Assignment Report

# LITERATURE REVIEW

*A brief description of the problem area and how you problem (the shipping records) fits in. A less brief description of the technical issues involved (with sources).*

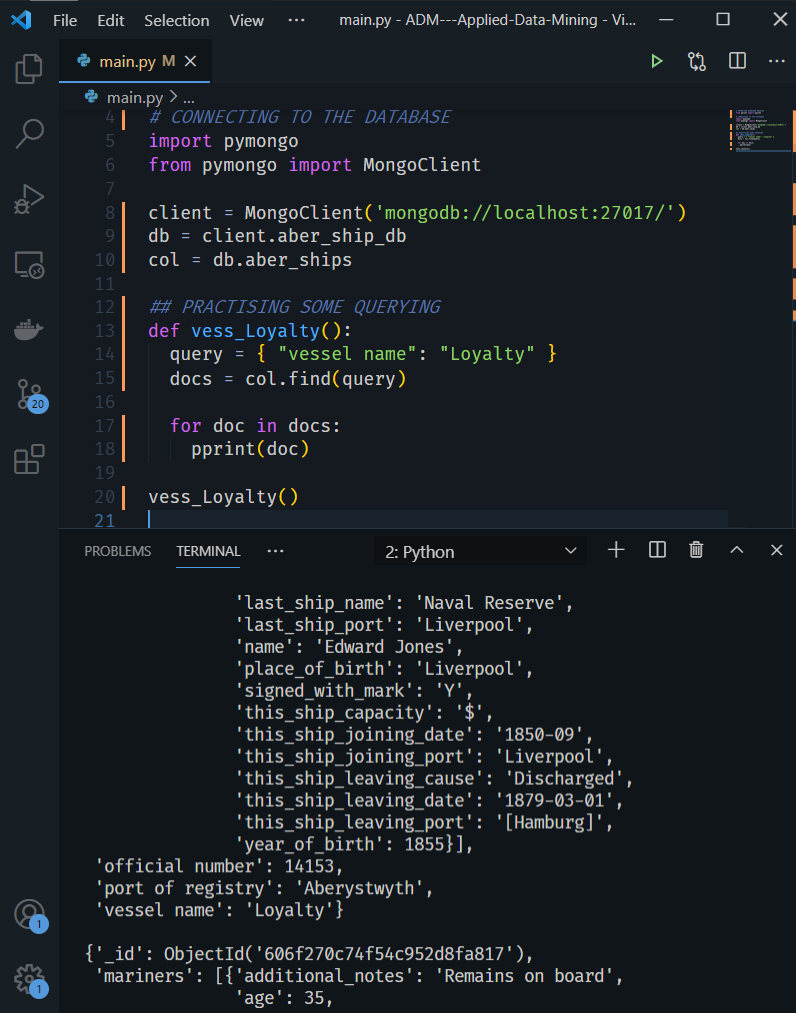
*At Masters level your report should not only describe your work but show an awareness of similar projects how your solution fits into a wider context. In this case the context is using databases for historical research. You need to describe other solutions to similar problems and how your work compares. Finding other similar works are a matter of your own research. Google Scholar is a good place to start.*

# METHODOLOGY

*A technical description of your solution with an explanation of the choices involved.*

## Test Functions

### Testing querying



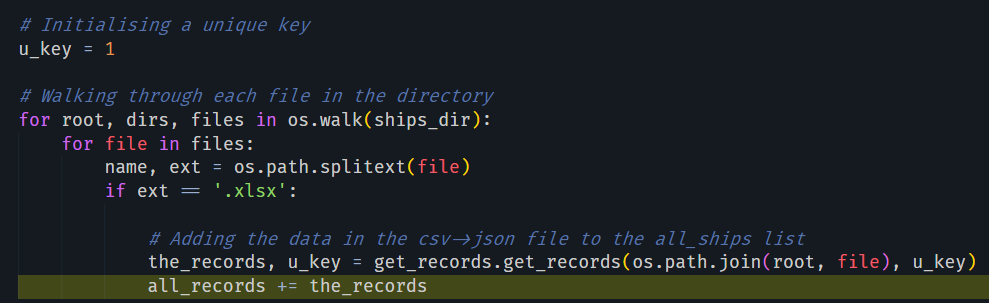
## Modifying the python files which relate to importing the data into mongo

### Original structure:



### Modifications:

* Creating a unique key (very difficult process which required a lot of debugging)
* Had issues with the unique key not being consistent across the python files, so each time the function was called (when the next file was selected), the unique key was being reset
* The solution was to return the unique key from the get\_records function, and then assign that in the transfer\_all\_ship\_data where the function is passed



### Testing the modifications:

Before I imported the data in with the modified importing python files, I wanted to check that my test data had imported correctly. As well as looking at Mongo Compass and seeing that the data appeared to be in order:

### Modified structure



I realised I could also count the number of mariner records (aka the number of rows in each excel worksheet and workbook, starting from the first mariner row), and compare that to the number of documents in my collection. That’s because each record = a new document in the collection, and therefore these numbers should match. **Check\_record\_count.py** is the code for how I achieved this, and prints out the results of the two counts (records and documents). When I ran this I saw the pleasing result that the numbers did in fact match:

And therefore I felt confident to import all of the data into my database with this new structure.

# RESULTS

*An analysis of your project from a technical standpoint. This should be evidence based.*

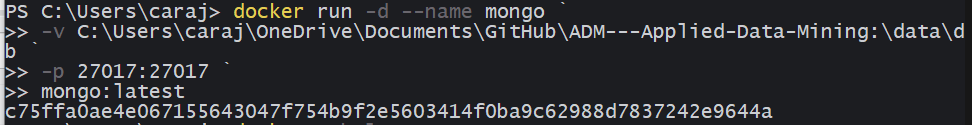
# DISCUSSION

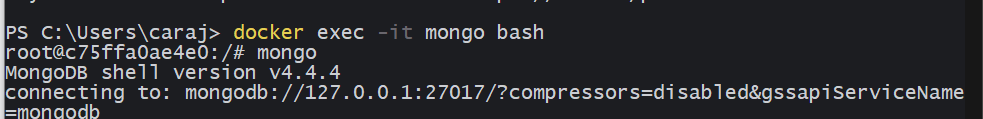
*An overall analysis of your project, this should summarise evidence from your results section, it can also be more subjective (but still evidence based).*

# SELF-REFLECTION

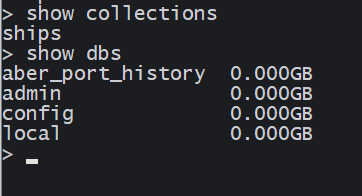
*Although it is not normally part of a peer-reviewed paper, it is university policy that you include a reflection section here. Reflection on your work has been shown to improve your learning in the long term. You should consult the marking schema and position your own work on it. You should also reflect on how you would do it differently if you were to repeat the exercise.*

#### Running a container with mongo

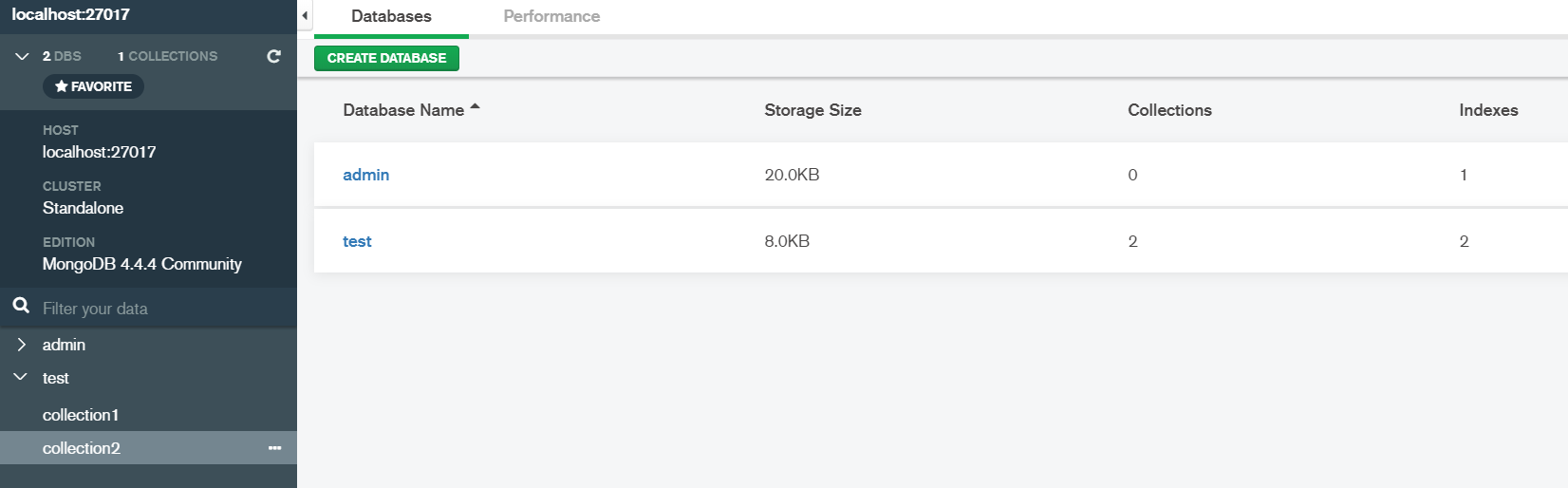


Entering the bash terminal within the container and then executing the mongo command to be able to perform mongo operations

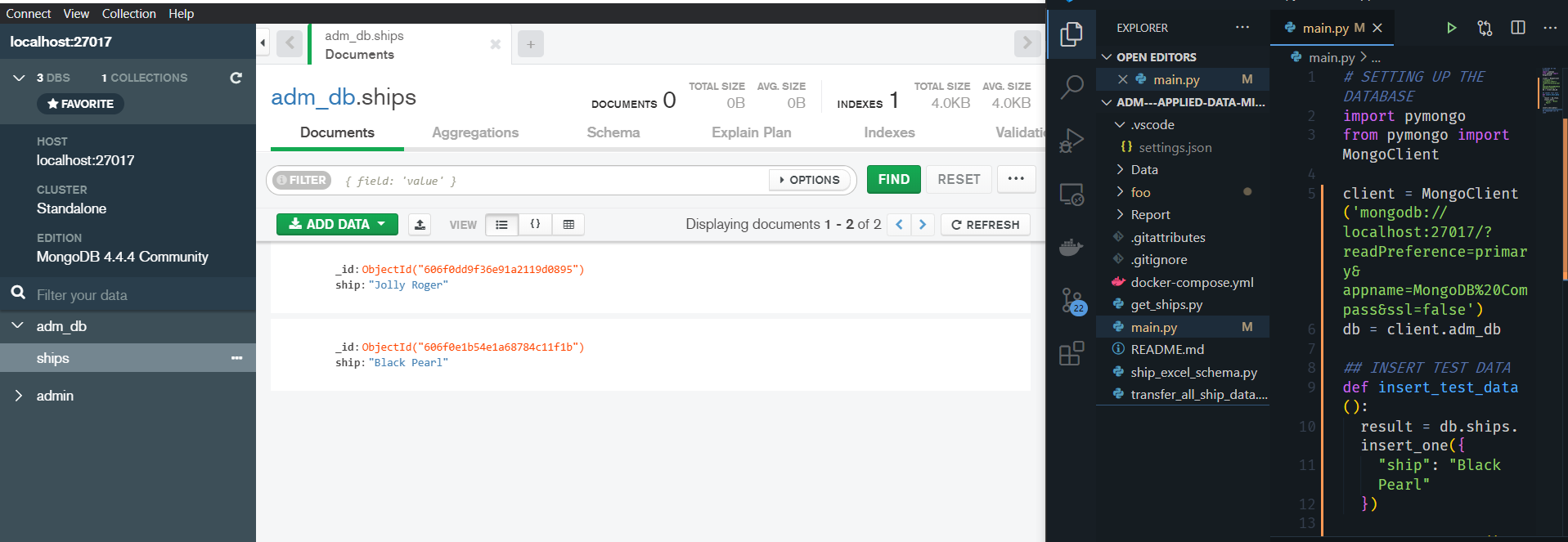
#### Executing mongo operations



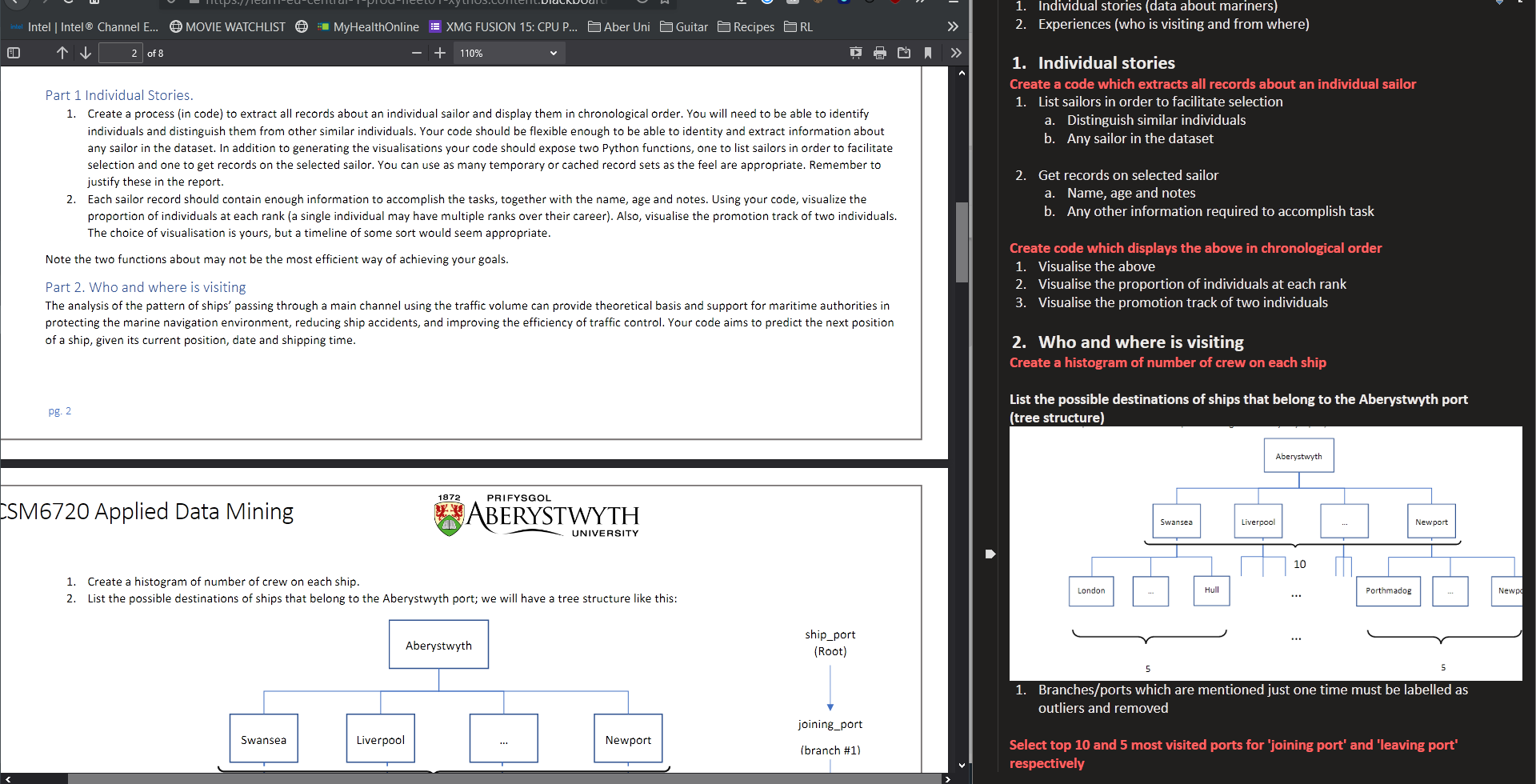
#### GUI Interface to quickly visualise mongo operations



#### Example of GUI interface visualising mongo operations



#### Converting the brief into requirements



Debugging was very useful