*Georgina Dangerfield*

*ST20190328*

*Reflective Report*

WRIT1, Task 4

Table of Contents

[Introduction 1](#_Toc102305413)

[The Role of Computational Techniques and Statistical Analysis in Healthcare 1](#_Toc102305414)

[Challenges of WRIT1 2](#_Toc102305415)

[General Reflection 2](#_Toc102305416)

[Way Forward 3](#_Toc102305417)

[References 4](#_Toc102305418)

# Introduction

For this assignment, I chose to look at AirBnB data for both London and Amsterdam, in order to answer the question “What can hosts do to maximise their chances of rental to increase income? How does this differ across Europe, if at all?” To answer this question, I broke this down further into sub-questions, namely:

1. What hosts' behaviours or profiles would influence AirBnB tenants reviews across Europe?

2. What words should hosts include in listings?

3. What features should hosts focus on to maximise booking potential?

I chose this question because, even two years after the COVID-19 pandemic began, certain areas of hospitality have still not reached pre-Covid, ‘normal’ levels of business. I completed my data analysis and analytical report while bearing in mind that hosts on AirBnB may still be facing reduced levels of booking, which could leave them out of pocket. My main aim was to demonstrate to hosts what to focus on to maximise their chances of their listings being rented.

# The Role of Computational Techniques and Statistical Analysis in Healthcare

According to [statistics released by the EPO in March this year](https://www.epo.org/news-events/news/2021/20210316.html), medical technology accounted for the largest proportion of European patent applications filed in 2020. (European Patent Office, 2021), while pharmaceuticals and biotechnology were the fastest growing.

An emerging trend in healthcare due to computational modelling is moving away from the “one size fits all” approach to therapy, and instead focusing on personalised treatments that are more effective for each patient, taking into account their individual genetics and physiological characteristics (Conroy, 2021). The Living Heart Project is one example of this, where a realistic computer simulation of a heart is being developed and applied, a so-called ‘digital twin’. Magnetic Resonance Imaging (MRI) data of the person's heart is collected and used to generate a virtual three-dimensional model of the heart. This 3D model is then turned into a realistic beating model of the biological heart utilising extensive mathematical modelling of electrical and mechanical tissue features. This has potential for being extremely important for the future of healthcare, as it will be able to replace human patients when designing and testing new drug therapies, as well as being able to predict how efficient different implant designs could be.

In recent years, it has also been possible for computational methods to predict chronic diseases in healthcare communities. In research from Induja and Raji (2019) created a decision tree model where they could predict the risk of cerebral stroke with up to 99% accuracy. This is extremely important as chronic diseases like cerebral stroke have the ability to result in the demise of a large number of human lives in a very unpredictable way, and this new model has the potential to reduce the impact that cerebral stroke can have on individuals and even save countless lives.

Computational modelling has also been useful in using fluid dynamics to reduce infection risk during the Covid-19 pandemic. It’s crucial to reduce the risks of airborne transmission, especially in hospitals, where there are potentially large numbers of infected patients in close proximity to hospital staff. Scientists worked with a hospital in France that needed to expand their capacity for Covid-19 patients, while keeping contaminated areas separate from the rest of the patients. The scientists came up with a detailed 3D model from using a 2D floor plan of the hospital wing, and by using virtual patients and fluid dynamics to investigate how virus particles would be expelled by these patients, the scientists found that they could contain the virus by opening and closing certain windows at different points in the hospital.

# Challenges of WRIT1

Overall, I found WRIT1 to be quite challenging in terms of creating a question and the data analysis process. I originally found it quite difficult to come up with a business question in Task 1 that I could use for Task 2, because I found it quite daunting to come up with just one question when I had access to so much data. I overcame this by deciding to come up with some sub-questions that were related, because I could them link them to form a bigger business question. Now that I have my question and have completed the data analysis that was required for Task 2, I am happy that this was the question I chose because I have managed to get plenty of visualisations, statistics and other general findings from my data analysis, all of which are useful in presenting back to AirBnB hosts to suggest to them how they can increase their booking potential as we come out of the pandemic.

Similarly, the fact that I had so much data at my fingertips meant that I struggled at times to come up with a clear and concise plan of what exactly I wanted to do for my data analysis. I feel as though I might have wasted a bit of time looking at visualisations that I didn’t end up including in my final code notebook. However, I am trying not to be too disheartened by this as it did give me the opportunity to try out things that I hadn’t used before, and although they weren’t used in this project, I now have the knowledge and understanding of the code to put them into practice in another project, either work or university related. If I was in this situation again, where I have such a large amount of data and many possibilities of how to analyse the data, I would start writing a plan before I started doing any coding. I would make sure to brainstorm all possible ideas, separating them into sections such as pre-processing, EDA, NLP, machine learning, etc, while thinking about which visualisations fit what I am trying to analyse the most effectively. I will then be able to visualise how I want my notebook to look as well as exactly what kind of data analysis I would like to undertake.

# General Reflection

Completing this module has benefitted me in many ways, allowing me to learn new skills and develop existing ones. One of my main strengths during this assignment has been my ability to debug my code that had errors; this is a skill which I have been developing over the past year as it is something that I come across every day at work. The fact that I can debug my code relatively quickly saved me quite a bit of time when creating the data analysis notebook.

Another strength of mine when completing this module has been my time management skills. Having learned from my mistakes in the first half of the year, where I started an assignment too late and ended up being extremely stressed the week leading up to the deadline, this time I took the opportunity to make a start on the assignment as soon as the brief was available. This meant I had ample time to work through each task, and the fact that I did waste a bit of time playing with different visualisations didn’t matter so much as I had plenty of time to spare.

As mentioned above, my indecisiveness is one of my biggest weaknesses. I am hoping this is something I can work on before my next assignments due at Christmas, as this will save me precious time and energy. I would say a weakness related to my data science skills is that I am still not as confident as I could be when creating maps because of my lack of experience with geospatial packages. I am aiming to work on this by making it one of my short-term goals at work, where I can practice using different packages such as geopandas, Arcpy and Folium.

# Way Forward

I thoroughly enjoyed completing this assignment as it gave me a good opportunity to develop my EDA and machine learning skills in Python. Perhaps if I had the chance to do a project like this again, I would use R rather than Python as it would give me the scope to improve in the language I am less confident at. I think R would be a good choice especially for when completing the EDA section of the task, as R has so many impressive packages such as ggplot, ggvis and lattice, some of which produce arguably better plots than Python. As stated above, I will make sure to plan my work better before I start coding, as this will benefit me in the long run and potentially save me some time.

I feel as though I have learned a lot about data science this year and I will be able to carry this forward to the last year of my degree and into my data science career. I plan to look into geospatial packages in more detail so I can have more practice creating maps in Python, and hopefully incorporate this into assignments in my third year. I will do this by watching YouTube videos and reading online articles, as I find these methods very effective when I am learning coding techniques.

# References

Conroy, R. (2021, September 22). *The Digital Healthcare Revolution: Computational Modelling - Intellectual Property Law*. Reddie & Grose. Retrieved 1 May 2022, from <https://www.reddie.co.uk/2021/09/22/the-digital-healthcare-revolution-computational-modelling/>

Dassault Systèmes®. (2021, September 21). *Air Flow Simulation Used by 3DEXPERIENCE Lab OPEN COVID-19 Community to Reduce Virus Propagation Risk in French Hospital*. Dassault SystÃ¨mes. https://www.3ds.com/newsroom/media-alerts/air-flow-simulation-used-3dexperience-lab-open-covid-19-community-reduce-virus-propagation-risk-french-hospital

European Patent Office. (2021, March 16). *EPO - Healthcare innovation main driver of European patent applications in 2020*. Copyright © 2007 European Patent Office. All Rights Reserved. Retrieved 1 May 2022, from https://www.epo.org/news-events/news/2021/20210316.html

Induja, S. N., & Raji, C. G. (2019, March 1). *Computational Methods for Predicting Chronic Disease in Healthcare Communities*. IEEE Conference Publication | IEEE Xplore. Retrieved 1 May 2022, from https://ieeexplore.ieee.org/abstract/document/8817044