

Analysis: preliminary dataset for business closures

Business closures

Page purpose: This page summarizes our approach to generating a dataset for business closures for Chinese provinces. It describes the process for generating preliminary datasets and also proposes next steps for more in-depth analysis

Analysis objective: build a dataset of POI closures and adaptations, including information about when these closures / adaptations were lifted

Takeaways:

- We find that the data in WHO on business closures is too sparse to use for our analyses and will not pursue this data source for business closure analyses. We also find that the Coronanet data is too sparse to use at this level of granularity
- In the process of exploring this data, we have created a mapping of key words into specific places of interest (saved below) that can hopefully be used across other datasources
- As a next step in this analysis, we should explore other datasources like the Chinese version of Yelp to assess when specific businesses closed and reopened. Our work creating the MassCPR policy database overview should allow us to validate these other datasets

Links:

Code	policy-database / notebooks / explore-china-policy-databases.R
Input data	policy-database / data / external
Output data	policy-database / notebooks / outputs

Table of contents:

- Business closures
- Data sources
- Data cleaning
- Output and next steps

Data sources

We explore two of the larger policy databases for use in determining the start and end dates of business closures across provinces and cities in China and find that neither is suitable for business

Data source	Github	Sponsoring organizations	Assessment
CoronaNet Research Project	https://github.com/saudiwin/corona_tscs	The project is supervised by senior political scientists from NYU Abu Dhabi, TU Munich and Yale University	✗ Not suitable for business closure analysis
WHO Public Health and Social Measures	FALSE	World Health Organization aggregation of ACAPS, Oxford, GPHIN, CSH Vienna, WHO IHR, CDC, JHU	✗ Not suitable for business closure analysis

Data cleaning

Categorized POIs - I find the POIs relatively easy to categorize by using the `targeted` variable which adds extra detail to the `who_measure` variable. Below are the categories I created

```

restaurant <- c("restaurant", "catering", "bar", "pub", "eater")
entertainment <- c("entertainment", "party", "karaoke", "hair",
"beauty", "salon",
"club", "cinema", "parlor", "bathhouse", "performance", "casino",
"recreation")
hotels <- c("hotel")
sports <- c("sport", "fitness", "gym")
markets <- c("market", "food", "convenience", "salmon", "outlet")
office_govt <- c("civil", "immigration", "facilit", "govt employees",
"government employees")
office_comm <- c("office", "workplace", "non-essential busi", "factori")
scenic <- c("disney", "religi", "church", "baseball", "event",
"cultural",
"leisure", "beach", "cultural", "museum", "playground", "tour",
"temple")
govt_service <- c("government service", "driving test", "librar",
"social welf")
all_emp_busi <- c("all employee", "all busine", "all worker")

```



Checking the measure stage variable - I manually checked the `measure_stage` variable against the `comments` variable to evaluate how the WHO dataset labels the stage of policy announcements - new announcement, modified announcement, finish announcement. My assessment of these spot checks leads me to believe that the variable is relatively accurate at labeling the “phase outs” of different practices.

The greater concern with using this variable to identify the end of each policy is with the comprehensiveness of policy announcements in our data source. While we expect most policy announcements to have a corresponding policy ending announcement, we find that there is not an even balance of “policy starts” and “policy ends”

Category	Policy stage	N policies
Gatherings	finish	7

Gatherings	new	24
Gatherings	update	20
Offi, busi, inst	finish	46
Offi, busi, inst	new	34
Offi, busi, inst	update	49
School	finish	42
School	new	20
School	update	22

Output and next steps

Based on this assessment, we decided not to proceed with WHO for business closure analyses. Going forward, we will explore other industry-specific sources to determine the start and end date of business closures and use our Mass CPR policy database collected internally for Beijing to validate these outputs