

它能将每个单词的第一个字母大写。Pandas 使用空格识别一个单词的结束位置和下一个单词的开始位置：

```
In [19] inspections["Name"].str.title().head()

Out [19] 0      Marriot Marquis Chicago
         1              Jets Pizza
         2              Room 1520
         3      Marriot Marquis Chicago
         4              Chartwells
         Name: Name, dtype: object
```

title 方法是处理地点、国家、城市和人的全名时的绝佳选择。

6.2 字符串切片

Risk 列中，每行的值都包括代表风险类型的编号(例如 “Risk 1”)和级别(例如 “High”)。首先了解该列中的内容：

```
In [20] inspections["Risk"].head()

Out [20]

0      Risk 1 (High)
1      Risk 2 (Medium)
2      Risk 3 (Low)
3      Risk 1 (High)
4      Risk 1 (High)
Name: Risk, dtype: object
```

假设要从每一行中提取数字形式的风险类型，考虑到数据集内的每一行的格式几乎相同，这个操作看起来很简单，但必须小心行事。在大型数据集内，总是存在各种出错的可能。

```
In [21] len(inspections)

Out [21] 153810
```

是否所有行的 Risk 列都遵循 “RiskNumber(RiskLevel)” 这种格式？可以通过调用 unique 方法找出答案，该方法返回一个由列的唯一值组成的 NumPy ndarray：

```
In [22] inspections["Risk"].unique()

Out [22] array(['Risk 1 (High)', 'Risk 2 (Medium)', 'Risk 3 (Low)', 'All',
              nan], dtype=object)
```

必须考虑两个额外的值：缺失值 NaN 和 “All” 字符串。如何处理这两个额外的值最终取决于企业的实际需要，即这些值是否重要，或者它们是否可以被删除。此处提出一个折中方案：删除缺失的 NaN 值，并将 “All” 值替换为 “Risk 4(Extreme)”。选择这种解决方案来确保所有 Risk 值具有一致的格式。

我们可以使用第 5 章介绍的 dropna 方法从 Series 中删除缺失值，向 subset 参数传递一个