

# Final Project - Step 1 (8 Points)

## PSTAT100: Data Science Concepts and Analysis

### Due Date

The deadline for this step is Friday, April 25.

### Formation of groups:

#### The Academy

- Luke Drushell (drushell)
- David Pacheco (dpacheco)
- Rafael Romero (romeroflores)
- Kai Barker (kaibarker)
- Diana Lee (dianadlee)
- Davian Valencia (davianvalencia)

## 1 Tentative Role Distribution Table

Name	Responsibilities
Rafael	Leader and Project Manager
Diana	Scribe
Davian	Data Mapper 1
Luke	Data Mapper 2
David	Comms and Scheduling
Kai	Data Cleaner and Reorganizer

## 2 Data Information

**Data Name/Title:** *Top 10,000 Popular Movies TMDb*

**Author/Owner:** *Mahesh Jadev*

**Date of Publication:** *June 13, 2023*

**Publication Venue:** *Kaggle*

**Retrieval Date:** *April 22, 2025*

**Link:** <https://www.kaggle.com/datasets/ursmaheshj/top-10000-popular-movies-tmdb-05-2023>

### 3 Initial Insights:

The data appears to be fairly cleaned already, with few null or NA values. This makes quite a lot of sense, as it is a list of the most popular movies on TheMovieDataBase, which is a well-respected and popular movie info crowdsourcing service.