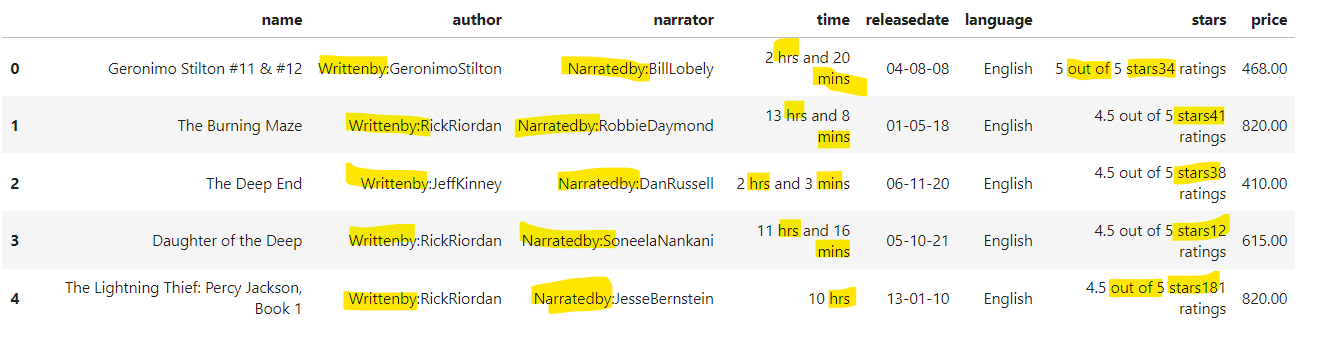
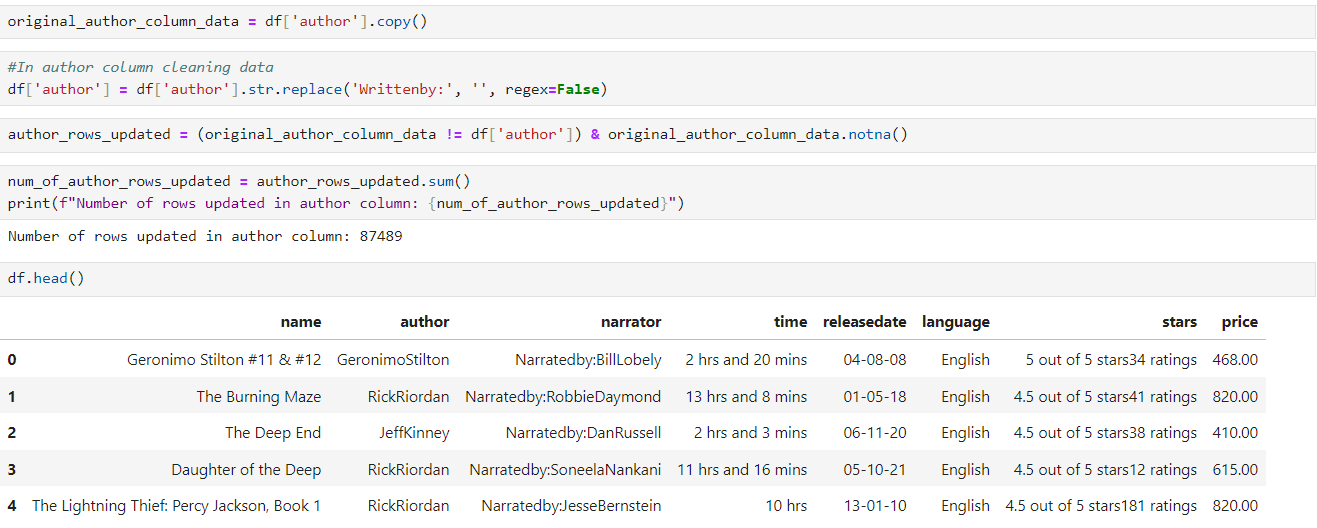
Step 13: As per the analysis this dataset doesn’t have missing values and scrapped data seems to not readable, we used data wrangling techniques to make it more readable and for “Not rated” yet data we changed values to “0” so that we can for graphs in future. Moreover, we changed the time format into minutes from “hours & minutes combination” to help with data for visualization.

Step 1: By visualization it seems data sets have impurities

Step2: we will clean the “author” column data first 

Step3: after that we will clean “narrator” column data same way

Step4: adding space in stars column between “5 out of 5 stars” and “34 ratings” to make it more readable 

Step5: used regex here above for adding space

Regex explanation

stars - to match the word

() - to capture the part of the matching

\d is for any digit between 0 to 9

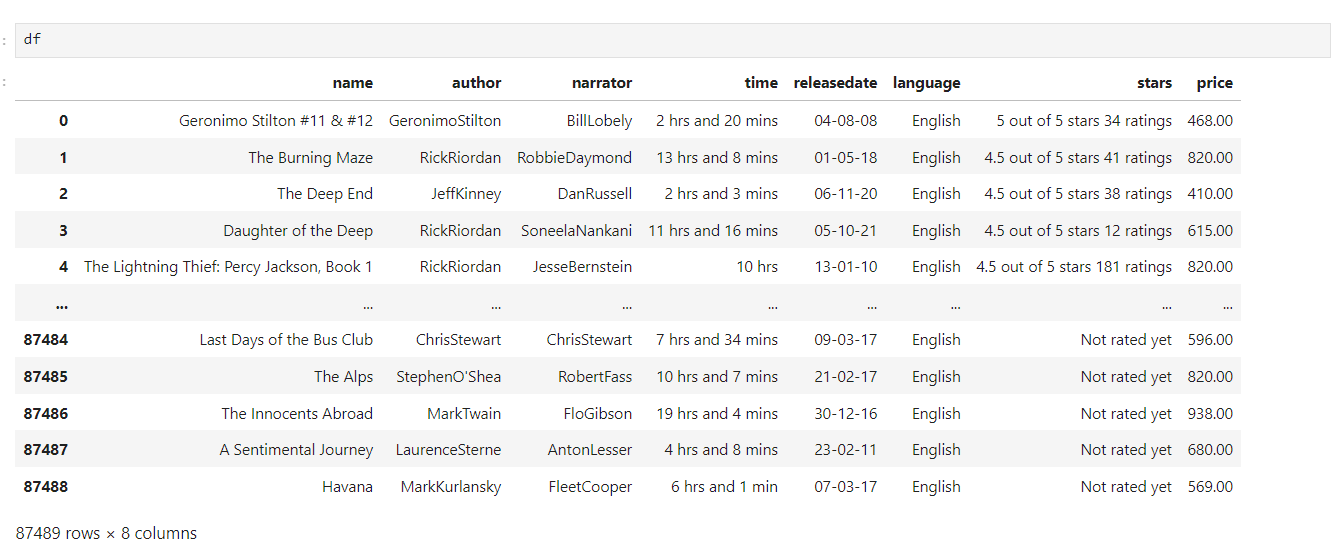
+ means one or more digits

\1: Refers to the first capture group "stars".

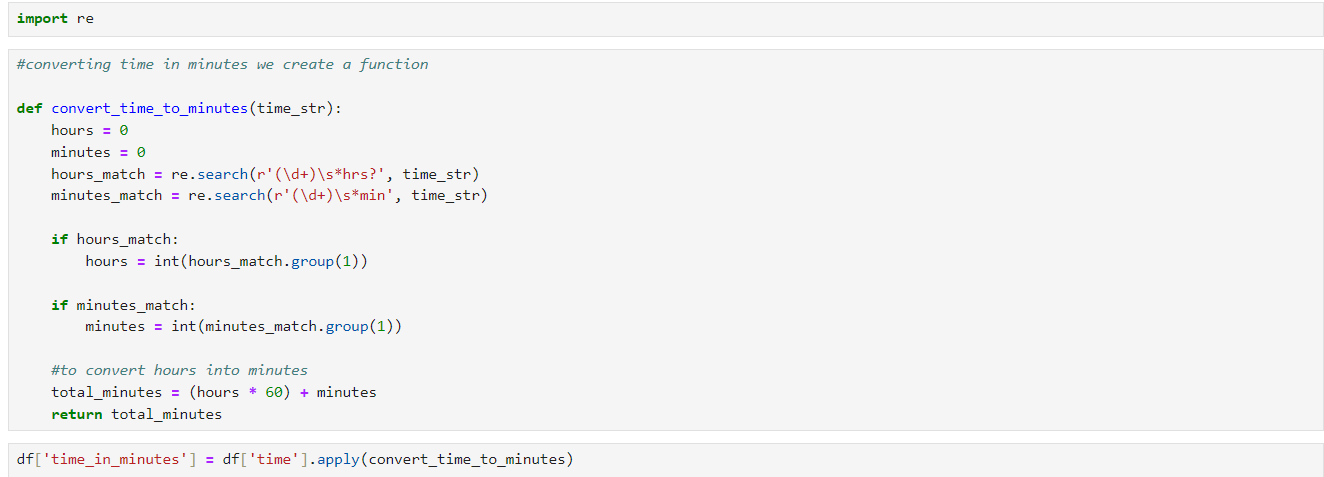
\2: Refers to numeric part (the digits after "stars")

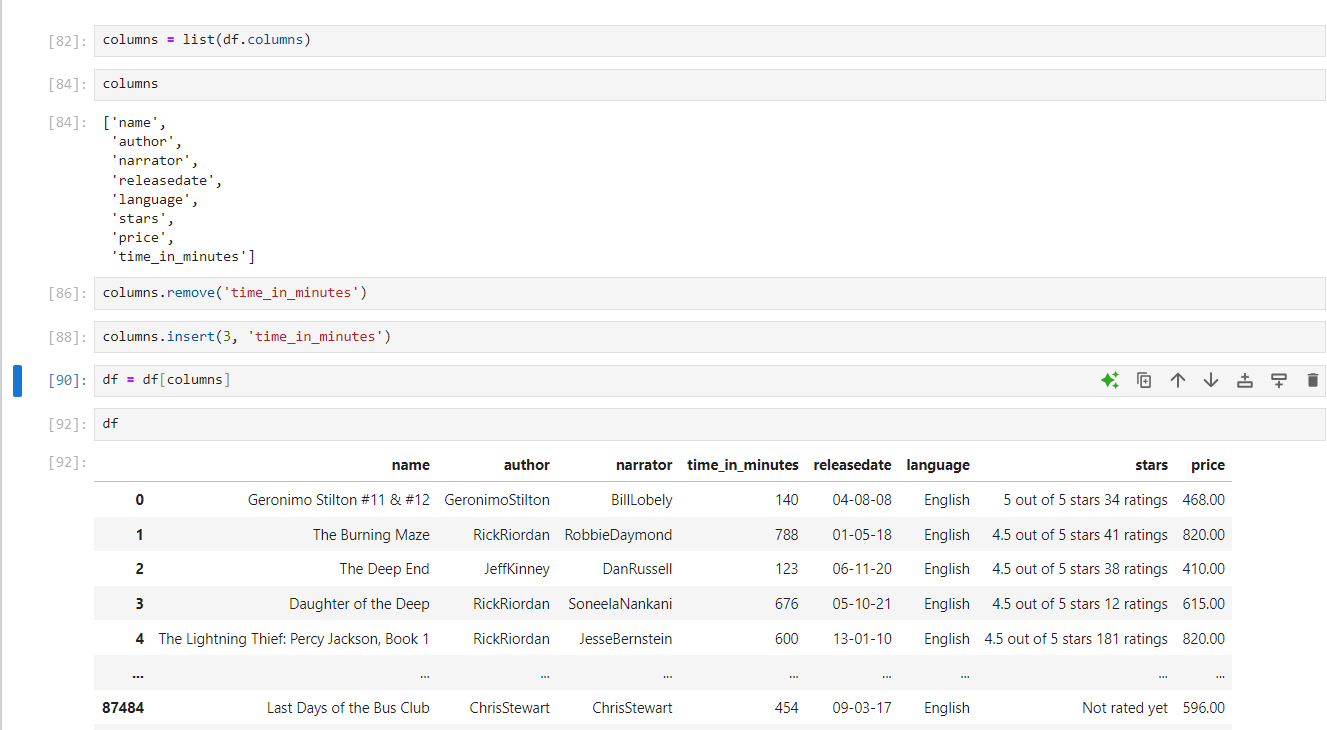
r'\1 \2' is telling pandas to keep the first part ("stars") and then insert a space.

Step6: data looks better to visualise however there seems to be some issues

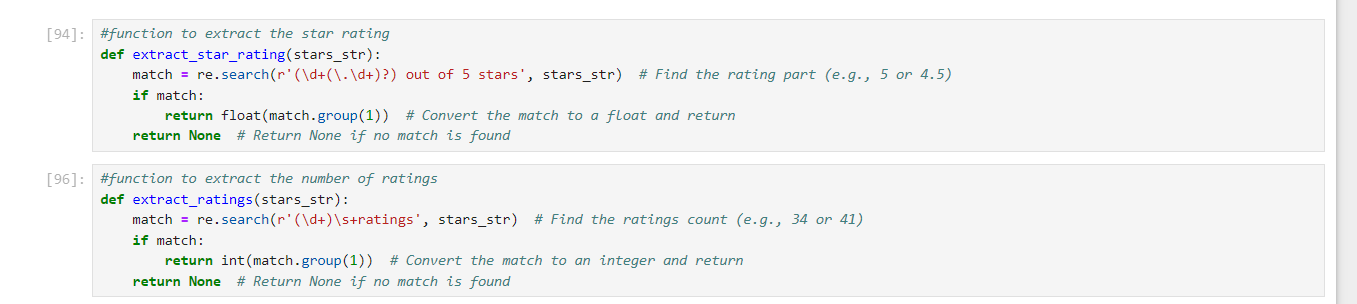


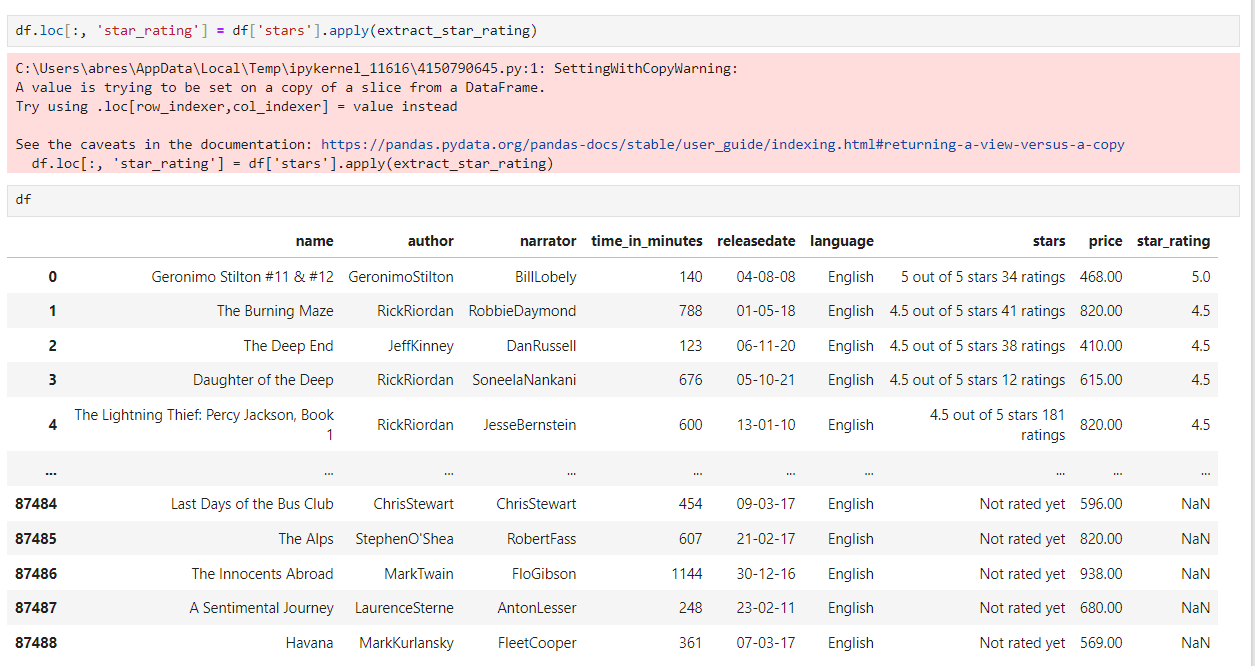
Step7: Create a function to change time from “hrs and minutes” to only “minutes” to make it more readable



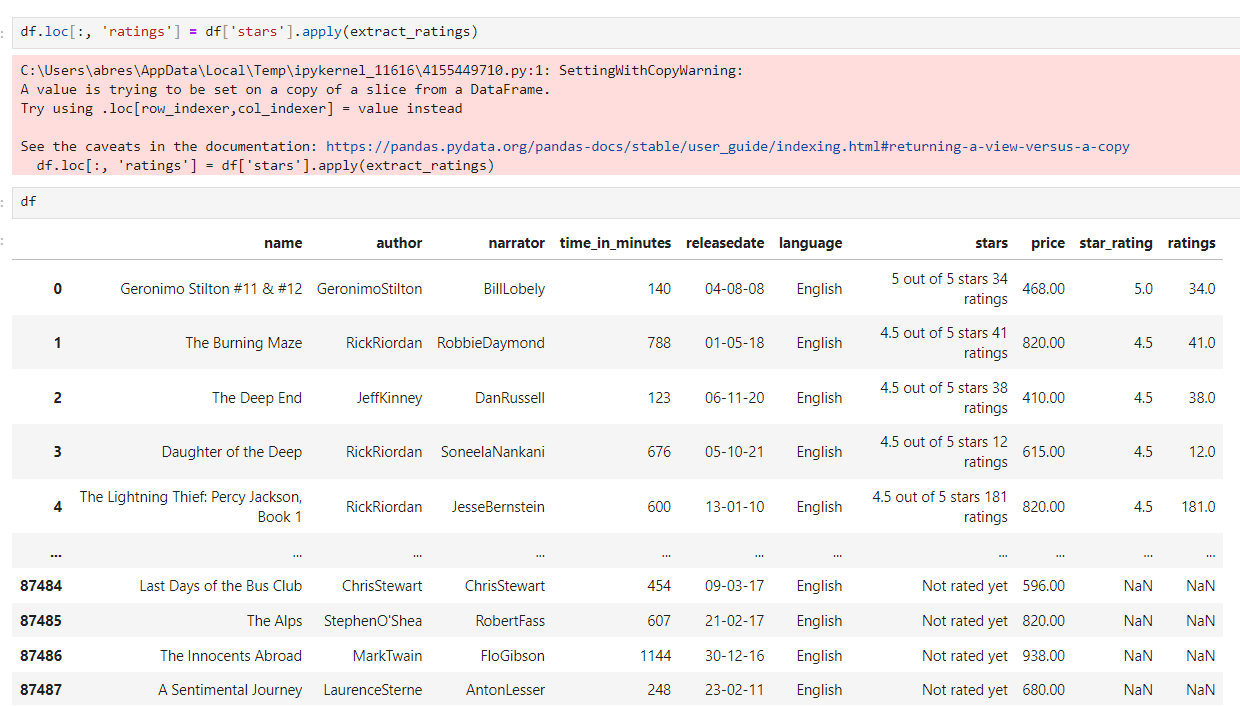
Step8: time in minutes now

Step9: Created two more functions to sort stars ratings to different columns



Step 10: added star\_rating column here

Step 11: Creating rating column too



Step 12: Removed “NaN” value and added “0” to make data more readable for “not rated yet” data and exported.

