



动手实验室 5:清理数据

预计所需时间： 45 分钟

在本实验中,首先您将学习如何处理不准确的数据、如何删除空行以及如何删除重复数据。接下来,您将学习如何更改文本的大小写、如何更改日期格式以及如何修剪数据中的空格。最后,您将学习如何使用 Excel 中的快速填充功能来帮助清理数据。

本实验室使用的软件

本课程中的教学视频使用完整的 Excel 桌面版本,因为它具有所有可用的产品功能,但对于动手实验,我们将使用免费的“Excel 网页版”版本,因为每个人都可以使用该版本。

虽然如果您有权访问此版本,则可以使用 Excel Desktop 软件,但建议您使用 Excel 网页版 进行动手实验,因为实验说明专门参考了此版本,并且在操作上存在一些细微差别。界面和可用功能。

本实验室使用的数据集

本实验中使用的数据集来自以下来源：<https://dataplatform.cloud.ibm.com/exchange/public/entry/view/f8ccaf607372882403a37d9019b3abf4>。该数据集由IBM发布,包含虚构的客户人口统计数据和销售数据。

我们在实验室中使用该数据集的修改子集,因此要成功遵循实验室说明,请使用实验室提供的数据集,而不是原始来源的数据集。

目标

完成本实验后,您将能够：

- 了解如何处理不相关或不准确的数据

- 删除空行和重复数据
- 更改文本大小写和日期格式
- 修剪数据中的空格
- 使用 Flash Fill 和函数来清理数据

练习 1:删除重复、不相关或不准确的数据

在本练习中,您将学习如何处理不准确的数据、如何删除空行以及如何删除重复数据。

任务A:检查拼写

1. 下载文件[Customer_demographics_and_sales_Lab5.xlsx](#),使用 Excel 网页版 上传并打开它。
2. 选择L 列 (CREDITCARD_TYPE),然后单击 “审阅”选项卡,然后选择 “拼写” 。
3. 单击正确的建议以更改拼写。
 - 注意 :进行拼写检查时不要更改 “jcb”拼写。练习 1 任务 D 需要 “jcb” 。
4. 关闭拼写窗格。

K	L
NUMBER	CREDITCARD_TYPE
1-8539	Master Card
1-8539	Master Card
1-8539	Master Card
173271	VISA
5-4321	American Expres
5-4321	American Expres
5-4321	American Expres
5-4321	American Expres
5-4321	American Expres
5-4321	American Expres
5-4321	American Expres
12037	Diners Club
12037	Diners Club
12037	Diners Club
11865	VISA
11865	VISA
4-7595	Diners Club

Spelling

Not in Dictionary

American Expres

Suggestions

Express

Expires

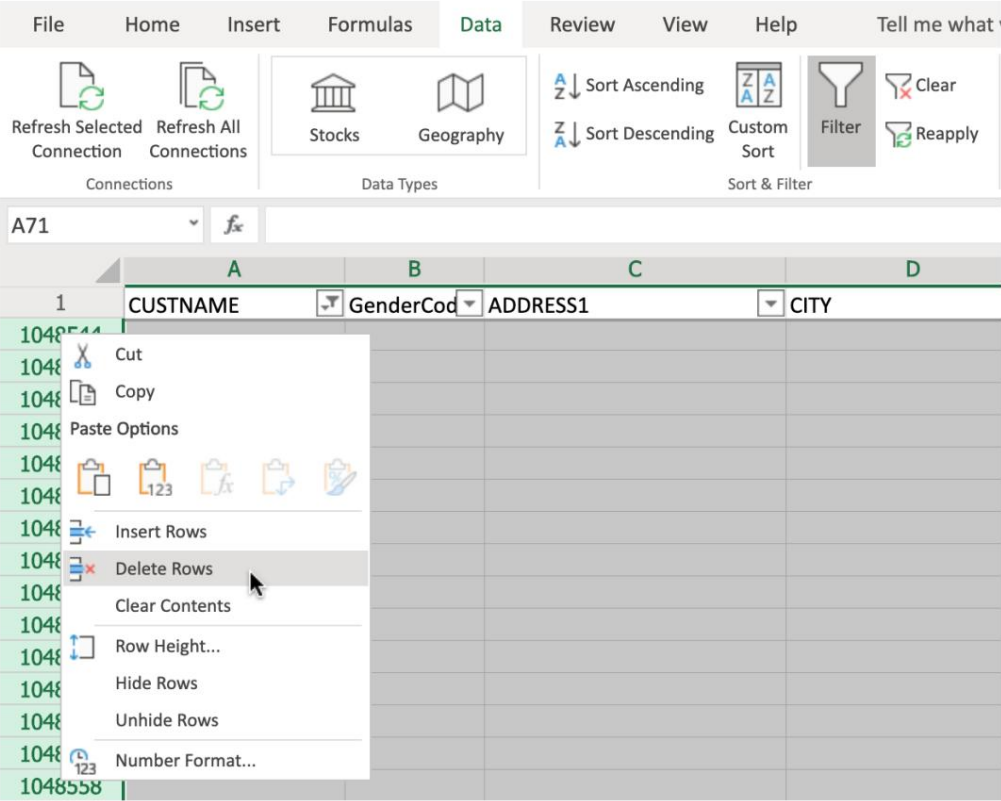
Expreso

Ignore

Ignore All

任务 B:删除空行

1. 按CTRL+HOME,然后按CTRL+SHIFT+END选择整个数据表。
2. 在数据选项卡上,单击过滤器。
3. 按CTRL+HOME,单击CUST_NAME列中的筛选器箭头,然后单击筛选器。
4. 单击全选复选框以取消选择全部。然后只选择空白,然后确定。
5. 选择第一行,然后按CTRL+SHIFT+END选择所有行。
6. 右键单击选定的行,然后单击删除行。
7. 最后,在“数据”选项卡上,单击“清除”,然后单击“过滤器”。



任务 C:删除重复行

1. 选择列T (ORDER_ID),因为 ORDER_ID 值是唯一的。
2. 在“开始”选项卡上,单击“条件格式”>“突出显示单元格规则”>“重复值”,然后单击“确定”。
3. 选择整个数据表 (CTRL+SHIFT+END)
4. 在“数据”选项卡上,单击“删除重复项”。
5. 在“删除重复项”对话框中,确保选中“选择所有列”并且还选中“我的数据有标题”,然后单击“确定”。

6. 在通知您找到并删除了多少重复值的弹出框中,单击“确定”。

Excel



5 duplicate values found and removed; 194 unique values remain.

[Give Feedback](#)

OK

任务 D:使用查找和替换来纠正拼写错误

1. 在“主页”选项卡上,单击“查找并选择”。
2. 单击“查找”。在查找内容中,输入jcb,然后单击查找全部。
3. 单击“替换”。
4. 在替换为中,键入JCB,单击全部替换,然后单击关闭图标。
5. 在“主页”选项卡上,单击“条件格式”>“清除规则”>“从整个工作表清除规则”。

Find

Find what:

jcb

Search Options

Within:

Sheet

Direction:

Down

☒ Match case
 ☒ Match entire cell contents

Results found (25)

Sheet	Cell	Value
Sheet1	L45	jcb
Sheet1	L46	jcb
Sheet1	L47	jcb
Sheet1	L49	jcb
Sheet1	L69	jcb
Sheet1	L70	jcb

Find Next

Find All

Replace...

练习 2:处理数据不一致

在本练习中,您将学习如何更改文本的大小写、如何更改日期格式以及如何修剪数据中的空格。

任务 A:使用 PROPER 函数将文本从大写更改为正确大小写

1. 选择第2 行,然后右键单击它并选择“插入行”。
2. 在单元格A2 中,键入=PROPER(A1)并按Enter。
3. 将鼠标悬停在单元格A2的右下角,然后将填充手柄拖到最后一列。

- 如果使用鼠标拖动太困难,请使用SHIFT+向右键选择第 2 行中的单元格,然后按F2将光标焦点放回到单元格A2 中,然后在按住CTRL的同时按Enter。

4. 选择第2 行,然后按CTRL+C。
5. 选择第1 行,右键单击并选择 “粘贴选项”> “值” 。
6. 选择第2 行,右键单击它并选择删除行。

任务 B:使用 UPPER 函数将文本从正确大小写更改为大写

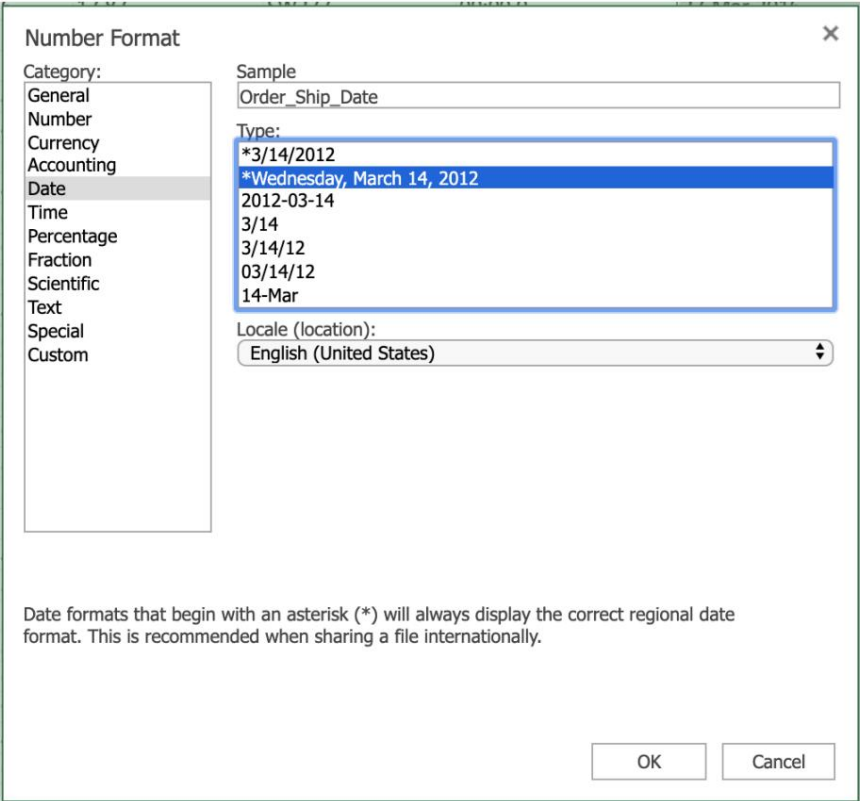
1. 选择AG (世代)列。然后右键单击并选择插入列。在单元格AG1 中,输入Generation。
2. 在单元格AG2 中,键入=UPPER(AH2)并按Enter 键。
3. 将鼠标悬停在单元格AG2的右下角上,然后双击填充手柄。
4. 选择AG 列,然后按CTRL+C。
5. 选择AH 列,右键单击并选择 “粘贴选项”> “值” 。
6. 选择AG 列,右键单击它并选择删除列。

任务 C:使用 LOWER 函数将文本从正确的大小写更改为小写

1. 选择AC 列 (T_Type) 。然后右键单击并选择插入列。在单元格AC1 中,键入T_Type。
2. 在单元格AC2 中,键入=LOWER(AD2)并按Enter 键。
3. 将鼠标悬停在单元格AC2的右下角上,然后双击填充手柄。
4. 选择AC 列,然后按CTRL+C。
5. 选择AD 列,右键单击并选择 “粘贴选项”> “值” 。
6. 选择AC 列,右键单击它并选择删除列。

任务 D:更改日期格式

1. 选择Z 列 (Order_Ship_Date)。
2. 在 “主页”选项卡的 “数字”组中,单击 “数字格式”> “更多数字格式” 。
3. 在 “类别”列表中,选择 “日期” 。
4. 在 “设置单元格格式”框中的 “区域设置”下,选择 “英语 (美国)” 。
5. 在 “类型”下,选择 “2012 年 3 月 14 日星期三” ,然后单击 “确定” 。



任务 E:使用查找和替换来修剪空格

1. 单击CTRL+HOME。
2. 使用 CTRL+SHIFT+END 选择所有数据。
3. 在“主页”选项卡上,单击“查找并选择”,然后单击“替换”。
4. 在“查找内容”中,键入2 个空格。在替换为中,键入1 个空格。
5. 单击“查找全部”,然后单击“全部替换”。
6. 单击关闭图标。

练习 3:更多用于清理数据的 Excel 功能

在本练习中,您将学习如何使用 Excel 中的快速填充功能来帮助清理数据。

任务 A:使用快速填充功能清理数据：

1. 选择A 列 (Cust_Name),单击鼠标右键并选择“插入列”。
2. 在单元格A1中输入Customer_Name ,然后按Enter 键。
3. 在单元格A2 中,输入Mr. Allen Perl ,然后按Enter。
4. 选择A 列 (Customer_Name),在“数据”选项卡上单击“快速填充”。
5. 单击“撤消”撤消此步骤。

如果您使用的是桌面版 Excel,则可以使用“文本分列”功能来执行下一个任务(有关说明,请参阅相应主题视频)。

如果您使用的是“Excel 网页版”(Excel 的在线版本),则“文本分列”功能不可用,但您可以使用函数实现相同的结果,如下步骤所示。

任务 B:使用 LEFT、RIGHT、LEN 和 SEARCH 函数清理数据：

1. 选择A 列 (Cust_Name),单击鼠标右键并选择“插入列”。
2. 再次选择A列,右键单击并选择“插入列”。
3. 在单元格A1 中,键入Customer_Firstname ,然后在单元格B1 中,键入Customer_Lastname。
4. 单击C1,然后在“开始”选项卡上单击“格式刷”,然后拖动到 A1 和 B1。
5. 双击A 列和B列之间的分隔线。
6. 在单元格A2中输入 =LEFT(C2, SEARCH(“ “,C2,1))并按Enter。
7. 在单元格B2中输入 =RIGHT(C2,LEN(C2)-SEARCH(“ “,C2,1))并按Enter。
8. 双击单元格 A2 上的填充手柄。
9. 双击单元格 B2 上的填充手柄。

恭喜!您已完成实验 5,并准备好学习下一个主题。

作者

- [桑迪普·萨哈·乔伊](#)。

其他贡献者

- [史蒂夫·瑞安](#)

变更日志

日期	版本	更改时间	2020-09-10	1.2	Steve	变更说明
Ryan Steve Ryan	ID/技术	审核通过				添加了软件/数据集信息
2020-07-07	1.1					
2020-07-01	1.0					
		Sandip Saha Joy	创建	初始	版本	