



Excel数据分析师突击——从入门到精通到项目实战

第6周

【声明】 本视频和幻灯片为炼数成金网络课程的教学资料，所有资料只能在课程内使用，不得在课程以外范围散播，违者将可能被追究法律和经济责任。

课程详情访问炼数成金培训网站

<http://edu.dataguru.cn>

关注炼数成金企业微信



■提供全面的数据价值资讯，涵盖商业智能与数据分析、大数据、企业信息化、数字化技术等，各种高性价比课程信息，赶紧掏出您的手机关注吧！



Excel实践——Excel的数据读入与整理

数据读入与整理，往往是数据分析最费时间的一步，如何高效完成这
个工作？

导入外部数据

◆ 导入文本数据



文本导入向导 - 第 1 步, 共 3 步

文本分列向导判定您的数据具有固定列宽。
若一切设置无误, 请单击“下一步”, 否则请选择最合适的数据类型。

原始数据类型
请选择最合适的文件类型:

☐ 分隔符号(D) - 用分隔字符, 如逗号或制表符分隔每个字段
☒ 固定宽度(W) - 每列字段加空格对齐

导入起始行(R): 1 文件原始格式(O): 936 : 简体中文(GB2312)

☐ 数据包含标题(M)。

预览文件 D:\data\aa-3rv.txt:

1	3.36350	3.47450	3.06920
2	4.38480	4.14700	6.71210
3	1.54080	1.21240	0.57110
4	1.74370	1.87230	2.12200
5	107.04110	11.07970	14.51810

取消 < 上一步(B) 下一步(N) > 完成(F)

◆ 导入文本数据

文本导入向导 - 第 2 步, 共 3 步

请设置字段宽度(列间隔)。

有箭头的垂直线为分列线。

要建立分列线, 请在要建立分列处单击鼠标。

要清除分列线, 请双击分列线。

要移动分列线位置, 请按住分列线并拖至指定位置。

数据预览(P)

3.36350	3.47450	3.06920
4.38480	4.14700	6.71210
1.54080	1.21240	0.57110
1.74370	1.87230	2.12200
107.04110	11.07970	14.51810

取消

< 上一步(B)

下一步(N) >

完成(F)

文本导入向导 - 第 2 步, 共 3 步

请设置分列数据所包含的分隔符号。在预览窗口内可看到分列的效果。

分隔符号

☒ Tab 键(T)

☐ 分号(M)

☐ 逗号(C)

☐ 空格(S)

☐ 其他(O):

☐ 连续分隔符号视为单个处理(R)

文本识别符号(Q):

数据预览(P)

3.36350	3.47450	3.06920
4.38480	4.14700	6.71210
1.54080	1.21240	0.57110
1.74370	1.87230	2.12200
107.04110	11.07970	14.51810

取消

< 上一步(B)

下一步(N) >

完成(F)

◆ 导入文本数据

文本导入向导 - 第 3 步, 共 3 步

使用此屏内容可选择各列, 并设置其数据格式。

列数据格式

☒ 常规(G)
☐ 文本(T)
☐ 日期(D): YMD
☐ 不导入此列(跳过)(I)

“常规”数据格式将数值转换成数字, 日期值会转换成日期, 其余数据则转换成文本。

高级(A)...

数据预览(P)

常规	常规	常规	常规
3.36350	3.47450	3.06920	
4.38480	4.14700	6.71210	
1.54080	1.21240	0.57110	
1.74370	1.87230	2.12200	
107.04110	11.07970	14.51810	

取消 < 上一步(B) 下一步(N) > 完成(F)

导入数据

请选择该数据在工作簿中的显示方式。

☒ 表(T)
☐ 数据透视图表(P)
☐ 数据透视图(C)
☐ 仅创建连接(O)

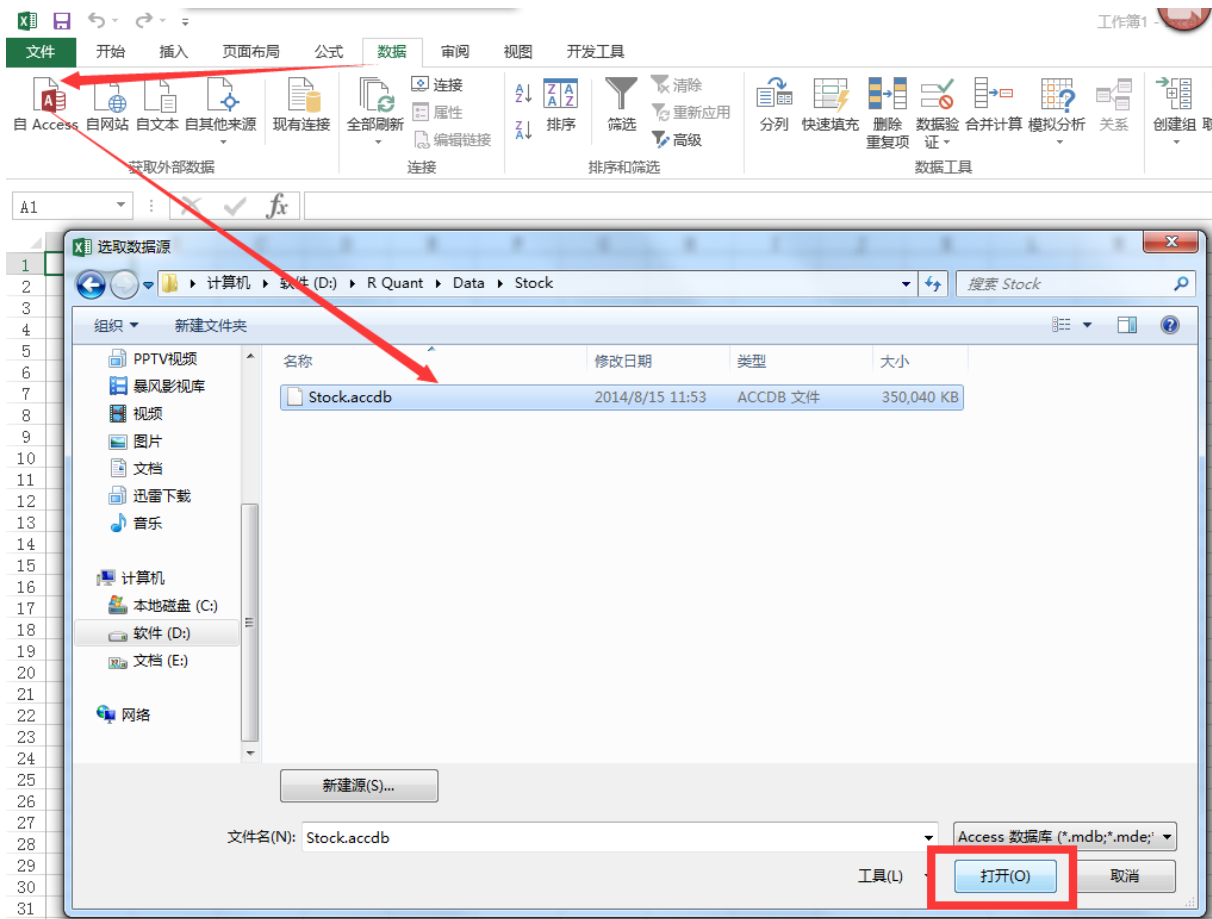
数据的放置位置

☒ 现有工作表(E):
=\$D\$145
☐ 新工作表(N)

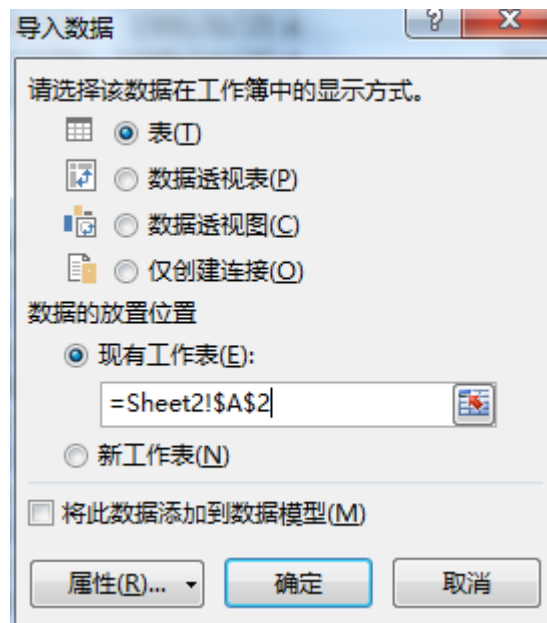
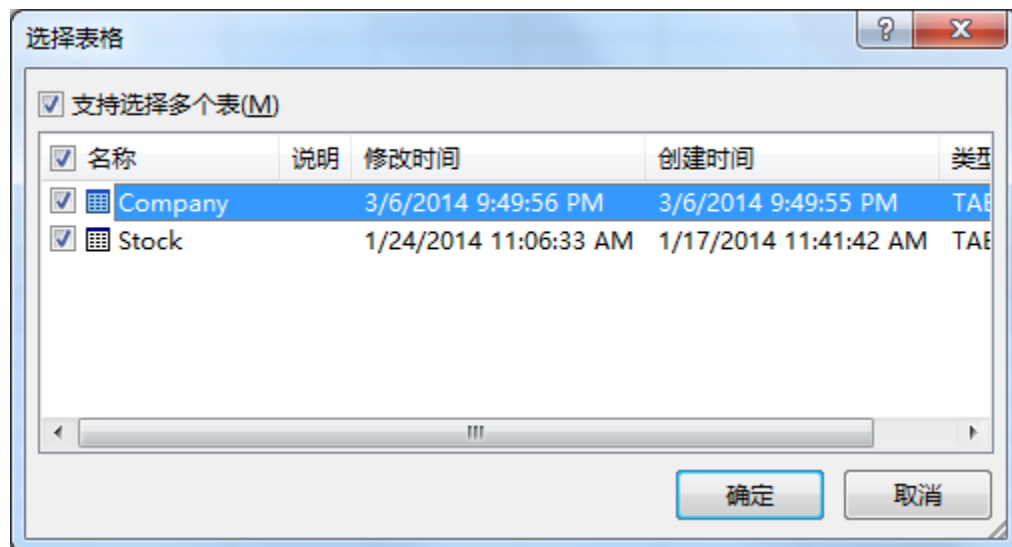
☐ 将此数据添加到数据模型(M)

属性(R)... 确定 取消

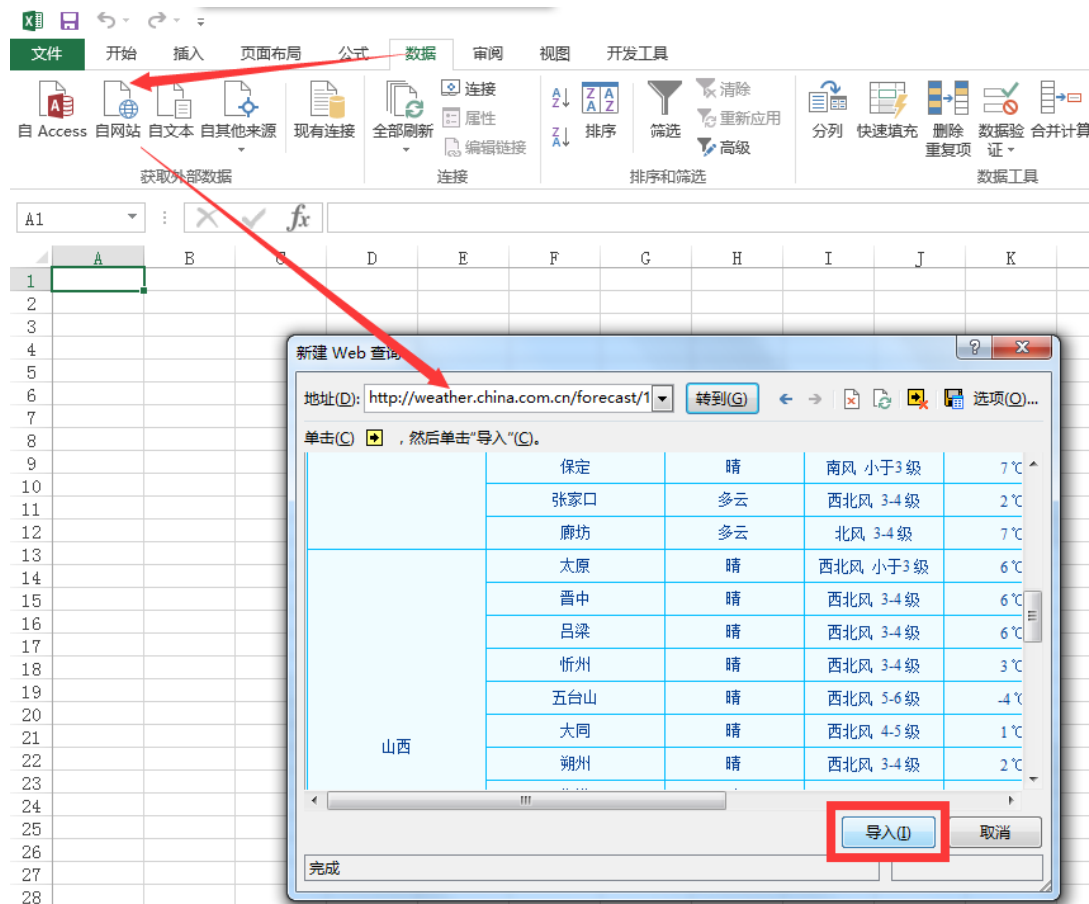
◆ 导入Access数据库数据



◆ 导入Access数据库数据



◆ 导入Internet网页数据



◆ 导入Internet网页数据

导入数据

数据的放置位置

☒ 现有工作表(E):

= \$A\$1

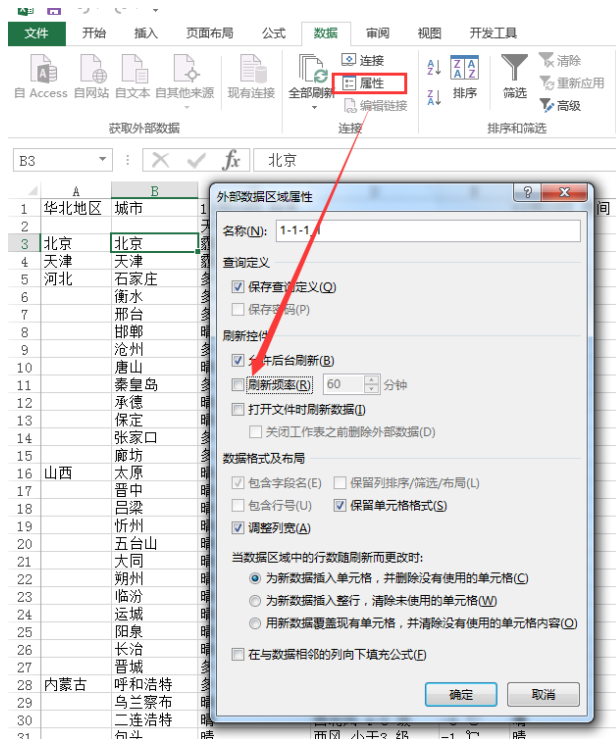
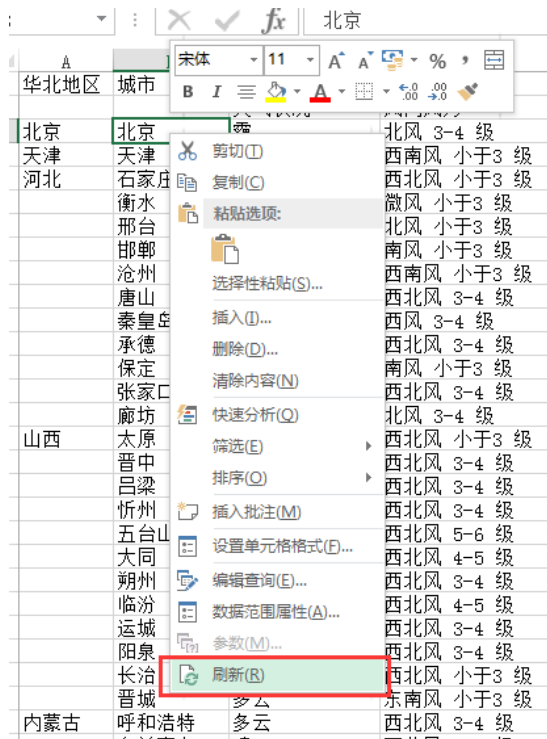
☐ 新工作表(N)

☐ 将此数据添加到数据模型(M)

属性(R)... 确定 取消

		天气状况	风向风力	最高温度	天气状况	风向风力	最低温度
北京	北京	霾	北风 3-4 级	8 ℃	晴	北风 3-4 级	-1 ℃
天津	天津	霾	西南风 小于3 级	7 ℃	霾	北风 4-5 级	-5 ℃
河北	石家庄	多云	西北风 小于3 级	7 ℃	多云	西北风 3-4 级	-2 ℃
	衡水	多云	微风 小于3 级	8 ℃	多云	北风 3-4 级	-4 ℃
	邢台	多云	北风 小于3 级	9 ℃	晴	北风 3-4 级	-1 ℃
	邯郸	晴	南风 小于3 级	11 ℃	晴	北风 3-4 级	0 ℃
	沧州	多云	西南风 小于3 级	6 ℃	多云	北风 3-4 级	-3 ℃
	唐山	晴	西北风 3-4 级	5 ℃	晴	西北风 3-4 级	-8 ℃
	秦皇岛	多云	西风 3-4 级	5 ℃	晴	北风 3-4 级	-5 ℃
	承德	晴	西北风 3-4 级	3 ℃	晴	西北风 3-4 级	-10 ℃
	保定	晴	南风 小于3 级	7 ℃	多云	北风 3-4 级	-5 ℃
	张家口	多云	西北风 3-4 级	2 ℃	多云	西北风 4-5 级	-11 ℃
	廊坊	多云	北风 3-4 级	7 ℃	晴	北风 3-4 级	-3 ℃
山西	太原	晴	西北风 小于3 级	6 ℃	晴	西北风 4-5 级	-10 ℃
	晋中	晴	西北风 3-4 级	6 ℃	晴	西北风 3-4 级	-9 ℃
	吕梁	晴	西北风 3-4 级	6 ℃	晴	西北风 3-4 级	-9 ℃
	忻州	晴	西北风 3-4 级	3 ℃	晴	西北风 3-4 级	-12 ℃
	五台山	晴	西北风 5-6 级	-4 ℃	晴	西北风 5-6 级	-14 ℃
	大同	晴	西北风 4-5 级	1 ℃	晴	西北风 3-4 级	-19 ℃
	朔州	晴	西北风 3-4 级	2 ℃	晴	西北风 3-4 级	-16 ℃
	临汾	晴	西北风 4-5 级	7 ℃	晴	西北风 4-5 级	-5 ℃
	运城	晴	西北风 3-4 级	7 ℃	多云	西北风 3-4 级	-5 ℃
	阳泉	晴	西北风 3-4 级	8 ℃	晴	西北风 3-4 级	-7 ℃
	长治	晴	西北风 小于3 级	6 ℃	晴	西北风 小于3 级	-7 ℃
	晋城	多云	东南风 小于3 级	9 ℃	晴	东南风 小于3 级	-4 ℃

◆ 根据网页内容更新工作表数据

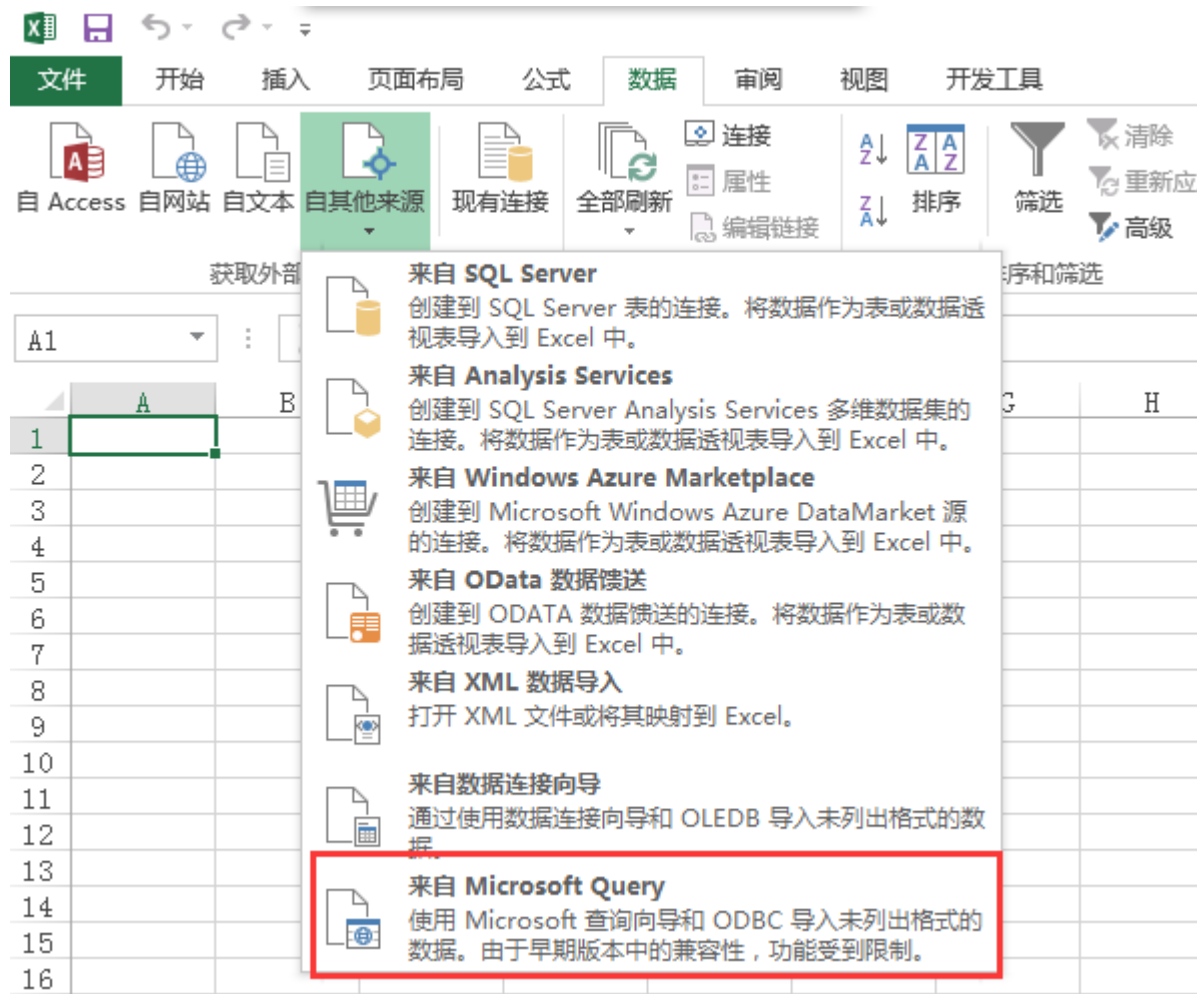


◆ 导入word数据

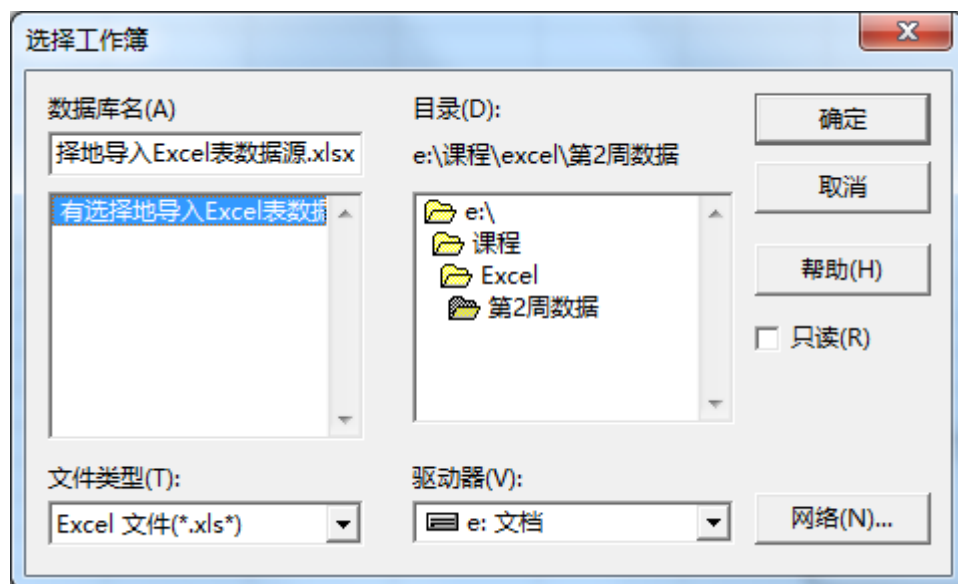
- 通过复制-粘贴
- 将Word文档另存为网页再导入网站数据



◆ 有选择地导入Excel表数据



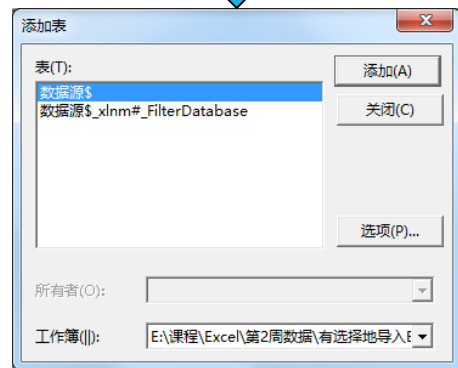
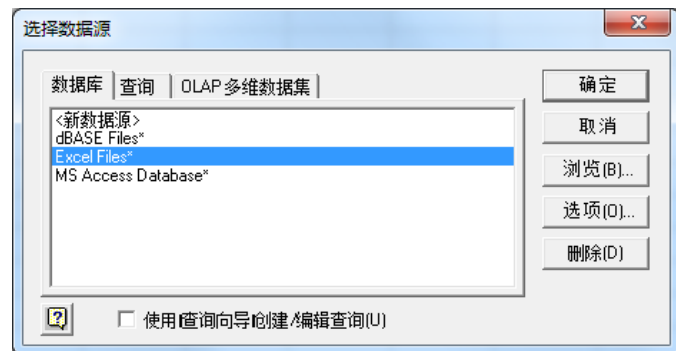
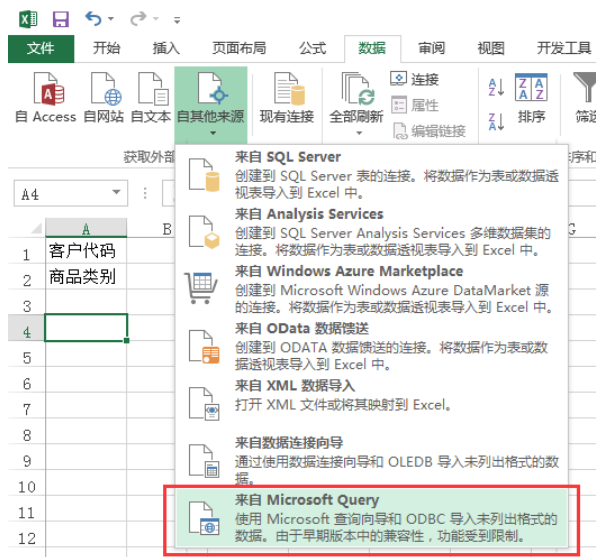
◆ 有选择地导入Excel表数据



	A	B	C	D	E	F	G
1	客户代码	商品类别	销售货号	款式号	数量	单价	销售金额
2	C000002	B	C005154-002	00577300RH	70	111.9888	7839.216
3	C000002	B	C005154-008	00577000LH	100	101.9502	10195.02
4	C000002	B	C005154-014	00576700LH	96	80.8632	7762.8672
5	C000002	B	C007581-001	00587105RL	48	85.9716	4126.6368
6	C000002	B	C007763-001	00587905RR	36	144.639	5207.004
7	C000002	B	C007769-001	00585205RR	24	84.7836	2034.8064
8	C000002	B	C007775-001	00585505RL	24	130.383	3129.192
9	C000002	B	C008329-002	00583305LL	2	168.0228	336.0456
10							
11							
12							
13							

导入外部数据

◆ 在导入数据时创建参数查询



客户代码	C000002					
商品类别	B					
客户代码	商品类别	销售货号	款式号	数量	单价	销售金额
C000002	B	C005154-002	00577300RH	70	111.9888	7839.216
C000002	B	C005154-008	00577000LH	100	101.9502	10195.02
C000002	B	C005154-014	00576700LH	96	80.8632	7762.8672
C000002	B	C007581-001	00587105RL	48	85.9716	4126.6368
C000002	B	C007763-001	00587905RR	36	144.639	5207.004
C000002	B	C007769-001	00585205RR	24	84.7836	2034.8064
C000002	B	C007775-001	00585505RL	24	130.383	3129.192
C000002	B	C008329-002	00583305LL	2	168.0228	336.0456



◆ 数据类型

- 数值
- 日期或时间
- 文本
- 公式
- 错误值与逻辑值

◆ 输入分数

◆ 特殊符号输入

◆ 指针移动

- 使用方向键控制指针移动
- 使用Enter键控制指针移动
- 使用Enter和shift+Enter在指定区域移动
- 使用Tab和shift+Tab在指定区域移动

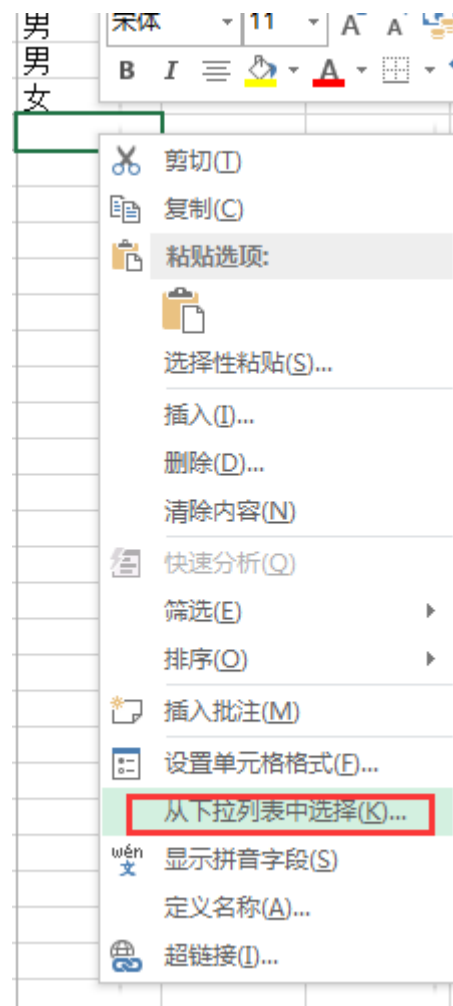
- ◆ 记忆式键入
- ◆ 记忆式键入的适用范围

编辑选项

- ☒ 按 Enter 键后移动所选内容(M)
 - 方向(I): 向下 ▾
- ☐ 自动插入小数点(D)
 - 位数(P): 2 ▴ ▾
- ☒ 启用填充柄和单元格拖放功能(D)
 - ☒ 覆盖单元格内容前发出警告(A)
- ☒ 允许直接在单元格内编辑(E)
- ☒ 扩展数据区域格式及公式(L)
- ☒ 允许自动百分比输入(I)
- ☒ 为单元格值启用记忆式键入(A)
 - ☒ 自动快速填充(F)
- ☐ 用智能鼠标缩放(Z)
- ☒ 发生可能会占用大量时间的操作时警告用户(I)
 - 受影响的单元格达到以下数目时(千)(U): 33,554 ▴ ▾
- ☒ 使用系统分隔符(U)
 - . / : ; ' " < > [] { } ~ , * & % ^ _ ` ~

◆ 从下拉列表中选择输入

- <Alt+下方向键>
- 右键-从下拉列表中选择



◆ 自动填充

- 数值填充
- 日期填充
- 文本填充
- 特殊文本填充
- 公式填充
- 时间序列预测
- 不同工作表填充

☒ 按 Enter 键后移动所选内容(M)

方向(D): 向下

☐ 自动插入小数点(D)

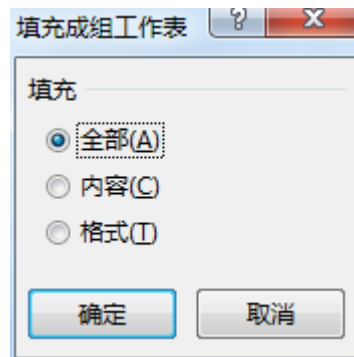
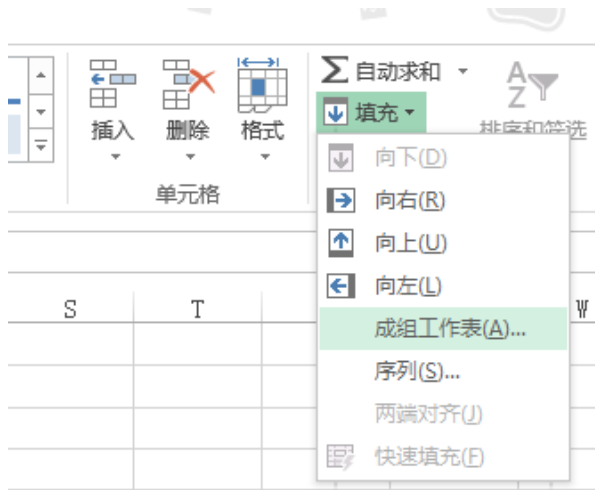
位数(P): 2

☒ 启用填充柄和单元格拖放功能(D)

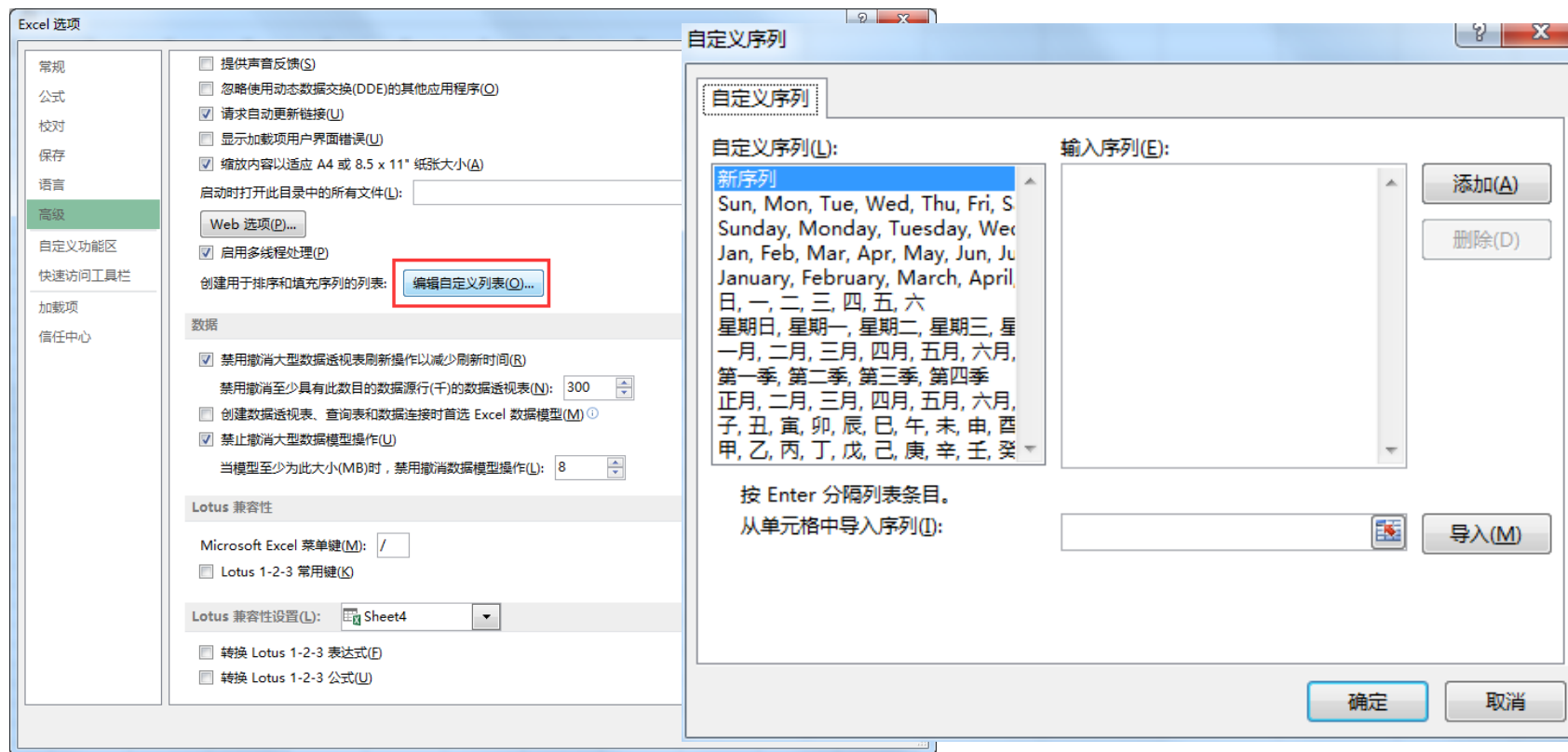
☒ 覆盖单元格内容前发出警告(A)

☒ 允许直接在单元格内编辑(E)

☒ 扩展数据区域格式及公式(L)



◆ 自定义序列



- ◆ 修改数据
- ◆ 移动数据
- ◆ 复制数据
- ◆ 删除数据
- ◆ 查找与替代

◆ 格式错乱的数据，怎么处理

A	B	C	D	E	F	G	H	
PersonID	FirstName	LastName	ZIP	Phone	CallID	Time		
127	^Alexia	#Rasmuss	(ID 127)	en#11422	#718-534-2403	#4252	#01/01/08 12:32	
98	^Brenden	#Rasmuss	(ID 98)	en#10015	#646-812-7298	#4253	#01/02/08 13:17	
94	^Beau	#Ware	(ID 94)	#10013	#646-391-9926	#4254	#01/02/08 14:54	
127	^Alexia	#Ra	(ID 127)	smussen#	11422	#718-534-2403	#4255	#01/03/08 15:32
156	^Jacoby	#Co	(ID 156)	ok#10045	#917-808-6186	#4256	#01/03/08 15:52	
139	^Quinten	#Hubba	(ID 139)	rd#10016	#917-687-2256	#4257	#01/04/08 13:19	
75	^Remington	#Farr	(ID 75)	ell#11693	#917-843-4608	#4258	#01/05/08 09:09	
91	^Pedro	#Pay	(ID 91)	ne#10005	#212-498-6862	#4259	#01/05/08 09:10	
152	^Elianna	#Me	(ID 152)	dina#10308	#646-819-2778	#4260	#01/05/08 09:41	
8	^Darren	#Burn	(ID 8)	s#10120	#212-220-3311	#4261	#01/05/08 16:11	
164	^Rebekah	#Be	(ID 164)	ard#10155	#212-700-7304	#4262	#01/07/08 11:23	
7	^Jaqueline	#Ewing	(ID 7)	#10106	#347-409-5439	#4263	#01/07/08 11:50	
105	^Porter	#Park	(ID 105)	s#11385	#718-483-5325	#4264	#01/07/08 13:31	
163	^Donavan	#Ingra	(ID 163)	m#11224	#718-543-2608	#4265	#01/08/08 14:28	
174	^Charles	#McCoy	(ID 174)	#11220	#347-312-4725	#4266	#01/08/08 15:48	
66	^Yaritza	#An	(ID 66)	draws#	11234	#718-746-4988	#4267	#01/09/08 13:57
61	^Rylie	#Ayala	(ID 61)	#10167	#917-745-5684	#4268	#01/09/08 15:10	
45	^Omari	#Rus	(ID 45)	h#11691	#646-516-3070	#4269	#01/09/08 16:29	
175	^Bridger	#Stephenson	(ID 175)	#11374	#718-792-7253	#4270	#01/10/08 09:41	
2	^Reynaldo	#Ha	(ID 2)	rris#11223	#646-382-3431	#4271	#01/11/08 08:06	
63	^Quentin	#Ra	(ID 63)	ndolph#	11207	#347-625-6847	#4272	#01/11/08 12:10
136	^Deborah	#La	(ID 136)	mb#10302	#718-408-4184	#4273	#01/11/08 12:46	
42	^Stanley	#Townse	(ID 42)	nd#10314	#212-775-3412	#4274	#01/12/08 16:05	
135	^Lena	#Rive	(ID 135)	ra#11361	#212-340-5502	#4275	#01/14/08 08:01	
177	^Aaliyah	#Stephens	(ID 177)	on#10169	#646-354-5674	#4276	#01/14/08 09:09	
186	^Quintin	#McConne	(ID 186)	ll#11413	#917-363-2471	#4277	#01/14/08 11:41	
63	^Quentin	#Rand	(ID 63)	olph#	11207	#347-625-6847	#4278	#01/14/08 15:10
10	^Aliya	#Hoffman	(ID 10)	#11366	#917-613-6421	#4279	#01/15/08 08:44	
90	^Janiah	#Ki	(ID 90)	dd#11220	#347-346-9229	#4280	#01/15/08 13:10	
131	^Desmond	#Knowl	(ID 131)	es#10122	#917-785-7273	#4281	#01/15/08 14:04	
142	^Harrison	#Du	(ID 142)	ke#10112	#718-245-5292	#4282	#01/15/08 16:33	

◆ 明确目标——想要什么数据与格式

全部拆分妥当后，可以按数据域进行排序、过滤，或将数据导入邮件合并程序、网页等。



必须得有电话号码……
这对销售团队至关重要！

编号	名	姓	电话
127	Alexia	Rasmussen	718-534-2403
98	Brenden	Rasmussen	646-812-7298
.....

这个ID域有用处，可以
确保数据的唯一性。

需要将姓名和电话号码分开。

◆ 利用分隔符区分数据列

文本分列向导 - 第 1 步, 共 3 步

文本分列向导判定您的数据具有分隔符。
若一切设置无误, 请单击“下一步”, 否则请选择最合适的数据类型。

原始数据类型

请选择最合适的文件类型:

☒ 分隔符号(D) - 用分隔字符, 如逗号或制表符分隔每个字段

☐ 固定宽度(W) - 每列字段加空格对齐

预览选定数据:

	PersonID	FirstName	LastName	ZIP	Phone	CallID	Time
1	127	Alexia	Rasmuss (ID 127)	en#11422	#718-534-2403	#4252	#01/01/08 12:32
3	98	Brenden	Rasmuss (ID 98)	en#10015	#646-812-7298	#4253	#01/02/08 13:17
4	94	Beau	Ware (ID 94)	#10013	#646-391-9926	#4254	#01/02/08 14:54
5	127	Alexia	Ra (ID 127)	smussen#11422	#718-534-2403	#4255	#01/03/08 15:32

取消 < 上一步(B) 下一步(N) > 完成(F)

文本分列向导 - 第 2 步, 共 3 步

请设置分列数据所包含的分隔符号。在预览窗口内可看到分列的效果。

分隔符号

☐ Tab 键(T)

☐ 分号(M)

☐ 逗号(C)

☐ 空格(S)

☒ 其他(O): #

☐ 连续分隔符号视为单个处理(R)

文本识别符号(Q): "

数据预览(P)

PersonID	FirstName	LastName	ZIP	Phone	CallID	Time
127	Alexia	Rasmuss (ID 127)	en	11422	718-534-2403	4252 01/01/08 12:32
98	Brenden	Rasmuss (ID 98)	en	10015	646-812-7298	4253 01/02/08 13:17
94	Beau	Ware (ID 94)		10013	646-391-9926	4254 01/02/08 14:54
127	Alexia	Ra (ID 127)	smussen	11422	718-534-2403	4255 01/03/08 15:32

取消 < 上一步(B) 下一步(N) > 完成(F)

◆ 清除多余的内容

fx =SUBSTITUTE(B2,"^","")

	B	C	D	E	F	G	H	I
D	FirstName	LastName	ZIP	Phone	CallID	Time		
27	^Alexia	Rasmuss(ID 127)en	11422	718-534-2403	4252	2001-1-8 12:32	Alexia	
98	^Brenden	Rasmuss(ID 98)en	10015	646-812-7298	4253	2001-2-8 13:17		
94	^Beau	Ware(ID 94)	10013	646-391-9926	4254	2001-2-8 14:54		
27	^Alexia	Ra(ID 127)smussen	11422	718-534-2403	4255	2001-3-8 15:32		
56	^Jacoby	Co(ID 156)ok	10045	917-808-6186	4256	2001-3-8 15:52		

fx =CONCATENATE(LEFT(C2,FIND("(",C2)-1),RIGHT(C2,LEN(C2)-FIND("(",C2)))

	B	C	D	E	F	G	H	I	J
D	FirstName	LastName	ZIP	Phone	CallID	Time			
27	^Alexia	Rasmuss(ID 127)en	11422	718-534-2403	4252	2001-1-8 12:32	Alexia	Rasmussen	
98	^Brenden	Rasmuss(ID 98)en	10015	646-812-7298	4253	2001-2-8 13:17			
94	^Beau	Ware(ID 94)	10013	646-391-9926	4254	2001-2-8 14:54			
27	^Alexia	Ra(ID 127)smussen	11422	718-534-2403	4255	2001-3-8 15:32			
56	^Jacoby	Co(ID 156)ok	10045	917-808-6186	4256	2001-3-8 15:52			
39	^Quinten	Hubba(ID 139)rd	10016	917-687-2256	4257	2001-4-8 13:19			

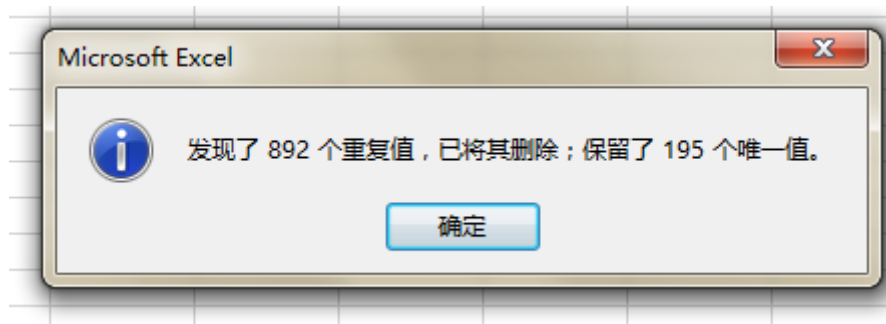
◆ 这样的结果满意吗？——不，有大量重复值

PersonID	FirstName	LastName	Phone
127	Alexia	Rasmussen	718-534-2403
98	Brenden	Rasmussen	646-812-7298
94	Beau	Ware	646-391-9926
127	Alexia	Rasmussen	718-534-2403
156	Jacoby	Cook	917-808-6186
139	Quinten	Hubbard	917-687-2256
75	Remington	Farrell	917-843-4608
91	Pedro	Payne	212-498-6862
152	Elianna	Medina	646-819-2778
8	Darren	Burns	212-220-3311
164	Rebekah	Beard	212-700-7304
7	Jaqueline	Ewing	347-409-5439
105	Porter	Parks	718-483-5325
163	Donavan	Ingram	718-543-2608
174	Charles	McCoy	347-312-4725
66	Yaritza	Andrews	718-746-4988
61	Rylie	Ayala	917-745-5684
45	Omari	Rush	646-516-3070
175	Bridger	Stephenson	718-792-7253
2	Reynaldo	Harris	646-382-3431
63	Quentin	Randolph	347-625-6847
136	Deborah	Lamb	718-408-4184
42	Stanley	Townsend	212-775-3412
135	Lena	Rivera	212-340-5502
177	Aaliyah	Stephenson	646-354-5674
186	Quintin	McConnell	917-363-2471
63	Quentin	Randolph	347-625-6847
10	Aliya	Hoffman	917-613-6421
90	Janiah	Kidd	347-346-9229
131	Desmond	Knowles	917-785-7273
142	Harrison	Duke	718-245-5292
94	Beau	Ware	646-391-9926
158	Alessandr	Casey	917-647-6620

排序后

PersonID	FirstName	LastName	Phone
1	Jaylen	Christensen	646-376-4466
1	Jaylen	Christensen	646-376-4466
1	Jaylen	Christensen	646-376-4466
1	Jaylen	Christensen	646-376-4466
1	Jaylen	Christensen	646-376-4466
1	Jaylen	Christensen	646-376-4466
1	Jaylen	Christensen	646-376-4466
2	Reynaldo	Harris	646-382-3431
2	Reynaldo	Harris	646-382-3431
2	Reynaldo	Harris	646-382-3431
2	Reynaldo	Harris	646-382-3431
2	Reynaldo	Harris	646-382-3431
2	Reynaldo	Harris	646-382-3431
2	Reynaldo	Harris	646-382-3431
2	Reynaldo	Harris	646-382-3431
3	Yandel	Franklin	718-510-6575
3	Yandel	Franklin	718-510-6575
3	Yandel	Franklin	718-510-6575
3	Yandel	Franklin	718-510-6575
4	Lia	Short	718-664-9574
4	Lia	Short	718-664-9574
4	Lia	Short	718-664-9574
4	Lia	Short	718-664-9574
4	Lia	Short	718-664-9574
4	Lia	Short	718-664-9574
5	Cael	Glover	646-623-7673

◆ 删除重复值



- ◆ **Dataguru（炼数成金）**是专业数据分析网站，提供教育，媒体，内容，社区，出版，数据分析业务等服务。我们的课程采用新兴的互联网教育形式，独创地发展了逆向收费式网络培训课程模式。既继承传统教育重学习氛围，重竞争压力的特点，同时又发挥互联网的威力打破时空限制，把天南地北志同道合的朋友组织在一起交流学习，使到原先孤立的学习个体组合成有组织的探索力量。并且把原先动辄成千上万的学习成本，直线下降至百元范围，造福大众。我们的目标是：低成本传播高价值知识，构架中国第一的网上知识流转阵地。
- ◆ 关于逆向收费式网络的详情，请看我们的培训网站 <http://edu.dataguru.cn>



Thanks

FAQ时间