

# Project 8: Going to the Movies

---

## Clarifications/Corrections

---

- Mar 14: the tests had questions 17 and 18 swapped, so please re-download the corrected test.py.
- Mar 25: q35 was corrected to ask about actors, not directors
- Mar 25: there is a `project.py` file; it's not really needed for this project, but feel free to use it if it's helpful. Stage 1 was modified to no longer code in `project.py`.
- Mar 26: correct example output for q34
- Apr 1: fixed test.py to accept multiple ways to break ordering ties for q34 and following

## Introduction

---

Having worked our way through soccer and hurricanes, we are now going to work on the IMDB Movies Dataset. A very exciting fortnight lies ahead where we find out some cool facts about our favorite movies, actors, and directors.

You'll hand in a `main.ipynb` file for this project; use the usual `#qN` format. Start by downloading the following files: `test.py`, `small_mapping.csv`, `small_movies.csv`, `mapping.csv`, and `movies.csv`.

## The Data

---

By stage 2, you will be mostly working mainly with `movies.csv` and `mapping.csv`. The `small_movies.csv` and `small_mapping.csv` have been provided to help you get your core logic working in stage 1 with some simpler data.

`small_movies.csv` and `movies.csv` have 6 columns: `title`, `year`, `rating`, `directors`, `actors`, and `genres`

Here are a few rows from `movies.csv`:

```
title,year,rating,directors,actors,genres
tt1931435,2013,5.6,nm0951698,nm0000134,"Comedy,Drama,Romance"
tt0242252,2001,6.1,nm0796124,"nm0048932,nm0000596,nm0004778","Drama,History,Romance"
tt0066811,1971,6.0,nm0125111,"nm0000621,nm0283499,nm0604702,nm0185281","Comedy,Family"
```

`small_mapping.csv` and `mapping.csv` have 2 columns: `id` and `name`

Here are a few rows from `mapping.csv`:

```
nm0000001,Fred Astaire  
nm0000004,John Belushi  
nm0000007,Humphrey Bogart  
tt0110997,The River wild
```

Each of those weird alphanumeric sequence is a unique identifier for either an actor or a director or a movie title.

## The Stages

---

This project is bigger than usual, so its broken into two parts, and you have more time. We recommend trying to complete stage 1 within one week so you have time for stage two.

- [Stage 1](#): combine the data from the movie and mapping files into a more useful format.
- [Stage 2](#): use the combined data to answer questions about movies, directors, and actors.