数据集说明

**一、Aberystwyth航运记录**

Aberystwyth航运记录是Excel电子表格的集合，其中包含有关在19XX年访问Aberystwyth的船只和海员的信息。请将其中的数据创建脚本并加载到MongoDB中。

**二、客户**

客户是Sylfaen Treftadaeth / Local Heritage Foundation的组织机构。他们对参观过Aberystwyth港的船只和海员的历史感兴趣。

三、具体要求（需求）

“对使用Aberystwyth港的海员的生活和经历感兴趣。相信港口大师们保存的船舶手写记录对于阐明这些过去的经历非常有帮助。

主要对访问过的人和他们从哪里来访问的感兴趣。要求对数据进行两次探索性分析，一是用于检索个人故事，二是查看该轮次中的经验。

1)个人故事。希望能够追踪个别海员在他们的生活和职业生涯中的生活情况。如他们所有工作生活中都是在同一艘船上还是经常更换船只和工作？他们是不断晋升还是保持同一职级？需要能够检索有关单个海员的所有记录并将其作为叙述内容呈现。

2)谁来港口参观。船只和海员从不同的地方抵达Aberystwyth港口。需要整理他们分别来自哪些地方的数据。哪些是比其他更受欢迎的来源和目的地？船舶是否每次使用相同的路线，或者它们是否遵循不同的路线？”

**四、可交付成果**

客户的需求相当广泛，不一定可以实现。用技术语言重新描述如下。

**首先，数据相当混乱，包含许多缺失的元素；为满足需求，任务的很大一部分是处理这些缺失的数据。**

**应该使用数据探索方法和可视化技术来预处理数据。**

**1、个人故事方面**

**1）创建一个进程以提取有关单个海员的所有记录并按时间顺序显示它们**。需能识别个体并将其与其他类似个体区分开来。**代码应该足够灵活，以便能够识别和提取有关数据集中任何海员的信息。**除了**生成可视化之外**，**代码还应该包括两个Python函数**，一个用于列出海员以便于选择，另一个用于获取所选海员的记录。可以使用尽可能多的临时或缓存来存放记录集。

**每个海员记录应包含足够的信息，如姓名，年龄和备注等。**

**2）编写代码创建以下可视化**

（1）每个职别的个人比例。 （一个人可能在其职业生涯中有多个职级）

（2）可视化两个或更多个人的晋升轨迹。**可视化的选择依照个人理解，但应体现晋升的时间顺序。**

**请注意，这两个功能未必是最好的，也可以发挥以实现更多的功能和可视化结果。**

**2、“谁来访问”方面**

对于每艘船只提取船舶已访问的港口并**创建以下可视化视图。**

1）**每艘船上的船员人数的直方图。**

2）**每个不属于Aberystwyth港的船舶的访问次数（由于每艘船有多名船员，这可能与提及次数不同）。可能有太多港口，至少应该选择10个以上。**

关于涉及的技术问题的简要说明。

解决方案的技术说明。

**五、开发工具**

**MongoDB queries，Javascript, Python编写**。 可使用图形或绘图库如 Matplotlib。

## 字段说明

|  |  |
| --- | --- |
| 'name' | 1姓名 |
| 'year\_of\_birth' | 2生日 |
| 'age' | 3年龄 |
| 'place\_of\_birth' | 4出生地 |
| 'home\_address' | 5住址 |
| 'last\_ship\_name' | 7上一个船名 |
| 'last\_ship\_port' | 8上一个港口 |
| 'last\_ship\_leaving\_date' | 9上一个船离开日期 |
| 'this\_ship\_joining\_date' | 10这个船的joining日期 |
| 'this\_ship\_joining\_port' | 11这个船的 |
| 'this\_ship\_capacity' | 12在这个船的职位 |
| 'this\_ship\_leaving\_date' | 13这个船离开的日期 |
| 'this\_ship\_leaving\_port' | 14离开的港口 |
| 'this\_ship\_leaving\_cause' | 15离开的原因 |
| 'signed\_with\_mark' | 16不知道咋翻译 |
| 'additional\_notes' | 17补充说明 |