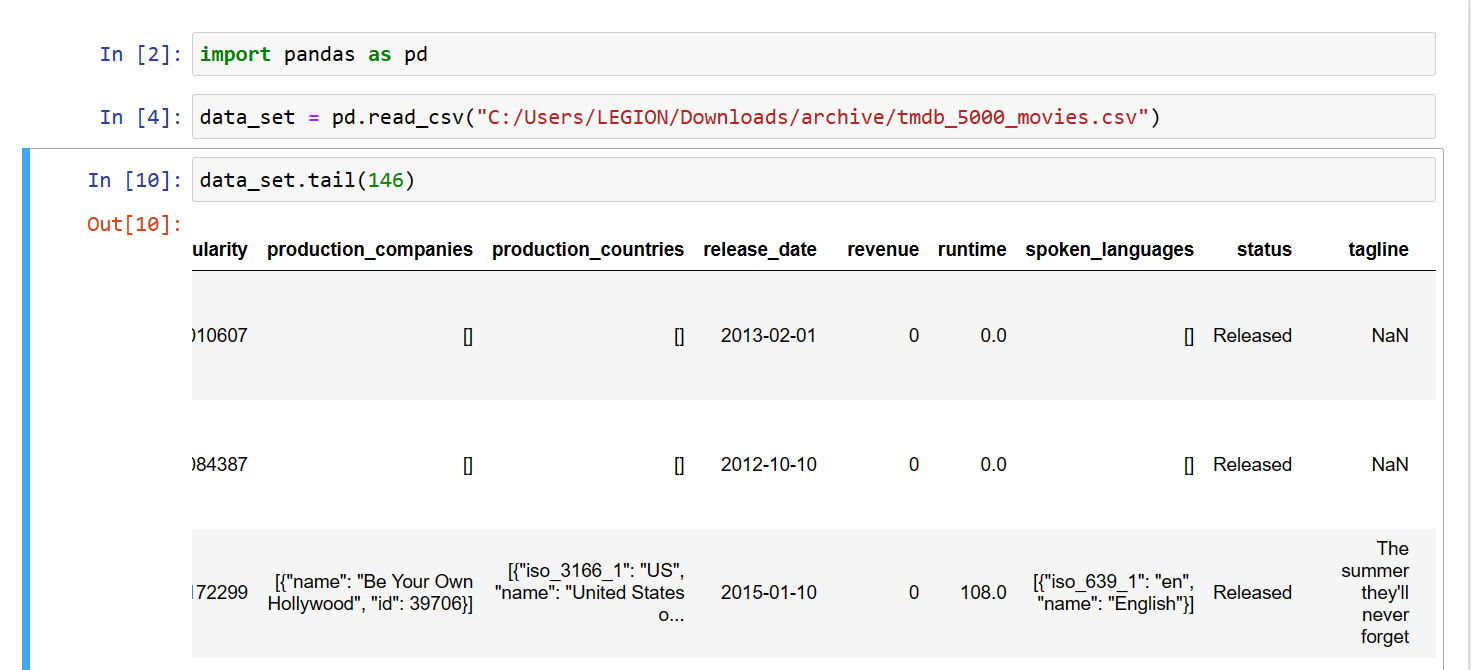
**Data mining assignment: Data Cleaning**

**Dataset:** TMDB 5000 Movie Dataset

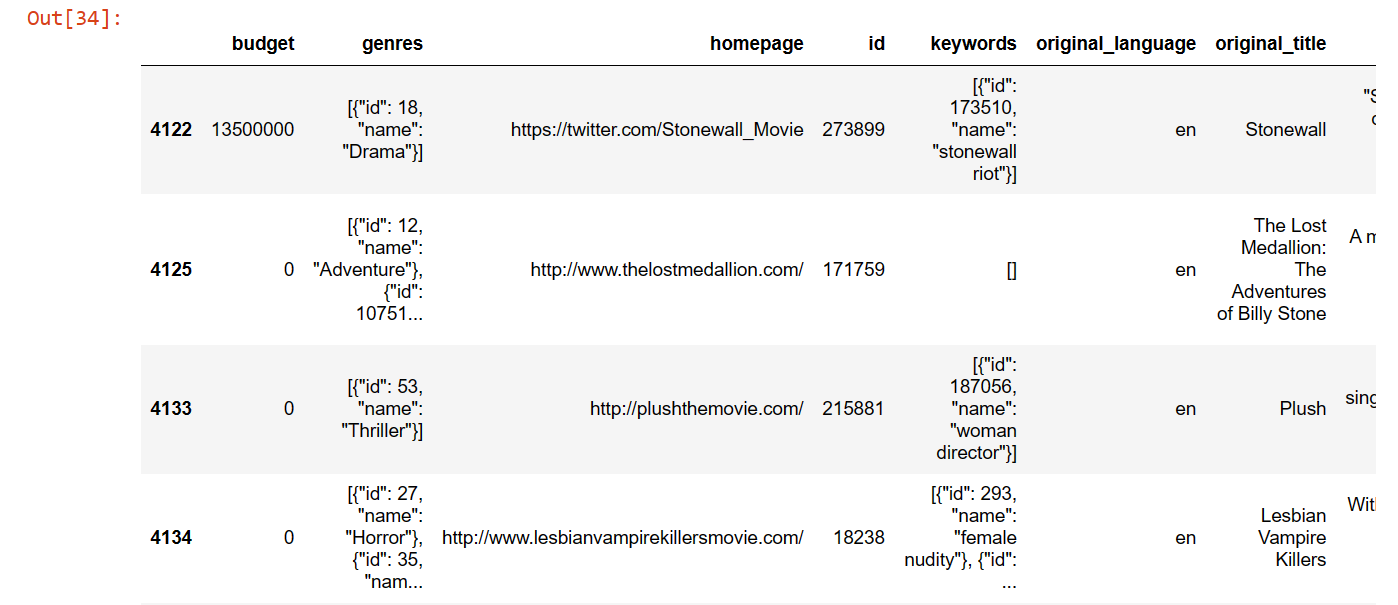
**Link:** [TMDB 5000 Movie Dataset (kaggle.com)](https://www.kaggle.com/datasets/tmdb/tmdb-movie-metadata/data?select=tmdb_5000_movies.csv)

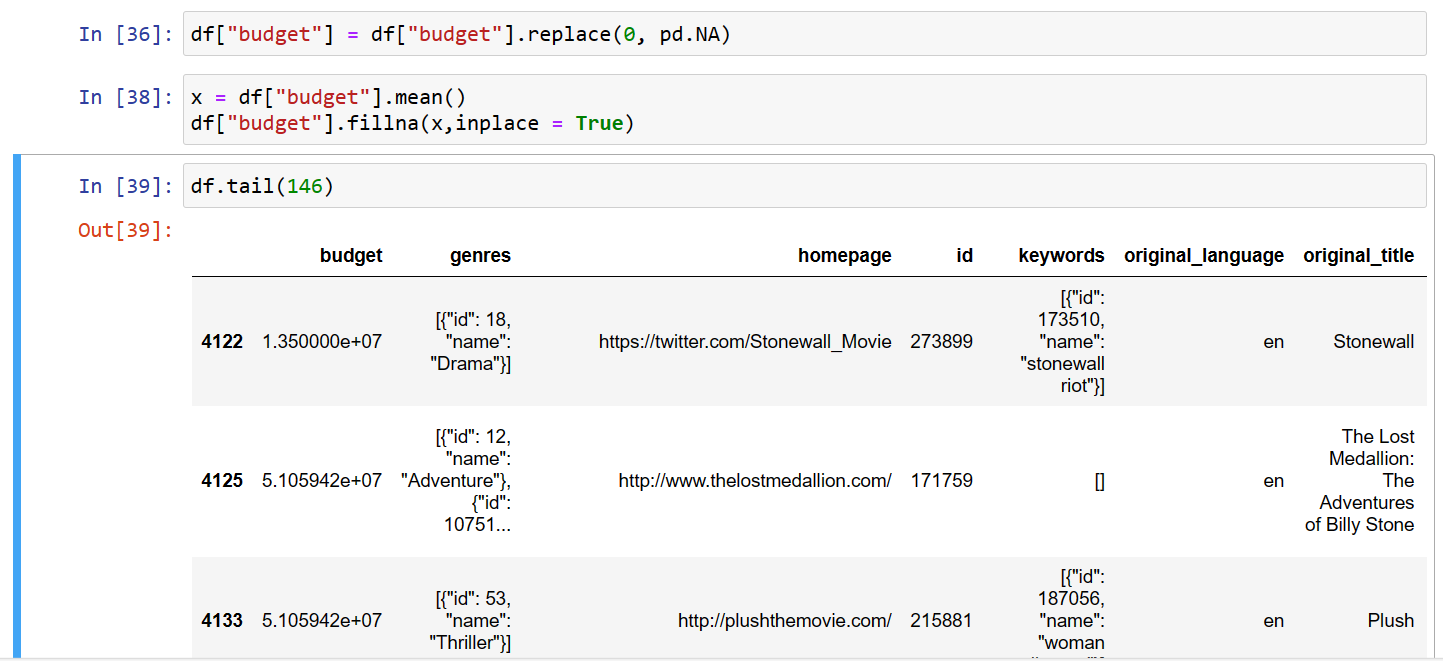
الطالب/ حامد شهاب الصبري

Upon reviewing the dataset, I have observed that the "runtime" column contains several instances where the value is 0. Given that the number of rows with a runtime value of 0 is relatively small, I have made the decision to remove these rows from the dataset.

I started by replacing the 0 with NaN and then deleting the rows that contain NaN in them:

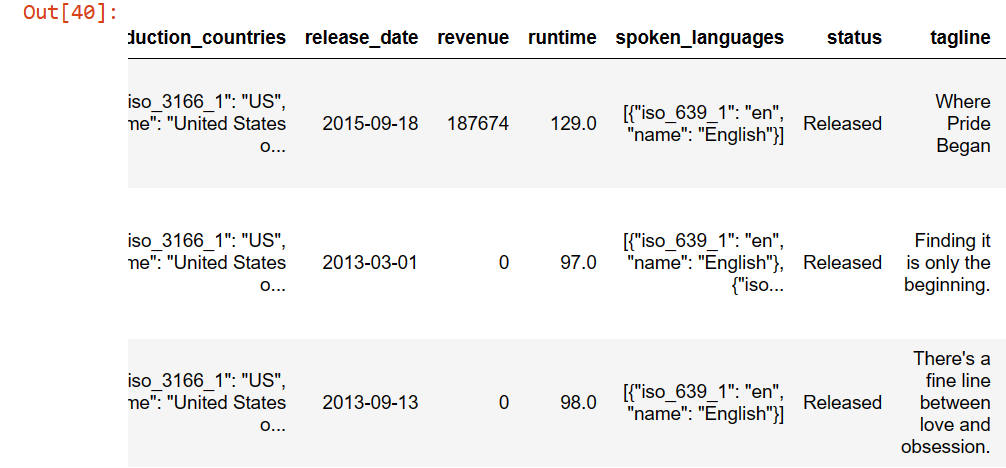
I also noticed that the budge and the revenue had a number of 0 values, so I decided to replace the budget zeros with the mean:

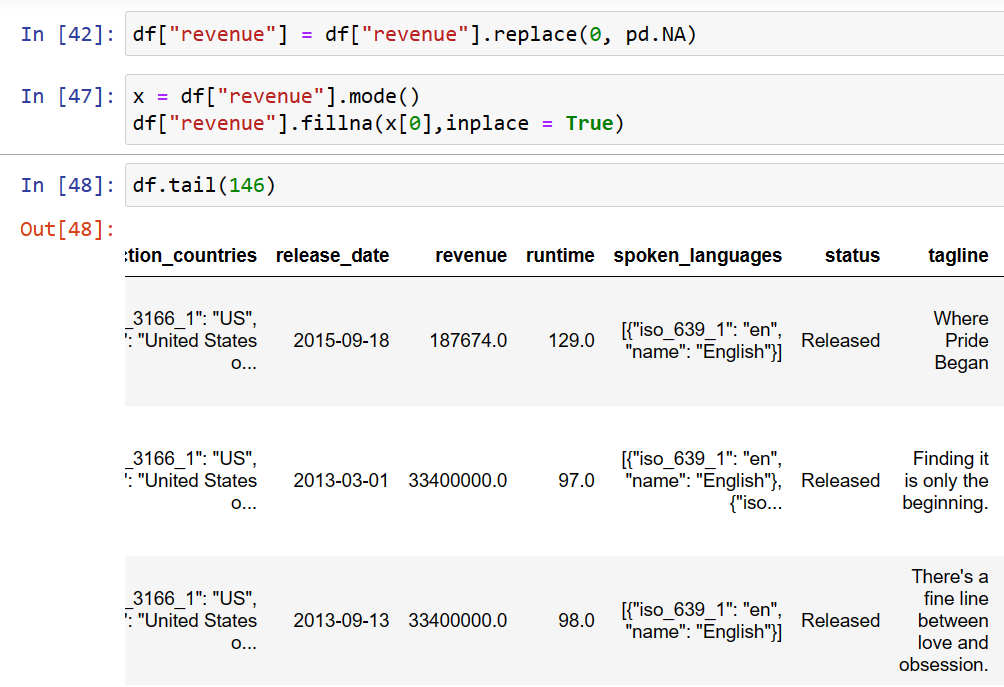
The budget before:

The budget after:

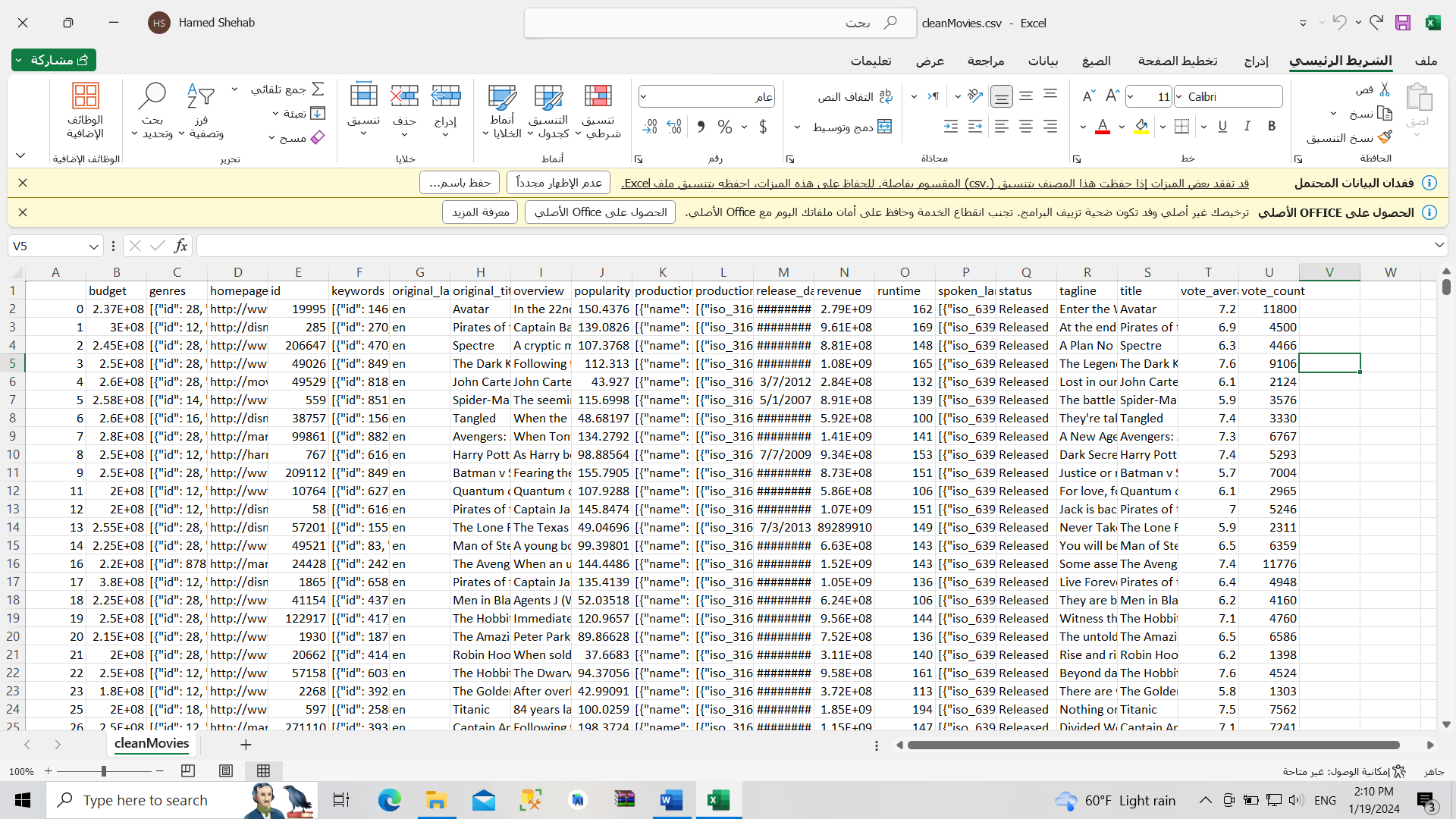
And I decided to replace the revenue zeros with the mode:

The revenue before:



The revenue after:

I also deleted the columns that were not needed

Data set before deleting columns:

Data set after deleting columns: