1. **Project Summary:**

*What is the problem?*

As Covid-19 has become a global pandemic, leaders at all levels of government, municipal, regional, and state, have had to make important policy decisions with limited information. While there remain many questions surrounding the nature of Covid-19 – its true infectiousness, how exactly it spreads, its effects on the human body, etc. – we have observed a high variation in outcomes amongst the countries of the world. Some countries have been ravaged by the virus, most famously Spain and Italy, but also Iran, and others have seemingly contained the virus in a swift, systematic manner, e.g., South Korea, Taiwan, Germany, among others.

*Why is it an important policy problem? (Provide citations if possible to support your claims of relevance. Examples could be government requests for comments, white papers by policy organizations, etc.)*

There are two main issues facing decision-makers in the Covid-19 era. First, it is been hard discern exactly which policy measures are most effective in saving lives and to what degree. Second, the adoption of public safety measures, most notably ‘social distancing’, has also brought about economic crises in nearly every country of the world, bringing about a tradeoff between public health and present and future economic stability.

The IMF has projected the world economy to contract 3% in 2020, comparatively worse than the 2008-09 financial crisis\*.

\* <https://www.imf.org/en/Publications/WEO/Issues/2020/04/14/World-Economic-Outlook-April-2020-The-Great-Lockdown-49306>

*Who is the audience for your report?*

The report is intended for policy makers from mayors to legislators, to heads of state, but is also of relevance to a concerned public desirous to educate itself. We intend to build a model that can evaluate the efficacy of specific policy measures in reducing death rates, given the social, health, geographic features of a country or region.

This report is intended to serve to inform decisionmakers and provide insights into the likely outcomes of available policy measures.

Leaders may also cite this report as a reason for them having taken specific actions rather than simply their intuition, and can cite real numbers, such as lives saved due to their decisions.

*What kinds of actions could be taken based on your results, and who is equipped to take those actions?*

Our model could inform a mayor about many lives social distancing could save and the negative economic tradeoffs of such action. It can could also inject empiricism into the policy debate surrounding easing of lockdown measures rathe than individuals/politicians relying on their gut.

*How will you validate whether your results might be relevant to your intended audience?*

We could evaluate the demand for such a model through open source competitions on Kaggle. We could also get in touch with local representatives or the mayor’s office to discuss the need for empirical models to inform policy measures.

1. **Data:**

*Describe the data you have and the data you’ll need to collect.*

Johns Hopkins has been tracking disaggregated data about infections, deaths, and recoveries by country and date. Furthermore, the dataset on the US is disaggregated even further, breaking down disease statistics by state and city. This data not only allows one to track the rates of infection, but also infer the date of a country’s first infection, which we can tie in with policy measures related to responsiveness.

Information on policy measures that have been taken will be vital, and we have already found a large dataset with country-level information from ACAPS, an NGO focused on the sharing of information pertaining to humanitarian crises. We plan to tie this in with the Johns Hopkins disease data as previously mentioned.

Information on country preparedness can be obtain in a neat format from World Bank Open Data datasets. They have aggregated relevant determinants of a country’s susceptibility to Covid-19 in the link below. Relevant indicators include number of hospital bed per capita and percent of smokers in the population. <http://datatopics.worldbank.org/universal-health-coverage/coronavirus/>

We will need to collect data on the economic impact of Covid-related policy measures. We will have to decide the relevant economic indicators that measures the economic well-being and stability of the country. The IMF has an [economic policy tracker](https://www.imf.org/en/Topics/imf-and-covid19/Policy-Responses-to-COVID-19) broken down by country that may be of use.

*Include some descriptive stats that show you have enough to solve the problem.*

??

*What analysis do you plan to perform or show on this data, to help inform your understanding of the data, as well as your design and selection of Machine Learning Models.*

As this is time-driven data, we can visualize the relationship between economic decline and policy measures taken, as well as the economic-human life tradeoff in visualizations to decide the factors that map together in the cleanest way.

1. **Machine Learning:**

*What type of machine learning problem is this? Are you developing a classification technique? Regression? Prediction? Clearly articulate the* ***learning*** *that your resulting models will enable.*

We plan to run a regression with the various features and policy measures.

Our model will be trained to take into account infection and economic trends and underlying development factors and a death to recovery ratio (a nice proxy that models the country’s ability to help citizens survive infection).

*What types of models will you apply? Justify your choice of models. Your considerations could include the nature of your dataset (types and nature of features, size of the data), the requirements for model training or testing (e.g., real-time classification), or any other considerations you might have.*

?

**5. Evaluation:**

*Describe your process for evaluating the models.*

MSE and R2?

*How will you validate the correctness of the models? Be as specific as possible about your evaluation techniques (e.g., out-of-sample errors, imbalanced training sets, etc.)*

We will explore using temporal holdouts to cross validate as we are working with time series data.

**6. Ethics:**

Briefly discuss the ethical implications of your work. Will your models suffer from any kind of bias? Will this potential bias impact the fairness of any proposed solutions?

Our project will be an interesting one because it cuts to the heart of an ethical debate – keep people in lockdown for long periods of time to save all possible lives or deconfine people gradually to alleviate the economic strain in exchange for a marginal loss of life. Our hope is that the model can provide estimations and statistics to this debate so tradeoffs can be evaluated with real numbers.

As is unavoidably the case, our model will suffer from various biases. If we are to evaluate every country and take their statistics at face value, our model will be biased due to inaccuracies in underreporting of deaths or infections. It may be that countries who are more prone to providing false statistics are also more prone to less efficient policy responses to disasters such as pandemics. We will take that into consideration and remove country data such as China, whose statistics are plausibly falsified and can through our model into further imbalance.

The unfortunate reality of underreported rates of infections and deaths will likely bias our model downwards when it comes to the life-saving effects of certain policy measures while leaving economic effects the same; i.e., the (less biased) economic effects will overstated relative to the biased health statistics.