I’ve decided that to keep track of the decisions I’m making and my thought process in general throughout this project, I’m going to keep this word document as a journal.

I’ll leave the documentation on the code and visualisations etc for other files.

So, after briefly reviewing the data by opening the csv files in excel I’m trying to decide what tools to use for the analysis and for the storage of the data too I suppose. The options I’m working with are;

* R
* MySQL
* Python

During the Google data analytics cert, I learned how to use R and the power of markdown files which I’m a big fan of. I love the fact you can embed inline code in chunks (which doesn’t have to be R code, it can be SQL, python etc) along with pics, vizzes and plain text and then best of all export it as a HTML, word doc or PDF file. The PDFs are nice because they’re created using LateX which I’m a big fan of because I really liked writing lab reports and assignments in college using LateX. I’ll be using R anyway to keep a markdown file.

But in terms of the storage of the data, it would be nice to create a database in MySQL and showcase my SQL skills because their kind of crucial for data analytics. However there’s really no need because the data isn’t real time data or being updated. The downside of MySQL is that, with large datasets you can’t really use the table import wizard because it takes a ridiculous amount of time so you need to write the SQL code for the creation of each table which takes a while. Compare that to python or R where you just import the file as a dataframe in one line of code. So do I use SQL to showcase SQL skills or chose the practical option?

If I’m using R to keep a markdown file should I just stick with that? I’ve way more experience with python and the pandas package though. Might be worth delving into jupyter into the future to solve this issue next time.