**ETL Project Proposal – Movies**

Group members: Hannah Duncan, Geoffrey Johnston, Ge (Melody) Yu, Vineet Sikri

Datasets to be used:

<https://www.kaggle.com/rounakbanik/the-movies-dataset?select=links_small.csv>

<https://www.kaggle.com/unanimad/the-oscar-award>

Proposed cleaning process:

1. For the Oscar file – we would like to keep only four categories from this dataset: picture, director, actors, actress.
2. From the Movies Dataset, we would like to use the credits.csv, movies\_metadat.csv, and ratings.csv.

Rough Breakdown of Tasks:

1. Preprocessing data (Geoffrey)
2. Extract (Geoffrey)
3. Transform (Hannah)
4. Load – we would like to load the final database into PostgreSQL (Vineet)
5. Write documentations and reports (Melody)