Lab 1: Why Databases?
Part I

Due date: Tuesday, April 11, 12:00pm (noon!)

Lab Assignment

Assignment Preparation

This is an individual assignment designed to accomplish two goals. First, it demonstrates the kinds of tasks that are typically accomplished by the database management systems, the roles DBMS play in data processing and delivery of results to end users. Second, it also tests your knowledge of Python, and potentially mildly encourages you to learn Python's declarative programming techniques for working with data.

The Task

The full lab assignment consists of two parts. The first part of the assignment is given to you now. The second part will be released on Thursday, January 19 and will consist of another set of 10 questions, but with somewhat different data.

The assignment consists of a Jupyter Python notebook shared with you on the course web page and an accompanying data file (a CSV file).

Data. The data you will be working with is one half of a popular Kaggle dataset documenting the effects of alcohol consumption on student learning. The full dataset contains two spreadsheets: one for a mathematics course in two Portuguese high schools, and one - for a Portuguese language class in the same high schools. We have selected the mathematics course spreadsheet for this lab assignment. The Python notebook contains the link to the original Kaggle page for the dataset which has explanations for each column in the spreadsheet. **Make sure you study this information before starting your work on the lab.**

Assignment. The Jupyter notebook shared with you loads the CSV file, renders it in three different ways (as a `pandas` data frame, as a numPy array, and as a Python list (of lists)). You can choose to work with any of the three data representations, on create your own (there are other ways to store 2-dimensional data in Python). The notebook contains some instructions that you need to read before working on the assignment. The assignment itself is 10 questions about the data contained in the CSV file. Each question needs to be answered by performing some computations over the contents of the CSV file - searching for information, and possibly using some of the found information to perform additional computations.

For each question you are given a Jupyter code cell to put your code in. The result of running the cell should be clear and consise output containing the information requested (either display the variable used to store the results, or pretty print the results). All questions shall be answered in isolation: that is, no variables used for answering one question shall be used to answer another. You can, however (and are encouraged to where appropriate), copy-paste the code.

Submission. When you are done, make sure that your notebook is properly saved. Put your name and email address at the top of the notebook. Submit using the `handin` command:

```
$ handin dekhtyar 365-lab01-1 <FILE>
```

The notebook contains submission instructions that allow you to submit directly from your Jupyter notebooks environment, without having to ssh into unix1.

Good Luck!