**Assignment Instructions**

On our AWS instance (Bar-Beer-Drinker) you will find is REAL data from Trip Advisor, from my home town of **Gdansk, Poland**.  It shows rankings of top 60 restaurants in Gdansk, their scores (average number of stars awarded by reviewers etc.

Below, we explain the meaning of individual columns:

**Restaurant (name)**Rank (how it was ranked by Trip Advisor, it is not known what their ranking function is!, it is NOT just SCORE, for sure)

**Score  (**average stars awarded by reviewers)

**User\_Name** -  name of the reviewer/alias

**Review\_Stars** - How many stars did this reviewer give to this restaurant

**Review\_Date** -  The date of the review

**User\_Reviews** - Total number of reviews this reviewer gave over his/her history of reviewing for hotels/resturants etc

**User\_Restaurant\_Reviews** - Same but just for restaurants

**User\_Helpful** - How many of the user reviews were marked as helpful by other reviewers  
  
You goal is to use sql aggregate functions and find some "interesting patterns" or may be even discover some hints towards how Trip Advisor is really ranking these restaurants.

To be ranked high by Tripadvisor is very important for the restaurants - often being ranked low means lots of lost revenue!

More and more tourists rely on tripadvisor reviews.  Are there are biased reviewers here may be?   Some who give high reviews to same restaurants or low reviews to others?  
  
USE SQL and show us how you used it to discover these hidden patterns. The whole point here is to use SQL, not R, not Python etc.....SQL is good for this and you should show is how.

Is there something fishy in the data? Can you help a restaurant to reverse engineer Tripadvisor ranking algorithm?  
  
Feel free to use plots if  this helps!  SUBMIT up to 10 **POWER POINT SLIDES** about what you discovered and why it is important  - since you may be asked to present in class.

The warm up SQL queries and answers you got should, on the other hand, be submitted as text file like for previous HW.  
  
Here are 3 warm-up queries which you also have to run and submit along with results:  
  
1. For each restaurant, how many 3-star reviews did each get? List these restaurants in tripadvisor rank order from top to bottom.  
2. Find average score of top 10 restaurants and compare it with average score of the bottom 10 restaurants according to trip advisor ranking.

Data has around 6,000+ tuples. If you want to do some joins and they take too long feel free to restrict your findings only to a subset to say, top 30 restaurants to reduce the size of data.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| RESTAURANT | RANK | SCORE | USER\_NAME | REVIEW\_STARS | REVIEW\_DATE | USER\_REVIEWS | USER\_RESTAURANT\_REVIEWS | USER\_HELPFUL\_VOTES | | |
| Ritz Restaurant | 1 | 4.5 | Iain U | 5 | 11/12/2015 | 60 | 33 | 29 |  |  |
| Ritz Restaurant | 1 | 4.5 | Darius S | 5 | 11/12/2015 | 2 |  | 4 |  |  |
| Ritz Restaurant | 1 | 4.5 | Jonathan A | 5 | 11/5/2015 | 4 | 3 | 3 |  |  |
| Ritz Restaurant | 1 | 4.5 | Juliesamari | 5 | 11/5/2015 | 39 | 23 | 14 |  |  |
| Ritz Restaurant | 1 | 4.5 | Novajane | 5 | 11/5/2015 | 6 | 6 | 3 |  |  |
| Ritz Restaurant | 1 | 4.5 | Jeanettehaga | 5 | 11/5/2015 | 3 |  | 1 |  |  |
| Ritz Restaurant | 1 | 4.5 | Knut Marius S | 5 | 11/5/2015 | 8 | 7 | 3 |  |  |
| Ritz Restaurant | 1 | 4.5 | malcjones | 5 | 10/29/2015 | 12 | 4 | 25 |  |  |
| Ritz Restaurant | 1 | 4.5 | Dai S | 5 | 10/29/2015 | 39 | 21 | 15 |  |  |
| Ritz Restaurant | 1 | 4.5 | Unni S | 5 | 10/29/2015 | 2 |  | 3 |  |  |
| Ritz Restaurant | 1 | 4.5 | Corniglia123 | 5 | 10/29/2015 | 2 |  | 2 |  |  |
| Ritz Restaurant | 1 | 4.5 | Jakub P | 5 | 10/22/2015 | 7 | 7 | 15 |  |  |
| Ritz Restaurant | 1 | 4.5 | Tori K | 5 | 10/22/2015 | 67 | 51 | 46 |  |  |
| Ritz Restaurant | 1 | 4.5 | j0annaa2015 | 5 | 10/22/2015 | 1 |  | 8 |  |  |