**Goal**

To create databases that can provide us the information of the movies that we can find in Netflix that have won an Oscar in any of these categories: Best Picture, Best Actor, Best Actress and Best Director.

**Extract (sources)**

In order to achieve our goal we used four databases that were obtained for the following links:

* <https://www.kaggle.com/shivamb/netflix-shows>
* <https://datahub.io/rufuspollock/oscars-nominees-and-winners#data>
* <https://www.openintro.org/data/index.php?data=oscars>
* https://cs.uwaterloo.ca/~s255khan/oscars.html#download

The original databases were found in csv format, you can find them in the Resources folder of our repository.

**Transform**

In order to make the join we had to do several transformations of the original databases.

At first we modified the Netflix database:

1. We filtered the database to keep just the movies.
2. We change the title of the movies to lowercase letters.

Then we modified the three tables that contain the categories of the Oscars that we are interested in. In general we had to do the next steps:

1. Filter the data from 1929 to 2015, since those are the years were we have complete information.
2. Filter the data to keep just the categories we’re interested in.
3. Delete the variables that we don’t need, such as the birthday of the actors and actresses.
4. Change the title to lowercase letters.
5. Concatenate the three databases.

All the specific steps we took can be found on the jupyter notebook ETL-team8.ipynb.

**Load**

On the folder you can find the tables created on SQL and the query used to join both tables and find the movies that have won an Oscar that are available in Netflix.

In team8\_project2.sql you can find the query to create the needed table. The tables we joined were netflix\_clean.csv and oscars\_clean.csv. The table that contains the result can be found as resultpgadmin.csv.