- **Find a problem worth solving, analyzing, or visualizing.**

  Creating a model to recommend 10 movies that have similar genres to a given movie title. If there are more than 10 movies with the same genres, the model will select and order the movies based on their ratings in descending order.

■ Process of creating a movie recommendation system:

1. Data and data delivery
   - Load movie csv and rating csv to two pandas.DataFrame.

2. Back end (ETL)
   - Firstly, splitting the movie_df title column into separate columns for title and year. Then filtering out movies with a release year greater than or equal to 2010.
   - That the ratings dataset grouped by 'movieId' with the average rating.
   - Finally, merge the 'movie_df' and 'rating_df' DataFrames together based on the 'movieId' column.

3. Create machine learning model by utilizing Scikit-learn package

   - We use train_test_split to split the dataset into train and test two sets.

   - Using a feature extraction technique CountVectorizer, we extract features from the ‘genres’ column in the training set. (example : there are two features in ‘Animation|Documentary’). Once the CountVectorizer is trained on the training set, we can use it to transform the genres of the movies in the testing set.

   - compute the pairwise cosine similarity between the genre vectors of movies in the training set using the cosine_similarity function. This allows us to measure the similarity between movies based on their genre features.