**ETL Project Proposal**

*Authors: Leeza Sergeeva, Alina Smirnova, Anna Burlyaeva*

**Idea:**

To create a pipeline to analyze the popularity of movies made in San Francisco. A movie’s popularity rate can be measured as a count of awards.

**Possible data sources:**

https://data.sfgov.org/Culture-and-Recreation/Film-Locations-in-San-Francisco/yitu-d5am

https://www.kaggle.com/theacademy/academy-awards

https://www.kaggle.com/carolzhangdc/imdb-5000-movie-dataset

https://www.kaggle.com/rounakbanik/the-movies-dataset#movies\_metadata.csv

**Possible types of transformation:** cleaning, joining, filtering, aggregating, etc.

**Final database can include:** movie\_name, year, actors, awards, gross\_income, budget, etc.