BDMM - 1st Project

April 24, 2020

1 Big Data Modeling and Management Assigment

1.1 The Beer project

As it was shown in classes, graph databases are a natural way of navegating distinct types of data. For this first project we will be taking a graph database to analyse beer and breweries!

For reference the dataset used for this project has been extracted from kaggle, released by Evan Hallmark. Even though the author does not present metada on the origin of the data it is probably a collection of open data from places like beeradvocate

Problem description Explore the database via python neo4j connector and/or the graphical tool in the NEO4J webpage. Answer the questions. Submit the results by following the instructions

Connection details to the neo4j database

```
Host: rhea.isegi.unl.pt:7475
Username: neo4j
Password: F3cfcrnvBev57KZ8mcMk78L9wHgJVZuJ
Connect URL : bolt://rhea.isegi.unl.pt:7687
```

Questions

0. **Example** Question *How many beers does the database contain?*
1. How many countries are accounted in this database?
2. Which entity holds the most reviews:
    1. Beer?
    2. Brewery?
    3. Country?
3. Find the user/users that have the most shared reviews(reviews or the same beers) with the user CTJman?
4. Which Portuguese brand has the most beers?
5. Out of those beers, which has the most reviews?
6. On average how many different beer styles does each brewery produce?
7. Which brewery produces the strongest beers according to ABV?
8. If I typically enjoy a beer due to its aroma and appearance, which beer style should I try?

9. If you had to pick 3 beers to recommend using only this data, which would you pick?

Groups Groups should have 3 to 4 people
You should register your group in the link: - https://liveeduisegiunl-my.sharepoint.com/:x:/g/personal/fpinheiro_novaims_unl_pt/EYtFTvuGwiRCsakM67f3C4oBrafEJ8M1juXlP7C

Submission Submission of the query results to be done to the redis database (explained on the first class), with the exception of question 9.
The following format is expected:

```
>>> redis.set("GROUP1:0", "358873")
```

This result should be the anwser of group 1 to question 0

The solutions with the code used to produce the results and respective explations should be uploaded to moodle. They should have a clear reference to the group, either on the file name or on the document itself. Preferably one Jupyter notebook per group.

Delivery date: Until the **midnight of May 7th**

Evaluation This will be 15% of the final grade.
Each solution will be evaluate on 2 components: correctness of results and simplicity of the solution. All code will go through plagiarism automated checks. Groups with the same code will undergo investigation.