Our data is centered on movies that are on popular tv streaming services. We retrieved our data from Kaggle. It provided us with data for four of the most popular tv streaming services. What was convenient about these four datasets were that all the information recorded were identical to each other. As a result, the data was essentially easy to join for comparability and proper analysis. Although the datasets recorded both movies and tv shows for each tv streaming service, we decided to only focus on movies because the data for it was larger compared to tv shows.

* Netflix ([Netflix Movies and TV Shows | Kaggle](https://www.kaggle.com/datasets/shivamb/netflix-shows))
* Amazon Prime ([Amazon Prime Movies and TV Shows | Kaggle](https://www.kaggle.com/datasets/shivamb/amazon-prime-movies-and-tv-shows?resource=download))
* Disney Plus ([Disney+ Movies and TV Shows | Kaggle](https://www.kaggle.com/datasets/shivamb/disney-movies-and-tv-shows))
* Hulu ([Hulu Movies and TV Shows | Kaggle](https://www.kaggle.com/datasets/shivamb/hulu-movies-and-tv-shows))

We both agreed that loading the data on PostgresSQL, a relational database, benefitted us the most due to its ability to not replicate duplicates and the syntax fits our style of coding the most. Transforming the data answered the essential question we asked ourselves: which tv streaming service is the best for your buck in regards to their catalog of movies. We used the following question to help answer our question.

1. Which tv streaming service has the most movies? Unsurprisingly, Netflix has the most movies.