Project 2 ETL – Zhuoli Ma, Rana Badri, Alysia Dugan

Extract:

<https://www.rottentomatoes.com/top/bestofrt/?year=2021>

<https://www.rottentomatoes.com/top/bestofrt/?year=2020>

<https://www.rottentomatoes.com/top/bestofrt/?year=2019>

<https://www.the-numbers.com/market/2021/top-grossing-movies>

Transform:

* In “the number” website, we copied and pasted the data into excel and convert to csv file.
* In the Rotten tomatoes webpage, we used pandas\_read\_html to load the data into Jupyter Notebook.
* Remove unnecessary string on the movie’s title using “replace” function.

Load:

* Used Jupyter Notebook to connect with pgAdmin.
* Created box\_office\_db in pgAdmin and xx
* Created two SQL statements for testing.