the code. Without the correct paths, the notebooks may fail to execute as expected.

A computer screen shot of a computer program

Description automatically generated

Figure 3 - Please ensure that the review\_path correctly point to the review files. Source: Author's own work.