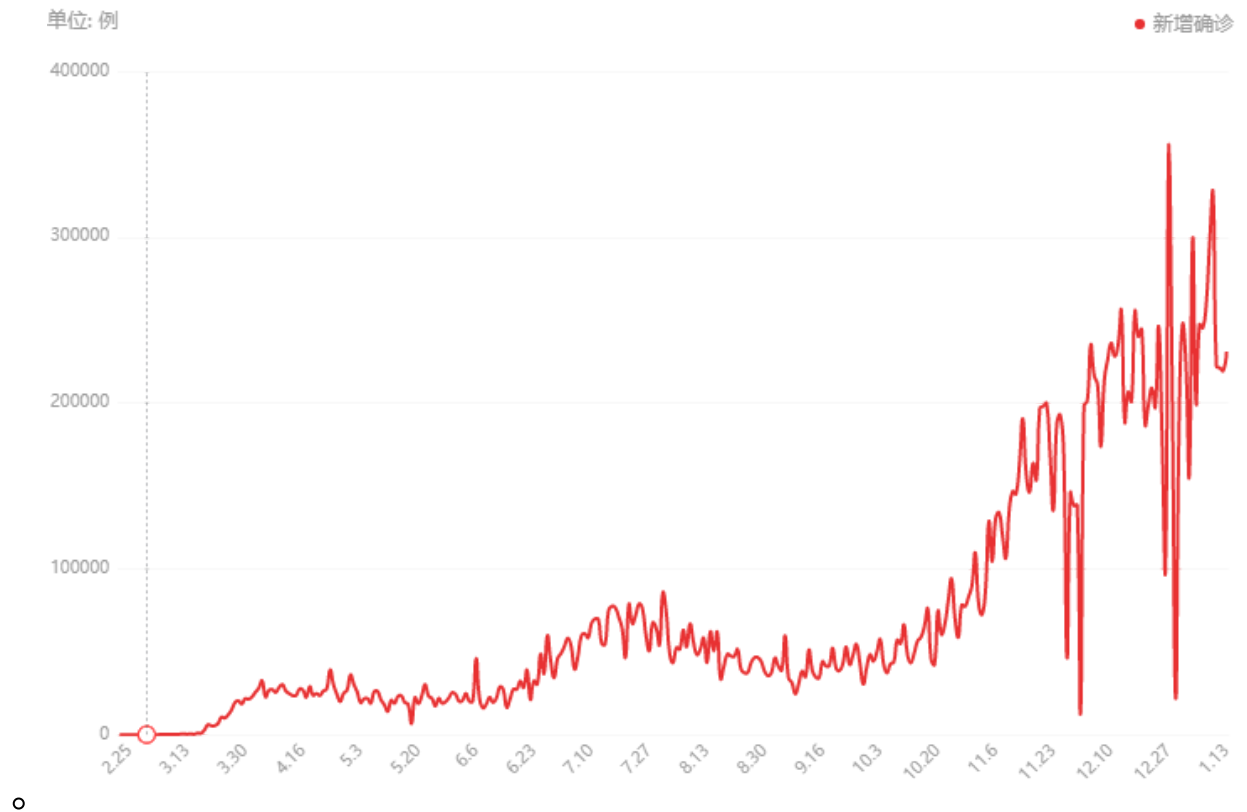


# 各国数据日期选取及kaggle 数据使用

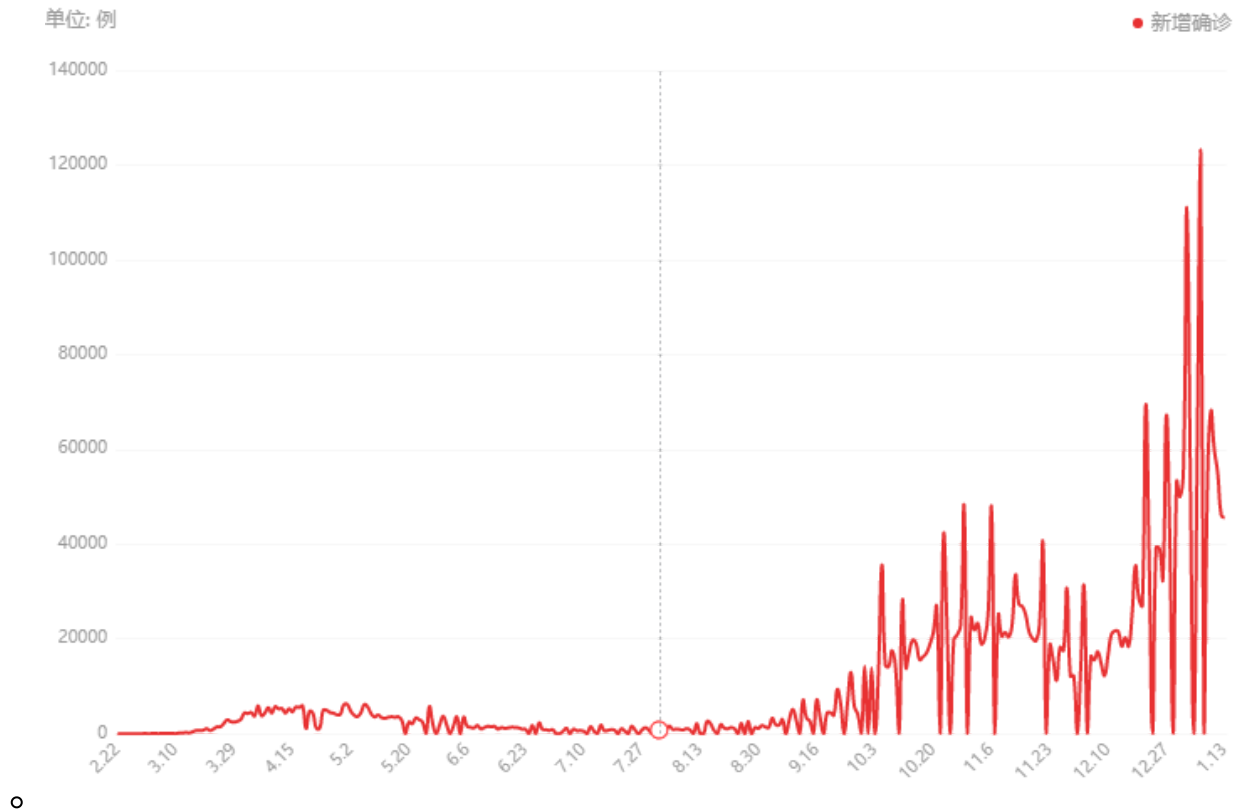
## 各国数据选取

图片来源于百度新增确诊趋势图

- 美国（预测峰值）：
  - 训练：2020.03.01 至 2020.12.31
  - 预测：2021.01.01 至 ?



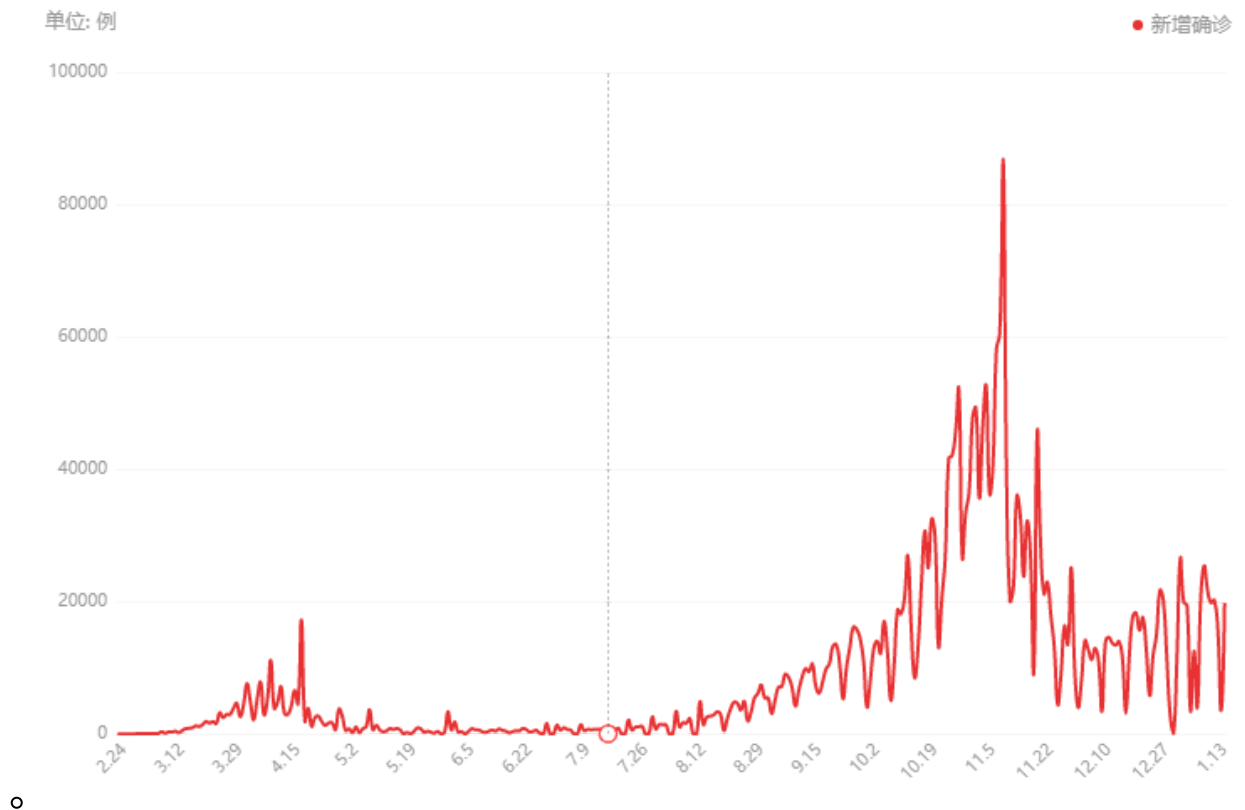
- 英国（预测第二个峰值）：
  - 训练：2020.08.01 至 2020.12.31
  - 预测：2021.01.01 至 ?



- 法国 (预测第三个峰值) :

- 训练: 2020.07.15 至 2020.12.31

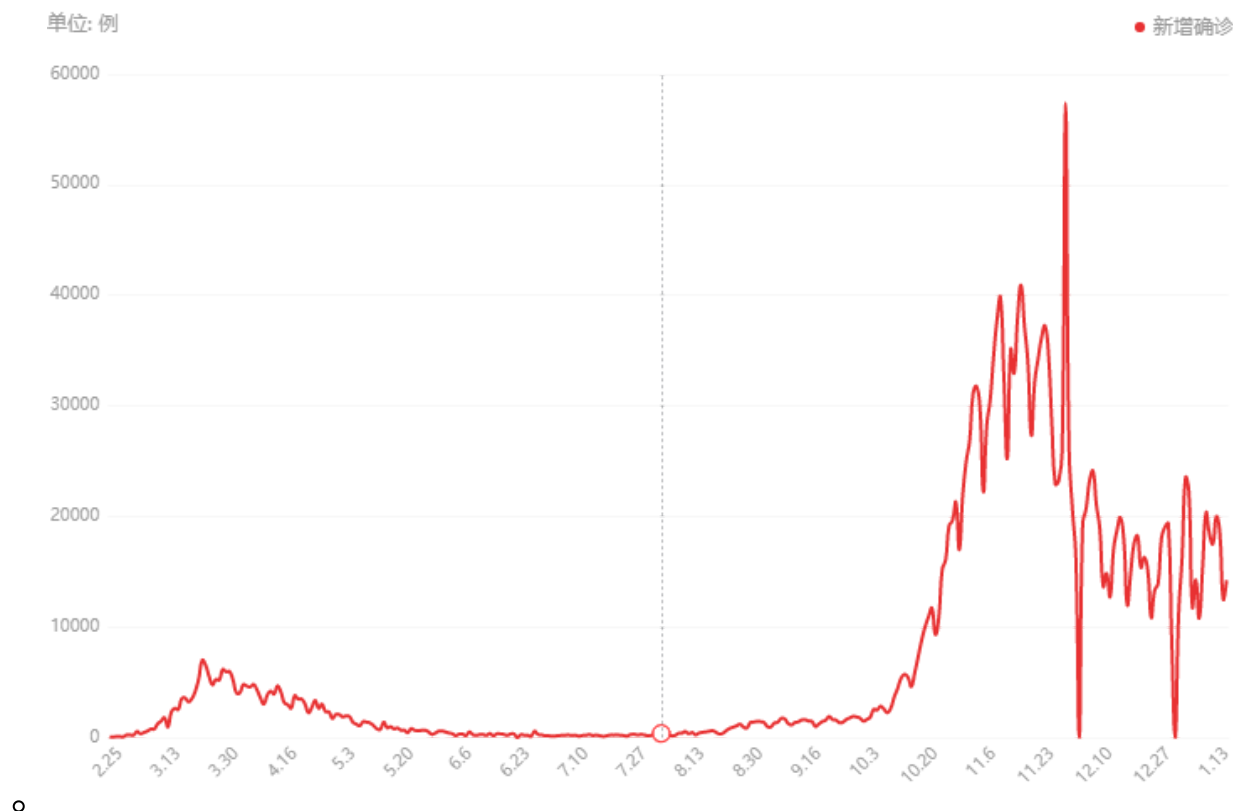
- 预测: 2021.01.01 至 ?



- 意大利 (预测是否会再次爆发) :

- 训练: 2020.08.01 至 2020.12.31

- 预测: 2021.01.01 至 ?



## John Hopkins 数据介绍

- 源文件 `./time_series_covid19_confirmed_global.csv`,  
`./time_series_covid19_deaths_global.csv`, `./time_series_covid19_recovered_global.csv`
  - 数据来源于 <https://github.com/CSSEGISandData/COVID-19.git>
  - 格式为 `csv` 文件, 分隔符为 `,`
  - 列变量名分别为: 省/州, 国家/地区, 纬度, 经度, 时间, 每行为一个国家或地区的数据 (部分国家没有子区域的数据)
  - 开始日期 `2020.01.22` 至今
- 调用脚本 `call.py`
  - 需安装 `pandas`
  - 函数 `get_name()`, 参数为 `Region=?`, 默认为 `'all'`
    - 输入参数为国家名 (`str`), 例如 `'France'`, 输出为省/州名称; 输入参数为 `'all'`, 输出为所有国家名称
    - 输出文件夹 `./name/`, 格式为 `csv`
  - 函数 `get_data()`
    - 参数 `Region=?`, 默认为 `'France'`
    - 参数 `level=?`, `'province'`: 输出省级数据; `'country'`: 输出国家整体数据, 默认为 `'country'`
    - 输出文件夹 `./data/`, 格式为 `csv`, 列变量名分别为: 日期, (累计) 确诊, 死亡, 治愈