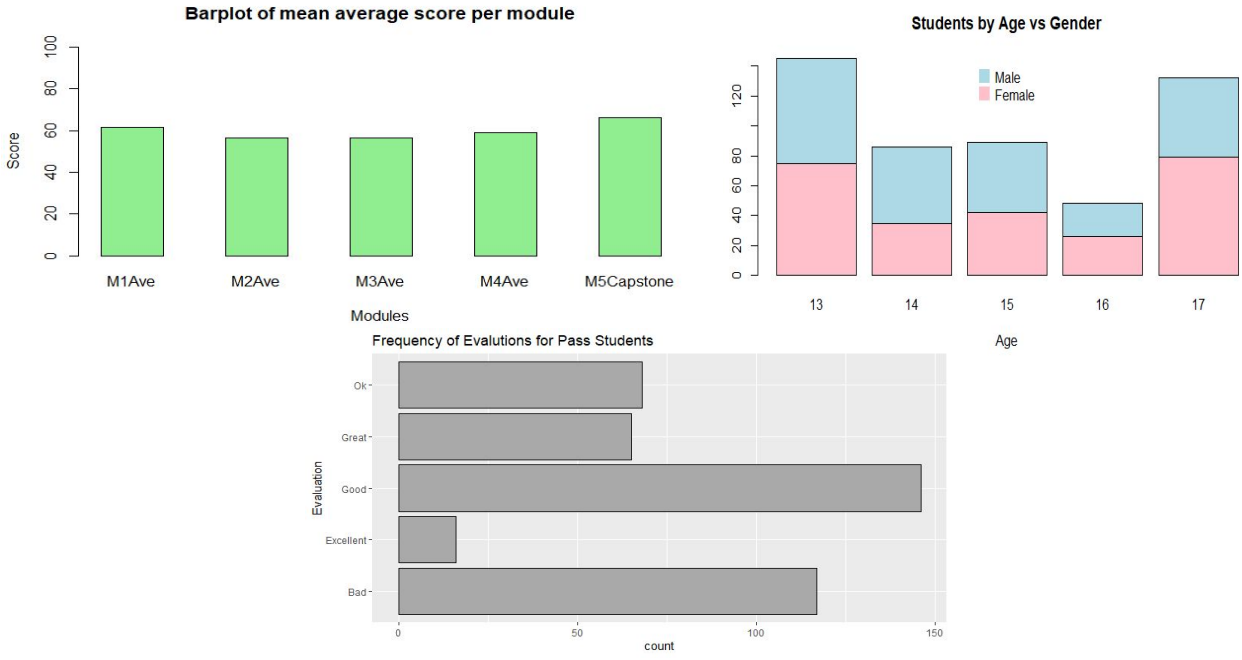
Purpose

This executive summary outlines current problematic trends with the *'Data Science for High School'* course and attempts to propose a solution that will remedify the issues, hence improving the quality of education students receive. In this context, it aims to further the fourth goal laid out by the United Nations Sustainable Development Goals: "Ensure inclusive and equitable education and promote lifelong learning opportunities for all." The approach taken will be to analyze the given data set and visualize using inherent R capabilities, imported libraries and techniques such as sentiment analysis.

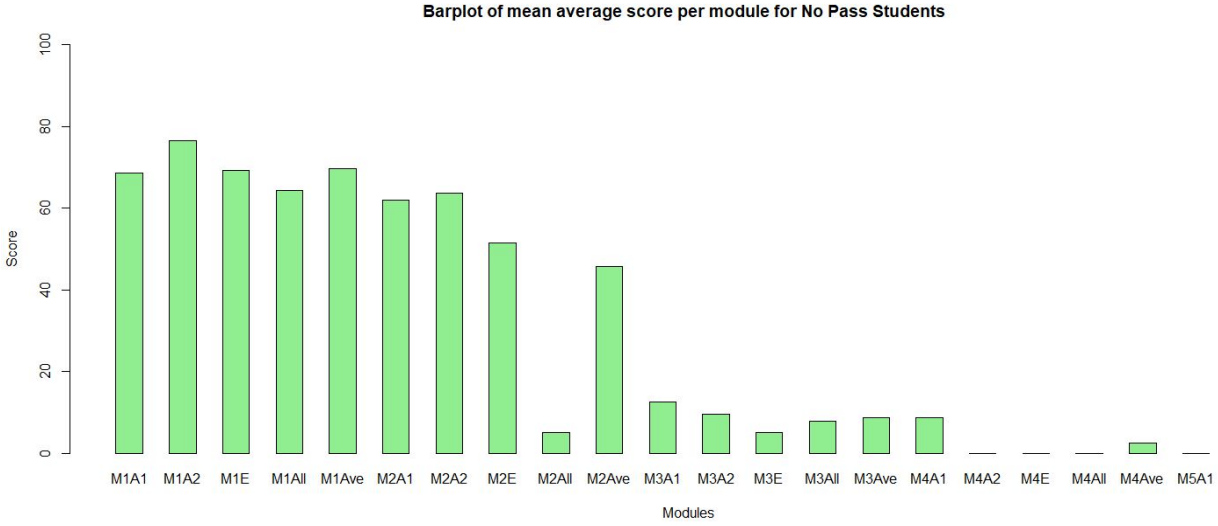
Observations

The initial analysis of the course data revealed several issues

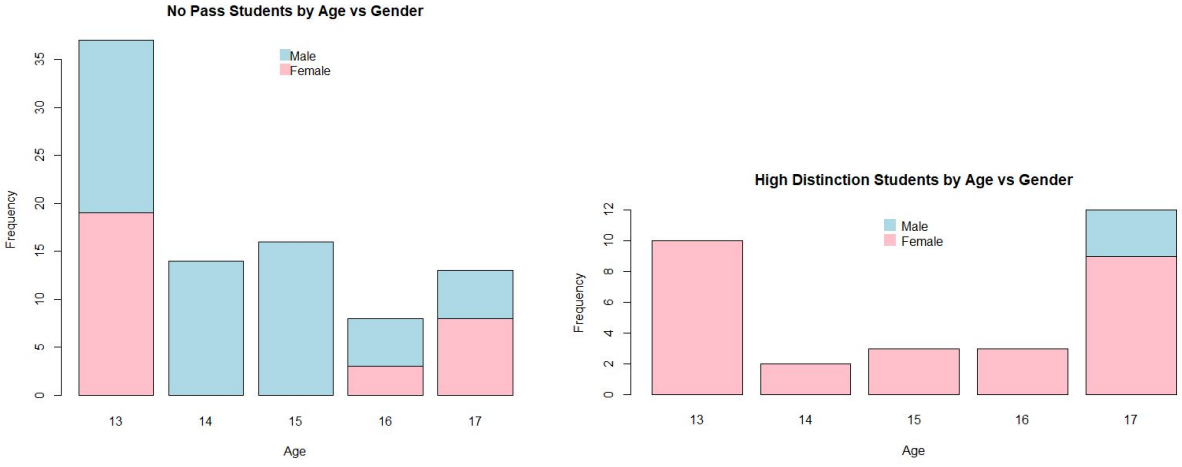
1. Modules, especially module 2 seem to be far too difficult as shown by the barely passing averages (median score of 57%). Only 6% of students that enroll are able to pass the class with high distinction (above 85%). Too many students are unsuccessful, and this also reflects with their general displeasement with how the course went for them. Students rate the course overwhelmingly negative and indifferent on their evaluations. Students are neither doing well academically in the course nor are they gaining a positive experience from the course. General sentiment on the topics covered in the course and the field of Data Science is positive, so the
2. Source of discontent for the students isn't coming from the content of the course, but rather how the content is delivered.



3. Failure to complete the assignments is a contributing factor to the failure of students rather than bad scores itself (below 50%) . This could be driven by lack of interest and/or engagement.



4. The drop rate of males taking the course is disproportionate to the ratio of males to females registering for the course.
5. The amount of males that receive high distinction is disproportionate to the ratio of males to females registering for the course.



6. A significant number of 13 year olds dropped the course. While the percentage dropped for all age groups except 17 year olds is similar as per the given data, it is important to note that the population size of 14, 15 and 16 year olds is not large and hence the similarity in percentage dropped could be coincidental. This is further compounded by the fact that the only other age

with a similar initial registration number is 17 year olds, and 17 year olds have a much lower drop rate. All of this suggests that there is a possibility of underlying causes that specifically affect 13 year old retention rate.

Problem Analysis

The following are deemed to be the possible reasons for these issues.

- The course is mostly self taught and skims through surface level topics. The quality of a student's learning seems to depend on how much they can utilize outside resources rather than being able to rely on the course itself.
- There are comments that explain what certain blocks of code together achieve to get the end result, but there is no clear focus on understanding syntax or what each line of code is achieving. When it comes to writing their own code, it requires consulting outside resources to learn how to code it on their own.
- Additionally, the instructions for the unit assignment are vague at times and require a lot of assumptions, and some hyperlinks in modules are dead links. Example code is outdated at times and doesn't run. For international students working across timezones, having to obtain clarifications from the instructor adds to the delay and sometimes disturbs the flow of study.
- While average grades across the board are pretty low, many students drop the course not solely because it is too difficult(as shown by relatively high scores in module 1), but because they seem to lose interest, or they do not see the value in completing the course.
- Assignments at times seem to lack clear connection to the reason many students take the course. What is being explored in the modules seems to be very different from the expectations from the assignments, which seems to discourage students from continuing further as they may feel misled.
- There are 2 probable causes to explain the high drop rates with 13 year olds.
    - They may not be interested in the subject as a whole and are just trying different things to see which one clicks. *This problem is not within our scope to fix.*
    - They could be interested but they might be finding the course too difficult for them to understand. *This problem is within our scope to fix.*

Proposed Solutions

The following are the recommendations to remedify the above stated problems with the course.

1. Introducing more interactive content such as video lectures and/or notes that clearly demonstrate what each line of code and each function specifically does
2. Students should be given coding exercises to build code from scratch, so that they can solidify their coding skills.
3. The hyperlinks can be completely removed and the information in those hyperlinks can be hosted on VHS servers so that there is no dependency on outside sources.
4. Clear explanations of the relation of assignments in certain units (specifically the final assignment in unit 3) to the bigger picture of data science and analytics should be provided, so that students understand the objective of the assignment.
5. Optional sections in each unit could be introduced to give more in-depth explanations of the concepts learned (for ex: statistical concepts). This would provide extra support to students who need it (specifically unit 2 so that those who drop early are potentially retained)

Lastly, the disproportionate number of males that score lower and have high drop rates is a difficult problem to solve. There isn't a clear reason as to why they are not as successful. It may just be coincidental that those who happen to do bad and/or drop are males, and that there is no causation between the two. Additionally, even assuming that there is a causation between the two, it would be difficult to identify a fix for this issue without knowing anything about the backgrounds of these students and their interests, as what works in engaging and retaining some males may push other males even further away. While there may be environmental, cultural or even biological factors that are potentially causing this disproportionate rate of dropping out, it would be extremely inaccurate to extrapolate data and determine a solution for future students with such a small sample size, limited knowledge over other potential factors, and no control over who gets admitted into the sample.

Conclusion

To summarize, the analysis identified the following problems:

1. Course is too difficult in current state
2. Students are too unsuccessful as they barely pass

3. Students lose interest and drop out
4. Males and 13 year olds perform worse
5. Students aren't gaining a positive experience

The analysis highlighted that not every student is the same, and while some students benefit from a more hands-off approach of teaching, other students require a more interactive and in-depth approach of teaching the same material to achieve similar levels of success.

The following proposed solutions solve these problems by improving course clarity.

1. Introduce interactive content
2. Introduce activities that strengthen coding skills
3. Clean up dead links
4. Clearly explain unit assignment connections to the bigger picture
5. Introduce optional sections that go into further detail on concepts learned for students that need extra support

The solutions listed above further the UN's fourth goal through improving the quality of education delivered to students, increasing retention rate of students, and promoting equality of education amongst all ages. When students will be able to learn the material easier, replicate it on their own, and achieve better proficiency in the course material, they will be more inclined to be engaged and learn. While these improvements won't make students job-ready, students will acquire a better grasp of foundational concepts in Data Science for pursuing advanced studies in the field in the future. This will result in students reporting a more positive learning experience, hence increasing the likelihood of students pursuing higher education in the field of Data Science and making a career for themselves in the field.