**1. 奖牌总数的计算：**

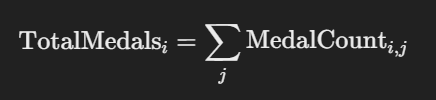
每个国家（NOC）在某一年的奖牌总数是通过金牌（Gold）、银牌（Silver）、铜牌（Bronze）数量的总和来计算的：

MedalCounti,j​=count of medals for country j in sport i

**2. 项目总奖牌数**

对于每个项目，我们通过对所有国家的奖牌数量求和，得到该项目的总奖牌数。

公式：

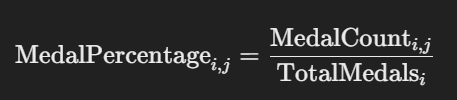


TotalMedalsi=∑jMedalCounti,j\text{TotalMedals}\_i = \sum\_{j} \text{MedalCount}\_{i,j}TotalMedalsi​=j∑​MedalCounti,j​

**3. 奖牌占比（Medal Percentage）**

每个国家在某个比赛项目中获得的奖牌占该项目总奖牌数的比例。

公式：



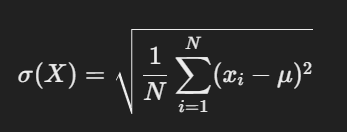
MedalPercentagei,j=MedalCounti,jTotalMedalsi\text{MedalPercentage}\_{i,j} = \frac{\text{MedalCount}\_{i,j}}{\text{TotalMedals}\_i}MedalPercentagei,j​=TotalMedalsi​MedalCounti,j​​

**4. 竞争激烈性的标准差（Competition Intensity）**

竞争激烈性通过计算每个项目中所有国家的奖牌占比的 **标准差** 来衡量。标准差越小，意味着奖牌分布越集中，竞争相对不那么激烈；标准差越大，意味着奖牌分布越均匀，竞争越激烈。

公式：

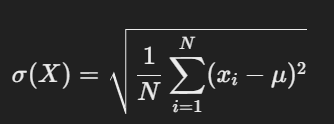
CompetitionIntensityi​=σ({MedalPercentagei,j​∣j∈all countries})



其中，σ表示标准差，MedalPercentagei,j是每个国家 j在比赛项目 i中的奖牌占比。

标准差的计算公式：

σ(X)=1N∑i=1N(xi−μ)2\sigma(X) = \sqrt{\frac{1}{N} \sum\_{i=1}^{N} (x\_i - \mu)^2}σ(X)=N1​i=1∑N​(xi​−μ)2​



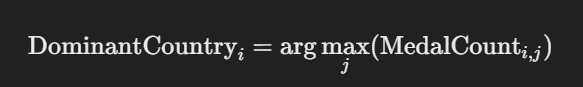
其中，X={x1,x2,...,xN}是一组数据，μ是数据的均值，N 是数据点的数量。

**5. 主导国家（Dominant Countries）**

每个比赛项目的主导国家是获得该项目最多奖牌的国家。

公式：

DominantCountryi=arg⁡max⁡j(MedalCounti,j)\text{DominantCountry}\_i = \arg\max\_j(\text{MedalCount}\_{i,j})DominantCountryi​=argjmax​(MedalCounti,j​)



其中， argmax 表示找到使得奖牌数量最大的国家。

**6. 垄断性项目（Monopolized Projects）**

一个项目被某个国家垄断的条件是该国家在该项目中获得的奖牌占该项目总奖牌数的比例超过 **5%**。

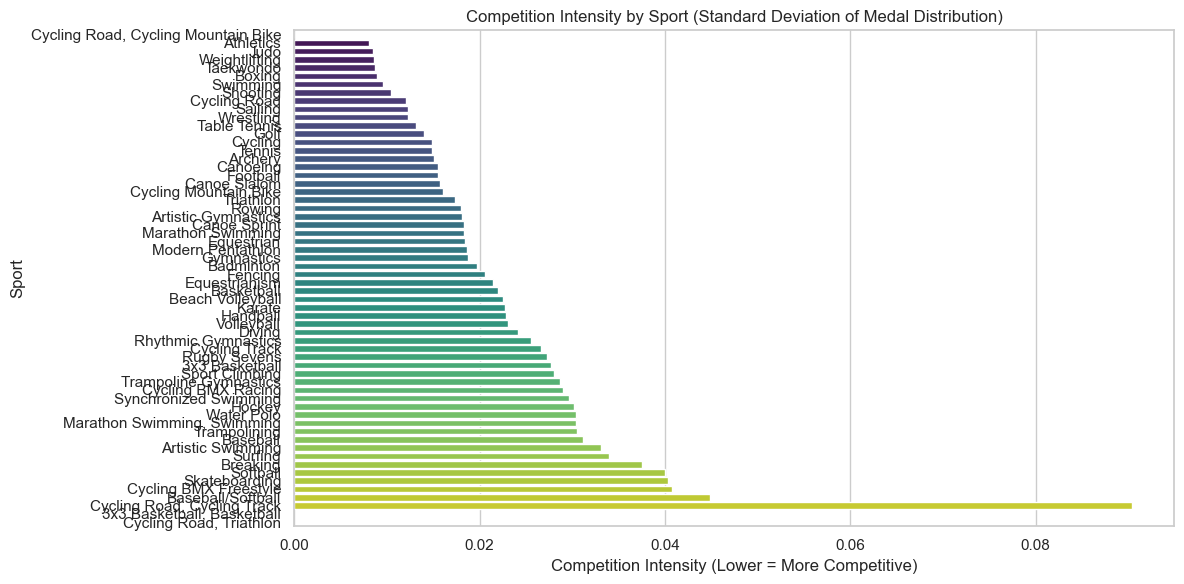
公式：

Monopolizedi if MedalPercentagei,j>0.05

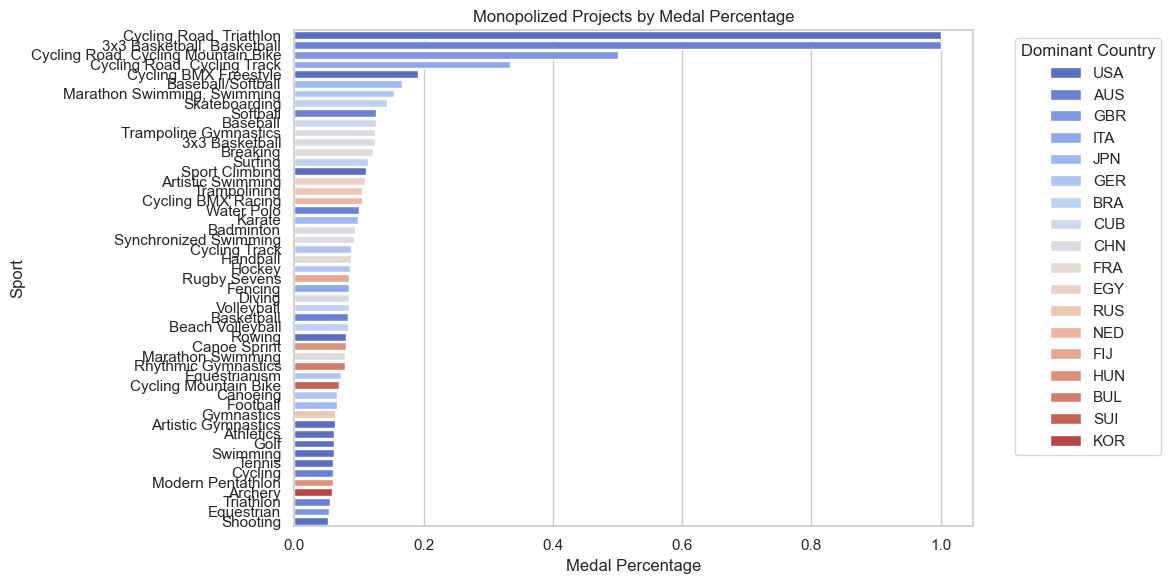
其中，MedalPercentagei,j\text{MedalPercentage}\_{i,j}MedalPercentagei,j​ 是国家 jjj 在项目 iii 中的奖牌比例。如果比例大于 5%，则该项目被该国家垄断。

在按照以上方法进行筛选之后，发现当某一个项目MedalPercentage为1的时候，也就是某一个项目被某一个国家完全垄断的时候，他的方差就会是0或者NaN。而且，多半这种项目也许开办的次数特别少，并且开办在很久之前，当前已经被废弃。

因此，我们再多增加了一个对于奖牌计算的筛选条件，也就是只考虑现代奥林匹克成熟之后，冷战结束之后的数据比较有代表性，因而只选取了2000悉尼奥运会之后的数据进行参考计算。

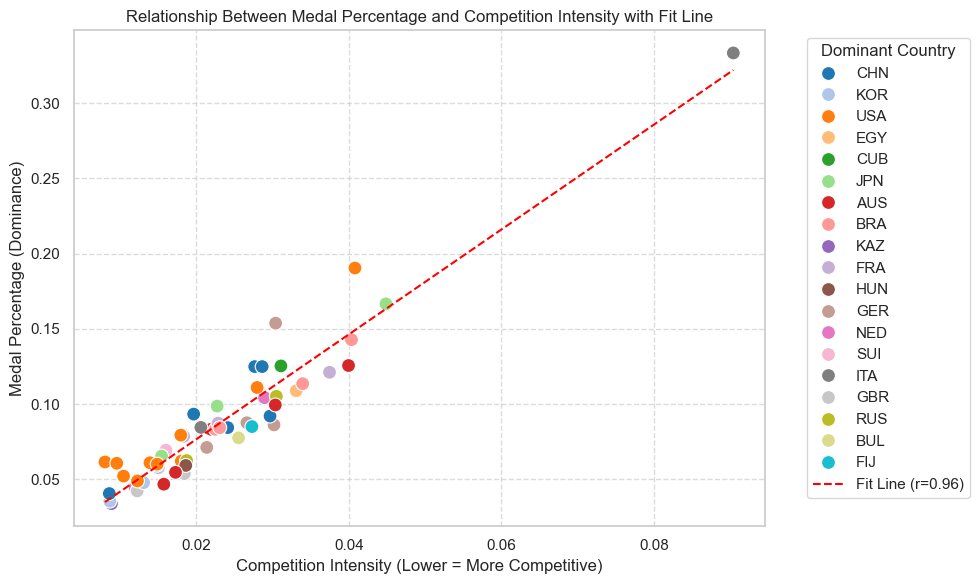


按照每个项目的竞争激烈程度排序得到结果。可以得知Athletics、Weightlifting、Swimming、Shooting、Cycling Road、Sailing等各个国家广泛参与的大项目各个国家都有获奖，竞争比较激烈。



按照每个项目的擅长的国家国家垄断程度排序得到结果。再去除总奖牌数量比较少没有参考价值的数据之后，可以得知JPA擅长Baseball/Softball和Karate（空手道）、CHN擅长Trampoline Gymnastics、Badminton、Synchronized Swimming和Diving

;USA擅长Sport Climbing、Rowing、Artistic Gymnastics、Athletics、Swimming和Tennis。这个结果也是相当符合大众认知和预计期望的，这也增加了我们数据处理与建模的可信程度。



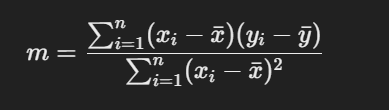
基于数据集 project\_analysis.csv 中的 MedalPercentage（奖牌比例）和 CompetitionIntensity（竞争强度）列生成了Relationship Between Medal Percentage and Competition Intensity with Fit Line。过滤掉了异常值和无效值，比如 NaN、无穷大、或不合理值。绘制了散点图和拟合直线。通过线性回归计算出最佳拟合直线，并将其绘制在散点图上，红色虚线表示回归结果。**R值**：标明了拟合直线的相关系数（r=0.96）

拟合直线可以表示为：

y=mx+b

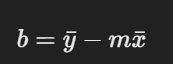
1. 斜率 (m)：

m=∑i=1n(xi−xˉ)(yi−yˉ)∑i=1n(xi−xˉ)2m = \frac{\sum\_{i=1}^n (x\_i - \bar{x})(y\_i - \bar{y})}{\sum\_{i=1}^n (x\_i - \bar{x})^2}m=∑i=1n​(xi​−xˉ)2∑i=1n​(xi​−xˉ)(yi​−yˉ​)​



1. 截距 (b)：

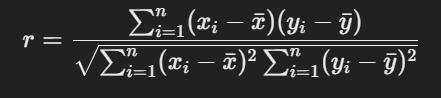
b=yˉ−mxˉb = \bar{y} - m \bar{x}b=yˉ​−mxˉ



**相关系数 r 的计算**

**相关系数 r 衡量 x 和 y之间的线性关系，其公式为：**

**r=∑i=1n(xi−xˉ)(yi−yˉ)∑i=1n(xi−xˉ)2∑i=1n(yi−yˉ)2r = \frac{\sum\_{i=1}^n (x\_i - \bar{x})(y\_i - \bar{y})}{\sqrt{\sum\_{i=1}^n (x\_i - \bar{x})^2 \sum\_{i=1}^n (y\_i - \bar{y})^2}}r=∑i=1n​(xi​−xˉ)2∑i=1n​(yi​−yˉ​)2​∑i=1n​(xi​−xˉ)(yi​−yˉ​)​**

****

**数据分布特点**：大部分点分布在竞争强度较低的区域（CompetitionIntensity < 0.05），奖牌比例也较低（MedalPercentage < 0.15）。

**结论**:表表明竞争强度越低，垄断国家奖牌比例越大。这可能是因为在高竞争环境下，表现更出色的国家在奖牌分配中占更大优势。