

MIGRATE NETFLIX AND OMBD DATA INTO PRODUCTION DATA BASE

Project Two

March 7^h, 2020

TEAM MEMBERS

- Jonpaul Mastropaolo (Data Architect)
- Tiger Cosmos (Data Starlord)
- Tiana Chan (Data Engineer)
- Clayton Skinner (Data Trash Panda)

TABLE OF CONTENTS

- Project Report
 - Data Sources
 - **E**xtract
 - **T**ransform
 - **L**oad
- Questions

DATA SOURCES

OMDB

Source: OMDB

Website: <http://www.omdbapi.com>

Netflix

CSV: [Kaggle](<https://www.kaggle.com/>)

EXTRACT

```
In [3]: # This cell contains the for loop to get all the data for movies in netflix in the omdb api

# Creating the list of movie titles to make the API calls on
filtered_movies = filtered_flix['title']

# initializing empty lists to append to in the loop
metascore = []
imdbrating = []
imdbvotes = []
boxoffice = []
title = []

# This is the api call for our OMDB dataframe
for item in filtered_movies:

    # If the loop doesn't work, we will skip the API call and list append function.
    try:
        # Creating the string for the URL to call the API from.
        movie = "t=" + item
        key = "&apikey="+jomdb
        url = 'http://www.omdbapi.com/?' + movie + key

        # Storing the response as a json object
        response = requests.get(url).json()

        # Append the items we care about from the json
        boxoffice.append(response['BoxOffice'])
        title.append(response['Title'])
        metascore.append(response['Metascore'])
        imdbrating.append(response['imdbRating'])
        imdbvotes.append(response['imdbVotes'])

    except:
        next
```

TRANSFORM

```
# Creating the DataFrame for the OMDb data
```

```
omdb_df = pd.DataFrame({'title':title,  
                        'metascore':metascore,  
                        'imdb_rating':imdbrating,  
                        'imdb_votes':imdbvotes,  
                        'box_office':boxoffice})
```

```
# Joining the two tables: the netflix table and the OMDb table
```

```
final_table = filtered_flix.join(omdb_df.set_index('title'),on = 'title', how = 'left')
```

```
# Doing away with the rows that contain "NaN" in any column.
```

```
final_table = final_table.dropna()
```

```
# Write final table to csv for use later
```

```
final_table.to_csv('final_table.csv',index = False, header = True, sep = ',')
```

```
final_table=pd.read_csv('final_table.csv')
```

```
final_table.head()
```

LOAD

```
In [3]: #connect to local data base  
rds_connection_string = "etl:etl@localhost:5432/ETLmovies_db"  
engine = create_engine(f'postgresql://{rds_connection_string}')
```

```
In [7]: #check for tables  
engine.table_names()
```

```
Out[7]: ['final_table_db']
```

```
In [6]: #upload data frame to database  
final_table.to_sql(name='final_table_db', con=engine, if_exists='append', index=False)
```

```
In [8]: #check the dataframe made it to the table  
pd.read_sql_query('select * from final_table_db', con=engine).head()
```

LOAD

ETLmovies_db/postgres@PostgreSQL 12 *

Query Editor Query History

```
1 select * from final_table_db;
2
```

Data Output Explain Messages Notifications

	show_id bigint	type text	title text	director text	cast text	country text	date_added text	release_year bigint	rating text	duration text	listed_in text	description text	metascore double precision
1	80060297	Movie	Manha...	Tom O'Brien	Tom O'...	United States	September 8, 20...	2014	TV-14	98 min	Comedies, I...	A filmmaker wo...	[nu
2	70304988	Movie	Stoneh...	Brad Ander...	Kate B...	United States	September 8, 20...	2014	PG-13	113 min	Horror Movi...	In 1899, a youn...	:
3	80057700	Movie	The Ru...	Austin Stark	Nicola...	United States	September 8, 20...	2015	R	90 min	Dramas, Ind...	A New Orleans ...	:
4	80045922	Movie	6 Years	Hannah Fidell	Taissa ...	United States	September 8, 20...	2015	NR	80 min	Dramas, Ind...	As a volatile yo...	:
5	80058026	Movie	Hell an...	Tom Gianas...	Nick S...	United States	September 6, 20...	2015	R	86 min	Action & Ad...	When best frien...	[nu
6	80147760	Movie	Carrie ...	Susan John...	Bel Po...	United States	September 5, 20...	2016	TV-14	98 min	Comedies, D...	A socially awkw...	:
7	80177405	Movie	Marc ...	Lynn Shelton	Marc ...	United States	September 5, 20...	2017	TV-MA	70 min	Stand-Up Co...	Battle-scarred s...	[nu
8	81078908	Movie	The W...	Brian Baugh	Caleb ...	United States	September 4, 20...	2019	PG	108 min	Dramas, Ro...	A teenage eque...	[nu
9	80201906	Movie	Black P...	Ryan Coogler	Chadwi...	United States	September 4, 20...	2018	PG-13	135 min	Action & Ad...	T'Challa, the su...	:
10	80128722	Movie	Gerald'...	Mike Flanagan	Carla G...	United States	September 29, 2...	2017	TV-MA	103 min	Horror Movi...	When her husba...	:
11	80104068	Movie	Our So...	Ritesh Batra	Robert ...	United States	September 29, 2...	2017	TV-14	103 min	Dramas, Ro...	After widowed ...	:
12	80019931	Movie	Monst...	William Lau	Karen ...	United States	September 29, 2...	2015	TV-Y7	72 min	Children & F...	Cleo is invited t...	[nu
13	80157072	Movie	Hold th...	Jeremy Sau...	Jeffrey...	United States	September 28, 2...	2018	TV-MA	126 min	Action & Ad...	In the grim Alas...	:
14	80231903	Movie	In the S...	Jim Mickle	Boyd H...	United States	September 27, 2...	2019	TV-MA	116 min	Dramas, Sci...	A Philadelphia d...	:

QUESTIONS

- Open discussion to Q&A from the class.

END