



華東師範大學

EAST CHINA NORMAL UNIVERSITY

华东师范大学数据科学与工程学院

数据科学与工程导论

期末大作业项目报告

课程名称：数据科学与工程导论

年级：2023 级

姓名：王志鸿

学号：10224507034

时间：2024.12.30

## 目录

一、数据集解释.....	3
1.1 数据集选取解释.....	3
1.2 数据集格式与内容解释.....	3
二、可视化分析.....	4
2.1 患病患者疾病严重程度可视化分析.....	4
2.2 各年龄类别下受 Covid-19 影响的男性和女性情况可视化分析.....	5
2.2.1 各年龄段男性和女性患者数量对比.....	6
2.2.2 整体趋势和结论.....	7
2.3 各属性之间相关性分析.....	7
2.3.1 症状之间的相关性.....	8
2.3.2 症状与年龄的相关性.....	8
2.3.3 症状与性别相关性.....	9
2.3.4 症状与病情严重程度的相关性.....	9
2.3.5 接触情况与其他变量的相关性.....	9
2.3.6 整体结论.....	9
2.4 各国受 Covid 影响人数可视化分析.....	10
2.4.1 各国受影响人数对比.....	10
2.4.2 整体结论与启示.....	11
2.5 不同国别与发热症状可视化分析.....	11
2.5.1 各国发热与无发热患者数量对比.....	12
2.5.2 整体结论与启示.....	13
2.6 出现症状国家与无症状国家的可视化分析.....	14
2.6.1 各国有症状与无症状患者数量对比.....	14
2.6.2 整体结论与启示.....	15
2.7 症状组合与病情严重程度的关系.....	15
2.7.1 Fever + Cough 组合, Cough + Difficulty-in-Breathing 组合, Fever + Difficulty-in-Breathing 组合 与 轻度 病情严重程度可视化分析 .....	16
2.7.2 Fever + Cough 组合, Cough + Difficulty-in-Breathing 组合, Fever + Difficulty-in-Breathing 组合 与 中度 病情严重程度可视化分析 .....	20
2.7.3 Fever + Cough 组合, Cough + Difficulty-in-Breathing 组合, Fever + Difficulty-in-Breathing 组合 与 重度 病情严重程度可视化分析 .....	24
三、总结.....	28

# 一、数据集解释

## 1.1 数据集选取解释

在众多与数据集中，我选择本 COVID-19 数据集进行研究主要基于以下原因。

尽管当前已处于后疫情时代，但新冠疫情给全球带来的巨大冲击仍历历在目，其复杂性和特殊性为未来公共卫生防控提供了宝贵的经验教训。

该数据集涵盖了丰富多样的关键信息，包含患者的疾病严重程度、年龄、性别、症状表现以及接触情况等多维度数据。这些详细信息全面反映了疫情在不同人群中的传播特点和影响差异，对于深入剖析疫情传播规律具有不可替代的作用。例如，通过分析不同年龄和性别的患者数据，可以清晰地揭示出特定人群在疫情中的易感性和感染后的症状差异，为针对性防控提供精准依据。

从疫情防控角度看，对本数据集的研究能够为未来可能出现的类似疫情防控提供直接且实用的参考。深入分析数据集中各国受影响人数及症状分布情况，有助于制定更具科学性和适应性的防控策略。

在医疗研究方面，数据集中症状之间以及症状与病情严重程度的相关性等信息，为医学研究人员深入探究疾病机制打开了新的窗口。通过研究这些相关性，能够帮助科研人员更好地理解新冠病毒在人体内的致病过程，进而为开发更有效的诊断方法、治疗药物和疫苗提供有力的数据支持。例如，发现某些症状组合与病情严重程度存在特定关联，可引导科研人员进一步研究其背后的生物学机制，为精准医疗奠定基础。

对于资源分配而言，分析患者疾病严重程度分布以及不同症状和人群的资源需求差异，能够使医疗资源在未来疫情中得到更合理、高效的配置。在疫情高峰期，确保医疗资源精准投向最需要的患者群体和医疗环节，避免资源浪费或短缺现象的发生，保障医疗系统的稳定运行。

## 1.2 数据集格式与内容解释

COVID-19 - symptom Checker 数据集的统计分析可以用表格形式表示，如下所示：

Fever	Tiredness	Dr-Cough	Difficulty	Sore-Throat	Mona_S	SymPains	Nasal-Cons	Runny-Nose	Diarrhea	Mona_Escape	Age_0-9	Age_10-19	Age_20-24	Age_25-29	Age_30-39	Age_40-49	Gender_F	Gender_M	Gender_T	Severity_1	Severity_2	Severity_3	Severity_4	Severity_5	Contact_1	Contact_2	Contact_3	Contact_4	Contact_5	Country					
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Iran				
0	1	1	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	Italy			
0	0	0	0	1	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Italy			
0	0	0	0	1	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Other			
1	1	1	1	1	0	0	0	0	0	0	0	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	Italy		
0	0	0	0	1	1	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	Other-EUR		
0	0	0	0	1	1	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	Other-EUR	
0	1	1	1	0	0	0	1	1	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	Germany	
1	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	Republic of Korean	
0	0	0	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	Italy
0	0	0	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	China
0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	Italy

除了 Range、Mean 和 Mode 等统计度量外，我还选择了 Standard Deviation 和 Median 这两个度量来分析数据集。指标说明如下：

- 范围：数据集的范围是其最大值和最小值之间的差值。
- 平均值：数据集的平均值，或算术平均值，是所有值的
- 模式：出现频率最高的号码，即出现频率最高的号码。总和除以值的总数。
- 标准差：它告诉我们提供的值与数据集中存在的属性的平均值相差多少。
- 中位数是数据集中间的值，这意味着 50%的数据点的值小于或等于中位数，50%的

数据点的值大于或等于中位数。

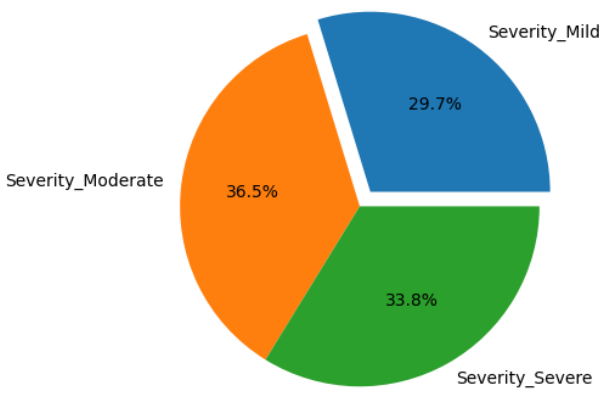
由于数据集中除了国家之外的所有属性值都是 0 或 1，表示一个人是否受到影响，因此对于症状列表，这是 False：“0”或 True：“1”，并且还表示年龄类别、性别和联系方式下的年龄，因此所有这些属性的范围为[1, 0]。对于所提供的数据集，疲倦、呼吸困难属性的标准差和平均值之间的差值接近于零。对于“严重轻度”、“严重中度”、“严重无”和“严重严重”属性，标准差与平均值的差值相同。最常见的数字是一些属性的模式，如干咳，喉咙痛，无症状，疼痛与中位数相同。

## 二、可视化分析

### 2.1 患病患者疾病严重程度可视化分析

下面的饼状图描述了数据集中患者的严重程度。严重性轻度、严重性中度和严重性严重是为此可视化评估的属性。在这种情况下，整个饼状图代表 100%，每种严重程度的比例用不同的颜色来描述，各自的百分比显示了给定个体的各种分类系列。

Pie Chart for visualizing the severity of the patients present in the dataset



从图 2 中，我们可以看见患者的具体患病严重程度：

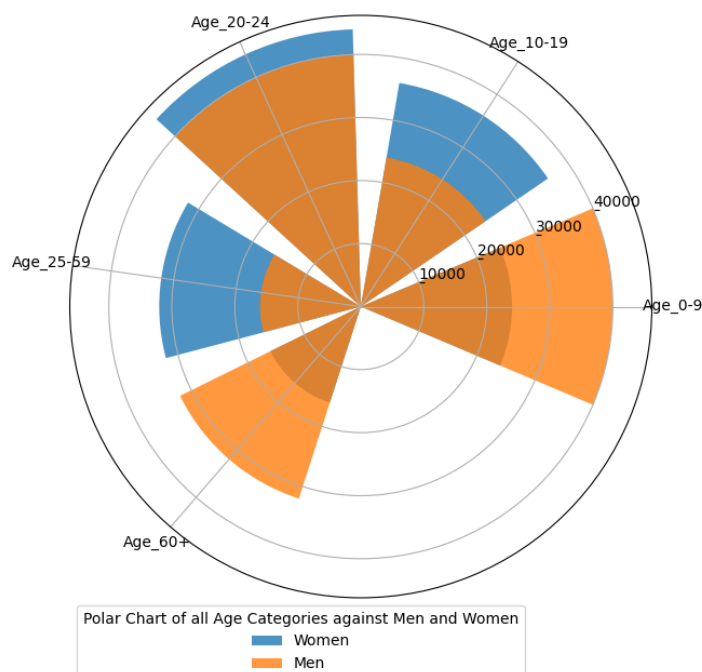
- **严重性中度 (Severity\_Moderate)**：占比最高，达到了 36.5%。这意味着在整个数据集中，有超过三分之一的患者病情处于中度状态。中度病情可能需要一定的医疗干预和关注，但患者的状况相对稳定，尚未发展到非常严重的程度。
- **严重性严重 (Severity\_Severe)**：占比为 33.8%，接近三分之一。这部分患者的病情较为严重，可能需要更密切的医疗监护和更积极的治疗措施，以防止病情进一步恶化，对医疗资源的需求也相对较大。

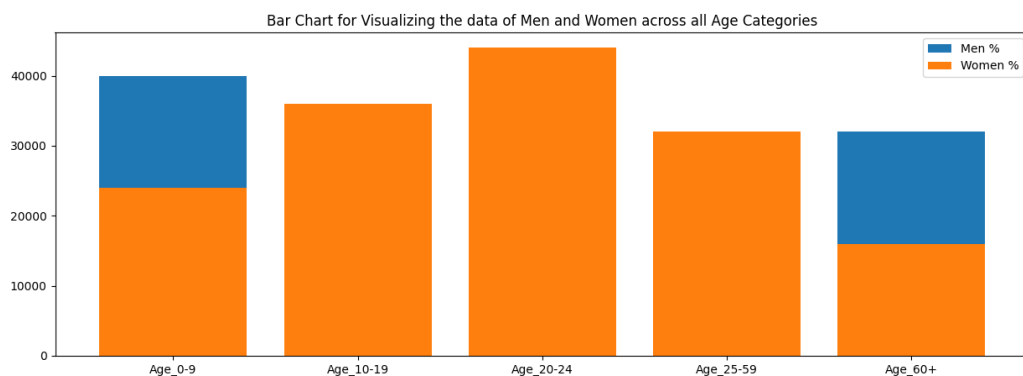
- **严重性轻度 (Severity\_Mild)**：占比 29.7%，是三种严重程度中占比最低的。这表明有接近三成的患者病情较轻，可能只需要一些基本的治疗和护理，身体能够较快恢复，对医疗资源的占用相对较少。

总体来看，数据集中患者的病情分布较为均匀，中度和重度患者的比例相对较高，这可能反映了当时疫情的严峻程度，对医疗系统的压力较大，需要合理分配医疗资源来应对不同严重程度的患者。同时，轻度患者也占有一定比例，说明疫情在不同个体身上的表现存在差异。这种分布情况对于医疗资源的规划、疫情防控策略的制定以及对患者的分类管理都具有重要的参考价值。

## 2.2 各年龄类别下受 Covid-19 影响的男性和女性情况可视化分析

下面的极坐标图以及条形统计图显示了每个年龄类别下受 Covid-19 影响的男性和女性人数。对于这个可视化，使用了以下属性：Age 0-9、Age 10-19、Age 20-24、Age 25-59、Age 60+、Gender Female、Gender Male。整个极地图代表了 100%的男性和女性，其中每个类别都用不同的颜色描绘，每个百分比值代表了所有年龄段的数字。





## 2.2.1 各年龄段男性和女性患者数量对比

### Age 0 - 9 (0 - 9 岁)

1. 男性（橙色）：在该年龄段中，男性患者数量相对较多，占据了较大的面积，从图中可以看出其数量可能接近或超过 40000。

2. 女性（蓝色）：女性患者数量明显少于男性，所占面积较小，可能在 24000 左右或更少。

0 - 9 岁：男性患者数量占据明显优势，远高于女性患者数量，呈现出男性高发的特点。

### Age 10 - 19 (10 - 19 岁)

1. 男性（橙色）：男性患者数量较少，面积小于女性，数量可能在 25000 左右。

2. 女性（蓝色）：女性患者数量相对较多，面积约为男性的 1.5 倍，数量可能在 35000 左右。

10 - 19 岁：女性患者数量反超男性，且数量差距较大，女性在该年龄段受影响程度相对更高。

### Age 20 - 24 (20 - 24 岁)

1. 男性（橙色）：男性患者数量较少，面积较小，数量可能在 45000 左右。

2. 女性（蓝色）：女性患者数量占据优势，数量可能在 43000 左右。

20 - 24 岁：男性和女性患者数量较为接近，但男性患者数量略高于女性，不过差距不大。

### Age 25 - 59 (25 - 59 岁)

1. 男性（橙色）：男性患者数量相对较少，面积较小，数量可能在 17000 左右。

2. 女性（蓝色）：女性患者数量较多，面积约为男性的两倍，数量可能在 32000 左右。

25 - 59 岁：女性患者数量显著多于男性，几乎是男性患者数量的两倍，显示出该年龄段女性受感染的情况更为突出。

### Age 60 + (60 岁及以上)

1. 男性（橙色）：男性患者数量较多，面积较大，数量可能在 30000 左右。

2. 女性（蓝色）：女性患者数量相对较少，面积约为男性的二分之一，数量可能在 15000 左右。

60 岁及以上：男性患者数量多于女性，且差距较为明显，男性在老年阶段受影响

相对更大。

## 2.2.2 整体趋势和结论

**性别差异影响显著：**不同年龄段中，男性和女性受 Covid-19 影响的程度存在明显差异，且这种差异在各年龄段的表现不同，说明性别因素可能与年龄等其他因素相互作用，共同影响感染情况。例如，在儿童阶段（0-9 岁）男性更易感染，而在青少年（10-19 岁）和中年（25-59 岁）阶段女性感染人数相对更多，老年阶段（60 岁及以上）又转变为男性感染人数较多。

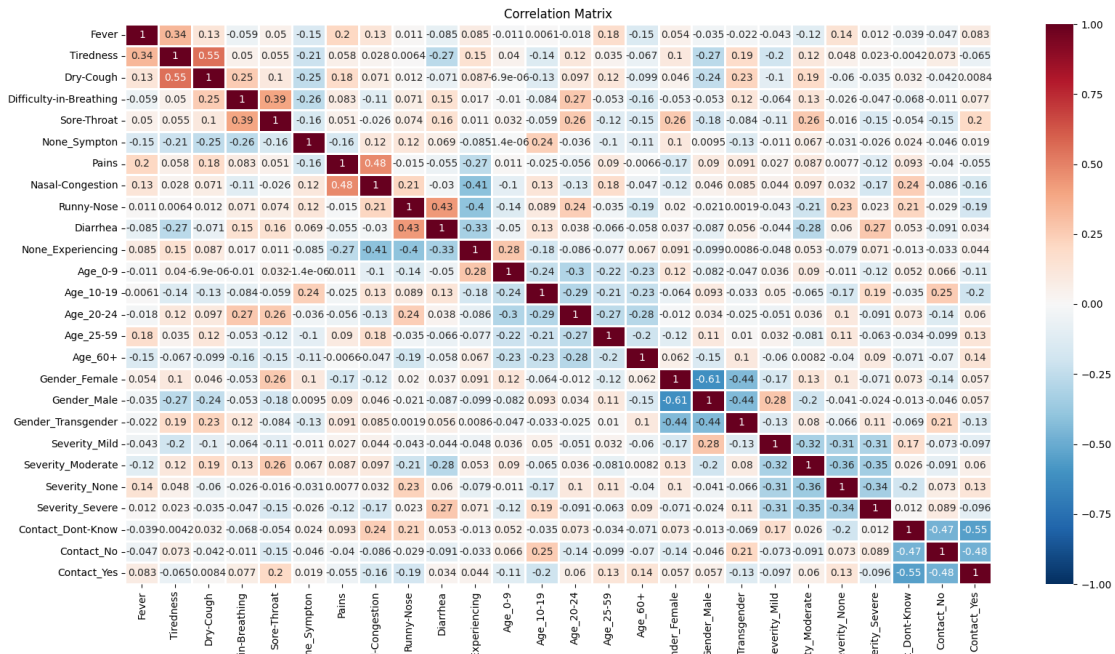
**防控措施需针对性：**基于这种趋势，疫情防控措施应根据不同年龄段的性别差异进行针对性调整。对于 0-9 岁男性高发的情况，需关注该年龄段男性的生活环境、活动场所等，加强相关防控；10-19 岁和 25-59 岁女性高发阶段，可针对女性在这些年龄段的常见活动场景（如学校、工作场所等）制定更细致的防控策略；对于 60 岁及以上男性相对高发，要考虑老年男性的身体特点和生活习惯，提供更适合的防护指导和医疗服务。

**进一步研究原因：**这些数据只是呈现了现象，要深入理解背后的原因，还需要进一步开展研究，包括不同性别在生理机能、免疫系统、生活方式、社交行为等方面的差异，以及这些因素如何与年龄相互作用影响感染几率，以便更精准地制定防控和干预措施。

该极坐标图和条形统计图为我们提供了一个直观的视角，帮助我们了解不同年龄段男性和女性受 Covid-19 影响的人数分布情况，这对于疫情防控、医疗资源分配以及进一步的医学研究都具有一定的参考价值。

## 2.3 各属性之间相关性分析

下图的相关矩阵展示了在我们的例子中变量或属性之间的关系。所有可能的值对之间的关系以矩阵形式呈现。



### 2.3.1 症状之间的相关性

**发热 (Fever) 与疲劳 (Tiredness) :** 相关系数为 0.34, 呈现出正相关。这意味着当患者出现发热症状时, 往往也更容易感到疲劳, 二者可能存在一定的生理关联, 例如身体在对抗感染时的应激反应会同时导致这两种症状的出现。

**咳嗽 (Dry-Cough) 与疲劳 (Tiredness) :** 相关系数为 0.55, 正相关性较强。说明咳嗽症状严重的患者, 疲劳感也可能更明显, 可能是因为持续咳嗽会消耗身体能量, 导致疲劳。

**呼吸困难 (Difficulty-in-Breathing) 与咳嗽 (Dry-Cough) :** 相关系数为 0.25, 正相关。表明咳嗽症状可能会在一定程度上引发或伴随呼吸困难, 这在呼吸道疾病中较为常见, 咳嗽可能影响呼吸道的通畅程度, 进而导致呼吸困难。

### 2.3.2 症状与年龄的相关性

**年龄 0 - 9 岁 (Age\_0 - 9) 与腹泻 (Diarrhea) :** 相关系数为 -0.05, 负相关。说明在 0 - 9 岁这个年龄段, 腹泻症状相对来说不常见, 可能与该年龄段儿童的消化系统发育特点、饮食等因素有关, 使其在感染新冠病毒后更不容易出现腹泻症状。

**年龄 10 - 19 岁 (Age\_10 - 19) 与流涕 (Runny-Nose) :** 相关系数为 0.089, 正相关。这意味着 10 - 19 岁年龄段的患者, 流涕症状相对突出, 可能与该年龄段人群的免疫系统特点或生活环境等因素有关, 导致流涕症状在这个年龄段的表现比其他年龄段明显。



### 2.3.3 症状与性别相关性

*女性 (Gender\_Female) 与流涕 (Runny-Nose)* : 相关系数为 0.02, 正相关。虽然相关性较弱, 但可能暗示女性患者在流涕症状方面相对男性患者来说可能更显著, 不过这一差异可能受到多种因素影响, 需要进一步研究验证。

*男性 (Gender\_Male) 与疲劳 (Tiredness)* : 相关系数为 -0.27, 负相关。表明男性患者在疲劳症状方面相对女性患者可能较轻, 这可能与男女在生理机能、激素水平等方面的差异有关, 但也不能排除其他因素的干扰。

### 2.3.4 症状与病情严重程度的相关性

*中度病情 (Severity\_Mild) 与鼻塞 (Nasal-Congestion)* : 相关系数为 0.044, 正相关。说明中度病情的患者, 出现了鼻塞症状, 并且可能还有其他因素也在影响病情的严重程度和症状表现。

*无病情 (None\_Sympton) 与呼吸困难 (Difficulty-in-Breathing)* : 相关系数为 -0.026, 负相关。这表明无病情的患者不是有很明显的呼吸困难的状况, 呼吸困难可能是病情严重程度的一个重要指标, 反映了肺部功能受到更严重的影响。

### 2.3.5 接触情况与其他变量的相关性

*有接触史 (Contact\_Yes) 与发热 (Fever)* : 相关系数为 0.083, 正相关。说明有接触史的患者更容易出现发热症状, 接触病毒可能增加了发热的风险, 但相关性较弱, 表明接触史并非发热的唯一决定因素。

*无接触史 (Contact\_No) 与无症状 (None\_Sympton)* : 相关系数为 -0.46, 负相关。这意味着没有接触史的人群中, 无症状的比例相对较低, 可能是因为没有接触病毒, 所以感染的几率较小, 从而表现出无症状的情况较少。

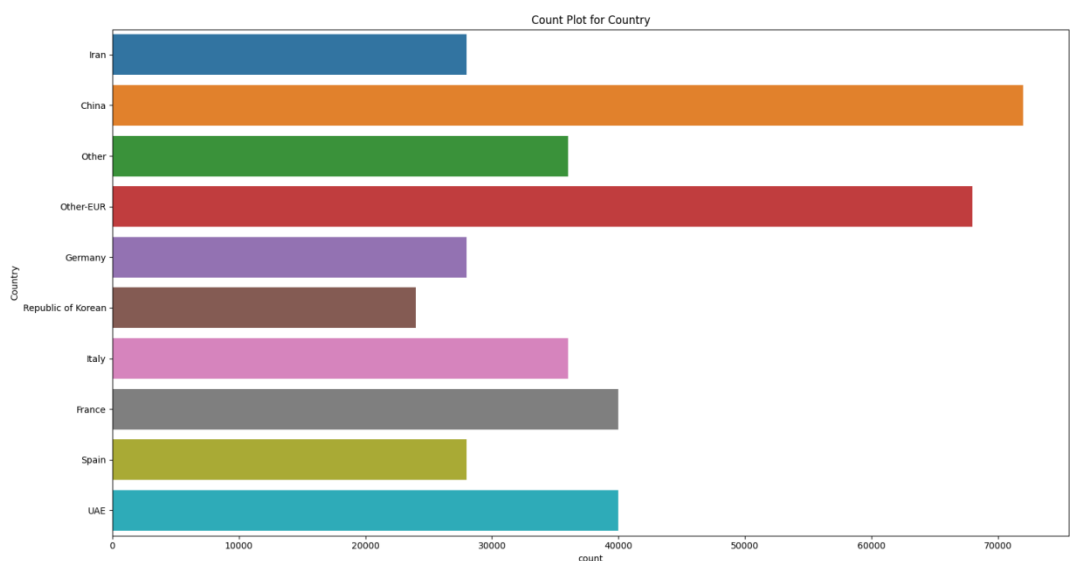
### 2.3.6 整体结论

该相关矩阵揭示了新冠患者各种症状、年龄、性别、病情严重程度以及接触史等因素之间的复杂关系。通过这些相关性, 可以初步了解不同变量之间的关联程度和方向, 为进一步研究新冠疾病的发病机制、症状表现以及风险因素等提供了数据支持和参考方

向。例如，在临床诊断和治疗中，可以根据症状之间的相关性，更全面地评估患者的病情；在疫情防控方面，根据接触史与症状的关系，有针对性地加强对有接触史人群的监测和防护措施等。但是，相关关系并不等同于因果关系，这些结果仅为进一步深入研究提供线索和启示。

## 2.4 各国受 Covid 影响人数可视化分析

该图显示了数据集中每个国家受 Covid 影响的人数。



### 2.4.1 各国受影响人数对比

中国 (*China*)：受 Covid 影响的人数最多，远远超过其他国家，数量 70000 余人。这可能反映了中国人口基数较大，以及在疫情初期受到较大冲击等因素。

其他欧洲国家 (Other - EUR)：受影响人数也较多，接近 65000，仅次于中国。欧洲国家众多，可能由于人口流动频繁、各国防控措施和医疗资源等方面的差异，导致整体受影响人数较高。

法国 (*France*)：受影响人数约为 40000 左右，在图中处于中等偏上水平。法国作为欧洲的重要国家，其人口规模、城市密度以及国际交流等因素可能对疫情传播产生了影响。

意大利 (*Italy*)：受影响人数约为 35000 左右，与法国较为接近。意大利在疫情期间也曾面临较为严峻的形势，可能与当地的人口结构、医疗系统等因素有关。

阿联酋 (*UAE*)：受影响人数约为 40000 左右，与法国相当。阿联酋作为一个国际

交流频繁的国家，可能由于大量的国际人员流动等因素导致了一定数量的感染病例。

*西班牙 (Spain)*：受影响人数约为 30000 左右，处于中等水平。西班牙的人口分布、旅游产业等可能对疫情传播有一定影响。

*德国 (Germany)*：受影响人数约为 30000 左右，与西班牙相近。德国的医疗水平较高，但人口规模较大且国际交流较多，也受到了疫情的影响。

*韩国 (Republic of Korean)*：受影响人数相对较少，约为 25000 左右。韩国在疫情防控方面采取了一些有效的措施，如快速检测和追踪等，可能对控制疫情传播起到了积极作用。

*伊朗 (Iran)*：受影响人数约为 30000 左右，处于中等水平。伊朗的地理位置、人口密度以及医疗资源等因素可能对疫情传播和防控产生了影响。

*其他 (Other)*：受影响人数约为 35000 左右，涵盖了除上述明确列出的国家之外的其他国家和地区，可能由于这些国家和地区的多样性，导致整体受影响人数处于中等水平。

## 2.4.2 整体结论与启示

**整体结论：**从图中可以看出，不同国家受 Covid 影响的人数存在较大差异，中国和一些欧洲国家受影响人数较多，而韩国等国家受影响相对较小。这反映了疫情在全球范围内的传播具有不均衡性，受到各国人口规模、地理环境、防控措施、医疗资源、国际交流等多种因素的综合影响。

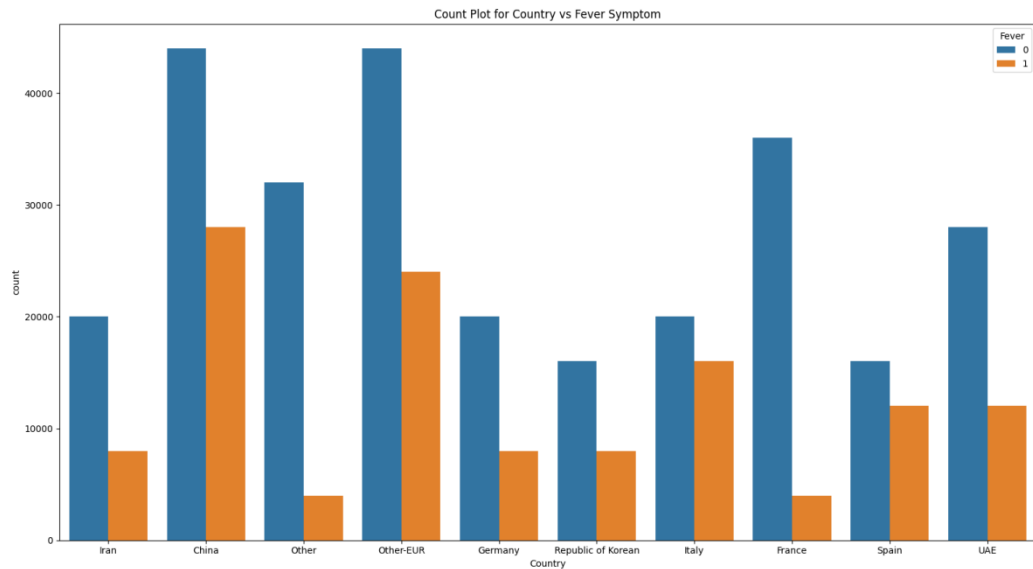
**启示：**对于受影响人数较多的国家，如中国和欧洲一些国家，其防控经验和教训具有重要的参考价值。这些国家在疫情防控中采取的大规模检测、社交隔离、医疗资源调配等措施，可以为其他国家提供借鉴。同时，受影响人数较少的国家，如韩国，其快速响应、精准防控等策略也值得学习和研究。此外，各国之间应加强国际合作与交流，共同应对疫情挑战，包括信息共享、医疗物资援助、疫苗研发与分配等方面，以更好地控制疫情的全球传播，保护全球公共卫生安全。

该国家计数图直观地展示了各国受 Covid 影响的人数分布情况，为进一步分析疫情的全球传播特点和各国防控效果提供了数据支持和参考依据。

## 2.5 不同国别与发热症状可视化分析

这张计数图描绘了数据集中受 Covid-19 影响并以发烧为症状的人数，以及每个国家没有发烧的人数。图表数据说明如下：

- Fever: 0 是受 Covid 影响但没有症状的人数
- Fever: 1 是受 Covid 影响但有症状的人数



### 2.5.1 各国发热与无发热患者数量对比

#### 伊朗 (Iran) :

无发热 (Fever: 0) : 人数约为 20000 左右, 占比较大。

有发热 (Fever: 1) : 人数约为 8000 左右, 明显少于无发热人数。这表明在伊朗, 受 Covid-19 影响但无发热症状的患者相对较多, 可能与当地的病毒株类型、人群体质等因素有关, 使得部分患者感染后发热症状不明显。

#### 中国 (China) :

无发热 (Fever: 0) : 人数超过 40000, 是所有国家中无发热人数最多的。

有发热 (Fever: 1) : 人数约为 28000 左右, 虽然有发热症状的患者数量也较多, 但与无发热人数相比仍有一定差距。这可能反映了中国在疫情防控过程中, 检测范围较广, 包括了大量无症状感染者, 同时也可能与中国人口基数大等因素有关。

#### 其他 (Other) :

无发热 (Fever: 0) : 人数约为 32000 左右。

有发热 (Fever: 1) : 人数较少, 约为 4000 左右。说明在这部分 “其他” 国家中, 无发热症状的患者占比较高, 可能这些国家的疫情特点、防控措施等因素导致了这种情况。

#### 其他欧洲国家 (Other - EUR) :

无发热 (Fever: 0) : 人数接近 45000, 是无发热人数较多的一类。

有发热 (Fever: 1) : 人数约为 24000 左右, 无发热人数远多于有发热人数。欧洲国家众多, 疫情情况复杂, 可能由于各国防控措施、检测力度以及病毒传播特点等因素, 使得无发热症状的患者在统计数据中占比较大。

#### 德国 (Germany) :

无发热 (Fever: 0) : 人数约为 20000 左右。

有发热 (Fever: 1) : 人数约为 8000 左右, 与伊朗情况类似, 无发热患者数量多于有发热患者数量。德国的医疗水平较高, 检测能力较强, 可能检测出了较多的无症状感染者, 同时也可能与当地的疫情防控措施和人群特点有关。

#### **韩国 (Republic of Korean) :**

无发热 (Fever: 0) : 人数约为 16000 左右。

有发热 (Fever: 1) : 人数约为 8000 左右, 无发热人数约为有发热人数的两倍。

韩国在疫情防控方面采取了快速检测和追踪等措施, 可能有助于发现更多的无症状感染者, 从而导致无发热患者数量相对较多。

#### **意大利 (Italy) :**

无发热 (Fever: 0) : 人数约为 20000 左右。

有发热 (Fever: 1) : 人数约为 16000 左右, 无发热人数略多于有发热人数。意大利在疫情期间受到了较大冲击, 可能由于病毒传播较广, 检测到了不同症状表现的患者, 不过整体上无发热患者仍占一定优势。

#### **法国 (France) :**

无发热 (Fever: 0) : 人数约为 36000 左右, 是无发热人数较多的国家之一。

有发热 (Fever: 1) : 人数较少, 约为 4000 左右, 无发热人数远多于有发热人数。

法国的疫情发展和防控措施等因素可能影响了患者的症状表现, 使得无发热症状的患者在统计中占比较高。

#### **西班牙 (Spain) :**

无发热 (Fever: 0) : 人数约为 16000 左右。

有发热 (Fever: 1) : 人数约为 12000 左右, 无发热人数略多于有发热人数。西班牙的疫情情况和防控措施等可能导致了这种患者症状分布情况, 与其他一些欧洲国家类似, 无发热患者占比较高。

#### **阿联酋 (UAE) :**

无发热 (Fever: 0) : 人数约为 28000 左右。

有发热 (Fever: 1) : 人数约为 12000 左右, 无发热人数多于有发热人数。阿联酋作为一个国际交流频繁的国家, 可能由于检测力度、人群特点等因素, 使得无发热症状的患者在统计数据中占比较大。

## **2.5.2 整体结论与启示**

**整体结论:** 从图中可以看出, 在大多数国家中, 受 Covid-19 影响但无发热症状的患者数量多于有发热症状的患者数量。这表明新冠病毒感染后的症状表现具有多样性, 发热并非所有患者的必然症状, 无症状感染者在疫情传播中可能起到了重要作用。

#### **启示:**

**防控措施方面:** 各国在疫情防控中不能仅依赖发热这一症状来筛查患者, 应加强对无症状感染者的检测和追踪, 例如扩大核酸检测范围、加强流行病学调查等, 以更全面地掌握疫情传播情况, 防止疫情扩散。

**公众认知方面:** 公众应认识到即使没有发热症状, 也可能感染了新冠病毒, 因此要保持良好的个人卫生习惯, 如勤洗手、戴口罩、保持社交距离等, 不能因自身无发热症状而放松警惕。

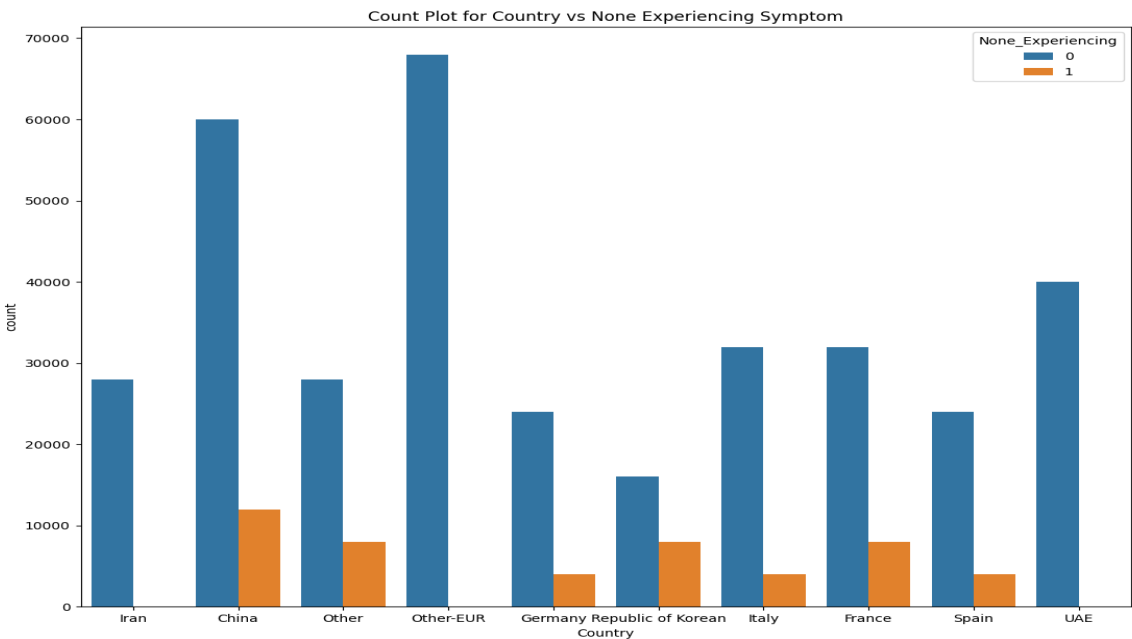
**医疗资源分配方面:** 由于无症状感染者数量较多, 医疗机构在资源分配上需要考虑到这一情况, 例如在隔离设施、检测试剂等方面做好相应准备, 同时也要关注无症状感染者可能出现的病情变化, 及时提供医疗支持。

该计数图为我们提供了不同国家患者发热症状的分布情况,有助于更深入地了解新冠疫情在全球范围内的特点和影响,为疫情防控提供了有价值的参考依据。

## 2.6 出现症状国家与无症状国家的可视化分析

此计数图描绘了数据集中受 Covid-19 影响但未出现任何症状的人数,并绘制了每个国家的情况。图表数据说明如下:

- None Experiencing: 0 是受 Covid 影响并出现症状的人数
- None Experiencing: 1 是受新冠病毒影响但未出现任何症状的人数



### 2.6.1 各国有症状与无症状患者数量对比

#### 伊朗 (Iran) :

有症状 (None Experiencing: 0) : 人数约为 28000 左右。有症状患者数量明显多于无症状患者数量,说明在伊朗受新冠病毒影响的人群中,出现症状的情况相对更为普遍,可能与当地的病毒传播特点、人群易感性等因素有关。

无症状 (None Experiencing: 1) : 人数为 0。可能是当地未做好该数据的检测工作

#### 中国 (China) :

有症状 (None Experiencing: 0) : 人数约为 60000 左右,是所有国家中有症状人数最多的。

无症状 (None Experiencing: 1) : 人数约为 12000 左右。中国人口基数大, 疫情初期受到较大冲击, 虽然有症状患者数量较多, 但也检测出了一定数量的无症状感染者, 这可能与中国广泛的检测范围和防控措施有关, 能够及时发现不同症状表现的患者。

**其他 (Other) :**

有症状 (None Experiencing: 0) : 人数约为 28000 左右。

无症状 (None Experiencing: 1) : 人数约为 8000 左右。在这部分 “其他” 国家中, 有症状患者数量多于无症状患者数量, 可能这些国家的疫情情况、防控力度等因素导致了这种症状分布差异。

**其他欧洲国家 (Other - EUR) :**

有症状 (None Experiencing: 0) : 人数约为 68000 左右。

无症状 (None Experiencing: 1) : 人数为 0, 同伊朗。

**德国 (Germany) :**

有症状 (None Experiencing: 0) : 人数约为 24000 左右。

无症状 (None Experiencing: 1) : 人数约为 4000 左右。德国有症状患者数量多于无症状患者数量, 不过两者差距相对较小, 可能与德国的医疗水平、检测能力以及防控措施等因素有关, 能够检测出一定比例的无症状感染者, 但整体上有症状患者仍占比较大。

**韩国 (Republic of Korean) :**

有症状 (None Experiencing: 0) : 人数约为 16000 左右。

无症状 (None Experiencing: 1) : 人数约为 6000 左右。韩国有症状患者数量多于无症状患者数量, 可能与韩国在疫情防控中采取的快速检测和追踪等措施有关, 这些措施有助于及时发现有症状患者, 但也检测出了一定数量的无症状感染者, 不过整体上有症状患者占主导。

**意大利 (Italy) :**

有症状 (None Experiencing: 0) : 人数约为 32000 左右。

无症状 (None Experiencing: 1) : 人数约为 8000 左右。意大利有症状患者数量明显多于无症状患者数量, 意大利在疫情期间受到了较大冲击, 可能由于病毒传播较广, 有症状患者更容易被发现和统计, 而无症状感染者在检测和统计中相对占比较小。

**法国 (France) :**

有症状 (None Experiencing: 0) : 人数约为 32000 左右。

无症状 (None Experiencing: 1) : 人数约为 8000 左右。与意大利类似, 法国有症状患者数量远多于无症状患者数量, 可能与法国的疫情发展和防控措施等因素有关, 导致有症状患者在统计数据中占比较高。

**西班牙 (Spain) :**

有症状 (None Experiencing: 0) : 人数约为 24000 左右。

无症状 (None Experiencing: 1) : 人数约为 4000 左右。西班牙有症状患者数量多于无症状患者数量, 可能与西班牙的疫情情况和防控措施等因素相关, 使得有症状患者更容易被识别和统计, 而无症状感染者相对较少。

**阿联酋 (UAE) :**

有症状 (None Experiencing: 0) : 人数约为 40000 左右。

无症状 (None Experiencing: 1) : 人数为 0 , 同伊朗

## 2.6.2 整体结论与启示

**整体结论：**从图中可以看出，各国受新冠病毒影响的人群中，有症状患者和无症状患者的数量分布存在差异。部分国家如中国、伊朗、意大利、法国、西班牙等，有症状患者数量较多；而“其他欧洲国家”无症状患者数量相对突出。这表明新冠病毒感染后的症状表现具有多样性，且不同国家的疫情特点和防控措施等因素会影响无症状感染者的检测和统计情况。

**启示：**

**防控策略方面：**各国应重视无症状感染者在疫情传播中的作用，不能仅依据症状来判断疫情传播情况。在防控措施上，要加强对无症状感染者的筛查和管理，例如在入境检测、社区筛查等环节扩大检测范围，采取有效的隔离和追踪措施，防止无症状感染者成为疫情传播的潜在源头。

**公众意识方面：**公众需要认识到无症状感染者的存在，即使自身没有明显症状，也可能感染病毒并传播给他人。因此，要继续保持良好的个人卫生习惯和社交距离，遵循防疫指南，共同维护公共卫生安全。

**医疗资源分配方面：**医疗机构在资源分配上要考虑到无症状感染者的数量和需求，例如在隔离设施、检测试剂、医护人员配备等方面做好相应准备，同时要加强对无症状感染者的监测和随访，以便及时发现病情变化并提供适当的医疗支持。

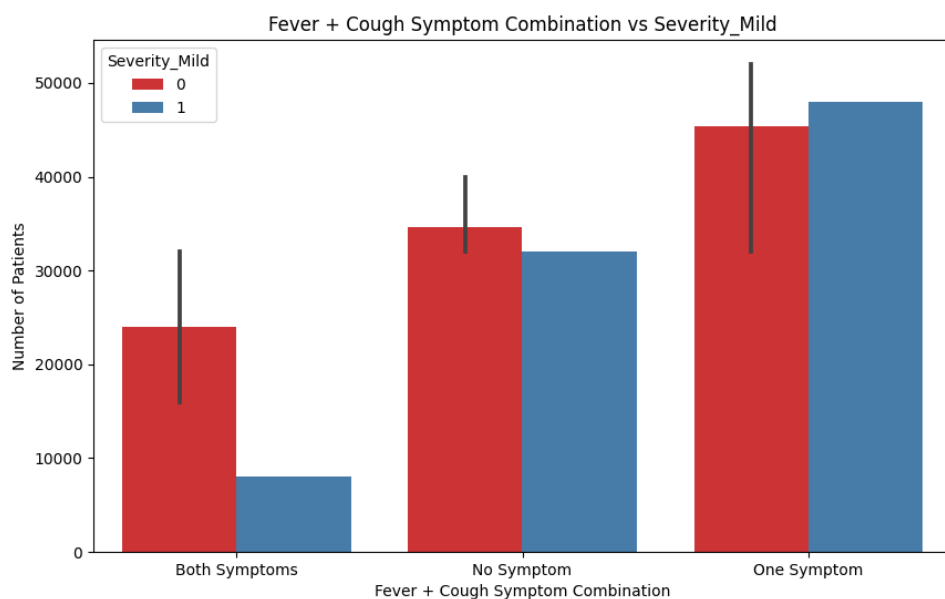
该计数图为我们提供了不同国家有症状和无症状感染者的分布情况，有助于更全面地了解新冠疫情在全球范围内的特点和影响，为各国制定和调整疫情防控策略提供了重要的数据参考和依据。

## 2.7 症状组合与病情严重程度的关系

不同症状的组合（如发热 + 咳嗽、咳嗽 + 呼吸困难等）与病情严重程度（轻度、中度、重度）之间是否存在特定的关联模式？通过分析这些模式，可以更好地预测患者的病情发展，为临床治疗提供更精准的依据。

### 2.7.1 Fever + Cough 组合，Cough + Difficulty-in-Breathing 组合，Fever + Difficulty-in-Breathing 组合 与 轻度 病情严重程度可视化分析



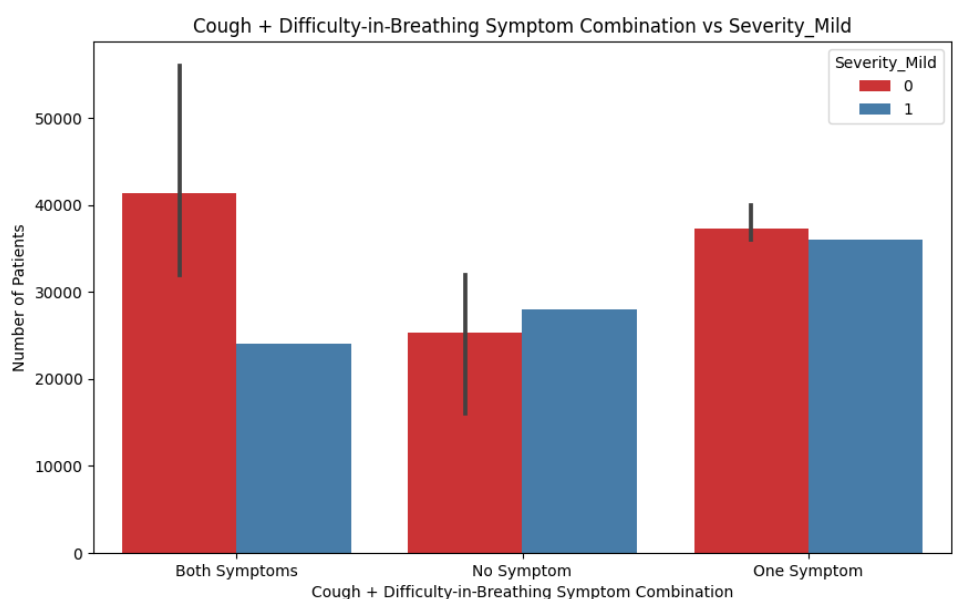


图一: Fever + Cough Symptom Combination vs Severity\_Mild

*Both Symptoms* (发热和咳嗽都有): Severity\_Mild 为 0 (可能表示非轻度病情) 的患者数量约为 24000, Severity\_Mild 为 1 (可能表示轻度病情) 的患者数量约为 8000。可以看出,同时有发热和咳嗽症状的患者中,非轻度病情的患者数量明显多于轻度病情的患者。

*No Symptom* (无发热和咳嗽症状): Severity\_Mild 为 0 的患者数量约为 34000, Severity\_Mild 为 1 的患者数量约为 32000。此症状组合下,非轻度病情和轻度病情的患者数量较为接近。

*One Symptom* (只有发热或咳嗽其中一种症状): Severity\_Mild 为 0 的患者数量约为 45000, Severity\_Mild 为 1 的患者数量约为 48000。该症状组合下,轻度病情的患者数量略多于非轻度病情的患者,但差距不大。

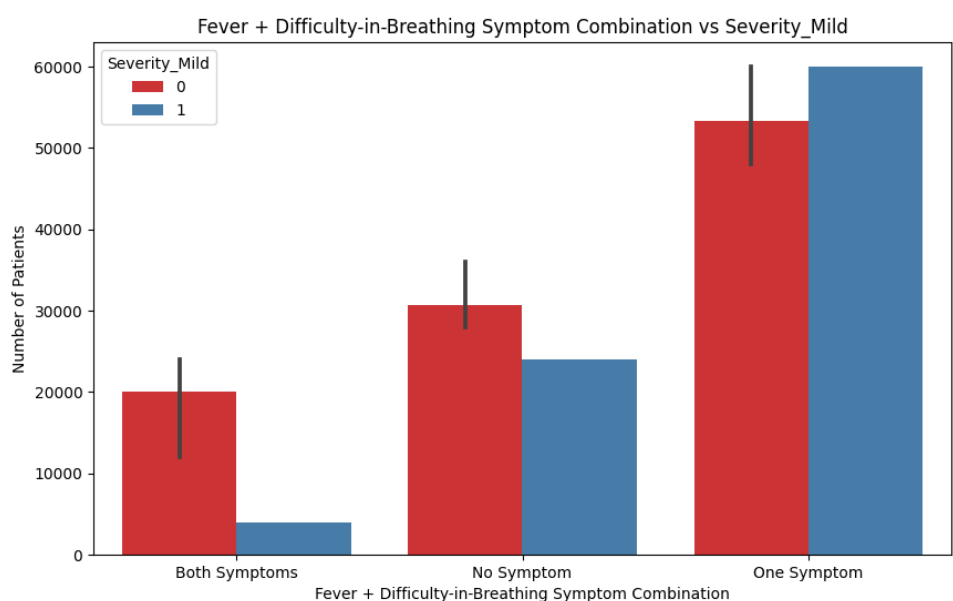


图二: Cough + Difficulty-in-Breathing Symptom Combination vs Severity\_Mild

*Both Symptoms* (咳嗽和呼吸困难都有): Severity\_Mild 为 0 的患者数量约为 41000, Severity\_Mild 为 1 的患者数量约为 24000。同时有咳嗽和呼吸困难症状的患者中, 非轻度病情的患者数量远多于轻度病情的患者。

*No Symptom* (无咳嗽和呼吸困难症状): Severity\_Mild 为 0 的患者数量约为 25000, Severity\_Mild 为 1 的患者数量约为 28000。此症状组合下, 轻度病情的患者数量略多于非轻度病情的患者, 但差距较小。

*One Symptom* (只有咳嗽或呼吸困难其中一种症状): Severity\_Mild 为 0 的患者数量约为 37000, Severity\_Mild 为 1 的患者数量约为 36000。该症状组合下, 非轻度病情和轻度病情的患者数量非常接近。



图三: Fever + Difficulty-in-Breathing Symptom Combination vs Severity\_Mild

*Both Symptoms* (发热和呼吸困难都有): Severity\_Mild 为 0 的患者数量约为 20000, Severity\_Mild 为 1 的患者数量约为 4000。同时有发热和呼吸困难症状的患者中,非轻度病情的患者数量大幅多于轻度病情的患者。

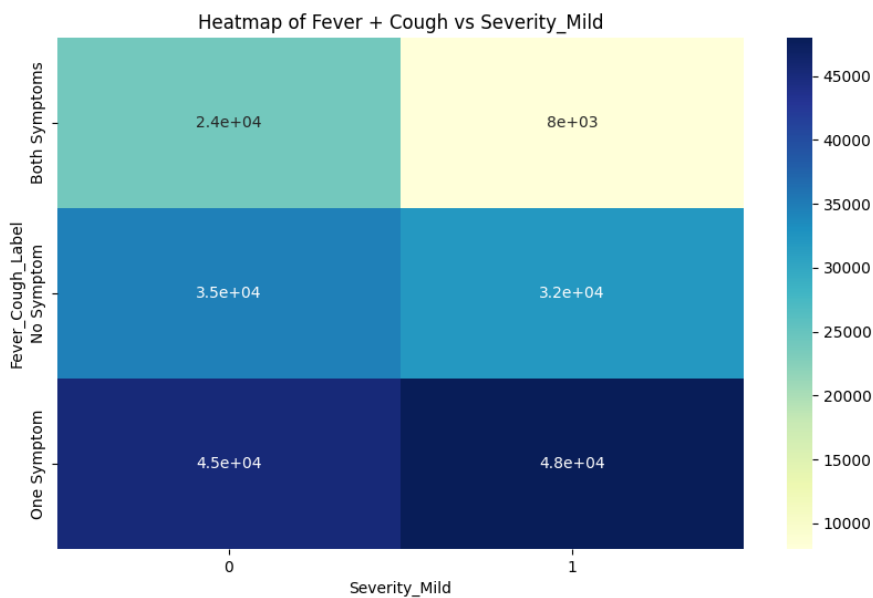
*No Symptom* (无发热和呼吸困难症状): Severity\_Mild 为 0 的患者数量约为 30000, Severity\_Mild 为 1 的患者数量约为 24000。此症状组合下,非轻度病情的患者数量略多于轻度病情的患者。

*One Symptom* (只有发热或呼吸困难其中一种症状): Severity\_Mild 为 0 的患者数量约为 53000, Severity\_Mild 为 1 的患者数量约为 60000。该症状组合下,轻度病情的患者数量稍多于非轻度病情的患者。

整体结论

从这三张图可以看出,不同的症状组合与轻度病情 (Severity\_Mild) 之间存在一定的关联。一般来说,同时出现两种症状 (如发热和咳嗽、咳嗽和呼吸困难、发热和呼吸困难) 的患者中,非轻度病情的患者数量相对较多;而只有一种症状或无相关症状的患者中,轻度病情和非轻度病情的患者数量分布相对较为均衡或差距较小。这表明某些症状组合可能与病情的严重程度有一定的相关性,在临床诊断和病情评估中,这些症状组合可以作为参考因素之一,但需要注意的是,这只是基于当前数据的分析,还需要结合更多的临床信息和研究来进一步验证和深入理解这种关系。

为此,我们用这三种情况与轻度病情相结合绘制热力图:



症状组合与轻度病情 (Severity\_Mild) 的关系

**Both Symptoms (发热和咳嗽都有):**

当 Severity\_Mild 为 0 (可能表示非轻度病情) 时,患者数量约为 (24000)。

当 Severity\_Mild 为 1 (可能表示轻度病情) 时,患者数量约为 (8000)。

可以看出，同时有发热和咳嗽症状的患者中，非轻度病情的患者数量明显多于轻度病情的患者，热力图中该区域颜色相对较浅，对应患者数量较少。

**No Symptom（无发热和咳嗽症状）：**

Severity\_Mild 为 0 的患者数量约为（35000）。

Severity\_Mild 为 1 的患者数量约为（32000）。

此症状组合下，非轻度病情和轻度病情的患者数量较为接近，热力图中该区域颜色相近，表明患者数量差异不大。

**One Symptom（只有发热或咳嗽其中一种症状）：**

Severity\_Mild 为 0 的患者数量约为（45000）。

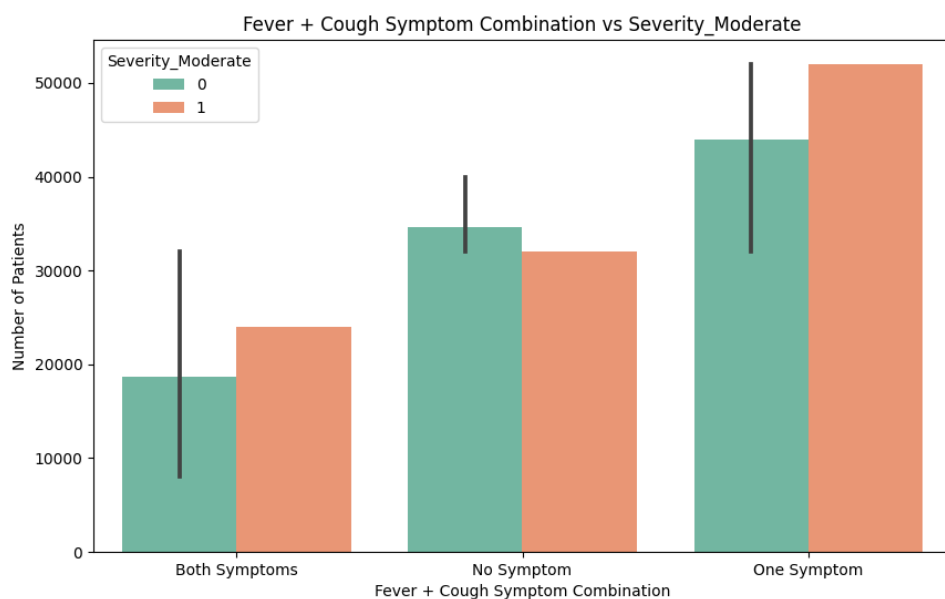
Severity\_Mild 为 1 的患者数量约为（48000）。

该症状组合下，轻度病情的患者数量略多于非轻度病情的患者，热力图中该区域颜色较深，对应患者数量较多，但两者差距不是非常大。

**整体结论**

从热力图可以直观地看出不同症状组合（Both Symptoms、No Symptom、One Symptom）下，轻度病情（Severity\_Mild 为 1）和非轻度病情（Severity\_Mild 为 0）的患者数量分布情况。颜色越深表示患者数量越多，颜色越浅表示患者数量越少。总体而言，同时有发热和咳嗽症状的患者中，非轻度病情占比较大；而只有一种症状或无相关症状的患者中，轻度病情和非轻度病情的患者数量分布相对较为均衡。这与之前柱状图的分析结果相呼应，进一步验证了症状组合与病情严重程度之间的关系，可为临床诊断和病情评估提供一定的参考依据，但仍需结合更多临床数据和研究进行深入分析。

## 2.7.2 Fever + Cough 组合，Cough + Difficulty-in-Breathing 组合，Fever + Difficulty-in-Breathing 组合 与 中度 病情严重程度可视化分析

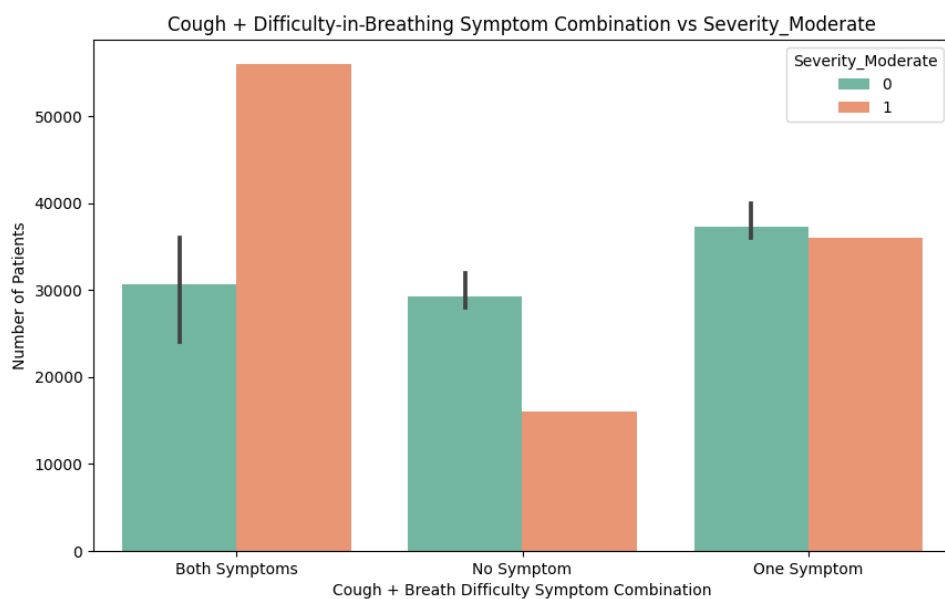


图一: Fever + Cough Symptom Combination vs Severity\_Severe

*Both Symptoms* (发热和咳嗽都有): Severity\_Severe 为 0 (可能表示非重度病情) 的患者数量约为 21000, Severity\_Severe 为 1 (可能表示重度病情) 的患者数量约为 16000。可以看出, 同时有发热和咳嗽症状的患者中, 非重度病情的患者数量略多于重度病情的患者。

*No Symptom* (无发热和咳嗽症状): Severity\_Severe 为 0 的患者数量约为 34000, Severity\_Severe 为 1 的患者数量约为 32000。此症状组合下, 非重度病情和重度病情的患者数量较为接近。

*One Symptom* (只有发热或咳嗽其中一种症状): Severity\_Severe 为 0 的患者数量约为 44000, Severity\_Severe 为 1 的患者数量约为 52000。该症状组合下, 重度病情的患者数量多于非重度病情的患者。

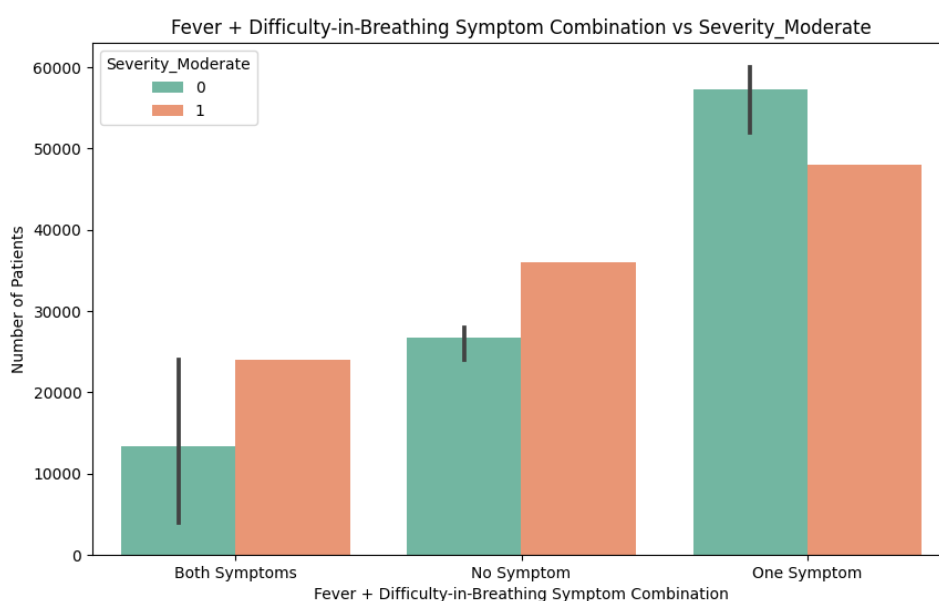


图二: Cough + Difficulty-in-Breathing Symptom Combination vs Severity\_Severe

*Both Symptoms* (咳嗽和呼吸困难都有): Severity\_Severe 为 0 的患者数量约为 39000, Severity\_Severe 为 1 的患者数量约为 32000。同时有咳嗽和呼吸困难症状的患者中, 非重度病情的患者数量多于重度病情的患者。

*No Symptom* (无咳嗽和呼吸困难症状): Severity\_Severe 为 0 的患者数量约为 25000, Severity\_Severe 为 1 的患者数量约为 28000。此症状组合下, 重度病情的患者数量略多于非重度病情的患者, 但差距较小。

*One Symptom* (只有咳嗽或呼吸困难其中一种症状): Severity\_Severe 为 0 的患者数量约为 36000, Severity\_Severe 为 1 的患者数量约为 40000。该症状组合下, 重度病情的患者数量稍多于非重度病情的患者。



图三: Fever + Difficulty-in-Breathing Symptom Combination vs Severity\_Severe

*Both Symptoms* (发热和呼吸困难都有): Severity\_Severe 为 0 的患者数量约为 17000, Severity\_Severe 为 1 的患者数量约为 12000。同时有发热和呼吸困难症状的患者中, 非重度病情的患者数量多于重度病情的患者。

*No Symptom* (无发热和呼吸困难症状): Severity\_Severe 为 0 的患者数量约为 29000, Severity\_Severe 为 1 的患者数量约为 28000。此症状组合下, 非重度病情和重度病情的患者数量非常接近。

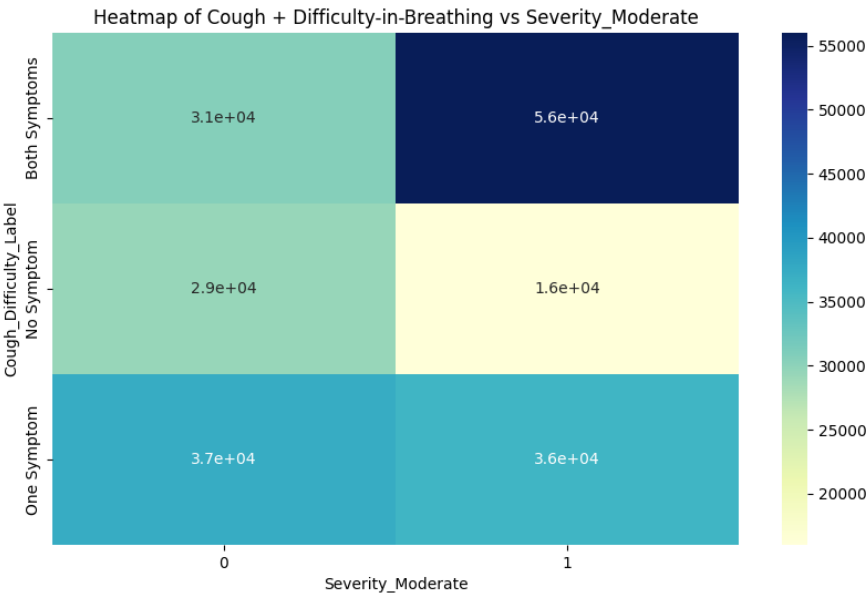
*One Symptom* (只有发热或呼吸困难其中一种症状): Severity\_Severe 为 0 的患者数量约为 53000, Severity\_Severe 为 1 的患者数量约为 60000。该症状组合下, 重度病情的患者数量多于非重度病情的患者。

### 整体结论

从这三张图可以看出, 不同的症状组合与重度病情 (Severity\_Severe) 之间存在一定的关联。一般来说, 同时出现两种症状 (如发热和咳嗽、咳嗽和呼吸困难、发热和呼吸困难) 的患者中, 非重度病情的患者数量相对较多或与重度病情的患者数量较为接近; 而只有一种

症状的患者中，重度病情的患者数量在某些情况下会多于非重度病情的患者。这表明症状组合可能对病情的严重程度有一定的影响，但这种关系并非绝对，在临床诊断和病情评估中，这些症状组合可以作为参考因素之一，但还需要结合更多的临床信息和检查结果来综合判断病情的严重程度。

为此，我们用这三种情况与轻度病情相结合绘制热力图：



症状组合与中度病情（Severity\_Moderate）的关系

**Both Symptoms（咳嗽和呼吸困难都有）：**

当 Severity\_Moderate 为 0（可能表示非中度病情）时，患者数量约为（31000）。

当 Severity\_Moderate 为 1（可能表示中度病情）时，患者数量约为（56000）。

可以看出，同时有咳嗽和呼吸困难症状的患者中，中度病情的患者数量明显多于非中度病情的患者，热力图中该区域颜色较深，对应患者数量较多。

**No Symptom（无咳嗽和呼吸困难症状）：**

Severity\_Moderate 为 0 的患者数量约为（29000）。

Severity\_Moderate 为 1 的患者数量约为（16000）。

此症状组合下，非中度病情的患者数量多于中度病情的患者，热力图中该区域颜色较浅，表明患者数量较少。

**One Symptom（只有咳嗽或呼吸困难其中一种症状）：**

Severity\_Moderate 为 0 的患者数量约为（37000）。

Severity\_Moderate 为 1 的患者数量约为（36000）。

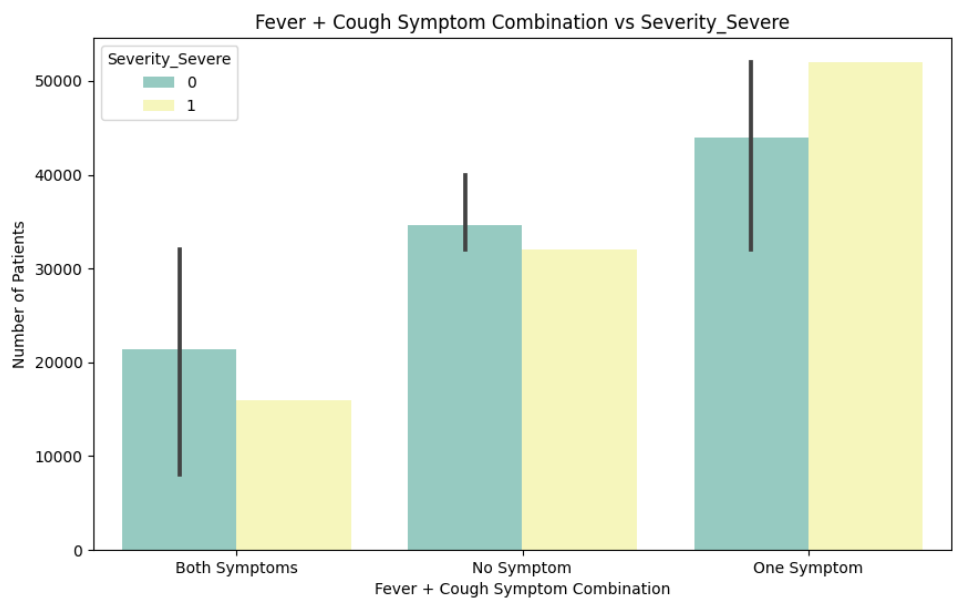
该症状组合下，非中度病情和中度病情的患者数量较为接近，热力图中这两个区域颜色相近，说明患者数量差异不大。

### 整体结论

从热力图可以直观地看出不同症状组合（Both Symptoms、No Symptom、One Symptom）下，中度病情（Severity\_Moderate 为 1）和非中度病情（Severity\_Moderate 为 0）的患

者数量分布情况。颜色越深表示患者数量越多，颜色越浅表示患者数量越少。总体而言，同时有咳嗽和呼吸困难症状的患者中，中度病情占比较大；而无相关症状的患者中，非中度病情占比较大；只有一种症状的患者中，中度病情和非中度病情的患者数量分布相对均衡。这为临床诊断和病情评估提供了一定的参考，提示症状组合可能与病情的中度程度存在一定关联，但仍需结合更多临床数据和研究进行深入分析。

2.7.3 Fever + Cough 组合，Cough + Difficulty-in-Breathing 组合，Fever + Difficulty-in-Breathing 组合 与 重度 病情严重程度可视化分析



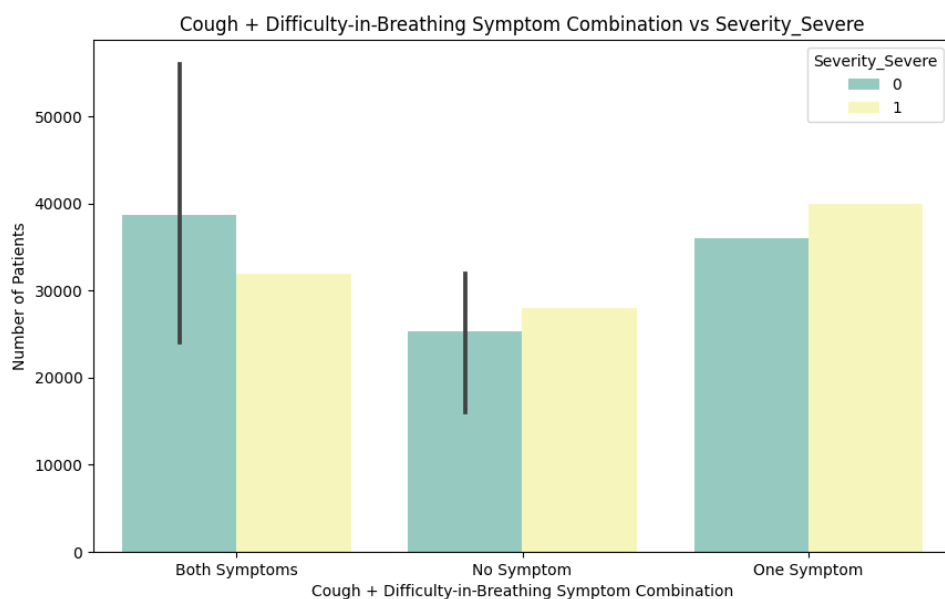
图一：Fever + Cough Symptom Combination vs Severity\_Severe

Both Symptoms（发热和咳嗽都有）：Severity\_Severe 为 0（可能表示非重度病情）的患者数量约为 21000。Severity\_Severe 为 1（可能表示重度病情）的患者数量约为 16000。可以看出，同时有发热和咳嗽症状的患者中，非重度病情的患者数量略多于重度病情的患者。

No Symptom（无发热和咳嗽症状）：Severity\_Severe 为 0 的患者数量约为 34000。Severity\_Severe 为 1 的患者数量约为 32000。此症状组合下，非重度病情和重度病情的患者数量较为接近。

One Symptom（只有发热或咳嗽其中一种症状）：Severity\_Severe 为 0 的患者数量约为 44000。Severity\_Severe 为 1 的患者数量约为 52000。该症状组合下，重度病情的患者数量多于非重度病情的患者。



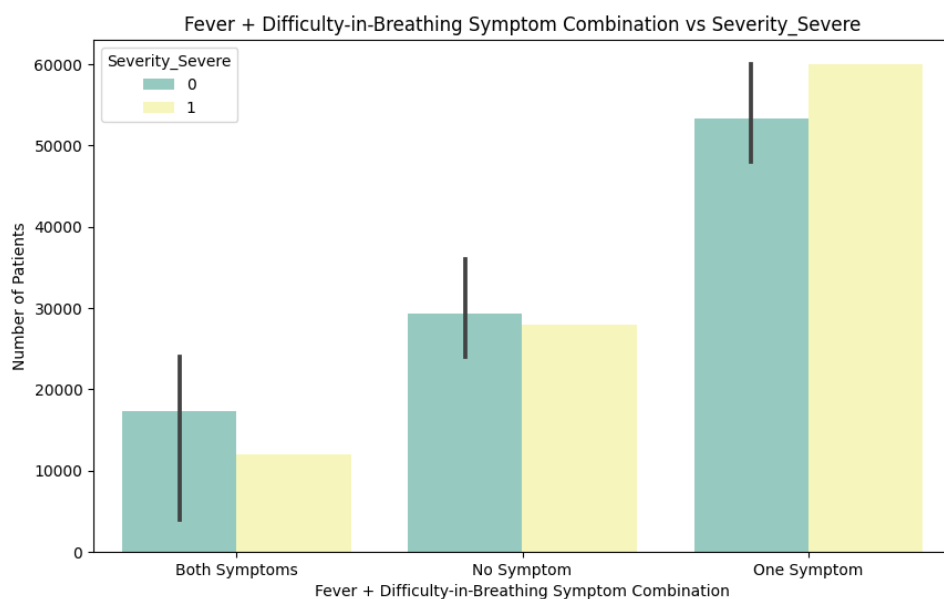


图二: Cough + Difficulty-in-Breathing Symptom Combination vs Severity\_Severe

*Both Symptoms* (咳嗽和呼吸困难都有): Severity\_Severe 为 0 的患者数量约为 39000。Severity\_Severe 为 1 的患者数量约为 32000。同时有咳嗽和呼吸困难症状的患者中,非重度病情的患者数量多于重度病情的患者。

*No Symptom* (无咳嗽和呼吸困难症状): Severity\_Severe 为 0 的患者数量约为 25000。Severity\_Severe 为 1 的患者数量约为 28000。此症状组合下,重度病情的患者数量略多于非重度病情的患者,但差距较小。

*One Symptom* (只有咳嗽或呼吸困难其中一种症状): Severity\_Severe 为 0 的患者数量约为 36000。Severity\_Severe 为 1 的患者数量约为 40000。该症状组合下,重度病情的患者数量稍多于非重度病情的患者。



**图三：Fever + Difficulty-in-Breathing Symptom Combination vs Severity\_Severe**

Both Symptoms(发热和呼吸困难都有):Severity\_Severe 为 0 的患者数量约为 17000。Severity\_Severe 为 1 的患者数量约为 12000。同时有发热和呼吸困难症状的患者中，非重度病情的患者数量多于重度病情的患者。

No Symptom(无发热和呼吸困难症状):Severity\_Severe 为 0 的患者数量约为 29000。Severity\_Severe 为 1 的患者数量约为 28000。此症状组合下，非重度病情和重度病情的患者数量非常接近。

One Symptom (只有发热或呼吸困难其中一种症状): Severity\_Severe 为 0 的患者数量约为 53000。Severity\_Severe 为 1 的患者数量约为 60000。该症状组合下，重度病情的患者数量多于非重度病情的患者。

### 整体结论

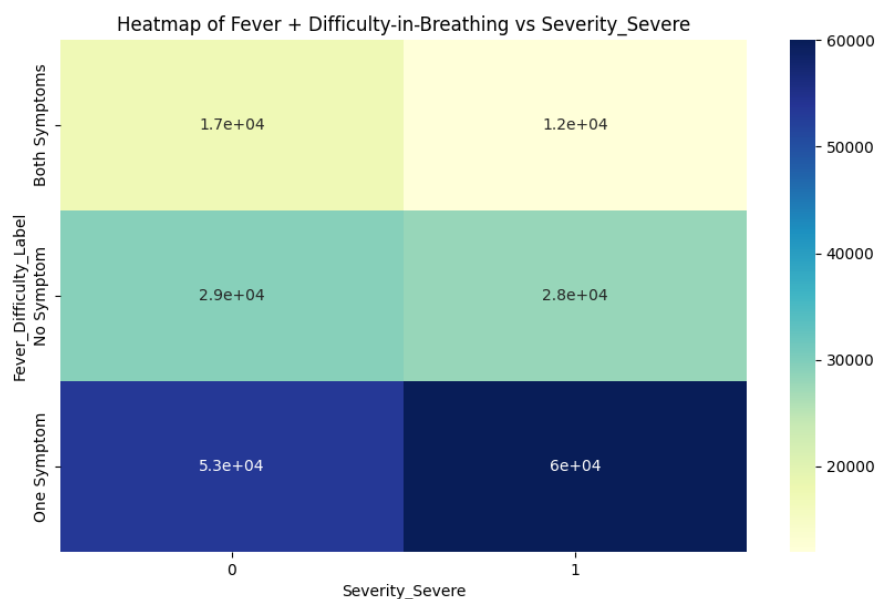
综合这三张图的分析，可以发现不同的症状组合与重度病情（Severity\_Severe）之间存在一定的关联，但这种关联并非绝对。

对于同时出现两种症状（如发热和咳嗽、咳嗽和呼吸困难、发热和呼吸困难）的患者，在某些情况下非重度病情的患者数量相对较多，而在其他情况下两者数量较为接近。

当只有一种症状时，在部分症状组合中重度病情的患者数量会多于非重度病情的患者。

这些结果提示临床医生在评估病情严重程度时，可以将症状组合作为参考因素之一，但不能仅依靠症状来判断病情的重度与否，还需要结合其他临床检查、患者的基础健康状况、病史等多方面的信息进行综合判断，以更准确地评估患者的病情严重程度并制定相应的治疗方案。

为此，我们用这三种情况与轻度病情相结合**绘制热力图**：



症状组合与重度病情（Severity\_Severe）的关系

**Both Symptoms（发热和呼吸困难都有）：**

当 Severity\_Severe 为 0（可能表示非重度病情）时，患者数量约为（17000）。

当 Severity\_Severe 为 1（可能表示重度病情）时，患者数量约为（12000）。

可以看出，同时有发热和呼吸困难症状的患者中，非重度病情的患者数量多于重度病情的患者，热力图中该区域颜色较浅，对应患者数量较少。

**No Symptom（无发热和呼吸困难症状）：**

Severity\_Severe 为 0 的患者数量约为（29000）。

Severity\_Severe 为 1 的患者数量约为（28000）。

此症状组合下，非重度病情和重度病情的患者数量非常接近，热力图中这两个区域颜色相近，说明患者数量差异不大。

**One Symptom（只有发热或呼吸困难其中一种症状）：**

Severity\_Severe 为 0 的患者数量约为（53000）。

Severity\_Severe 为 1 的患者数量约为（60000）。

该症状组合下，重度病情的患者数量多于非重度病情的患者，热力图中该区域颜色较深，对应患者数量较多。

## 整体结论

从热力图可以直观地看出不同症状组合（Both Symptoms、No Symptom、One Symptom）下，重度病情（Severity\_Severe 为 1）和非重度病情（Severity\_Severe 为 0）的患者数量分布情况。颜色越深表示患者数量越多，颜色越浅表示患者数量越少。总体而言，同时有发热和呼吸困难症状的患者中，非重度病情占比较大；而无相关症状的患者中，非重度病情和重度病情的患者数量分布相对均衡；只有一种症状的患者中，重度病情的患者数量相对较多。这为临床诊断和病情评估提供了一定的参考，提示症状组合可能与病情的重度程度存在一定关联，但仍需结合更多临床数据和研究进行深

### 三、总结

本项目围绕 COVID-19 数据集展开深入分析,通过多种可视化手段揭示了疫情相关数据的多方面特征与规律,为未来再次出现类似情形时的疫情防控、医疗研究及资源分配等提供了有价值的参考。

在患者疾病严重程度方面,数据集中中度病情患者占比 36.5%,重度占 33.8%,轻度占 29.7%,表明疫情下患者病情分布相对均衡,对医疗资源分配提出了精细化要求,需确保各层级医疗资源能合理应对不同严重程度患者的救治需求,尤其要重视中度和重度患者的医疗保障。

从年龄和性别维度分析,各年龄段男女受疫情影响差异显著。如 0 - 9 岁男性患者数量突出,而 10 - 19 岁及 25 - 59 岁女性患者相对较多,60 岁及以上男性又成为高发群体。这凸显了疫情防控需精准到不同年龄段的性别差异,针对特定人群活动场景和生理特点优化防控策略,如加强 0 - 9 岁男性活动场所卫生管理,在学校和工作场所为 10 - 19 岁、25 - 59 岁女性制定专门防控措施,为 60 岁及以上男性提供更贴合的医疗与防护服务,并深入探究性别与年龄交互影响感染机制,为防控提供科学依据。

属性相关性分析中,症状之间、症状与年龄、性别、病情严重程度及接触情况等变量关系复杂多样。症状关联为临床诊断提供线索,如发热伴疲劳、咳嗽伴疲劳等症状组合提示病情发展可能,医生可据此更全面评估患者;接触史与症状相关性助力防控,加强对有接触史人群发热症状监测意义重大。但相关关系不等同因果,需深入研究明确内在机制,为精准医疗与防控奠定基础。

全球范围内,各国受疫情影响人数差异明显,中国和部分欧洲国家人数众多,韩国相对较少,这受人口规模、地理、防控及医疗资源等综合因素影响。各国应相互学习借鉴防控经验,如中国的大规模防控体系、韩国的快速精准策略,在国际合作中共同提升全球疫情应对能力,包括共享疫情数据、协调医疗物资调配与疫苗合作,形成全球抗疫合力。

在症状表现上,多数国家无症状和无发热患者占比不容忽视,改变了仅依赖症状筛查患者的传统认知。防控中需强化无症状感染者检测追踪,利用核酸检测与流调结合,防止疫情隐匿传播;公众应强化个人防护意识,持续保持卫生习惯;医疗机构要合理配置资源,预留无症状感染者管理所需人力、物力,完善监测随访机制,及时干预病情变化。

症状组合与病情严重程度关联分析表明,不同症状组合对病情判断有参考价值。如发热咳嗽、咳嗽呼吸困难等组合在不同病情程度中呈现特定患者数量分布规律,为临床病情评估增添依据,但鉴于其复杂性,务必结合临床全方位检查、患者基础健康和病史信息综合判断,确保诊断治疗精准性,推动医学研究深入探索症状与病情关系本质,优化诊疗方案。

最后,希望这份数据科学与工程导论的期末可视化作业可以为未来再次出现类似情形时的疫情防控、医疗研究及资源分配等提供有价值的参考,真正做到有意义、有价值的可视化研究。