Project Description

> PS: This report is based on the information I collected on 2021/11/29, the infomation may be updated over time.

Introduction

For the project description, it will include the motivation of the project, brief of data sources, analysis of the data and the conclusion drawn from the project's results.

Motivation

Since the Chinese Spring festival is approaching, my family and I plan to travel. Currently, we are still experiencing the covid 19 pandemic all over the world. Therefore, before traveling, I need to verify the epidemic risk level of the destinations and only travel to low-risk areas. Traveling to areas with other risk levels may cause the personal health code to become a red code, which may result in home isolation and inconvenience in daily travel. After determining which cities are safe to travel to, I will decide the final destination based on the specific weather conditions in these cities.

Brief of Data Sources

I used three data sources in total.

The first data source is scraping a travel website to get the most visited cities in China. There are 15 cities in total and these cities are the keys to combine the another two data sources together.

```
-----City Info-----
   Area
0    重庆
1    上海
2    北京
3    武汉
4    西安
5    成都
6    广州
7    贵阳
8    天津
9    杭州
10   昆明
11   南京
12   宁波
13   苏州
14   青岛
```

The second data source is a public API which can be used to access the daily covid information of all provinces and city areas in China. This public API can be used freely and without private key. I used this data source to get the risk levels for the destinations I got from the first data source. Since the risk levels infomation(grade column) are not available to all areas in China, thus I will make decision base on the information provided by Now_confirmed and Suspect columns.

```
-----Covid Info-----
      Area  Updated_Today  Now_Confirmed  Suspect          Grade
0       台湾           True           1998        0  Not_Available
1    地区待确认           True           1998        0  Not_Available
2       辽宁           True            157        0  Not_Available
3       大连          False            152        0          部分中风险
4     境外输入          False              5        0  Not_Available
5       丹东          False              0        0  Not_Available
6    地区待确认          False              0        0  Not_Available
7       营口          False              0        0          全部低风险
8       铁岭          False              0        0  Not_Available
9       抚顺          False              0        0  Not_Available
..     ...            ...            ...      ...            ...
496     天水          False              0        0          全部低风险
497     陇南          False              0        0  Not_Available
498   境外输入          False              0        0  Not_Available
499     定西          False              0        0  Not_Available
500     平凉          False              0        0  Not_Available
501     庆阳          False              0        0  Not_Available
502     白银          False              0        0  Not_Available
503    甘南州          False              0        0  Not_Available
504     临夏          False              0        0  Not_Available
505     金昌          False              0        0  Not_Available

[506 rows x 5 columns]
```

The third data source is a public weather API, and I used this data source to get the daily weather for the cities I got from the first data source.

```
-----Weather Info-----
   Area Day_Weather Night_Weather Max_tmp Min_tmp        Date
0    重庆          小雨            多云      13       9  2021-11-29
1    上海          小雨            小雨      18      10  2021-11-29
2    北京          小雨            多云       5      -1  2021-11-29
3    武汉          多云             晴      18       4  2021-11-29
4    西安          小雨             晴      11       0  2021-11-29
5    成都           阴            多云      12       6  2021-11-29
6    广州          多云            多云      25      15  2021-11-29
7    贵阳          小雨             阴      10       5  2021-11-29
8    天津         雨夹雪             阴       9       1  2021-11-29
9    杭州          小雨            小雨      16      10  2021-11-29
10   昆明          多云            小雨      17       5  2021-11-29
11   南京          多云            多云      20       8  2021-11-29
12   宁波          小雨            小雨      18      12  2021-11-29
13   苏州        小到中雨          小到中雨      19      11  2021-11-29
14   青岛          小雨            小雨      13       5  2021-11-29
```

In order to show all the data in a more intuitive appearance, I merged these three dataframe together on areas and got the summary dataframe.

```
-----Info Summary-----
   Area  Updated_Today  Now_Confirmed  Suspect          Grade Day_Weather Night_Weather Max_tmp Min_tmp        Date
0    昆明          False              0        0  Not_Available          多云            小雨      17       5  2021-11-29
1    上海           True             72        0  Not_Available          小雨            小雨      18      10  2021-11-29
2    广州          False              0        0          全部低风险          多云            多云      25      15  2021-11-29
3    杭州          False              1        0          全部低风险          小雨            小雨      16      10  2021-11-29
4    宁波          False              0        0  Not_Available          小雨            小雨      18      12  2021-11-29
5    青岛          False              0        0  Not_Available          小雨            小雨      13       5  2021-11-29
6    成都          False             25        0          全部低风险           阴            多云      12       6  2021-11-29
7    天津           True             23        0  Not_Available         雨夹雪             阴       9       1  2021-11-29
8    南京          False              0        0  Not_Available          多云            多云      20       8  2021-11-29
9    苏州          False              0        0  Not_Available        小到中雨          小到中雨      19      11  2021-11-29
10   北京          False             10        0  Not_Available          小雨            多云       5      -1  2021-11-29
11   西安          False              0        0  Not_Available          小雨             晴      11       0  2021-11-29
12   重庆           True              5        0  Not_Available          小雨            多云      13       9  2021-11-29
13   武汉          False              0        0          全部低风险          多云             晴      18       4  2021-11-29
14   贵阳          False              0        0  Not_Available          小雨             阴      10       5  2021-11-29
```

Project Analysis

For this part, I decided to use the summary dataframe for analysis. In total, I got 15 cities from the website. There are only 4 cities ("广州", "杭州", "成都", "武汉") have covid grade level which are low-risk areas ("全部低风险"). Thus these cities are safe to travel currently. For another 11 cities, I infer they're safe to travel base on both the numbers showing on the Now_Confirmed and Suspect columns are 0, which means no one is currently infected with covid. Therefore, there are 7 safe cities include "昆明", "宁波", "青岛", "南京", "苏州", "西安", "贵阳". The next step is making the final decision based on the weather condition on those cities. I want to travel to the warm places where the minimum temperature is greater than or equal to 10 degree. In summary, the safe and warm cities are "广州", "杭州", "宁波", "苏州".

Conclusion

In conclusion, there are 4 cities which are both safe (showing "全部低风险" in Grade column or "0" in both Now_Confirmed & Suspect columns) and warm enough for traveling (Min_tmp >= 10 degree). Because we will have more than two weeks break during the Spring Festival, we will choose two or three cities for traveling.

```
-----Info Summary-----
   Area  Updated_Today  Now_Confirmed  Suspect          Grade Day_Weather Night_Weather Max_tmp Min_tmp        Date
0    昆明          False              0        0  Not_Available          多云            小雨      17       5  2021-11-29
1    上海           True             72        0  Not_Available          小雨            小雨      18      10  2021-11-29
2    广州          False              0        0         全部低风险          多云            多云      25      15  2021-11-29
3    杭州          False              1        0         全部低风险          小雨            小雨      16      10  2021-11-29
4    宁波          False              0        0  Not_Available          小雨            小雨      18      12  2021-11-29
5    青岛          False              0        0  Not_Available          小雨            小雨      13       5  2021-11-29
6    成都          False             25        0         全部低风险           阴            多云      12       6  2021-11-29
7    天津           True             23        0  Not_Available         雨夹雪             阴       9       1  2021-11-29
8    南京          False              0        0  Not_Available          多云            多云      20       8  2021-11-29
9    苏州          False              0        0  Not_Available         小到中雨           小到中雨      19      11  2021-11-29
10   北京          False             10        0  Not_Available          小雨            多云       5      -1  2021-11-29
11   西安          False              0        0  Not_Available          小雨             晴      11       0  2021-11-29
12   重庆           True              5        0  Not_Available          小雨            多云      13       9  2021-11-29
13   武汉          False              0        0         全部低风险          多云             晴      18       4  2021-11-29
14   贵阳          False              0        0  Not_Available          小雨             阴      10       5  2021-11-29
```

Reference

1. MAIGOO. 2020年中国十大热门旅游城市 国内人气最旺十大旅游城市

https://www.maigoo.com/goomai/173274.html

2. Tencent. “新冠全国疫情”.

https://www.free-api.com/doc/438

3. 京东万象. 中国和世界天气预报.

Private key：75e6937a5e116e06c6b6c7cd1980dce0

https://way.jd.com/he/freeweather?city=+city&appkey=personal key