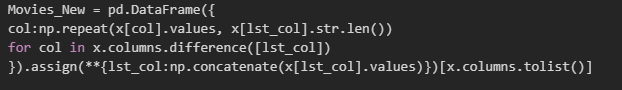
ETL PROJECT

E - Extracted 4 IMDb csvs from Kaggle dataset.

(Resource folder with original dataset)

T - Cleaned and cleared data using python, pandas etc:

Dropped irrelevant columns of things we didn’t need or were lacking in data, made dates easier to access, made genres for demos easier to access, etc.

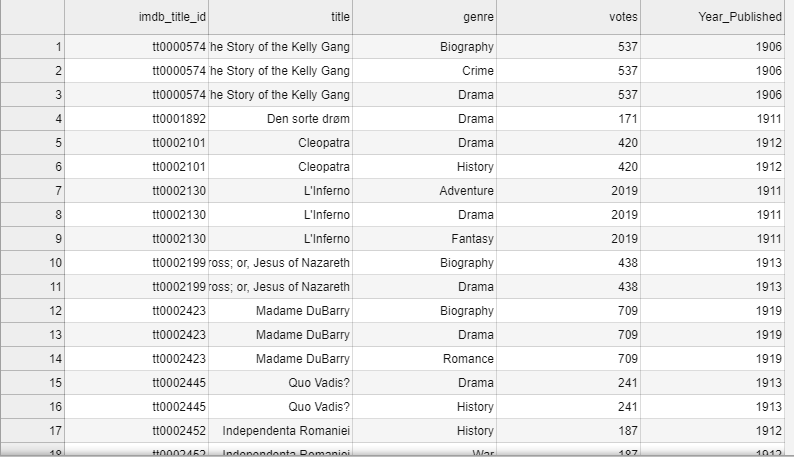




Before sample:

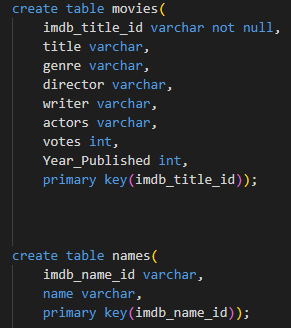
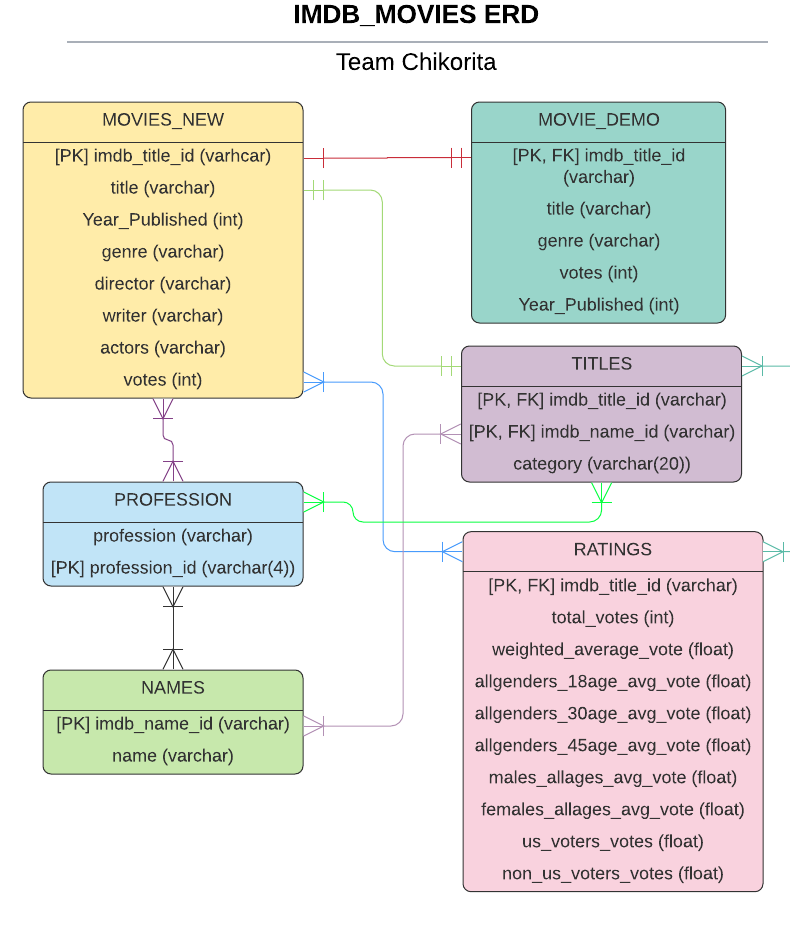


After sample:

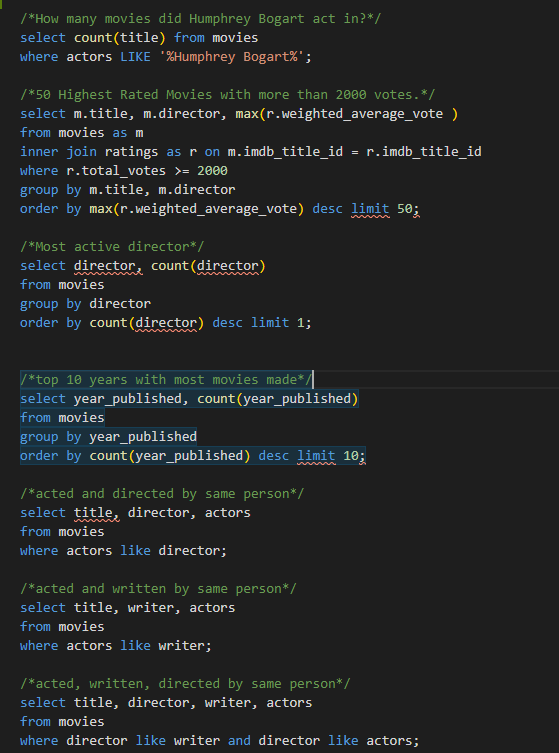


(Analysis.ipynb)

L - Saved changed dataframes back to csvs, created SQL schema with primary and foreign keys, created tables and imported data, performed multiple and various queries including joins and aggregation to show usefulness and depth.  
Why PostgreSQL? Already a relational format, easy to query with complexity, and relate tables to each other with pk/fk. Common usage. + it has a dark theme.

 (SQLSchema.SQL)

(SQLQueries.SQL sample)



(QueryResults/Figs. sample)

