

***Book***

***Recommender***

***review and***

***ideas***



# ***Fast review – key findings***

- Provided code used only inline function calls
- There were no methods/functions or class
- In some place the code did not have enough comments and was quite confusing
- There was also a problem with names of variables – variables should have more general (but still understandable) names. Names in the given code were too specific.
- Original code did not function on itself.
- There was no input for user asking for his/her own book.
- There was no main() method.
- There are much more specific comments in the .py files



# ***The largest problem – Correlation func.***

- The largest problem I found in the original code is the correlation „function“
- This part of the code was not in a function/method even though I think it should have been
- It accepted only hard-coded title of a book (LOTR I)
- It was supposed to go through a list of titles but!:
  - + There was only 1 book in the list
  - + There was no check if it even exists
  - + I think the approach was over-complicated and memory consuming



# ***Improving the code – My approach I***

- In the first place I re-coded most of the code into functions and some even into separate class called „BookRecommender“

- Logical parts of the code were reworked into functions/methods:

`BookRecommender.load_data(data loading from files, separator)`

`BookRecommender.get_unique_readers(from data, by title)`

`BookRecommender.create_pivot(creating pivot from final data set)`

`get_pivot_for_title(book_title)`

`correlator(pivot_table, book_title)`



## ***Improving the code – My approach II***

- I tried to rewrite the code as much as possible so it would accept any given book on the input.
- I changed names of the variables so they are more „generic/general“ and they do not refer to a specific book as in the original code.
- The most important I rewrote the code into `main()` function.
- I created simple input procedure so user can try to find his/her own book.



# ***Persisting problems***

- The largest persisting problem is that most of the books have lesser number of reviews than is the set threshold (=8). It makes it hard to compute correlation for these book and the programme may crash. (This may be easily repaired by printing out statement „Not enough data for correlation“)
- The data set is very limited – there are not variable like genre etc. I would suggest using some newer data set. For learning purposes it may come for example from kaggle.com.
- The data are not clean enough – there is a problem when merging two original datasets.
- The data could be loaded in batches.



# ***Ideas for futher improvement***

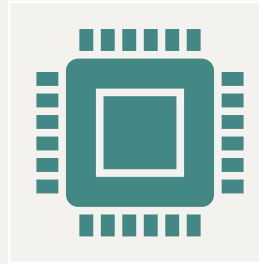
- We should consider using different data set.
- Maybe adding function for searching the series of the titles so a person who provides only a partial title may still find his/her book
- Is correlation really enough? Maybe we can create something more relative (combination of rating, number of readers, correlation...).
- Removing the problem with the books with small number of reviews (should not take so much time)
- GUI or front-end – I did not manage to do the front-end for this app but it should be so hard using Flask.
- Another form of output – such as saving into .csv or as API (for online communication).



# ***My own opinion***



Rewriting the code  
took me something  
around 6 hours



Firstly I did not  
completely understood  
the code leading to the  
computing of correlation  
but once I figured it  
out it was quite easy to  
rewrite the code.



I found assessment like  
this very useful and  
even if I do not get  
the position I applied  
for I think this task  
enriched me 😊







***Petr Martinek***

