Microsoft Movie Studio

Overview

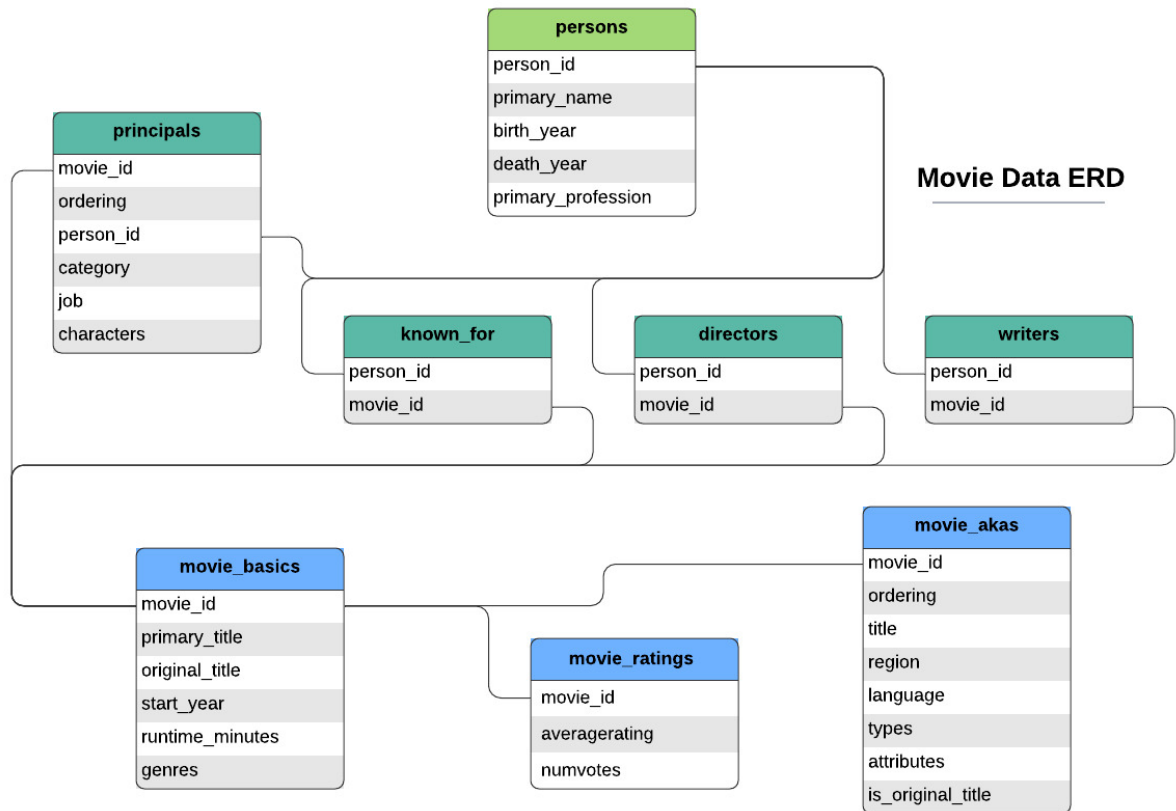Flatiron School Data Science Phase 1 project.

Business Problem

Microsoft sees all the big companies creating original video content and they want to get in on the fun. They have decided to create a new movie studio, but they don't know anything about creating movies. You are charged with exploring what types of films are currently doing the best at the box office. You must then translate those findings into actionable insights that the head of Microsoft's new movie studio can use to help decide what type of films to create.

Data Used

data from the following sources can be found in the 'zippedData' folder

```
* [IMDB](https://www.imdb.com/) - im.db.zip
* [Box Office Mojo](https://www.boxofficemojo.com/) - bom.movie_gro
ss.csv.gz
* [The Numbers](https://www.the-numbers.com/) - tn.movie_budgets.cs
v.gz
```

Movie Data ERD

persons
- person_id
- primary_name
- birth_year
- death_year
- primary_profession

principals
- movie_id
- ordering
- person_id
- category
- job
- characters

Movie Data ERD

known_for
- person_id
- movie_id

directors
- person_id
- movie_id

writers
- person_id
- movie_id

movie_basics
- movie_id
- primary_title
- original_title
- start_year
- runtime_minutes
- genres

movie_ratings
- movie_id
- averagerating
- numvotes

movie_akas
- movie_id
- ordering
- title
- region
- language
- types
- attributes
- is_original_title

"

Deliverables

- [Non-Technical Presentation ()](())
- [GitHub Repository (https://github.com/ceflynn/dsc-phase-1-project-v2-4)](https://github.com/ceflynn/dsc-phase-1-project-v2-4)
- [Jupyter Notebook (https://github.com/ceflynn/dsc-phase-1-project-v2-4/blob/master/student.ipynb)](https://github.com/ceflynn/dsc-phase-1-project-v2-4/blob/master/student.ipynb)

Action Steps

- **Recommendation 1** - Compete with Big Studios
    - Create dataframe from BoxOfficeMovie.com
    - Data Cleaning

        ```
        * Missing values and NaN
        * Commas in 'foreign_gross' - Get Values from website
        * Change columns in int64 for calculations
        ```

        - Find and graph total gross
    - Create dataframe from the-numbers.com
    - Data cleaning

```
          * Get rid of '$' and ',' from 'production_budget' and 'worl
      dwide_gross
         * Change those columns to int64
         * Create column and calculate profit
         * Drop unecessary columns
         * Change 'release_date' to 'year' and only have the 4 digit
      year
```

- Merge data with studio data
- Get full studio names
- Create data visualization for movie studio Profit
- **Recommendation 2** - Create Multi-Genre Movies
  - unzip db
  - Create dataframe
  - Join tables (movie_basics and movie_ratings)
  - Get popular and highly rated movies
  - Data cleaning
  - Create boolean multi-genre column
  - Data visualization using value_counts()
- **Recommendation 3** - Focus on Biographies
  - Get list of Genres from the dataframe in Rec 2
  - Create boolean column for each Drama
  - Count values for genre columns
  - Create data visualization for genre_counts
  - Find average rating for each genre for all Movies
  - Data cleaning
  - Use most common genres
  - Get the mean for each genre
  - Create data visualization for genre ratings

Repository

- images
  - awesome.gif
  - microsoftlogo.png
  - movie_data_erd.jpeg
- pdf
  - microsoft-pres.pdf
- zippedData
  - bom.movie_gross.csv.gz
  - im.db.zip
  - rt.movie_info.tsv.gz
  - rt.reviews.tsv.gz
  - tmdb.movies.csv.gz
  - tn.movie_budgets.csv.gz
- student.ipynb
- CONTRIBUTING.md
- Instructions.md
- LICENSE.md

- README.md