SOURCE AND REFERENCE EVALUATION

This is an evaluation of the sources and references used during the EPQ. They have been divided into 2 sections, one for the references used in the EPQ Diary and one for the references used in the EPQ Dissertation. There may be overlaps for the same.

# EPQ DIARY SOURCE AND REFERENCE EVALUATION

Brownlee, J. (2018, March 2). *How to Calculate Principal Component Analysis (PCA) from Scratch in Python*. Retrieved August 6, 2020, from Machine Learning Mastery: https://machinelearningmastery.com/calculate-principal-component-analysis-scratch-python/

* This is a reliable source as the author is a machine learning specialist who seems to be highly knowledgeable in data science and machine learning
* The information used from this source is very helpful as it gave an easy to follow guide on implementing the algorithm Principal Component Analysis which was then used in the project
* Overall, highly useful and relevant resource

Button, T. (2020, May 21). *MEI Introduction to Data Science 1-2*. Retrieved July 8, 2020, from YouTube: https://www.youtube.com/watch?v=ip5TNIiEV8Y

* This is one of multiple videos by the same author and they have the same significance and thus share an evaluation
* These are a set of videos on YouTube made by the exam board MEI so are highly reliable
* The concepts are basic but provide a good foundation for data science
* They were not closely relevant to the project however these videos were used before a concrete understanding of the project idea was formed
* Thus, they helped to delve deeper into data science and the video on CRISP-DM or the life cycle of a data science project was very helpful

Gellin. (2020).

* I cannot seem to locate this reference and not a lot of information about it was recorded when this reference was made
* This reference is only used to support the opinion that health workers should get the vaccine first, which is not used anywhere in the core project research
* So although there is little information about this reference, it does not impact the project in any major way

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning.* MIT Press.

* These authors are highly distinguished in their fields and this book is a very detailed and trusted source of information
* The main significance of this source was to help understand the mathematics behind the algorithm PCA which was used in the project
* Thus, even though it may not have been essential to the project, it was a helpful and relevant resource

Google Research. (2020).

* Google Colaboratory is an online runtime environment for python which is what I used for the research and analysis in this project
* This source simply describes Google Colab and based on experience using it, it is true and reliable
* Very useful resource

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., . . . Oliphant, T. E. (2020, August). Array programming with NumPy. *Nature, 585*(7825), 357-362. doi:10.1038/s41586-020-2649-2

* This is the source for the NumPy library which was mainly used for the algorithm PCA in python
* Very helpful and essential to project

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering, 9*(3), 90-95. doi:10.1109/MCSE.2007.55

* This is the source for the Matplotlib library which was used to create all the graphs in python
* Very helpful and essential to project

Isha Berry, J.-P. R. (2020, April 14). *Open access epidemiologic data and an interactive dashboard to monitor the COVID-19 outbreak in Canada*. Retrieved August 11, 2020, from cmaj: https://www.cmaj.ca/content/192/15/E420

* This is the source for the Canada Individual Cases and Deaths Dataset which is one of the datasets that were analyzed
* Very essential and useful for the project
* The data was collated from government health authorities and news media, entered manually and coordinated by the University of Toronto
* Thus, the data seems highly reliable and fit to use

Jolliffe, I. T., & Cadima, J. (2016, April 13). Principal component analysis: a review and recent developments. *The Royal Society*. doi:https://doi.org/10.1098/rsta.2015.0202

* This source speaks to the merits of the algorithm PCA used in the project for analysis of one of the datasets
* This is from an article by The Royal Society hence it seems very reliable
* This was helpful as it established the credibility of using PCA for dimensionality reduction of data even though there was some information loss

NYC Health. (2020, April 14). *Daily Data Summary*. Retrieved August 10, 2020, from NYC Government: https://www1.nyc.gov/assets/doh/downloads/pdf/imm/covid-19-daily-data-summary-deaths-04152020-1.pdf

* This data was collated by the New York City government, providing information about COVID-19 cases and deaths in relation to age an underlying health conditions
* This was important because it provided insight into how age must not be the only determining factor for a vaccination priority and thus allowed the project to take a different turn in the conclusion
* Thus, reliable and essential source

NYC Health. (2020, May 12). *Daily Data Summary*. Retrieved August 10, 2020, from NYC Government: https://www1.nyc.gov/assets/doh/downloads/pdf/imm/covid-19-daily-data-summary-deaths-05132020-1.pdf

* Same as above

Pedregosa, F., Varoquaux, G., Gramfort, A. a., Thirion, B., Grisel, O., Blondel, M., . . . Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research, 12*(2825-2830).

* This is the source for the Scikit-learn library which was used to help implement PCA algorithm in python
* Very helpful and essential to project

The pandas development team. (2020, February). pandas-dev/pandas: Pandas. *1.0.3*. Zenodo. doi:10.5281/zenodo.3509134

* This is the source for the panas library which was used for dataset manipulation in python
* Very helpful and crucial to project

University of Oxford. (2020, July 29). *COVID Conversations: Mathematical Epidemiology*. Retrieved July 29, 2020, from YouTube: https://www.youtube.com/watch?app=desktop&v=Bar6KJKJNAE

* This video talk about epidemiology and the mathematics behind
* Helped me understand epidemic reproduction rates and how they are calculated as well as their drawbacks and important points
* Was from University of Oxford
* Thus, very reliable and helpful to project

Waskom, M. L. (2017, September). mwaskom/seaborn: v0.8.1. Zenodo. doi:https://doi.org/10.5281/zenodo.883859

* This is the source for the seaborn library which was used to create all the graphs in python and make them look more attractive
* Very helpful and essential to project

WebMD. (2020). *How Does Coronavirus Spread?* Retrieved August 12, 2020, from WebMD: https://www.webmd.com/lung/coronavirus-transmission-overview#2

* Very helpful source that explained the meaning of the term “Community spread” for COVID-19
* It was very useful as it allowed to understand an important part of the data which was labelled “Community” but did not explain the term itself, thus helped in drawing more informed conclusions from the data
* Seems fairly reliable as other sources on the internet agree on the definition of that term

# EPQ DISSERTATION SOURCE AND REFERENCE EVALUATION

Berry, I., Soucy, J.-P. R., Tuite, A., & Fisman, D. (2020, April 14). Open access epidemiologic data and an interactive dashboard to monitor the COVID-19 outbreak in Canada. *CMAJ, 192*(15). doi:https://doi.org/10.1503/cmaj.75262

* This is the source for the Canada Individual Cases and Deaths Dataset which is one of the datasets that were analyzed
* Very essential and useful for the project
* The data was collated from government health authorities and news media, entered manually and coordinated by the University of Toronto
* Thus, the data seems highly reliable and fit to use

Brownlee, J. (2018, March 2). *How to Calculate Principal Component Analysis (PCA) from Scratch in Python*. Retrieved August 6, 2020, from Machine Learning Mastery: https://machinelearningmastery.com/calculate-principal-component-analysis-scratch-python/

* This is a reliable source as the author is a machine learning specialist who seems to be highly knowledgeable in data science and machine learning
* The information used from this source is very helpful as it gave an easy to follow guide on implementing the algorithm Principal Component Analysis which was then used in the project
* Overall, highly useful and relevant resource

Gellin. (2020).

* I cannot seem to locate this reference and not a lot of information about it was recorded when this reference was made
* This reference is only used to support the opinion that health workers should get the vaccine first, which is not used anywhere in the core project research
* So although there is little information about this reference, it does not impact the project in any major way

Hasell, J., Mathieu, E., Beltekian, D., Appel, C., Gavrilov, D., Giattino, C., . . . Roser, M. (2020). A cross-country database of COVID-19 testing. *Scientific Data, 7*(345). doi:https://doi.org/10.1038/s41597-020-00688-8

* This is the source for the COVID-19 Case Demographics Dataset which is one of the datasets that were analyzed
* Very essential and useful for the project
* The database consists of official data on the number of COVID-19 diagnostic tests performed over time across 94 countries and relies on figures published in official sources, including press releases, government websites, dedicated dashboards, and social media accounts of national authorities
* The database is also updated regularly and verified with sources provided
* Thus, the data seems highly reliable and fit to use

Isha Berry, J.-P. R. (2020, April 14). *Open access epidemiologic data and an interactive dashboard to monitor the COVID-19 outbreak in Canada*. Retrieved August 11, 2020, from cmaj: https://www.cmaj.ca/content/192/15/E420

* This is the source for the Canada Individual Cases and Deaths Dataset which is one of the datasets that were analyzed
* Very essential and useful for the project
* The data was collated from government health authorities and news media, entered manually and coordinated by the University of Toronto
* Thus, the data seems highly reliable and fit to use

NYC Health. (2020, April 14). *Daily Data Summary*. Retrieved August 10, 2020, from NYC Government: https://www1.nyc.gov/assets/doh/downloads/pdf/imm/covid-19-daily-data-summary-deaths-04152020-1.pdf

* This data was collated by the New York City government, providing information about COVID-19 cases and deaths in relation to age an underlying health conditions
* This was important because it provided insight into how age must not be the only determining factor for a vaccination priority and thus allowed the project to take a different turn in the conclusion
* Thus, reliable and essential source

NYC Health. (2020, May 12). *Daily Data Summary*. Retrieved August 10, 2020, from NYC Government: https://www1.nyc.gov/assets/doh/downloads/pdf/imm/covid-19-daily-data-summary-deaths-05132020-1.pdf

* Same as above