Project Proposal – Aaron Kohn

Topic

Determine the effect of government orders on the spread of Covid-19.

Business Problem

Through out the duration of the Covid-19 pandemic many rules and regulations have been implemented to try to limit the negative effects of the virus. This study will try to quantify the effect of some of the regulations put into effect. What are the effects of specific regulations, and which are most effective?

Datasets

The data includes timeseries information on all measures for Covid-19 on the county level. The data is made available by Covid Act Now. The data regarding regulations can be retrieved from the CDC/Policy + Surveillance. The data is a day-by-day record for each county and the level of the mandate implemented. There are separate data sets for each of the regulations.

Methods

For this project, I plan on following the CRISP-DM methodology framework. The purpose of this project is to determine the interventions that are effective against the spread of the Covid-19 virus. As a starting point the goal will be to determine the effect of specific regulations, and which are most effective.

Working with timeseries data will require some testing and transforming of the data to se if there is seasonality or autocorrelation within the data. This will determine the exact models and transformations required for the data. As a starting point autoregressive models will be used to see if they are effective. Other models may be tried if deemed necessary.

For verification purposes timeseries train test splits will be preformed to determine the effectiveness of the model's prediction power. Furthermore, a random subset of counties will be withheld to see if the model can duplicate its results on those as well.

Ethical Considerations

Unfortunately, there has been some politicization around Covid-19 regulations. It is important to ensure that an open mind and fair process is followed throughout the project. Findings and trends should be kept in perspective and a willingness to explore all avenues is important. The data being used is assumed to be accurate, there have been accusations of incorrect reporting techniques in some areas. The numbers will be used as they are reported as they need to be assumed the most accurate available.

Challenges/Issues

Dealing with timeseries data for all counties will result in a very large dataset. There may be issues with dealing with such large amount of data with the limited resources available. Subsets of the data may need to be used. Care will be needed to ensure that accurate representation of counties is present with in the selected data.

Cleaning and transforming the data to be able to apply the necessary models will require care to ensure that the data remains accurate. Properly merging data sets and matching information from different sources requires measure to prevent errors within the data.

References

CDC, COVID-19 Community Intervention and At-Risk Task Force, Monitoring and Evaluation Team & CDC, Center for State, Tribal, Local, and Territorial Support, Public Health Law Program, "State and Territorial COVID-19 Orders and Proclamations for Individuals to Stay Home," (August 15, 2021).

CDC, COVID-19 Community Intervention and Critical Populations Task Force, Monitoring and Evaluation Team & CDC, Center for State, Tribal, Local, and Territorial Support, Public Health Law Program, "State and Territorial COVID-19 Orders and Proclamations Banning Gatherings," (August 15, 2021).

CDC, COVID-19 Community Intervention & Critical Populations Task Force, Monitoring & Evaluation Team, Mitigation Policy Analysis Unit and the CDC, Center for State, Tribal, Local, and Territorial Support, Public Health Law Program, "State and Territorial COVID-19 Orders and Proclamations Closing and Reopening Restaurants" (August 15, 2021).

CDC, COVID-19 Community Intervention & Critical Populations Task Force, Monitoring & Evaluation Team, Mitigation Policy Analysis Unit, the CDC, Center for State, Tribal, Local, and Territorial Support, Public Health Law Program, and Max Gakh, Assistant Professor, School of Public Health, University of Nevada, Las Vegas, "U.S. State and Territorial Orders Requiring Masks in Public," (August 15, 2021).

Jason Brownlee (August 6, 2018), 11 classical time series forecasting methods in python (cheat sheet). Retrieved from https://machinelearningmastery.com/time-series-forecasting-methods-in-python-cheat-sheet/

Keita Miyaki (August 15, 2019), Time series split with scikit-learn. Retrieved from https://medium.com/keita-starts-data-science/time-series-split-with-scikit-learn-74f5be38489e