Research design updated

Post-processing the below variables

|  |  |  |
| --- | --- | --- |
| **Dimension** | **Variable** | **Definition** |
| Spatial | GEOID | ID for table join |
| CountyName | County Name |
| STATEcode | State code |
| STATEname | State name |
| Temporal | Date | May 1 – Sep 30, 2016 to 2022 by day |
| Holiday | 0-nonholiday; 1-holiday |
| Year | 2016 to 2022 |
| Month | 1-12 |
| Week | 1-52 (weekly order for each year) |
| Weekend | 0-non-weekend; 1-weekend |
| Sentiment | SentimentScore | A relative measure of score to indicate sentiment level generated by natural language processing tools based on geotweets collected from 2016 to 2022 by Harvard CGA |
| Heatwave | Heat Count | 0-not affected by heatwave; 1-affected by heatwave |
| Weather / climate | Humidity | Level of humidity (we may not use this) |
| Precipitation | Level of rainfall (to be replaced by rainfall anomaly) |
| Solar | Level of solar radiation (we may not use this) |
| Demographic and socioeconomic | Vulnerability index (THEMES) | US CDC vulnerability index as a whole (by county but time-invariant, bi-year available in 2020, 2018 and 2016 but applied to one following year for 2021-2022, 2019, and 2017) |
| VI1 | US CDC vulnerability index in Theme 1 (socioeconomic status; total five themes) (we don’t use this) |
| VI2 | US CDC vulnerability index in Theme 2 (racial and ethnic minority status) (we don’t use this) |
| VI3 | US CDC vulnerability index in Theme 3 (household characteristics) (we don’t use this) |
| VI4 | US CDC vulnerability index in Theme 4 (housing type/transportation) (we don’t use this) |
| Pandemic | Stringency index | Index of policy stringency provided by Oxford policy tracker, only available for 2020 and 2021(we don’t use this) |
| Government response index | Index of government response provided by Oxford policy tracker, only available for 2020 and 2021(we don’t use this) |
| Containment health index | Index of containment health provided by Oxford policy tracker, only available for 2020 and 2021(we don’t use this) |
| Economic support index | Index of economic support provided by Oxford policy tracker, only available for 2020 and 2021(we don’t use this) |
| Total index | Add up the above four |
| Air pollution | Air pollution raw | Raw measure of air pollution (aerosol optical depth (AOD) (we don’t use this) |
| Air pollution interpolate | Derived measure of air pollution, interpolated spatially and temporarily |

Variables to be used in modelling

|  |  |
| --- | --- |
| **Dependent Variable:** | Sentiment |
| **Independent Variables in BASE MODEL** |  |
| Demographic and socioeconomic indicator | Vulnerability index (THEMES) |
| Pandemic indicator | Total index (add up four indexes) |
| Spatial fixed effect | STATE and County (use polygon shapefiles) |
| Temporal fixed effect | Holiday |
|  | Year |
|  | Month |
|  | Week |
|  | Weekend |
|  | Event |
| Below variables to be added in DLNM, respectively |  |
| Air pollution indicator | Air pollution interpolate |
| Heatwave indicator | Heat Count |
| Weather/climate indicator | Precipitation anomality |

We still use the nationwide at the county level as the whole to start from the simple linear modelling given it is not accurate to aggregate the heat data to the state level.

1. Start from simple linear models

BASE Model: for the entire US by county; You can try to add in the below spatial/temporal/fixed effects using back/forward entry to find out the best combinations.

Sentiment ~ Spatial effect (input of state polygon and county polygon) + temporal effect (year, month, week, weekend, and holiday) + fixed effect (vulnerability, events)

1. ADVANCED Model with no lag:

BASE MODEL + heatwave

BASE MODEL + air pollution

BASE MODEL + rainfall

BASE MODEL + heatwave + air pollution

BASE MODEL + heatwave + rainfall

BASE MODEL + air pollution + rainfall

BASE MODEL + heatwave + air pollution + rainfall

1. After select the above best model then add DLNM: ADVANCED Model with DLNM

DLNM with lag 1-14 days