Module 2 assignment - Reflective piece

Introduction

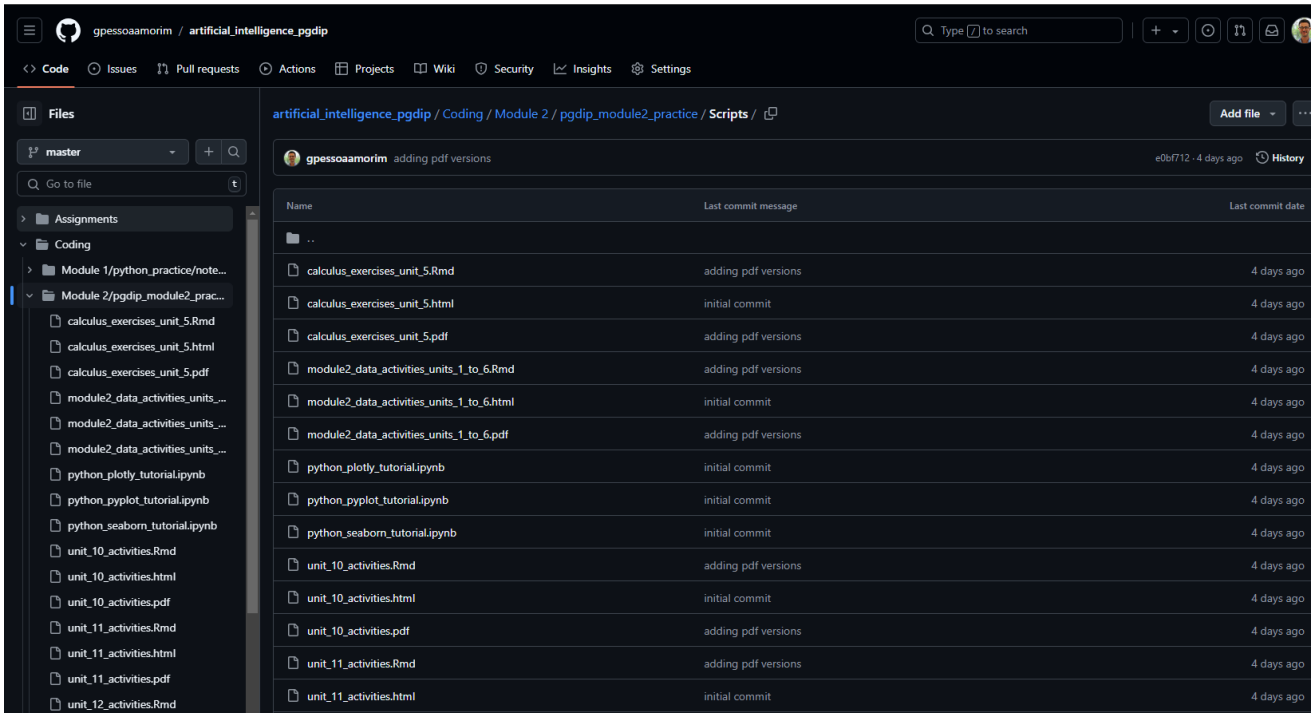
The learning objectives for this module were to explore theoretical concepts underlying computing and statistical analysis, understand and apply computer programming and data science approaches, and learn to interpret outputs generated using these techniques, while building an understanding of real-life applications and challenges related to these tools, and applying a reflective and independent approach to the learning process.

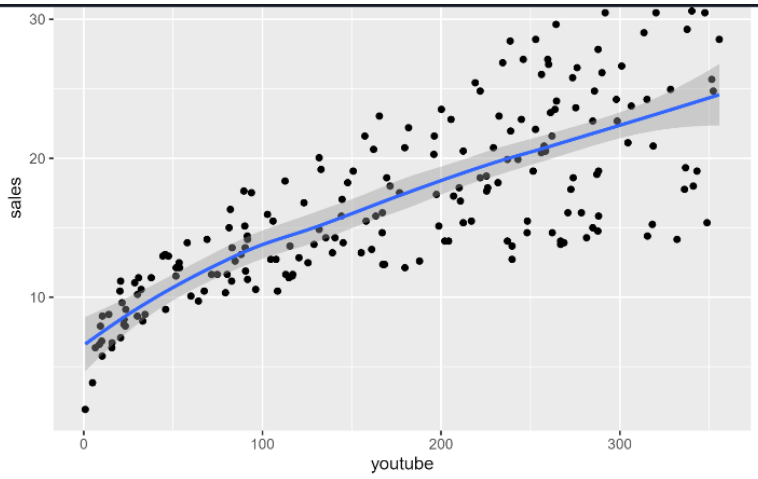
In this piece, I will use Rolfe et al.'s approach to critical reflection (Rolfe et al., 2001) together with Gibbs learning cycle theory as a general framework (Gibbs, 1988), by describing the module outcomes (WHAT, or Description), discussing my learning journey (SO WHAT, or Feelings/Evaluation/Analysis), and finally reflecting upon the learning outcomes achieved and how to apply them going forward (NOW WHAT or Conclusion/Action Plan).

Activities performed and learning outcomes ("WHAT")

The main focus of this module was on employing R statistical software to analyse and extract information from data, building on from a review of theoretical concepts of statistics and data science supported by the recommended learning. Some of the practical tasks undertaken in R included installing and running the software, loading and saving files, conducting simple mathematical operations and data transformations, and sub-setting and visualising data. This knowledge was then applied for inferential statistics, through hypothesis testing (including interpretation of p-values and statistical significance) and regression analyses.

For this purpose, I studied the recommended learning materials and worked through all the formative activities related to data analysis. I produced my work using R notebooks, which allowed me to save code along with the corresponding outputs, and shared this in my ePortfolio on GitHub (https://github.com/gpessoaamorim/artificial_intelligence_pgdip). The screenshots below the ePortfolio, an example output, and a view from the Git terminal.



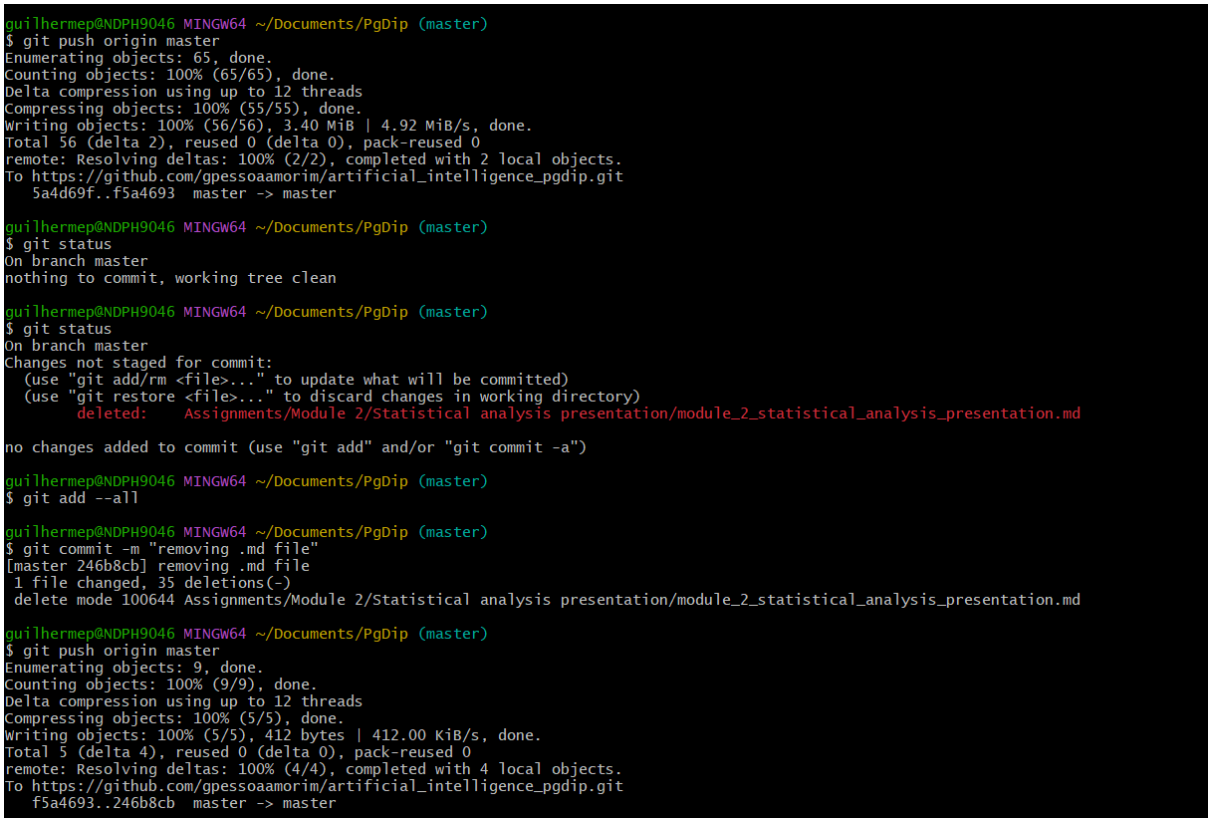


The scatter plot indicates a linear relationship between both variables ### Correlation

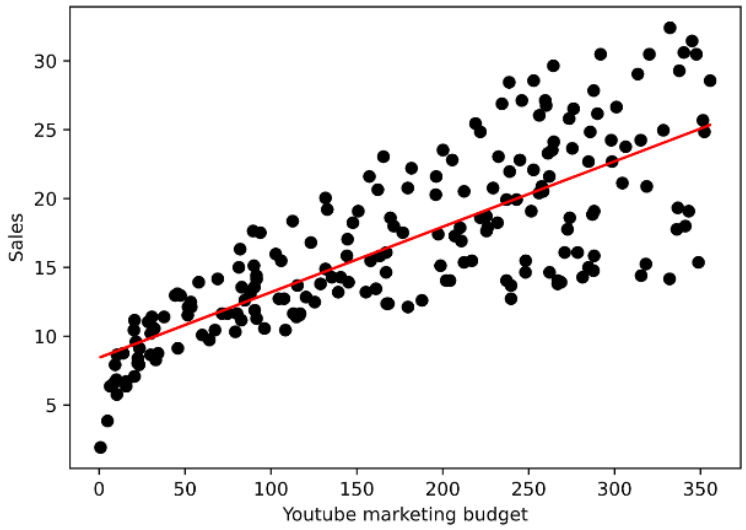
```
cor.test(marketing$youtube, marketing$sales)
```

```
##
##  Pearson's product-moment correlation
##
## data:  marketing$youtube and marketing$sales
## t = 17.668, df = 198, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.7218201 0.8308014
## sample estimates:
##       cor
## 0.7822244
```

From the correlation coefficient, we can say that there is a strong, positive, statistically significant association between budget spent on Youtube marketing and sales



Following the suggestions at the beginning of the module, I also explored coding in Python by replicating all activities conducted in R (as shown below).



Correlation

```
stats.pearsonr(data.youtube, data.sales)
```

```
## PearsonRResult(statistic=0.7822244248616065, pvalue=1.467389700194595e-42)
```

Linear regression model

```
x=np.array(data.youtube).reshape(-1, 1)

y=np.array(data.sales).reshape(-1, 1)

model = LinearRegression().fit(x, y)

r_sq = model.score(x, y)
print(f"intercept: {model.intercept_}")
```

Personal reflection on learning journey ("SO WHAT")

I have a PhD in health data science and am quite comfortable using R for most of the tasks described, particularly in relation to data manipulation, producing summary and contingency tablets, and visualisation. However, I do not routinely work with hypothesis testing or regression, and found it helpful to take some time to review the theoretical mathematical concepts and their practical application, namely running statistical tests and interpreting p-values and confidence intervals, which I feel now quite comfortable doing (including confirming test assumptions). I also tried to make the most of this module by tailoring my learning journey to my own experience and objectives, in particular developing data science and programming techniques. For this purpose, I expanded my R skills by exploring the corresponding Python commands and delving into the differences between the two languages, made extensive use of R markdown techniques to produce comprehensive and engaging reports, and deepened my understanding of Git by using the BASH module to host my code and outputs on GitHub (instead of the graphic user interface which I had used previously). Learning Python was a challenging task as the syntax and debugging mechanisms are quite different from R, which required frequent online searches and constant trial and error, but equipped with valuable hands-on experience. Similarly, transitioning to Git terminal was very difficult in the beginning, particularly as I was trying to use it to host all my outputs from both this module and previous ones in the ePortfolio. This required careful masking using .gitignore and learning a number of new commands, but I was eventually able to find suitable solutions through online searching.

The tailored learning approach I chose was motivated by a desire to personalise my programming learning, which has been previously shown to improve educational outcomes (Pane et al., 2017; Inthanon and Wised, 2024; Juniarni et al., 2024). I am also an avid coder and take immense joy taking a hands-on approach to develop my coding skills and solving complex data analysis and visualisation tasks, an approach that also facilitates learning, and generates wider cognitive benefits (Von Hausswolff et al., 2020; Scherer et al., 2021). I also needed to maximise the return from time invested in this module as it coincided with an extremely challenging period in my personal life, due to family illness, a sudden decision to move abroad, job interviews, and additional workload in my regular job. This situation created immense time management challenges, requiring me to work in intense bursts of activity at the beginning and end of the module, and not allowing

time to engage in the collaborative discussions and seminars, exploring the supplementary reading, or engaging with the module tutor for feedback (as I had done in the previous module). Nonetheless, I made very good use of the activities offered and successfully achieved the specified learning outcomes, while receiving excellent feedback from the assignment already submitted (mathematics test, graded with distinction).

Learning and changed actions ("NOW WHAT")

This module provided an intense learning experience, generated both by reviewing theoretical concepts and their application in the recommended literature (especially Bruce et al. 2020), and the comprehensive practical implementation and exploration I performed using R and Python, driven by a clear goal of personal growth. In particular, I have increased my confidence running statistical tests in R, initiated a process of knowledge transfer in data manipulation and analysis from R to Python, and developed a structured approach to producing data analysis reports and hosting them online, all through independent and autonomous learning. My journey was also improved by the assignments conducted throughout this module, namely the statistical presentation and this piece, which allowed me to delve into my learning journey and extract a number of insightful reflections using Rolfe's and Gibbs' models to structure and contextualize my progress.

Going forward, I intend to put the lessons learnt during this module into practice both in future models and in my day job, where I am transitioning to a data science and engineering role within the pharmaceutical industry. I am particularly looking forward to start developing automated data processing pipelines, machine learning algorithms, and interactive dashboards using Python, together with R for specific data analysis and visualisation projects. I also aim to start using Git and GitHub more intensively and on a daily basis to establish good version control practice, collaborate with others, and showcase my work.

Conclusion

I have thoroughly enjoyed the learning journey undertaken during this module, despite the many challenges faced. I have also found both the theoretical and practical learning

obtained helpful in supporting my progress in a data science career, and am looking forward to applying this in next modules and throughout my career.

References

Bruce, P., Bruce, A. & Gedeck, P. (2020) *Practical statistics for data scientists: 50+ essential concepts using R and Python*. O'Reilly Media.

Gibbs, G. (1988) *Learning by doing: a guide to teaching and learning methods*. Oxford: Further Education Unit, Oxford Polytechnic.

Inthanon, W. & Wised, S. (2024) 'Tailoring Education: A Comprehensive Review of Personalized Learning Approaches Based on Individual Strengths, Needs, Skills, and Interests', *Journal of Education and Learning Reviews*, 1(5), pp. 35–46. Available from: https://doi.org/10.60027/jelr.2024.779.

Juniarni, C., Sodikin, M., Akhyar, A., Almujahid, A. & Asvio, N. (2024) 'The Importance of Personalized Learning: How to Tailor Education to the Individual Needs of Students', *Education Studies and Teaching Journal (EDUTECH)*, 1(1). Available from: https://doi.org/10.62207/agxhq160.

Pane, J.F., Steiner, E.D., Baird, M.D., Hamilton, L.S. & Pane, J.D. (2017) *Informing Progress - Insights on Personalized Learning Implementation and Effects*. RAND. Available from: https://www.rand.org/pubs/research_reports/RR2042.html.

Rolfe, G., Freshwater, D. & Jasper, M. (2001) *Critical reflection in nursing and the helping professions: a user's guide*. Basingstoke.

Scherer, R., Siddiq, F. & Sánchez-Scherer, B. (2021) 'Some Evidence on the Cognitive Benefits of Learning to Code', *Frontiers in Psychology*, 12, p. 559424. Available from: https://doi.org/10.3389/fpsyg.2021.559424.

Von Hausswolff, K., Eckerdal, A. & Thuné, M. (2020) 'Learning to program hands-on: a controlled study', in *Koli Calling '20: Proceedings of the 20th Koli Calling International Conference on Computing Education Research*. *Koli Calling '20: 20th Koli Calling International Conference on Computing Education Research*, Koli Finland: ACM, pp. 1–10. Available from: https://doi.org/10.1145/3428029.3428058.