**Effectiveness of State Policy Actions to Address COVID-19**

**Plan & Detail**

**Team Members:**

Kacper Malysa (KM) · Mingxi Xia (MX) · Roman Sydorchuk (RS) · Spencer Sumner (SS) · Yurun Liu (YL)

| **General** | **Task** | **Description** | **Subtasks** | **Assigned to** | | | | | **Range** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| KM | MX | RS | SS | YL |
| **Data Cleanup** | Main Dataset Review | Review of the dataset which includes information about state policies over-time. | * Github Review * KFF Review * JHU review |  |  |  |  |  | 11/2-11/7 |
| Secondary Dataset Review | Review of the dataset which includes information about infections, deaths, hospitalizations, and different covid-related statistics over time. | * Death Rates * Hospitality Rates * Infection Rates |  |  |  |  |  | 11/2-11/7 |
| Useless Data Review | Figuring out what to do with useless data case-by-case. | * Identify useless data cases * Figure out what to do with each case |  |  |  |  |  | 11/7-11/11 |
| Handling of Useless Data | Handle the useless data as discussed in the review | * Duplicates * Error data * Unknown Values |  |  |  |  |  | 11/7-11/11 |
| **Data Analysis** | Combining All Datasets | Take our current datasets that we have compiled, and overlap and combine them together so that we know how to use all of our data together. | * Look at each dataset * See how they can interact with each other |  |  |  |  |  | 11/11-11/15 |
| Variable Identification | Finalize input features and output variables, bring up any possible issues to the team | * Input features * Output variables |  |  |  |  |  | 11/11-11/15 |
| Representations for Each Feature | Consider the best representation for each feature and consider how can we approach each one of them for the remainder of the project | * Linear * Parametric * Non-parametric |  |  |  |  |  | 11/14-11/18 |
| Data Visualizations | Plot death rates/hospitalizations/infection rates across the USA over time, mark federal responses on graph. Plot output variables by state for comparison. | * Plot Different Rates * Make State and country-wide plots |  |  |  |  |  | 11/14-11/18 |
| **Midpoint Reassessment** | Catch-Up | Assess current progress and see if we need to pivot in any way towards a different goal. Make sure that what we are trying to accomplish is possible with our data. | * Possible Pivot * Pros and Cons of different approaches |  |  |  |  |  | 11/18-11/20 |
| Team Assessment | Do a midpoint assessment of the team to make sure that everyone is doing equal amounts of work, and that everyone feels good about where we are with our project. | * Split work evenly * Make sure team is ready for data modeling |  |  |  |  |  | 11/18-11/20 |
| Plan Out Model Construction | Prepare the data and search for any possible issues with the upcoming model construction, to make the team aware of any roadblocks that might occur. | * Pick out models * Any modeling related roadblocks are discussed |  |  |  |  |  | 11/20-11/22 |
| Plan out Dimension Reduction | Prepare the data and search for any possible issues with the upcoming dimensions reduction, to make the team aware of any roadblocks that might occur. | * Identify possible issues for dimension reduction * Make sure the team is on the same page |  |  |  |  |  | 11/20-11/22 |
| **Data Modelling** | Define Train and Test Set | Split our data into train and test sets to prepare for model application and selection process | * Split the data * Identify possible issues |  |  |  |  |  | 11/22-11/26 |
| Choose Model | From the multiple choices we are leaving open to ourselves, choose the best model that works with our data, and that results in accurate output. | * Simple/Multiple linear regression * Logistic regression * KNN |  |  |  |  |  | 11/22-11/26 |
| Variable Reduction / Validation | Mix every feature in the model and remove the weak ones, to allow for the best results and the most accurate modeling. | * Mixing of features from models |  |  |  |  |  | 11/25-11/29 |
| Identify The Best Models | Identifying the best models to use in the final report. Weigh pros and cons of each approach, and see which one of the models allows us for not the most convenient to us, but the most honest and accurate results. | * Compare models * Select the best model |  |  |  |  |  | 11/25-11/29 |
| **Project Report & Presentation** | Final Changes | Polish what we currently have, and collect all the work and notes from the entire semester into one place for easier final report writing. | * Organization * Polish |  |  |  |  |  | 11/27-11/30 |
| Summary of All Work | Finalize calculations, data wrangling and any other related work. Allow time for the entire team to summarize what we have done, and to look over what is going to be our final narrative in the project report | * Everyone looks back at what they did * Final team assessment is done |  |  |  |  |  | 11/27-11/30 |
| Final Visualizations | Once all the data is completed and we have our results, create graphs, plots, and other visual aids to make the final report easier to understand for anyone. | * Graphs * Plots * Visualizations |  |  |  |  |  | 11/29-12/02 |
| Final Report & Presentation | Finalize the report & presentation, and allow time for the entire team to look over it multiple times to allow for any final checks. | * Final Report * Presentation |  |  |  |  |  | 11/29-12/02 |

**Gantt Chart**

