

Project 2: Group Work

1. How many Danish language movies are in the database?

```
SELECT COUNT(*)
FROM movie
WHERE language="Danish";
```

```
+-----+
| COUNT(*) |
+-----+
|      419 |
+-----+
```

Before index: 1.66 sec

After index: 0.00 sec

2. For each year, what is the total number of votes casted for all movies produced that year?

```
SELECT year, SUM(imdbVotes)
FROM movie
GROUP BY year;
```

```
+-----+-----+
| year | SUM(imdbVotes) |
+-----+-----+
| 1888 |          3825 |
| 1890 |           963 |
| 1891 |          1236 |
| 1892 |           358 |
| 1893 |           790 |
| 1894 |          7423 |
| 1895 |          9980 |
| 1896 |         13931 |
| 1897 |          2976 |
| 1898 |          4321 |
| 1899 |          1979 |
| 1900 |          3979 |
| 1901 |          5329 |
| 1902 |         13327 |
| 1903 |         14169 |
| 1904 |          4887 |
| 1905 |          3361 |
| 1906 |          3597 |
| 1907 |          3166 |
| 1908 |          3224 |
| 1909 |          4167 |
```

Project 2: Group Work

Group Fight Club

1910	4553
1911	2224
1912	7198
1913	5963
1914	21609
1915	26812
1916	22583
1917	13764
1918	15437
1919	25515
1920	47461
1921	56768
1922	76115
1923	29488
1924	52264
1925	102626
1926	60520
1927	129105
1928	89372
1929	84572
1930	101197
1931	254188
1932	180266
1933	215253
1934	160389
1935	189159
1936	226665
1937	223108
1938	211481
1939	539669
1940	460076
1941	505172
1942	484971
1943	198557
1944	306530
1945	241880
1946	457290
1947	228431
1948	380013
1949	319746
1950	479455
1951	458710
1952	376344
1953	474915
1954	717862
1955	529199
1956	450072
1957	794051
1958	507063
1959	770428
1960	774684
1961	541846
1962	826567
1963	690372

Project 2: Group Work

Group Fight Club

1964	885511
1965	640772
1966	840379
1967	806408
1968	1116911
1969	685213
1970	591246
1971	1178587
1972	1342327
1973	1220833
1974	1298563
1975	1559823
1976	1252216
1977	1379074
1978	1200022
1979	1741864
1980	2023343
1981	1661699
1982	2075497
1983	1889586
1984	2645872
1985	2501151
1986	2758365
1987	3138509
1988	3192192
1989	3377561
1990	3834953
1991	3382756
1992	3788834
1993	4806487
1994	6719609
1995	6600236
1996	5496440
1997	7603741
1998	7863270
1999	10257878
2000	9168448
2001	11048966
2002	11103390
2003	11430018
2004	13287732
2005	13182780
2006	14248999
2007	14961292
2008	14387256
2009	13132136
2010	12573962
2011	12036633
2012	3662518
+-----+	

Before index: 0.06

After index: 0.06

3. How many actors and directors have a first name starting with C?

We assume that an actor/director is unique (in that we don't count them multiple times, if they appear in multiple movies).

```
SELECT COUNT(DISTINCT person.id)
FROM person
INNER JOIN involved ON person.id=involved.personId
WHERE name LIKE 'C%'
AND (role='actor' OR role='director')
ORDER BY name;
```

```
+-----+
| COUNT(DISTINCT person.id) |
+-----+
|                10660 |
+-----+
```

Before index: 46.34 sec

After index: 0.61 sec

4. What are the names and birth years for all actors in Pulp Fiction? Your query should list the actors in increasing order of birth year.

We assume that ordering by birth year breaks ties by month, date etc. (birth date)

```
SELECT name, YEAR(birthdate)
FROM person
INNER JOIN involved on person.id=involved.personId
INNER JOIN movie ON movie.id=involved.movieId
WHERE movie.title='Pulp Fiction'
AND role='actor'
ORDER BY birthdate;
```

```
+-----+-----+
| name          | YEAR(birthdate) |
+-----+-----+
| Caleb Allen   | NULL            |
| Rian Turner   | NULL            |
| Don Blakely   | NULL            |
| Brad Blumenthal | NULL            |
| Patricio Calderon | NULL            |
```

Project 2: Group Work

Group Fight Club

Emmett Clark		NULL	
Angel Star Jones		NULL	
Linda Kaye		NULL	
Sophie Griffiths		NULL	
Bronagh Gallagher		NULL	
Peggy Greene		NULL	
Samadu Jackson		NULL	
Stephen Hibbert		NULL	
Devin Marcus Miller		NULL	
Venessia Valentino		NULL	
Glendon Rich		NULL	
Richard Ruth		NULL	
Terry Roth		NULL	
Emil Sitka		1914	
Harvey Keitel		1939	
Rene Beard		1941	
Christopher Walken		1943	
Joseph Pilato		1949	
Brenda Hillhouse		1953	
John Travolta		1954	
Bruce Willis		1955	
Amanda Plummer		1957	
Lawrence Bender		1957	
Steve Buscemi		1957	
Karen Maruyama		1958	
Ving Rhames		1959	
Duane Whitaker		1959	
Rosanna Arquette		1959	
Julia Sweeney		1959	
Kathy Griffin		1960	
Eric Stoltz		1961	
Michael Gilden		1962	
Quentin Tarantino		1963	
Frank Whaley		1963	
Maria de Medeiros		1965	
Burr Steers		1965	
Phil LaMarr		1967	
Alexis Arquette		1969	
Uma Thurman		1970	
+-----+-----+			

Before index: 1.59 sec

After index: 0.00 sec

5. Which movies have either John Travolta or Uma Thurman, but not both starred in?

```
SELECT title FROM involved
INNER JOIN person ON involved.personId=person.id
INNER JOIN movie ON movie.id = movieId
WHERE name IN ('John Travolta', 'Uma Thurman')
GROUP BY movieId
HAVING COUNT(*) = 1;
```

title
"20 to 1"
"20/20"
"Ant & Dec's Saturday Night Takeaway"
"Behind the Music"
"Biography"
"Breakfast"
"CBS News Sunday Morning"
"El hormiguero"
"Emergency!"
"Fat Actress"
"Fear Factor"
"Gomorron"
"Howard Stern"
"I Love the '80s"
"Independent Lens"
"Inside the Actors Studio"
"Kirstie Alley's Big Life"
"Larry King Live"
"Last Call with Carson Daly"
"Late Night with Conan O'Brien"
"Late Night with Jimmy Fallon"
"Le grand journal de Canal+"
"Live from Studio Five"
"MADtv"
"Medical Center"
"Parkinson"
"Rove Live"
"Saturday Night Live"
"Smash"
"Taff"
"Tavis Smiley"
"The 7PM Project"
"The City"
"The Daily Show"
"The Dodge Dancing Party"
"The Drew Carey Show"
"The Howard Stern Radio Show"

Project 2: Group Work

Group Fight Club

"The Jay Leno Show"	
"The Merv Griffin Show"	
"The O'Reilly Factor"	
"The Rookies"	
"The Tonight Show Starring Johnny Carson"	
"The Tonight Show with Conan O'Brien"	
"This Morning"	
"Today"	
"Top of the Pops"	
"Unscripted"	
"Welcome Back, Kotter"	
A Civil Action	
A Love Song for Bobby Long	
A Month by the Lake	
All the Love You Cannes!	
Austin Powers in Goldmember	
Basic	
Batman & Robin	
Battlefield Earth	
Beautiful Girls	
Bel Ami	
Blow Out	
Bolt	
Boris and Natasha	
Broken Arrow	
Carrie	
Ceremony	
Chains of Gold	
Chelsea Walls	
Dangerous Liaisons	
Domestic Disturbance	
Even Cowgirls Get the Blues	
Eyes of an Angel	
Face/Off	
Final Analysis	
From Paris with Love	
Gattaca	
Get Shorty	
Grease	
Hairspray	
Henry & June	
Jennifer Eight	
Johnny Be Good	
Junket Whore	
Kaze no tani no Naushika	
Kill Bill: Vol. 1	
Kill Bill: Vol. 2	
Kiss Daddy Goodnight	
Ladder 49	
Les Misérables	
Lonely Hearts	
Look Who's Talking	
Look Who's Talking Now	
Look Who's Talking Too	

Project 2: Group Work

Group Fight Club

Lucky Numbers	
Mad City	
Mad Dog and Glory	
Magnificent Desolation: Walking on the Moon 3D	
Michael	
Mission Zero	
Moment by Moment	
My Super Ex-Girlfriend	
Off the Menu: The Last Days of Chasen's	
Orientation: A Scientology Information Film	
Paul Williams Still Alive	
Paycheck	
Percy Jackson & the Olympians: The Lightning Thief	
Perfect	
Phenomenon	
Primary Colors	
Prime	
Religulous	
Saturday Night Fever	
Savages	
She's So Lovely	
Shout	
Staying Alive	
Sweet and Lowdown	
Swordfish	
Tape	
That's Dancing!	
The Accidental Husband	
The Adventures of Baron Munchausen	
The Avengers	
The Devil's Rain	
The Experts	
The General's Daughter	
The Golden Bowl	
The Guru	
The Life Before Her Eyes	
The Naked Brothers Band: The Movie	
The Producers	
The Punisher	
The Taking of Pelham 1 2 3	
The Thin Red Line	
The Truth About Cats & Dogs	
Two of a Kind	
Urban Cowboy	
Valentino: The Last Emperor	
Vatel	
Welcome to Hollywood	
Where the Heart Is	
White Man's Burden	
Wild Hogs	
+-----+	

Before index: 1.57

After index: 0.00

6. Which actors in Pulp Fiction have never, before or after, starred in the same movie as one of the other actors in Pulp Fiction?

7. What are the titles and years of all movies since 1980 where John Travolta costarred with Samuel L Jackson?

We assume that costarring in a moving involves the possibility of being both actor and director.

```
SELECT title, movie.year
FROM person
INNER JOIN involved ON involved.personId=person.id
INNER JOIN movie ON movie.id=involved.movieId
WHERE year >= 1980
AND name in ('Samadu Jackson', 'John Travolta')
GROUP BY title, year
HAVING count(distinct name) = 2
ORDER BY year;
```

title	year
"Entertainment Tonight"	1981
"Biography"	1987
"Live with Regis and Kathie Lee"	1988
"The Charlie Rose Show"	1991
"The Tonight Show with Jay Leno"	1992
"Gomorrion"	1992
"HBO First Look"	1992
"Late Show with David Letterman"	1993
"GMTV"	1993
"Inside the Actors Studio"	1994
Pulp Fiction	1994
"The O'Reilly Factor"	1996
"The Rosie O'Donnell Show"	1996
Off the Menu: The Last Days of Chasen's	1997
"The View"	1997
You're Still Not Fooling Anybody	1997
"The Howard Stern Radio Show"	1998
"Friday Night with Jonathan Ross"	2001
Basic	2003
"Ellen: The Ellen DeGeneres Show"	2003
"Le grand journal de Canal+"	2004

"Tavis Smiley"	2004
Boffo! Tinseltown's Bombs and Blockbusters	2006
"The Tonight Show with Conan O'Brien"	2009

+-----+

Before index: 1.28

After index: 0.01

8. What are the top-5 highest rated movies from the 1990s according to IMDB users?

We assume that the "movies" refers to the movies' titles.

```
SELECT title FROM movie
INNER JOIN ratings ON movieId=id
WHERE year >= 1990
AND year < 2000
ORDER BY rating
DESC
LIMIT 5;
```

+-----+
title
+-----+
The Matrix
American Beauty
Pulp Fiction
The Sixth Sense
Léon
+-----+

Before index: 0.03

After index: 0.00

9. In 1994, what was the average IMDB rating of a movie for each language?

The movie "The Lion King" didn't have a language associated with it.

```
SELECT avg(rating), language FROM ratings
JOIN movie ON movie.id=ratings.movieId
WHERE year=1994
GROUP BY language;
```

```
+-----+-----+
| avg(rating) | language |
+-----+-----+
|      7.0000 |         |
|      8.3043 | English  |
|      9.0000 | French   |
+-----+-----+
```

Before index: 0.02

After index: 0.00

10. Which movie starring John Travolta has the highest IMDB user rating?

```
SELECT title
FROM involved
INNER JOIN person ON personId=person.id
INNER JOIN movie ON movieId=movie.id
WHERE name='John Travolta'
ORDER BY imdbRank DESC
LIMIT 1;
```

```
+-----+
| title      |
+-----+
| Pulp Fiction |
+-----+
```

Before index: 0.86

After index: 0.00

11. How many actresses did not live at the same time as Charles Chaplin?

```
SELECT COUNT(DISTINCT p2.id)
FROM person AS p1, person AS p2
INNER JOIN involved AS inv ON inv.personId=p2.id
WHERE p1.name='Charles Chaplin'
AND inv.role='actor'
AND (p2.deathdate < p1.birthdate OR p2.birthdate > p1.deathdate)
AND p2.gender='f';
```

```
+-----+
| COUNT(*) |
+-----+
|      5405 |
+-----+
```

Before index: 1.55

After index: 0.34

12. Who was the shortest actor to star in an action movie?

```
SELECT name, height
FROM involved
INNER JOIN genre ON genre.movieId=involved.movieId
INNER JOIN person ON person.id=personId
WHERE genre='Action'
AND role='actor'
AND height IS NOT NULL
GROUP BY personId
ORDER BY height
LIMIT 1;
```

```
+-----+-----+
| name           | height |
+-----+-----+
| Crystal the Monkey |      48 |
+-----+-----+
```

Before index: 4.30

After index: 2.23

13. What is the average rating of movies from each genre?

```
SELECT genre, avg(rating) AS AverageRating
FROM movie
INNER JOIN genre ON movie.id=genre.movieId
INNER JOIN ratings ON ratings.movieId=movie.id
GROUP BY genre.genre
ORDER BY AverageRating DESC;
```

```
+-----+-----+
| genre      | AverageRating |
+-----+-----+
| Crime      | 8.3950        |
| War        | 8.2692        |
| Biography  | 8.1333        |
| History    | 8.1250        |
| Mystery    | 8.0896        |
| Drama      | 7.8423        |
| Thriller   | 7.7544        |
| Action     | 7.5348        |
| Sci-Fi     | 7.4710        |
| Adventure  | 7.4479        |
| Family     | 7.3143        |
| Animation  | 7.3143        |
| Fantasy    | 7.2685        |
| Comedy     | 7.2381        |
| Romance    | 7.1707        |
| Musical    | 7.0000        |
+-----+-----+
```

Before index: 0.00

After index: 0.00

14. For each genre, what is the number of student ratings of a movie from that genre? List only genres with at least 10 ratings.

Our assumption is that a 'student' means someone with an ITU-email.

```
SELECT * FROM (
    SELECT genre, SUM(rating) AS RatingSum
    FROM movie
    INNER JOIN genre
    ON movie.id=genre.movieId
    INNER JOIN ratings ON ratings.movieId=movie.id
    WHERE user LIKE"%@itu.dk%"
    GROUP BY genre.genre ORDER BY RatingSum
) AS DerivedTable
```

```
WHERE RatingSum >= 10;
```

genre	RatingSum
History	33
Biography	49
War	109
Animation	137
Family	137
Romance	166
Comedy	191
Mystery	249
Fantasy	378
Crime	475
Sci-Fi	483
Thriller	596
Adventure	658
Action	790
Drama	873

Before index: 0.01

After index: 0.01

15. Which movie has the largest number of 2-link references?
(If A refers to B, and B refers to C, then we say that A has a 2-link reference, through B, to C. If there are several paths leading from A to C, we count all of them.)

```
SELECT movie.title, COUNT(*)
FROM movieref AS mr1, movieref AS mr2, movie
WHERE mr1.fromId = movie.id
AND mr1.toId = mr2.fromId
GROUP BY mr1.fromId
ORDER BY COUNT(*)
DESC LIMIT 1;
```

title	COUNT(*)
"Saturday Night Live"	38834

Before index: 2.04

After index: 1.93

16. How many actors have also been active as director of at least one movie?

We assume that the a person only appears once in the involved table for each role/movieId combination.

```
SELECT COUNT(DISTINCT personId)
FROM (
    SELECT personId FROM involved
    GROUP BY personId, movieId
    HAVING COUNT(*) > 1
) AS actorDirectors;
```

```
+-----+
| COUNT(DISTINCT personId) |
+-----+
|                2951      |
+-----+
```

Before index: 8.79

After index: 1.00

17. Which two genres are most often linked to the same movie? (Note that each movie has a set of genres.)

```
SELECT g1.genre, g2.genre, COUNT(*)
FROM genre AS g1, genre AS g2
WHERE g1.movieId=g2.movieId
AND g1.genre < g2.genre
GROUP BY g1.genre, g2.genre
ORDER BY COUNT(*)
DESC LIMIT 1;
```

```
+-----+-----+-----+
| genre | genre | COUNT(*) |
+-----+-----+-----+
| Drama | Romance |      5837 |
+-----+-----+-----+
```

Before index: 0.46

After index: 0.44

Question 2: Indexing

```
CREATE INDEX i_personId ON involved(personId);  
CREATE INDEX i_movieId ON involved(movieId);  
CREATE INDEX i_pmId ON involved(personId, movieId);
```

```
CREATE INDEX i_movieId ON ratings(movieId);  
CREATE INDEX i_rating ON ratings(rating);
```

```
CREATE INDEX i_language ON movie(language);  
CREATE INDEX i_height ON person(height);  
CREATE INDEX i_genre ON genre(genre);
```