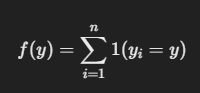
7 问题4: 分析新的奥运项目设置对奖牌数的影响

由于summerOly\_programs.csv文件中每一个字段都只是包含一个运动大类（Discipline）。而不是包含所有小项，因此无法根据该文件计算出每一年增加的所有小项，只能计算出每一届增加的运动大类（Discipline）。

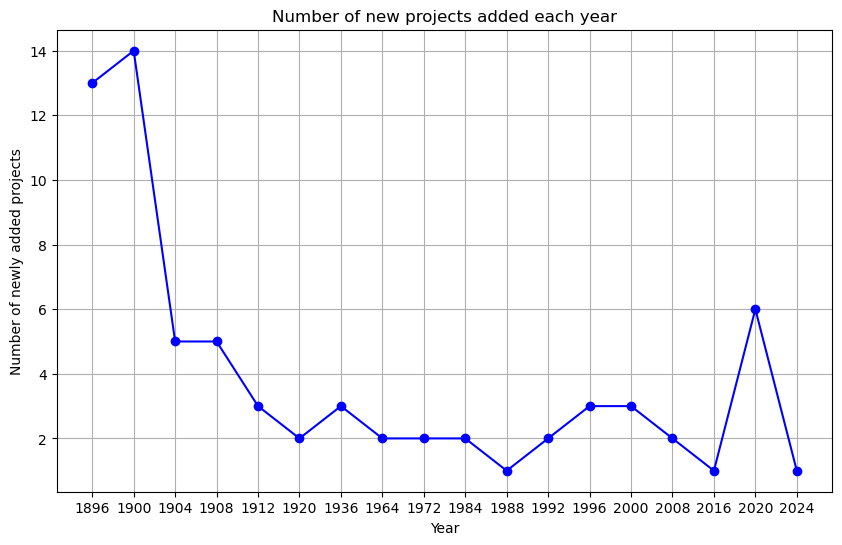
下面开始统计**每一届新增的项目数量**

7.1 **统计每一届新增的项目数量**：

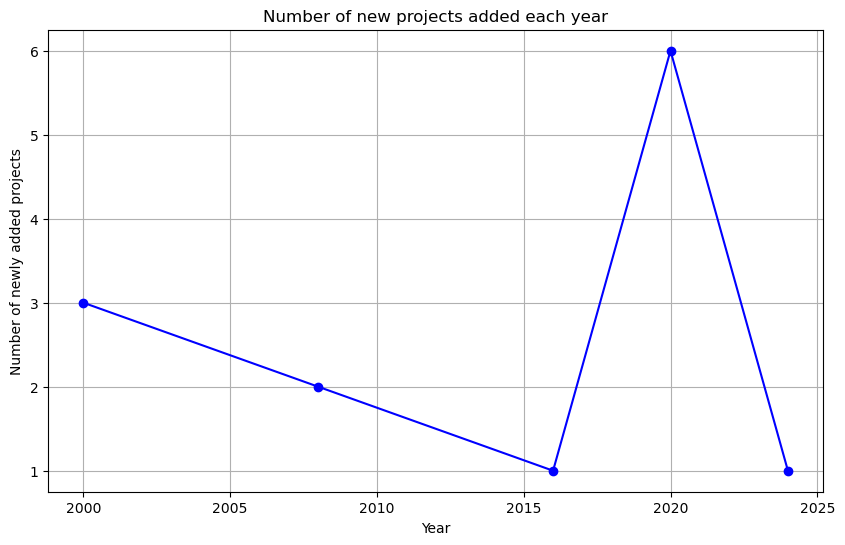
* 使用 value\_counts() 来统计每个年份新增项目的数量。该方法的数学原理是计数每个年份在 Year 列中出现的次数。
* value\_counts() 计算每年新增项目的频次，其实是**频数统计**，具体公式为： f(y)=∑i=1n1(yi=y)f(y) = \sum\_{i=1}^{n} \mathbb{1}(y\_i = y)f(y)=i=1∑n​1(yi​=y)
* 
* 其中 f(y)是年份 y的出现次数，1 是指示函数，表示当 yi=y 时值为1，否则为0。

**折线图绘制**：

* 绘制折线图使用了 matplotlib.pyplot.plot()，通过年份和新增项目数量的关系展示趋势。这里的数学原理主要是数据可视化：通过折线图连接每年的数据点，从而展示数据的趋势。
* 该步骤主要是通过坐标系展示每年新增项目的数量。



由于过早的数据没有意义，直接选取2000之后的数据开始计算新增项目对于东道主奖牌数量的影响。



7.2求解新增项目所获奖国家

 **分组操作（GroupBy）**：使用 groupby 方法对数据进行分组，以便对每个组执行聚合操作。

 **计数操作（Counting）**：在分组后，使用 size 或 value\_counts 方法计算每个组的大小，即每个组中元素的数量。

通过分组计数的思想求解新增项目的获奖国家。

7.3 模型建立

7.3.1**Data Merging (Inner Join)**:

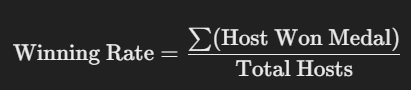
Merged Data=Data1⋈on Year​Data2

7.3.2**Boolean Expression for Host Wins**:

* Concept: A boolean comparison is used to check if the country that won the medal is the same as the host country. This results in a binary value (True/False or 1/0), which is used for further aggregation.

7.3.3**Mean/Proportion Calculation:**

 Winning Rate=Total Hosts∑(Host Won Medal)​

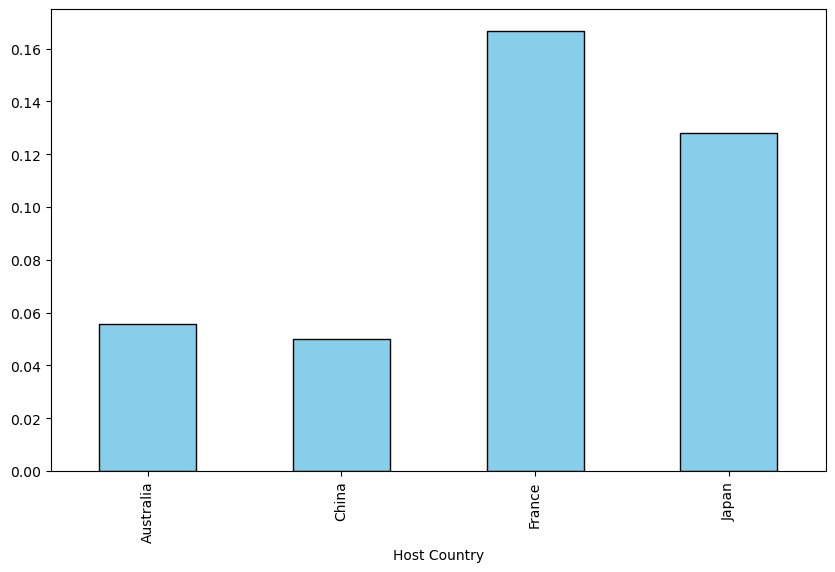
 

*  Concept: This calculates the proportion of times each host country won medals. It sums up the values of "Host Won Medal" (which is 1 for True and 0 for False) and divides by the total number of host occurrences, yielding the winning rate.

7.3.4**GroupBy and Aggregation**:

* Concept: The "groupby" operation groups the data by host country and then computes the mean of "Host Won Medal" for each group. This effectively calculates the percentage of times each host country won medals across the relevant sports and years.

7.4 模型求解



通过分析结果，可以得出几乎所有东道主都通过新增对于自己有利的项目来获得奖牌数量上的优势。2000年后，2020东京奥运会和2024法国奥运会上东道主在新增项目上的奖牌获得率甚至远超过10%。2000悉尼奥运会和2008年北京奥运会上，澳大利亚和中国也在新增项目上获得了超过5%的奖牌获得率。这个部分远远大于这些国家在其他所有项目上的奖牌获得率。

因此可以得出结论，几乎所有东道主都通过新增对于自己有利的项目来获得奖牌数量上的优势。但是这也与一个国家的整体体育水平有较大关系，2016年里约奥运会和2004巴塞罗那奥运会中巴西和西班牙并没有因此获得更多奖牌，因为这些国家的整体体育水平并没有比上述提到的四个国家体育水平高。

那么当一个国家需要提升国家总体体育水平时，可以通过申办奥运会的方法。研究表明，主办国通过充分发挥主场优势，显著提升了金牌总数、总奖牌榜排名以及运动员参赛规模，进而增强了本国在冬季运动项目上的竞技实力。