CSC 370 — Database Systems
Fall 2016
Assignment No. 5

Note 1 **This assignment is to be done individually**

Note 2 Working with other people is prohibited.

- Due date: Oct 28, 2016, at the beginning of the class.
- This assignment is worth 1% of your total course mark.
- Summit in paper your queries, and their corresponding relational algebra.
- Submit electronically the SQL queries in a single **text** file.

Objectives

After completing this assignment, you will have experience:

- Use aggregation.

Your task, should you choose to accept it

1. Answer the following questions, both in relational algebra, and SQL. **nRelational algebra queries should match SQL**. For SQL queries provide the query and the result. One query per question. Your query should only use the information provided in the question. **You cannot use the LIMIT keyword in your queries.**

   1. Let us analyze TV shows. In this case, we want to list the TV-shows with episodes that have ratings in at least 4 different seasons. To limit the number of tuples, let us narrow our search to those TV shows that have an average rank of their episodes > 8.5 (only the episodes) and have an average number of votes for their episodes of 1,000 votes or more. List the id of their production (episodeof), their average rank, their average number of votes, the number of episodes with ratings and the number of different seasons with ratings. Order the result by average rank descending first and in case of collisions by average number of votes descending second. Hint: use count(distinct attr) to count the number of different values of a given attribute.

```
| episodeof                     |            arank |                avotes | cepisodes | cseasons |
|-------------------------------+------------------+-----------------------+-----------+----------|
| "Person of Interest" (2011)   | 9.13980582524272 | 2047.6310679611650485 |       103 |        5 |
| "Breaking Bad" (2008)         | 9.01935483870967 |    11124.709677419355 |        62 |        5 |
| "Game of Thrones" (2011)      | 8.99833333333333 |    19718.333333333333 |        60 |        6 |
| "Sherlock" (2010)             | 8.94166666666667 |    15931.333333333333 |        12 |        4 |
| "Suits" (2011)                | 8.85357142857143 | 1318.0000000000000000 |        84 |        6 |
| "The Wire" (2002)             | 8.84333333333333 | 1422.1166666666666667 |        60 |        5 |
| "House of Cards" (2013)       | 8.72884615384616 | 2628.5000000000000000 |        52 |        4 |
| "Supernatural" (2005)         | 8.71611570247934 | 2059.5123966942148760 |       242 |       11 |
| "House M.D." (2004)           | 8.69318181818182 | 1439.4602272727272727 |       176 |        8 |
| "Sons of Anarchy" (2008)      | 8.69239130434782 | 1145.6195652173913043 |        92 |        7 |
| "Prison Break" (2005)         | 8.68395061728395 | 1629.8518518518518519 |        81 |        4 |
| "Lost" (2004)                 | 8.67692307692308 | 2819.5811965811965812 |       117 |        6 |
| "The Sopranos" (1999)         | 8.67209302325582 | 1574.4186046511627907 |        86 |        6 |
| "Dexter" (2006)               | 8.64166666666667 | 2802.8750000000000000 |        96 |        8 |
| "Luther" (2010)               | 8.59411764705882 | 1103.4705882352941176 |        17 |        4 |
| "Boardwalk Empire" (2010)     | 8.55087719298245 | 1095.0877192982456140 |        57 |        5 |
| "Friends" (1994)              | 8.54152542372881 | 1642.8898305084745763 |       236 |       10 |
| "Vikings" (2013)              | 8.53589743589743 | 1740.2051282051282051 |        39 |        4 |
| "Mad Men" (2007)                | 8.53152173913044 | 1100.8804347826086957 |        92 |         7 |
| "Arrow" (2012)                  | 8.52717391304348 | 3053.7934782608695652 |        92 |         4 |
| "Once Upon a Time" (2011)       |  8.5212389380531 | 1000.0530973451327434 |       113 |         5 |
| "Arrested Development" (2003) | 8.51470588235294 | 1062.6176470588235294 |        68 |         4 |
(22 rows)
```

2. Of the movies with at least 50,000 votes, list the one(s) with the highest rank. List its title, year, rank, and votes.

```
|          title           | year | rank |  votes
--------------------------+------+------+---------
 The Shawshank Redemption | 1994 |  9.3 | 1698604
(1 row)
```

3. This query is restricted to movies with a rank of at least 8 and at least 50,000 votes. Find the **pid** of persons who have been in at least 10 of these movies. List their **pid**, number of such movies, and average rating of such movies. Order by average rank.

```
| pid                | count |              avg |
|--------------------+-------+------------------|
| Tovey, Arthur      |    10 |             8.27 |
| Lynn, Sherry (I)   |    11 | 8.27272727272727 |
| Jackson, Samuel L. |    10 |             8.28 |
| Ratzenberger, John |    11 | 8.29090909090909 |
| Flowers, Bess      |    12 | 8.30833333333333 |
| De Niro, Robert    |    10 |             8.36 |
(6 rows)
```

4. For movies with at least 50000 votes, and rank of more than 8 list the person (or persons) that has appeared in the most of such movies, the id of the movie, their billing, and their character. Result should contain pid, id, billing, and character.

```
          pid           |                          id                           | billing |        character
------------------------+-------------------------------------------------------+---------+--------------------------
 Ratzenberger, John (I) | Finding Nemo (2003)                                   |      24 | Fish School
 Ratzenberger, John (I) | Gandhi (1982)                                         |     124 | American Lieutenant
 Ratzenberger, John (I) | Inside Out (2015/I)                                   |      17 | Fritz
 Ratzenberger, John (I) | Monsters, Inc. (2001)                                 |       8 | The Abominable Snowman
 Ratzenberger, John (I) | Sen to Chihiro no kamikakushi (2001)                  |      56 | Assistant Manager
 Ratzenberger, John (I) | Star Wars: Episode V - The Empire Strikes Back (1980) |      31 | Rebel Force Major Derlin
 Ratzenberger, John (I) | Toy Story (1995)                                      |       6 | Hamm
 Ratzenberger, John (I) | Toy Story 3 (2010)                                    |       8 | Hamm
 Ratzenberger, John (I) | Up (2009)                                             |       7 | Construction Foreman Tom
 Ratzenberger, John (I) | WALL·E (2008)                                        |       6 | John
(10 rows)
```

5. For this question consider only movies with at least 50,000 votes. Movies with at least 40,000 votes and a rank > 8.4 are usually very good; let us call these movies *good movies* Some directors are really good, others are lucky. For every director who has directed at least 8 movies (regardless rating), but had directed at least one *good movie*, display his/her pid, the total number of movies made, the percentage of movies with rank > 8, the number of movies with rank > 8.4, their average ranking, the number of movies less or equal to this ranking, and their average, and compute the difference between the average of the good ones minus the average of the rest. Order by the difference (descending) first, and in the case of same difference, by percentage of *good movies* (descending). Your result should contain 8 columns: pid, number of movies made (total), total of good ones (goodones), their average rank (avggoodones), the number of other movies (rest), their average rank (avgrest) and the difference of agvgoodones minus avgrest. Note the formatting of the output, you can achieve it by using the `to_char` function in postgresql.

```
          pid           |  prop  | total | goodones | avggoodones | rest | avgrest | diff
------------------------+--------+-------+----------+-------------+------+---------+-------
 Coppola, Francis Ford  | 33.3% |     9 |        3 |  8.9        |    6 |  7.2    |   1.7
 Zemeckis, Robert       | 13.3% |    15 |        2 |  8.7        |   13 |  7.1    |   1.6
 Scott, Ridley          | 11.1% |    18 |        2 |  8.5        |   16 |  7.0    |   1.5
 Singer, Bryan          | 12.5% |     8 |        1 |  8.6        |    7 |  7.1    |   1.5
 Jackson, Peter (I)     | 27.3% |    11 |        3 |  8.8        |    8 |  7.4    |   1.4
 Spielberg, Steven      | 12.0% |    25 |        3 |  8.7        |   22 |  7.4    |   1.2
 Fincher, David         | 20.0% |    10 |        2 |  8.7        |    8 |  7.5    |   1.2
 Nolan, Christopher (I) | 66.7% |     9 |        6 |  8.7        |    3 |  7.7    |   1.0
 Tarantino, Quentin     | 16.7% |    12 |        2 |  8.7        |   10 |  7.8    |    .9
 Scorsese, Martin (I)   | 13.3% |    15 |        2 |  8.6        |   13 |  7.7    |    .9
 Miyazaki, Hayao        | 10.0% |    10 |        1 |  8.6        |    9 |  8.0    |    .6
 Hitchcock, Alfred (I)  | 14.3% |    14 |        2 |  8.5        |   12 |  8.0    |    .5
 Kubrick, Stanley       | 18.2% |    11 |        2 |  8.5        |    9 |  8.0    |    .5
(13 rows)
```

What to submit

In paper submit your the Relational Algebra, the SQL queries. Electronically, submit your SQL queries and the results.