MSc Project Proposal

# Coursework 2 of Module IS4S706

|  |  |  |
| --- | --- | --- |
| **Enrolment number** | 17076749 – Mark Baber | |
| **MSc award title** | MSc Data Science | |
| **Project title** | GIS and the COVID-19 Pandemic. | |
|  | |  |
| **Main aim of the project** | |  |
| To develop a GIS solution for modelling and predicting the spread of COVID-19. | | |
|  | | |
| **Project objectives** | |  |
| 1. To explore existing approaches to modelling and mapping the pandemic worldwide. 2. To determine whether GIS technology can be useful in managing the spread of COVID-19. 3. To understand which COVID-19 factors will be important in modelling and prediction. 4. To identify which modelling and prediction algorithms best reflect the spread of COVID-19. | | |
|  | | |
| **Project description** | |  |
| 1 – Lit Review– Geographical Information Systems (GIS) A Geographical Information System (GIS) is a framework for gathering, managing, manipulating, and analysing data (ESRI, 2021). It is made up of a mix of several different areas which includes, computer cartography, database management, computer-aided design (CAD) and remote sensing (Maguire, 1991). This framework can be used to enable problem solving and decision-making processes, along with digitally restoring old maps and visualising spatial data (Agrawal & Gupta, 2017).  The use of GIS was becoming more and more popular are Unix operating systems within the 90’s due to the processing power of ‘modern’ computers (Maguire, 1991), which only requires standard input, storage and output devices.  One of the biggest benefits to using a GIS is the sheer mass of data which is being created thanks to the use of Smart Phones, IoT and smart cities, with the volume of geospatial data increasing 20% each year (Franch-Pardo et al., 2020). – GIS and Epidemiology The use of GIS within the medical area was not a dominant use, yet this did not stop the developers applying GIS to look at health issues locally, regionally, and globally (Cromley, 2003). This opened a new way for the medical area to explore their data and to track the spread of disease (Cromley, 2003) whilst also being used for decision making on matters such as ‘where would be best for a new clinic?’ (Jacquez, 2000). This would enable the users of GIS to easily visualise this on a map whilst adding the location data with a heat map of higher risk areas of diseases (Jacquez, 2000). 1.3 – GIS and the COVID-19 Pandemic Whilst the use of GIS has been explored with diseases, it is not surprising to see how quickly GIS was used to develop online portal of the recent COVID-19 pandemic. With companies like World Health Organisation (WHO) and John Hopkins University (JHU) developing their own portals and giving everyone a visualised look at the pandemic.  There is already academic work out there which looks at how GIS is used around the Covid pandemic, along with the challenges that they are facing (Zhou et al., 2020), there does not seem to be many papers which are adding modelling and prediction to the area as of this time, though there are a few who are using statistical probabilistic modelling for the virus (Bherwani et al, 2021).  When looking at the first few weeks of the virus, many countries were trying to understand geographic modelling, but several of their findings were inadequate (Mollalo et al, 2020). This involved using over 35 different types of variables to explain the spatial variability of the disease and finding that income inequality was a major factor in explaining the spread of Covid (Martellucci et al, 2021, Mollalo et al, 2020).  This would suggest more work could be done on the use of GIS with the Covid pandemic, which could look to touch on decision making (Sarwar, et al, 2020) which will better understand high risk area, looking at predictive modelling for the spread of data with time series geospatial datasets, which is where this project would come in. 2 - Methodology The ontological position of this work is objectivism, and the epistemological position of the work is positivism meaning data can only be collect based on observed phenomena such as someone being recorded as sick. This research is there for quantitative in nature, using the deductive approach the hypothesis will emerge from a review of the literature.  Although there are many different types of data analysis including, Machine learning, neural networks and so on, GIS will be used here because understanding location will be beneficial, as it is people who spread the virus.  There are different types of GIS software out there but for the purpose of this project, we will look at the 2 popular types of GIS packages. The 2 packages which will be explored are Desktop GIS and Web GIS, both technologies have a range of advantages and disadvantages. These will be explored below:   |  |  |  | | --- | --- | --- | | **Technology** | **Advantages** | **Disadvantages** | | Desktop GIS | A lot of free options available.  Good speed and performance.  Good community and documentation for a lot of the software packages. | Will need to be installed on all computers who needed access to it.  If using a proprietary software package, could become expensive. | | Web GIS | All processing takes place in the ‘cloud’ and accessed via a web browser.  Access to ‘portals’ which is a framework for sharing and using apps, data and maps.  Software is installed in a central place (database and maps) and can be accessed by the whole team.  Can use multiple data sources. | Could become expensive to develop the software.  Data costs can be very expensive.  Importing mass data could become difficult and time consuming. |   When looking at the advantages and disadvantages explored above, for this type of project there would need to be a web-based GIS system which could be setup within the cloud. This would allow access from any location if there a browser, with the ability to easily collaborate with other enthusiasts, along with the fact we could easily link datasets from outside sources such as Gov.uk, Data.gov, ONS and more.  There are multiple types of data collection for a project like this, primary and secondary. Whilst it would be great to be able to get primary data from users or people who want to contribute, it is highly recommended to use secondary datasets (open-source data). This would allow us to use the same if not close to the same datasets which other companies have used when developing a GIS portal. | | |
|  | | |
| **Ethical considerations** | |  |
| When working with data, it can often become an ethical issue if the data is not anonymised. Especially when looking at a person’s point of view, which could easily be met with backlash. Although this project would mainly focus on secondary data, it is important to be mindful of the ethical issues.  As stated by the BCS code of conduct (BCS, 2019), a professional should:  Only undertake to do work or provide a service that is within your professional competence.  NOT claim any level of competence that you do not possess.  Develop your professional knowledge, skills, and competence on a continuing basis, maintaining awareness of technological developments, procedures, and standards that are relevant to your field.  Ensure that you have the knowledge and understanding of Legislation\* and that you comply with such Legislation, in carrying out your professional responsibilities.  Respect and value alternative viewpoints and, seek, accept, and offer honest criticisms of work.  Avoid injuring others, their property, reputation, or employment by false or malicious or negligent action or inaction.  Reject and will not make any offer of bribery or unethical inducement. | | |
|  | | |
| **References** | |  |
| Agrawal, S. and Gupta, R.D., 2017. Web GIS and its architecture: a review. Arabian Journal of Geosciences, 10(23), pp.1-13.  Bherwani, H., Anjum, S., Kumar, S., Gautam, S., Gupta, A., Kumbhare, H., Anshul, A. and Kumar, R., 2021. Understanding COVID-19 transmission through Bayesian probabilistic modelling and GIS-based Voronoi approach: a policy perspective. Environment, Development and Sustainability, 23(4), pp.5846-5864.  Cromley, E.K., 2003. GIS and disease. Annual review of public health, 24(1), pp.7-24.  ESRI (2021) *What is GIS?* Available at: <https://www.esri.com/en-us/what-is-gis/overview> (Accessed 27/04/2021)  Franch-Pardo, I., Napoletano, B.M., Rosete-Verges, F. and Billa, L., 2020. Spatial analysis and GIS in the study of COVID-19. A review. Science of The Total Environment, 739, p.140033.  Jacquez, G.M., 2000. Spatial analysis in epidemiology: Nascent science or a failure of GIS?. Journal of Geographical Systems, 2(1), pp.91-97.  Maguire, D.J., 1991. An overview and definition of GIS. Geographical information systems: Principles and applications, 1, pp.9-20.  Martellucci, C.A., Sah, R., Rabaan, A.A., Dhama, K., Casalone, C., Arteaga-Livias, K., Sawano, T., Ozaki, A., Bhandari, D., Higuchi, A. and Kotera, Y., 2020. Changes in the spatial distribution of COVID-19 incidence in Italy using GIS-based maps. Annals of Clinical Microbiology and Antimicrobials, 19(1), pp.1-4.  Mollalo, A., Vahedi, B. and Rivera, K.M., 2020. GIS-based spatial modeling of COVID-19 incidence rate in the continental United States. Science of the total environment, 728, p.138884.  Sarwar, S., Waheed, R., Sarwar, S. and Khan, A., 2020. COVID-19 challenges to Pakistan: Is GIS analysis useful to draw solutions?. Science of the Total Environment, 730, p.139089.  Zhou, C., Su, F., Pei, T., Zhang, A., Du, Y., Luo, B., Cao, Z., Wang, J., Yuan, W., Zhu, Y. and Song, C., 2020. COVID-19: challenges to GIS with big data. Geography and sustainability, 1(1), pp.77-87. | | |
|  | | |
| **Project plan** | |  |
| Whilst some people are good at planning and estimating their work, this is something of a weakness of mine as I often work on things for longer than is needed. Below is an estimate of my initial project plan.    (Figure 1 – Spreadsheet of Project plan)    (Figure 2 – Gantt Chart of Project plan)  Both charts were created with an open-source piece of software called ProjectLibre. Whilst it is important to keep to a plan a lot of time plans can change. This plan has included a few floats between each task to allow for any issues. ***Total Days: 28 – This gave me a lot of free time.*** | | |