Please develop and optimize 3 machine learning models (logistic regression, decision tree and random forest) to predict those who have diabetes (gh >= 6.5%) using Python in a Jupyter notebook. Thereafter, please interpret and compare the fine-tuned models.

The data dictionary may be found at https://hbiostat.org/data/repo/nhgh
The dataset may be found at https://hbiostat.org/data (scroll down to NHANES glycohemoglobin data and download the nhgh.tsv file)

Your solution may include the following:

- Convert data to a tidy format
- Export to normalized tables in an SQLite3 database
- Use SQL statements (using from within pandas is acceptable) to retrieve the data needed for each visualization

Place your notebook with reproducible results in a public GitHub repository. By reproducible, we mean that when we run all cells in the notebook, the same results should be recreated each time.

We are looking for evidence of the following basic data science skills:

- Idiomatic python
- Jupyter notebooks
- Literate programming
- Experience with version control
- Data processing skills
- Relational database knowledge
- Data visualization skills
- Machine learning knowledge
- Modelling skills

You may expect to present your results during the interview followed by a short Q&A, if you are invited.