Software Development Evaluation Problem

This exercise is to help us understand how you go about solving a problem, and what methodologies and techniques you used to solve it.

When you come in for your interview you will be asked to give a presentation on this project. What you talk about in this is up to you. We are looking to see how you solve problems, and how you present findings to a group.

Problem Description

We have included a dataset that represents achievement in secondary education in a school. The data attributes include student grades, demographic, social and school related features. The dataset is a comma delimited file (CSV). Its attributes can be defined as follows:

1. **school** - student's school (binary: 'GP' - Gabriel Pereira or 'MS' - Mousinho da Silveira)
2. **sex** - student's sex (binary: 'F' - female or 'M' - male)
3. **age** - student's age (numeric: from 15 to 22)
4. **address** - student's home address type (binary: 'U' - urban or 'R' - rural)
5. **famsize** - family size (binary: 'LE3' - less or equal to 3 or 'GT3' - greater than 3)
6. **Pstatus** - parent's cohabitation status (binary: 'T' - living together or 'A' - apart)
7. **Medu** - mother's education (numeric: 0 - none, 1 - primary education (4th grade), 2 - 5th to 9th grade, 3 - secondary education or 4 - higher education)
8. **Fedu** - father's education (numeric: 0 - none, 1 - primary education (4th grade), 2 - 5th to 9th grade, 3 - secondary education or 4 - higher education)
9. **Mjob** - mother's job (nominal: 'teacher', 'health' care related, civil 'services' (e.g. administrative or police), 'at_home' or 'other')
10. **Fjob** - father's job (nominal: 'teacher', 'health' care related, civil 'services' (e.g. administrative or police), 'at_home' or 'other')
11. **reason** - reason to choose this school (nominal: close to 'home', school 'reputation', 'course' preference or 'other')
12. **guardian** - student's guardian (nominal: 'mother', 'father' or 'other')
13. **traveltime** - home to school travel time (numeric: 1 - <15 min., 2 - 15 to 30 min., 3 - 30 min. to 1 hour, or 4 - >1 hour)
14. **studytime** - weekly study time (numeric: 1 - <2 hours, 2 - 2 to 5 hours, 3 - 5 to 10 hours, or 4 - >10 hours)
15. **failures** - number of past class failures (numeric: n if 1<=n<3, else 4)
16. **schoolsup** - extra educational support (binary: yes or no)
17. **famsup** - family educational support (binary: yes or no)
18. **paid** - extra paid classes within the course subject (Math or Portuguese) (binary: yes or no)
19. **activities** - extra-curricular activities (binary: yes or no)
20. **nursery** - attended nursery school (binary: yes or no)
21. **higher** - wants to take higher education (binary: yes or no)
22. **internet** - Internet access at home (binary: yes or no)
23. **romantic** - with a romantic relationship (binary: yes or no)
24. **famrel** - quality of family relationships (numeric: from 1 - very bad to 5 - excellent)
25. **freetime** - free time after school (numeric: from 1 - very low to 5 - very high)
26. **goout** - going out with friends (numeric: from 1 - very low to 5 - very high)
27. **Dalc** - workday alcohol consumption (numeric: from 1 - very low to 5 - very high)
28. **Walc** - weekend alcohol consumption (numeric: from 1 - very low to 5 - very high)
29. **health** - current health status (numeric: from 1 - very bad to 5 - very good)
30. **absences** - number of school absences (numeric: from 0 to 93)
31. **G1** - first period grade (numeric: from 0 to 20)
32. **G2** - second period grade (numeric: from 0 to 20)
33. **G3** - final grade (numeric: from 0 to 20, output target)

Requirements for the Program

- We want to be able to query the dataset and ask the following questions and see what the average final grade is based upon a choice of

  1. did you have internet access
  2. number past failures
  3. your study time
  1. absences

  So, if I select "internet access" I should see a result that gives average with internet access, and average without internet access.

- We would like to be able to see a graph of the travel time to the overall final grade achieved (Did a longer commute translate in to a difference in grades?)

- A graph that would correlate overall health to final grade (See Example 1)
  - Bonus if you can show, or explain to us a way we can see the G1, G2, and G3 Grades in a graph corresponding to overall health of the student
- Anything else you find interesting to show us (what else can you infer from the data you see?)

Additional Information

- You can use any programming tools you are most comfortable with. Be prepared to explain your choice of tools.
- We do not require you to input this data into a database, but it is a plus if you know how to accomplish this. Want to build a web front end , great!. A console application, fine. How you accomplish this task is up to you.
- I recommend you bring your laptop if available to show your work on. If you don't have a laptop then put your work on a thumb drive, or on github. You will be provided a windows 8 machine to show us your results.
- A projector screen will be provided for you to connect to for your presentation
- When you arrive for your onsite interview you will be given the opportunity to present your results to the group that will be interviewing you. We understand that we are asking you to develop a software solution on top of your already hectic schedule. We are looking for how you approached the problem, and how you communicate how you solved it. We will not do an in depth analysis of your code, but we welcome you to show us your work and what you did.
- You will also have a whiteboard if you wish to explain anything in detail to us.

We hope this exercise is fun and a learning experience for you. We designed it to be thought through and developed over a few hours of coding at your favorite coffee shop.

Example 1.   An example chart based on type of data in dataset.
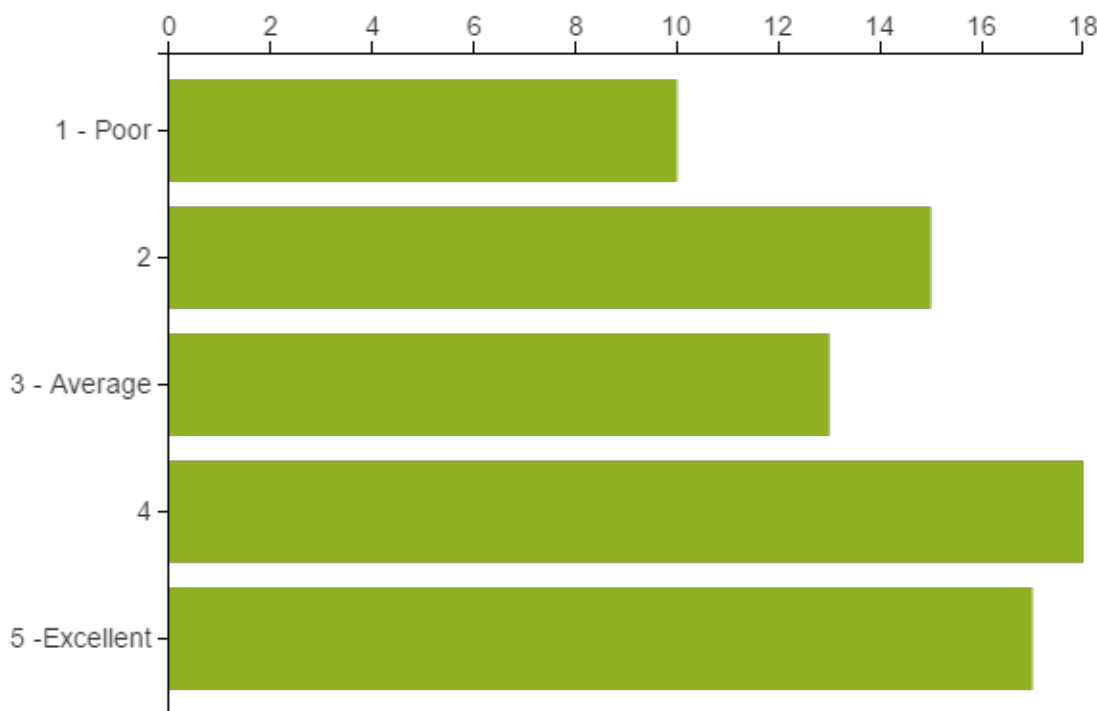
	health	Average Grade
1	1 - Poor	10
2	2	15
3	3 - Average	13

| 4 | 4 | 18 |
| 5 | 5 -Excellent | 17 |

References:

https://archive.ics.uci.edu/ml/datasets/Student+Performance#