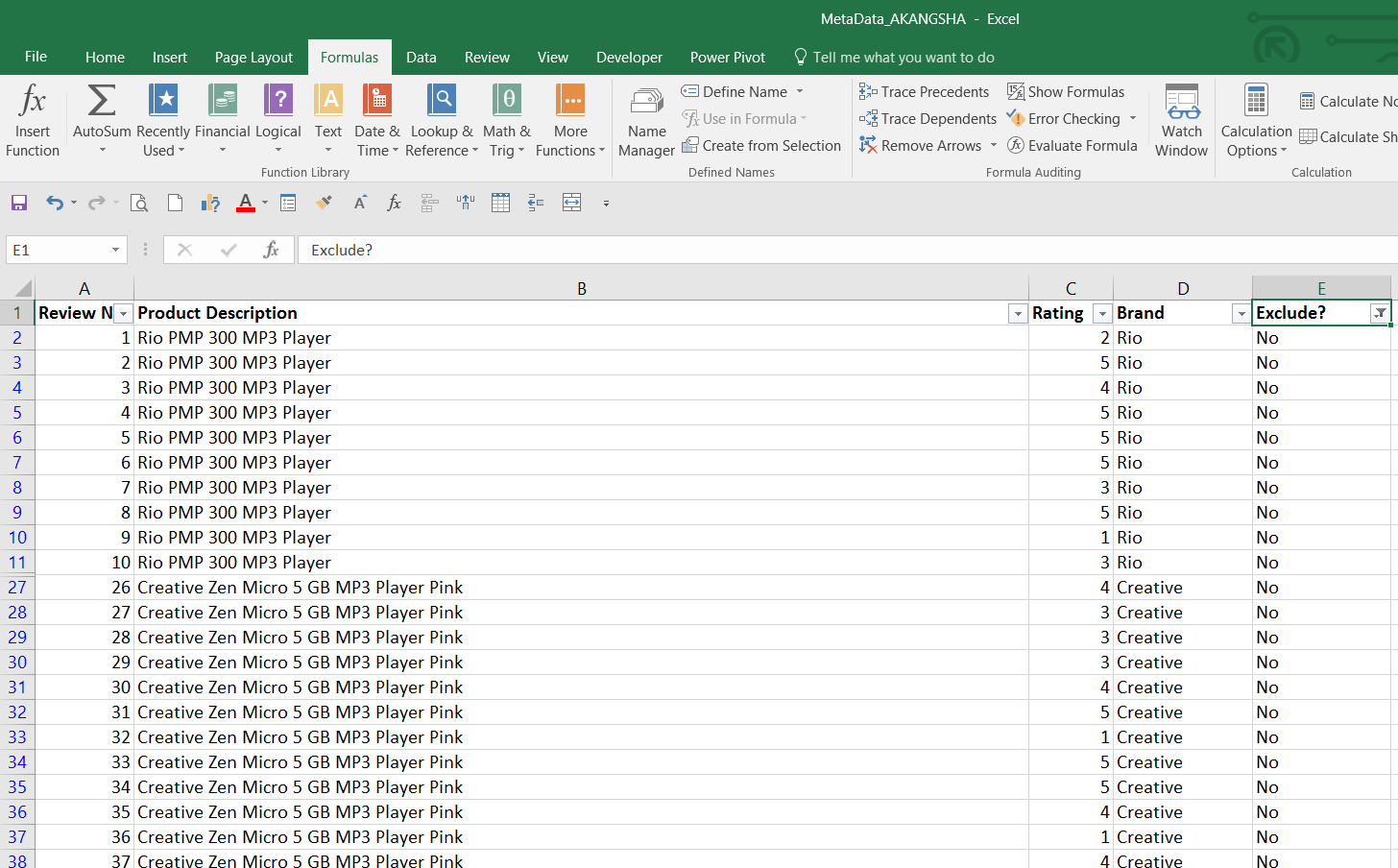
**Data Preparation**

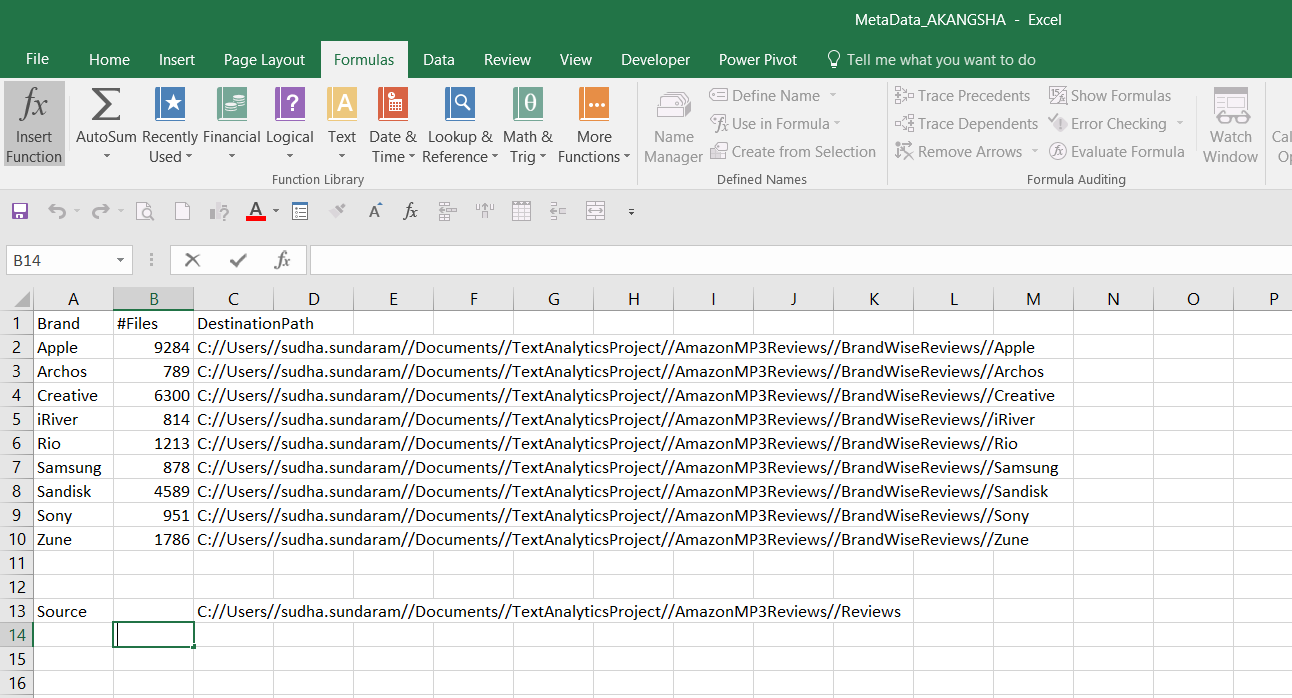
**Steps involved in shortlisting brands to be used for analysis**

**Steps involved in creating brand wise corpora**

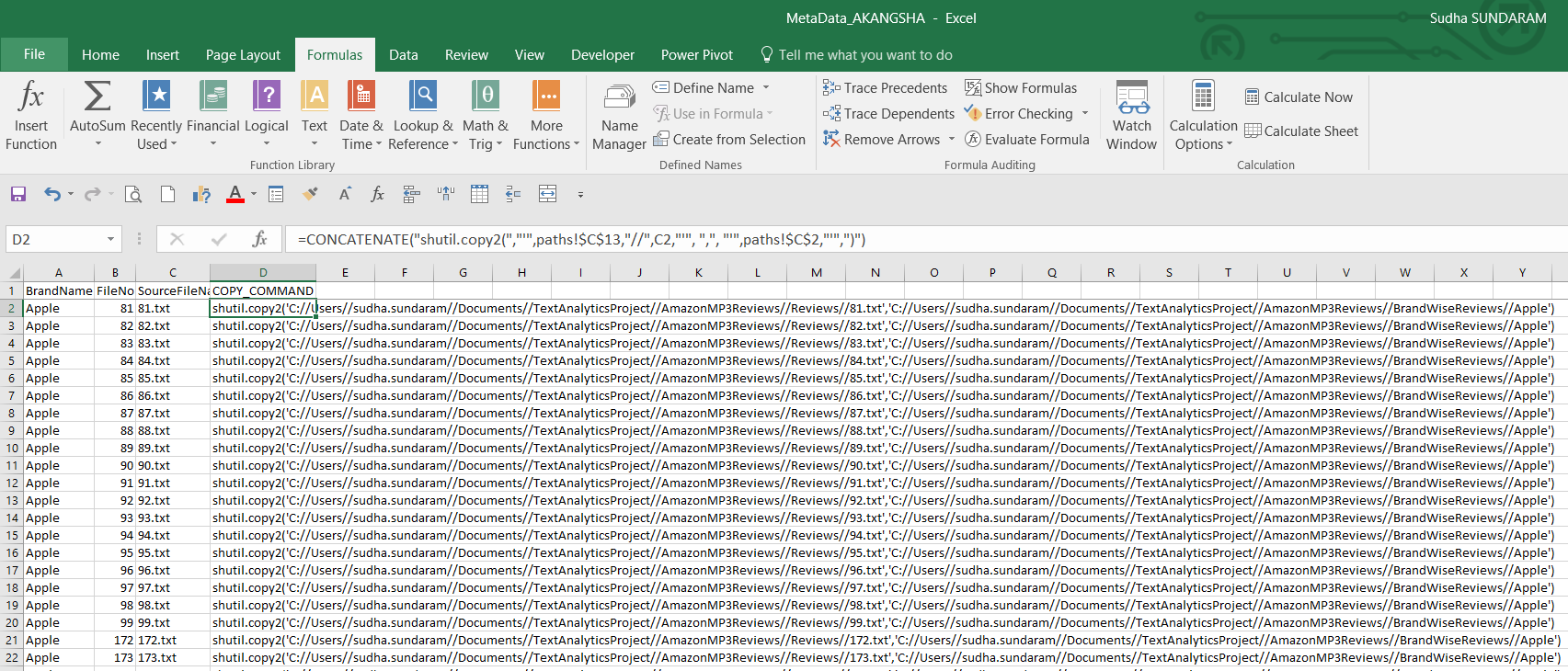
* After the data cleaning steps, have been undertaken for the brand names and the nine brands with sufficient number of reviews has been shortlisted, we need to create separate folders with the review texts pertaining to the brands
* We utilize the summary in excel created as part of the data cleaning activity – a snapshot of which is shown below



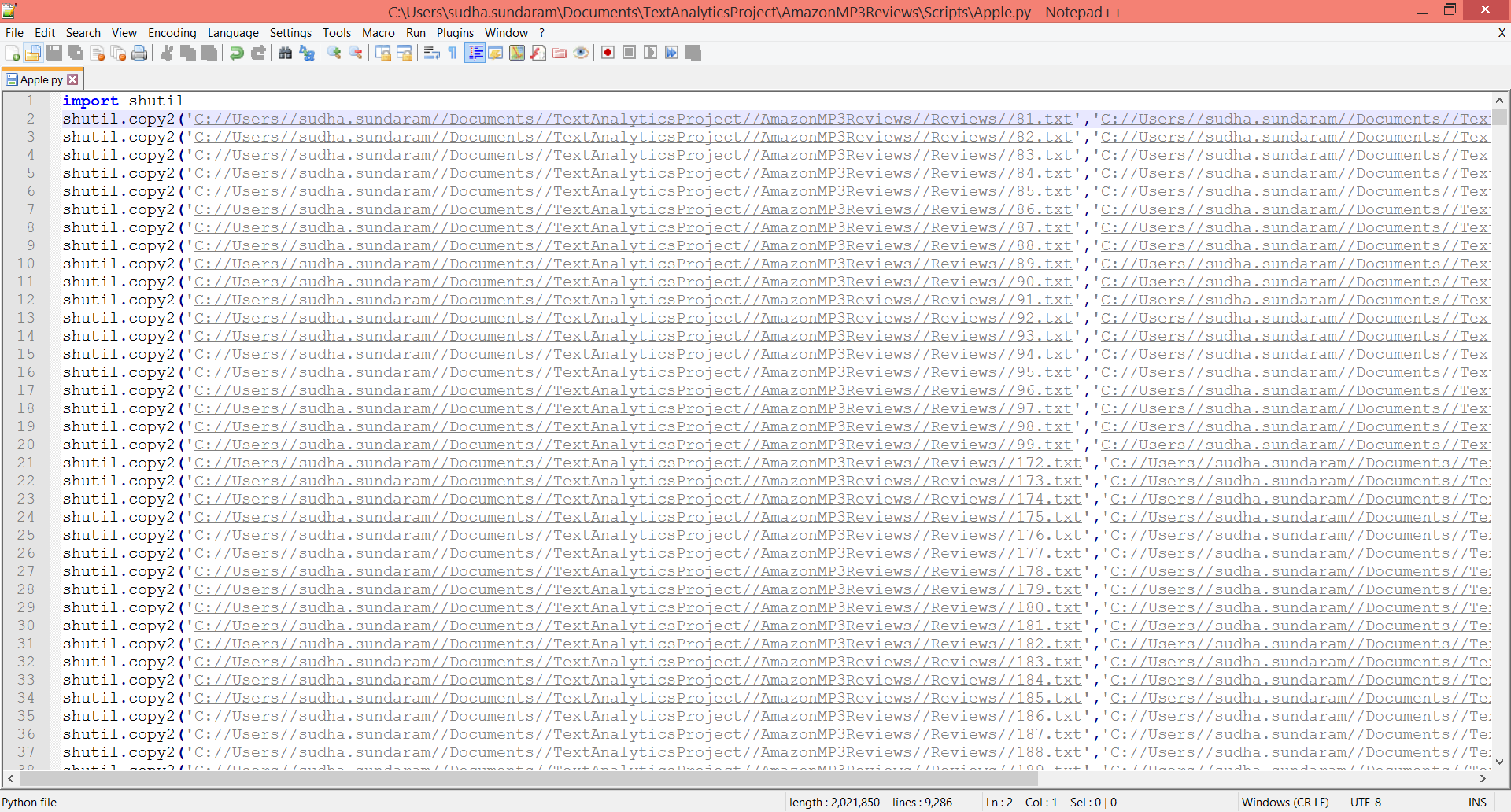
* A semi-automated process was used for writing the review text files to their corresponding brand folders.
* First, 9 folders for each brand were created and their folder paths are noted.
* The source path is also noted (all the 31,000 review text files are in this folder)



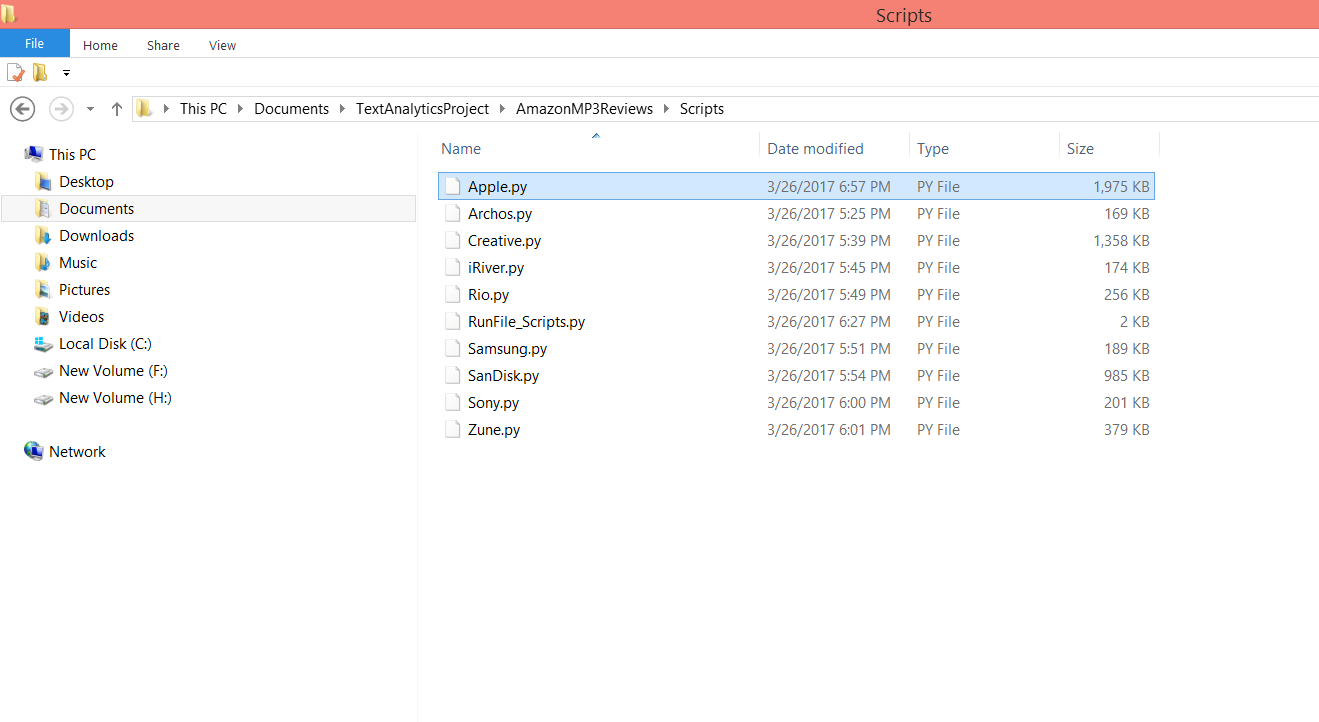
* For the first brand, Apple, we create a code in Python for copying a review text file from the source folder to destination folder. This code needs to replicated for each individual file and hence we use Excel functions such as Concatenate and copy-paste to generate codes for all the files



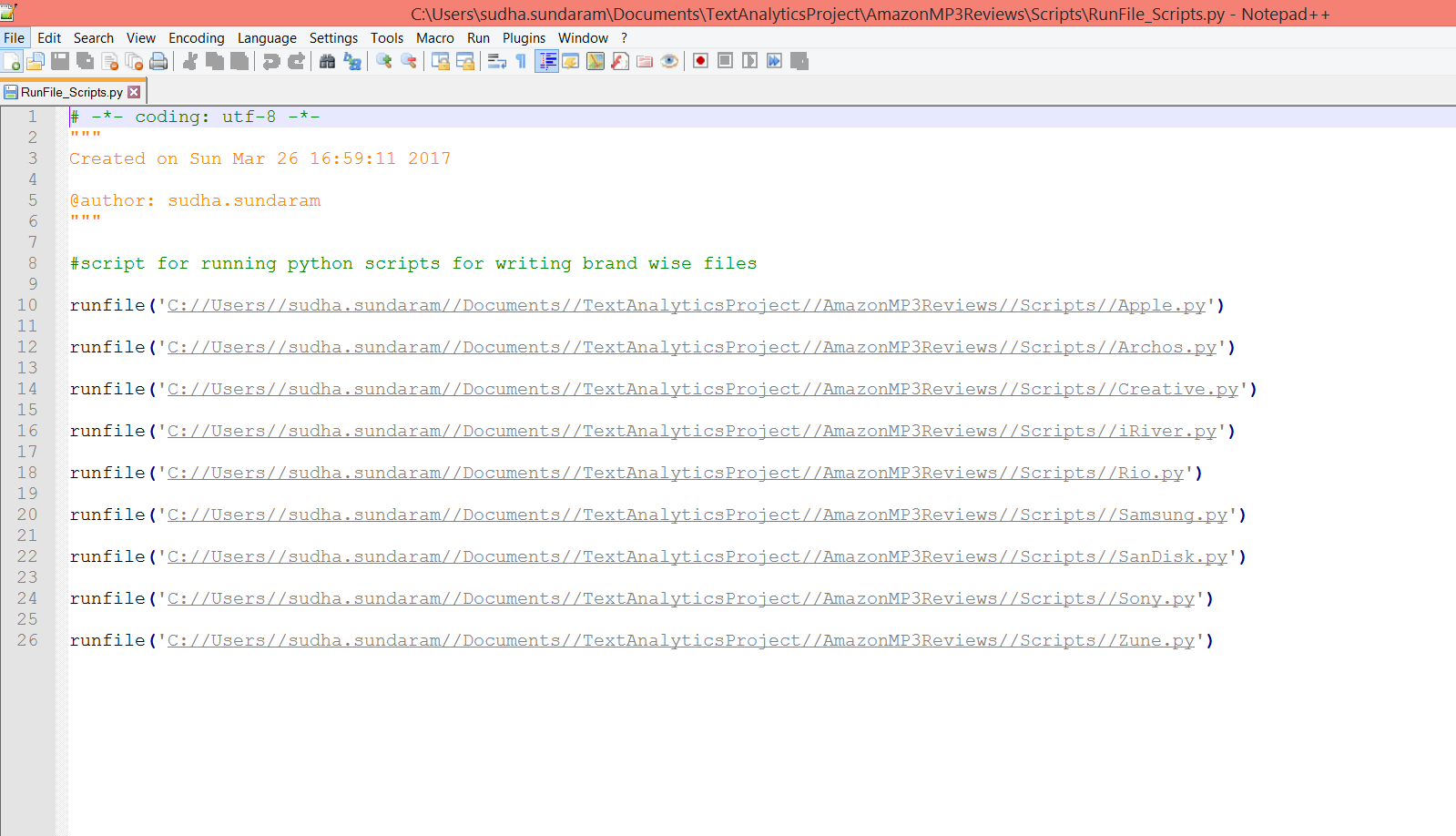
* The code created for each file is then saved in a .py script for each brand – for example – see below the contents of the Apple.py script



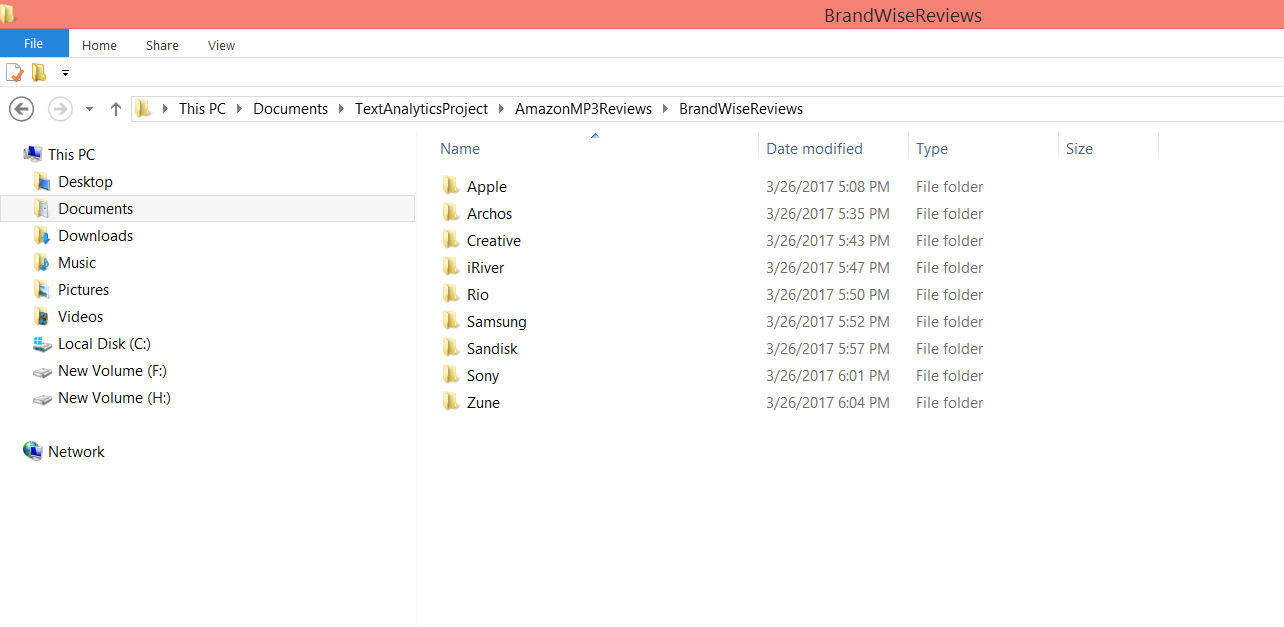
* Python scripts created for each brand shown below.



* These 9 scripts were then run in another script called RunFile\_Scripts.py one after the other.



* Brand wise folders having the review texts of each brand have been created



* We have also created a folder called “CombinedReviews” for containing the 26,605 reviews of all the brands put together.