

Databases for data analytics

<https://github.com/evidencebp/databases-course/>

Aggregations and Grouping

Aggregations allow us to get statistics on sets, as a set

```
# Aggregations
select
count(*) as cnt
, count(distinct name) as movies_names
, max(m.rank) as rank_max
, min(m.rank) as rank_min
, avg(m.rank) as rank_avg
, std(m.rank) as rank_std
from
imdb_ajs.movies as m
;
```

In class exercises

- Number of movies per role
- Movies pairs of the same director
- Movies of director distribution (ignore directors without movies)

Exercises

- Movies pairs with at least 10 common roles, and the roles (concatenated)
 - Use only the pair whose first id is lower than 1,000 to restrict size
- Actors per movie distribution, with zero actors
- The interesting question - find an interesting question about movies and answer it with SQL