- COVID19全球检测数据
 - 简介
 - 指导
 - 练习
 - 1. 理解数据
 - 2. 数据清洗
 - 3. 提取具有最多COVID-19病例的前十个国家
 - 4. 确定最高的阳性病例占测试病例的比例
 - 5. 确定顶部三个国家进行的最高阳性病例占测试次数的比例
 - 6. 显示结果
 - 7. 结论

COVID19全球检测数据

简介

这个数据集 tested_worldwide.csv 来自 Kaggle。这个数据集记录了随时间进行的检测次数,对于帮助理解每天报告的病例数以及理解COVID-19在每个国家的真实传播情况非常重要。

指导

- 1. 使用 Copilot Chat 在你的项目中创建一个新的笔记本。使用命令 /newnotebook 并将其命名为 "COVID19 Worldwide Testing Data"。
 - 2. 使用 Copilot 和 Copilot Chat 来完成练习并支持你的学习。

练习

我们的分析试图回答这个问题: **哪些国家报告的阳性病例数量在进行的检测次数方面最高? **

1. 理解数据

- 1.1. 导入数据集并显示数据框的前5行。
- 1.2. 显示数据框中的行数和列数。
- 1.3. 显示每列的数据类型。
- 1.4. 显示每列中的缺失值数量。
- 1.5. 显示每列中的唯一值数量。

2. 数据清洗

- 2.1. 删除分析不需要的列。
- 2.2. 重命名列名以使其更易读。
- 2.3. 删除包含缺失值的行。
- 2.4. 将列的数据类型转换为适当的类型。
- 2.5. 显示每列中的缺失值数量。

3. 提取具有最多COVID-19病例的前十个国家

- 3.1. 创建一个包含每个国家总阳性病例数量的新数据框。
- 3.2. 将数据框按总阳性病例数量降序排列。
- 3.3. 显示具有最多阳性病例的前十个国家。

4. 确定最高的阳性病例占测试病例的比例

- 4.1. 创建一个包含每个国家进行的总测试次数的新数据框。
- 4.2. 将数据框按进行的总测试次数降序排列。
- 4.3. 显示进行了最多测试的前十个国家。

5. 确定顶部三个国家进行的最高阳性病例占测试次数的比例

- 5.1. 合并前面步骤中创建的两个数据框。
- 5.2. 创建一个包含阳性病例数与进行的测试次数比率的新列。
- 5.3. 将数据框按阳性病例数与进行的测试次数比率降序排列。
- 5.4. 显示阳性病例数与进行的测试次数比率最高的前三个国家。

6. 显示结果

- 6.1. 显示一个图表,展示阳性病例数与进行的测试次数比率最高的前三个国家。
- 6.2. 显示一个图表,展示具有最多阳性病例的前十个国家。
- 6.3. 显示一个图表,展示进行了最多测试的前十个国家。

7. 结论

- 7.1. 你的结论是什么?
- 7.2. 分析的局限性是什么?
- 7.3. 你将采取哪些步骤来改善这个分析?