Proagrica Coding Challenge

Congratulations on progressing to the next stage of the Proagrica Data Science interviews. Below you will find an exploration/modelling challenge that you will need to complete within the allotted time and then sent back to [rob.mansfield@proagrica.com](mailto:rob.mansfield@proagrica.com).

Please include all code, plots, summary statistics and model parameters that are produced and try to present code in such a way that it is understandable to someone else without you being there. We expect to see examples of exploration, modelling and validation in the code and any associated plots.

A local library has been having problems with the late return of books. In order to address this they have asked that you produce a model to help to predict if a book will be returned late. They can then use this to target reminders and advise staff when issues may occur. They have provided you with the following three tables and asked for the most accurate model possible.

If you feel you do not know the full requirements or dataset specifications from the client you may guess but please justify/explain any such assumptions.

* Books
  + UID – Unique Identifier for books
  + publication\_date – Date the book was published
  + subject – General topic
  + title – Title
  + author – Author
  + length – Number of Pages
  + rental\_period – Number days book is allowed out for
* Rentals
  + book – Book UID
  + borrower – Borrower UID
  + out – Date borrowed
  + returned – Date returned
  + UID – Unique identifier for rentals
* Borrowers
  + UID – Unique identifier for borrowers
  + DoB – Date of birth
  + address – Anonymised postcode (last two letters removed)
  + ethnic\_group – ethnic group code
  + gender – Gender of borrower
  + join\_date – Date they joined the library