

## Заключительный проект в Power Query: объединяем все вместе

Это фрагмент книги [Гил Равив. Power Query в Excel и Power BI: сбор, объединение и преобразование данных.](#)

[Предыдущий раздел](#)

[К содержанию](#)

Примите поздравления. Вы приступаете к последней главе. Пришло время подтвердить ваши навыки по обработке данных и проверить возможности в заключительном проекте.

### *Отчет о доходах компании Wide World Importers*

Вы – специалист по данным в компании Wide World Importers. После атаки хакеров вас попросили восстановить отчет о доходах компании на основе старых данных. Ваш предшественник вел отчеты в трех файлах Excel за 2015-2017 годы. В каждом из них содержалось 12 листов – по одному на каждый месяц. А выручка 2018 года была экспортирована в формате csv. Нужно импортировать данные с 36 листов, объединить их с данными 2018 года и создать единый отчет.

Загрузите папку C14E01 и целевую таблицу в файле C14E01 - Goal.xlsx. Изучите файлы 2015.xlsx, 2016.xlsx и 2017.xlsx. Каждый лист содержит сводную таблицу формата 2\*3:

	A	B	C	D	E	F	G	H	I
1			ColorName	SupplierCategoryName	SupplierName				
2			Black					Blue	
3			Clothing Supplier	Novelty Goods Supplier		Packaging Toy Suppl	Clothing S	Packaging Toy	
4	Customer	CityName	Fabrikam, Inc.	Graphic Design Institut	The Phone Compan	Litware, Ir	Northwin	Fabrikam, Litware, Ir	Nor
5	Computer	Beaver Ba	4471,2	134,55				195,5	
6		Bombay Beach						431,25	
7		Caruthers	616,4	59,8				138	2208
8		Conesus L	2235,6	104,65					
9		Creelsboro		134,55					
10		Darling	201,25	104,65					
11		Flowella	1490,4	164,45					
12		Greig		104,65			379,5	430,1	
13		Humansville				2208			
14		Konnerl	1107,45	119,6	212,75		2949,75		2461

Рис. 1. Исходные данные за 2015–2017 гг. в файлах Excel

В строках представлены категории и города клиентов, а в столбцах – цвет продукта, категория и наименование поставщика.

Данные о доходах за 2018 году находятся в файле со значениями, разделенными запятыми (CSV):



```
Attribute, Value
Date, 2018 Jan
Color, Black
Supplier Category, Clothing Supplier
Supplier Name, "Fabrikam, Inc."
Customer Category, Computer Store
City, Basin
Revenue, 1987.2
Date, 2018 Jan
Color, Black
Supplier Category, Clothing Supplier
Supplier Name, "Fabrikam, Inc."
Customer Category, Computer Store
City, Biorka
Revenue, 3726
Date, 2018 Jan
Color, Black
Supplier Category, Clothing Supplier
```

Рис. 2. Исходные данные за 2018 г. в файле \*.csv

Попробуйте решить проблему самостоятельно. Потратьте пару часов, пытаясь это сделать. Если вам потребуется помощь, можно прочитать текст ниже.

Не знаете, как поступить? Вы можете разбить задачу на этапы:

1. Импортируйте все файлы Excel из папки.
2. Создайте новый столбец с 36 строками объектов *Table* (по таблице на лист в трех файлах).
3. Для каждого листа удалите первую строку, столбец и строку итогов.
4. Примените функцию *FnUnpivotSummarizedTable*.
5. Импортируйте файл 2018.csv и выполните сведение столбца.
6. Соедините доходы за четыре года. Обратите внимание, что даты представлены по-разному.

### Часть 1. Подготовка таблиц

Начнем с преобразования сводных помесечных данных за 2015-2017 гг. в таблицу фактов.

Откройте новую книгу Excel. Пройдите *Данные* → *Получить данные* → *Из файла* → *Из папки*.

Выберите путь для папки *C:\Data\C14\C14E01* и щелкните *Открыть*. В следующем окне щелкните *Преобразовать данные*. Откроется окно редактора Power Query.

Выберите элемент управления фильтром столбца *Name* и снимите галочку с файла 2018.csv.

Щелкните *Ок*. Переименуйте запрос из *C14E01* в *Revenues*. В заголовке столбца *Content* щелкните на значке *Объединить файлы*. В окне *Объединить файлы* выделите папку *Параметр 1 [12]* и щелкните *Ок*. Power Query автоматически создаст ряд запросов:

Запросы [5]

- Преобразовать файл из Revenues [2]
- Вспомогательные запросы [3]
- Параметр1 (Пример файла)
- Пример файла
- Преобразовать файл
- Преобразовать пример файла
- Другие запросы [1]
- Revenues

Формула: = Table.TransformColumnTypes("#Столбец расширенной таблицы1",{"Source.Name", type tex

	Source.Name	Name	Data	Item	
1	2015.xlsx	Jan	Table	Jan	She
2	2015.xlsx	Feb	Table	Feb	She
3	2015.xlsx	Mar	Table	Mar	She
4	2015.xlsx	Apr	Table	Apr	She
5	2015.xlsx	May	Table	May	She
6	2015.xlsx	Jun	Table	Jun	She
7	2015.xlsx	Jul	Table	Jul	She
8	2015.xlsx	Aug	Table	Aug	She
9	2015.xlsx	Sep	Table	Sep	She
10	2015.xlsx	Oct	Table	Oct	She
11	2015.xlsx	Nov	Table	Nov	She

Рис. 3. Импорт всех листов из всех Excel-файлов из папки

Убедитесь, что активен запрос *Revenues* и удалите последние три столбца. Удалите расширение *xlsx* из столбца *Source.Name* для извлечения года и переименуйте столбец в *Year*. (Имеется несколько способов удалить расширение файла. Если вы не уверены, как это сделать, вернитесь к [главе 3](#).) Переименуйте столбец с *Name* на *Month*. В столбце *Data* находятся объекты *Table*. Если кликнуть левой кнопкой мыши на пустое пространство справа от слова *Table*, в нижней части окна увидим предварительный просмотр таблицы:

fx = Table.RenameColumns("#Замененное значение",{{"Name", "Month"}})

Year	Month	Data
2015	Jan	Table
2015	Feb	Table
2015	Mar	Table
2015	Apr	Table
2015	May	Table

Column1	Column2	Column3	Column4	Column5	Column6	Column7
	null	null	ColorName	SupplierCategoryName	SupplierName	null
	null	null	Black	null	null	null
	null	null	Clothing Supplier	Novelty Goods Supplier	null	Packaging Supplier
CustomerCategoryName	CityName	Fabrikam, Inc.	Graphic Design Institute	The Phone Company	Litware, Inc.	Northwind
Computer Store	Beaver Bay	4471,2	134,55	null	null	null
	null	Bombay Beach	null	null	null	null
	null	Caruthers	616,4	59,8	null	null
	null	Conesus Lake	2235,6	104,65	null	null
	null	Creelsboro	null	134,55	null	null
	null	Darling	201,25	104,65	null	null
	null	Flowella	1490,4	164,45	null	null
	null	Greig	null	104,65	null	null

Рис. 4. Первая строка каждой таблицы не несет информации и мешает отмене сведения столбца. Следует удалить первую строку и общие итоги в последнем столбце и последней строке. Пройдите *Добавление столбца* → *Настраиваемый столбец*. Настройте окно *Настраиваемый столбец*:

### Настраиваемый столбец

Добавьте столбец, который вычисляется на основании других столбцов.

Имя нового столбца

Настраиваемая формула столбца  ⓘ

Доступные столбцы

- Year
- Month
- Data

<< Вставить

Сведения о формулах Power Query

✔ Синтаксические ошибки не обнаружены.

Рис. 5. Параметры окна *Настраиваемый столбец*

Нажмите Ok. Новый столбец возвращает объект *Table* такой же, как и в столбце *Data* но без первой строки. Можете проверить, кликнув левой кнопкой мыши на пустое пространство справа от слова *Table* в столбце *Skip First Row*. Удалите столбец *Data*. Пройдите *Добавление столбца* → *Настраиваемый столбец*. В окне *Настраиваемый столбец* в поле *Имя нового столбца* введите *No Grand Total Column*. В поле *Формула пользовательского столбца* введите:

```
= Table.RemoveColumns(
    [Skip First Row], {
        List.Last(Table.ColumnNames([Skip First Row]))
    }
)
```

)

Эта формула удаляет последний столбец. С помощью функции Table.ColumnNames вы получаете имена столбцов таблицы в столбце *Skip First Row*, а List.Last возвращает имя последнего столбца. Опять же, можете проверить, что из всех таблиц удалены столбцы *Column19*, содержавшие общие итоги.

Удалите столбец *Skip First Row*. Пройдите *Добавление столбца* → *Настраиваемый столбец*. В окне *Настраиваемый столбец* в поле *Имя нового столбца* введите *Summarized Table*. В поле *Формула пользовательского столбца* введите:

```
= Table.RemoveLastN([No Grand Total Column], 1)
```

Проверить удаление последней строки не удастся, так как предварительный просмотр таблицы показывает лишь несколько первых строк. Удалите столбец *No Grand Total Column*.

Альтернативный способ удаления первой строки, столбца и строки итогов состоит в написании пользовательской функции. Находясь в редакторе Power Query пройдите *Главная* → *Создать источник* → *Другие источники* → *Пустой запрос*. В строку формул введите код:

```
(inputTable) =>
    let
        SkipFirstRow = Table.Skip(inputTable, 1),
        NoGrandTotalColumn = Table.RemoveColumns(
            SkipFirstRow, {
                List.Last(Table.ColumnNames(SkipFirstRow))
            }
        ),
        Result = Table.RemoveLastN(NoGrandTotalColumn, 1)
    in
        Result
```

Назовите запрос *FnCleanSummarizedTable*. Создайте дубль запроса *Revenues*. В запросе *Revenues (2)* удалите шаги от *Добавлен пользовательский объект* до последнего. Пройдите *Добавление столбца* → *Вызвать настраиваемую функцию*:

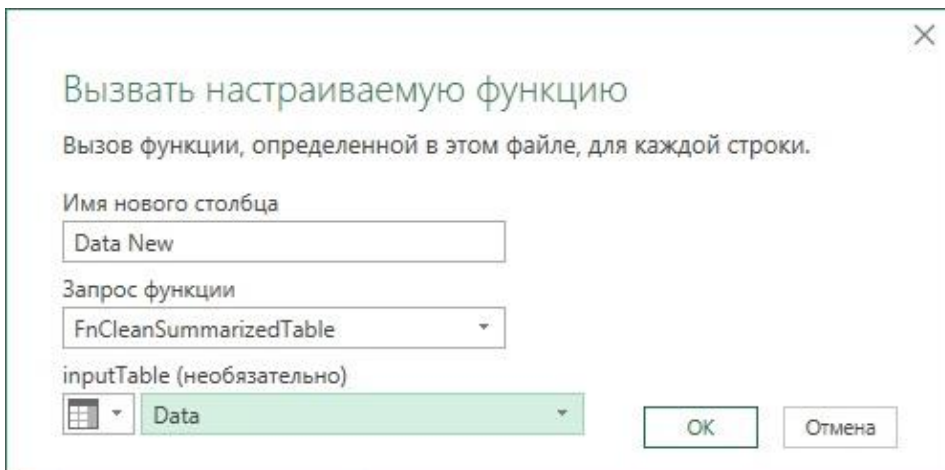


Рис. 6. Удаление первой строки и итогов с помощью пользовательской функции

## Часть 2. Вызов функции отмены сведения столбца

Для отмены сведения столбцов применим пользовательскую функцию *FnUnpivotSummarizedTable*, рассмотренную [ранее](#). Пройдите *Главная* → *Заккрыть и загрузить в...* → *Только создать подключение*. Сохраните файл под именем *C14E01 - Result.xlsx*. Загрузите файл *C07E03 - Solution.xlsx* с сайта [Microsoft Press](#) или файл *C07E02 - Result.xlsx* из [архива](#). Откройте файл Excel. На панели *Запросы и подключения* щелкните правой кнопкой мыши на *FnUnpivotSummarizedTable* и выберите *Копировать*. Вернитесь в файл *C14E01 - Result.xlsx*. Откройте редактор PQ. Правой кнопкой мыши щелкните где-либо в фоновой области панели *Запросы* и выберите *Вставить*.

В качестве альтернативы, находясь в редакторе PQ файла C14E01 - Result.xlsx, пройдите *Главная* → *Создать источник* → *Другие источники* → *Пустой запрос*. А затем *Главная* → *Расширенный редактор*. Выделите и удалите всё содержимое в окне *Расширенный редактор*. Вставьте код:

```
let
```

```
    Источник = (Source, RowFields, ColumnFields, ValueField)=>
```

```
let
```

```
    #"Filled Down" = Table.FillDown(Source, List.FirstN(Table.ColumnNames(Source),  
List.Count(RowFields) - 1)),
```

```
    #"Merged Columns" = Table.CombineColumns(#"Filled Down",  
List.FirstN(Table.ColumnNames(#"Filled Down"), List.Count(RowFields)),  
Combiner.CombineTextByDelimiter(":", QuoteStyle.None),"Merged"),
```

```
    #"Transposed Table" = Table.Transpose(#"Merged Columns"),
```

```
    #"Filled Down1" = Table.FillDown(#"Transposed Table",List.FirstN(Table.ColumnNames(#"Transposed  
Table"), List.Count(ColumnFields)-1)),
```

```
    #"Promoted Headers" = Table.PromoteHeaders(#"Filled Down1", [PromoteAllScalars=true]),
```

```
    #"Unpivoted Other Columns" = Table.UnpivotOtherColumns(#"Promoted  
Headers",List.FirstN(Table.ColumnNames(#"Promoted Headers"), List.Count(ColumnFields)) ,  
"Attribute", ValueField),
```

```
    #"Split Column by Delimiter" = Table.SplitColumn(#"Unpivoted Other Columns", "Attribute",  
Splitter.SplitTextByDelimiter(":", QuoteStyle.Csv), RowFields),
```

```
    #"Renamed Columns" = Table.RenameColumns(#"Split Column by Delimiter",  
List.Zip({List.FirstN(Table.ColumnNames(#"Split Column by Delimiter"), List.Count(ColumnFields)),  
ColumnFields})),
```

```
    #"Changed Type" = Table.TransformColumnTypes(#"Renamed Columns",{ValueField, type number})
```

```
in
```

```
    #"Changed Type"
```

```
in
```

```
    Источник
```

Нажмите *Готово*. Переименуйте *Запрос1* в *FnUnpivotSummarizedTable*.

Итак, вы тем или иным способом создали функцию *FnUnpivotSummarizedTable*. На панели *Запросы* выберите *Revenues*. Пройдите *Добавление столбца* → *Настраиваемый столбец*. В окне *Настраиваемый столбец* введите имя столбца *Unpivoted* и код:

```
= FnUnpivotSummarizedTable([Summarized Table], {"Customer Category", "City"}, {"Color", "Supplier  
Category", "Supplier Name"}, "Revenue")
```

Удалите столбец *Summarized Table*. Разверните столбец *Unpivoted*. На панели *Развернуть* отмените флажок *Использовать имя исходного столбца в качестве префикса* и щелкните *Ок*. В результате в панели предварительного просмотра можно увидеть, что для всех рабочих листов операция отмены свертывания столбцов выполнена корректно.



Year	Month	Color	Supplier Category	Supplier Name	Customer Category	City	Revenue
2015	Jan	Black	Clothing Supplier	Fabrikam, Inc.	Computer Store	Beaver Bay	4471,2
2015	Jan	Black	Clothing Supplier	Fabrikam, Inc.	Computer Store	Caruthers	616,4
2015	Jan	Black	Clothing Supplier	Fabrikam, Inc.	Computer Store	Coneus Lake	2235,6
2015	Jan	Black	Clothing Supplier	Fabrikam, Inc.	Computer Store	Darling	201,25
2015	Jan	Black	Clothing Supplier	Fabrikam, Inc.	Computer Store	Flowella	1490,4
2015	Jan	Black	Clothing Supplier	Fabrikam, Inc.	Computer Store	Kopperl	1107,45
2015	Jan	Black	Clothing Supplier	Fabrikam, Inc.	Computer Store	La Bolt	2235,6
2015	Jan	Black	Clothing Supplier	Fabrikam, Inc.	Computer Store	Lompoc	281,75
2015	Jan	Black	Clothing Supplier	Fabrikam, Inc.	Computer Store	Nitro	1242
2015	Jan	Black	Clothing Supplier	Fabrikam, Inc.	Computer Store	Panton	2318,4
2015	Jan	Black	Clothing Supplier	Fabrikam, Inc.	Computer Store	Pastura	859,05
2015	Jan	Black	Clothing Supplier	Fabrikam, Inc.	Computer Store	Phelps City	281,75
2015	Jan	Black	Clothing Supplier	Fabrikam, Inc.	Computer Store	Switz City	40,25
2015	Jan	Black	Clothing Supplier	Fabrikam, Inc.	Computer Store	Tolchester Beach	1242
2015	Jan	Black	Clothing Supplier	Fabrikam, Inc.	Computer Store	Wounded Knee	110,4

Рис. 7. Таблица доходов за 2015-2017 гг.

Теперь можно обработать доходы за 2018.

### Часть 3. Последовательность свертывания столбцов для доходов 2018 года

Займемся файлом 2018.csv. Этот файл содержит разделенные запятыми значения пар атрибут/значение, и можно предположить, что каждая многострочная запись о доходах начинается с атрибута *Date* (см. рис. 2).

Находясь в редакторе PQ пройдите *Главная* → *Создать источник* → *Файл* → *Текстовый или CSV-файл*. Выберите файл 2018.csv и нажмите *Импорт*. В окне 2018.CSV нажмите *Ок*. Переименуйте запрос в *2018 Revenues*. Пройдите *Преобразование* → *Использовать первую строку в качестве заголовка*. Пройдите *Добавление столбца* → *Столбец индекса*. Пройдите *Добавление столбца* → *Условный столбец*. Настройте окно:

**Добавление условного столбца**

Добавьте условный столбец, который вычисляется из других столбцов или значений.

Имя нового столбца: Row ID

Если	Имя столбца	Оператор	Значение	То	Вывод
Если	Attribute	равно	Date	To	Индекс

Добавить предложение

В противном случае

ABC 123	null
---------	------

ОК Отмена

Рис. 8 Окно *Добавление условного столбца*

Щелкните *Ок*. Выберите столбец *Row ID* и пройдите *Преобразование* → *Заполнить* → *Вниз*. Удалите столбец *Индекс*. Выберите столбец *Attribute* и пройдите *Преобразование* → *Столбец сведения*. В окне *Столбец сведения* в поле *Столбец значений* выберите *Value*. Раскройте раздел *Расширенные параметры*. В поле *Функция агрегированного значения* выберите *Не агрегировать*. Щелкните *Ок*. Удалите столбец *Row ID*.

Теперь можно объединить результаты за 2015-2017 гг. и за 2018 г.

### Часть 4. Объединение доходов

Теперь имеются два запроса, причем *Revenues* включает столбцы *Year* и *Month*, а *2018 Revenues* – столбец *Date*. Выберите запрос *Revenues*. Выделите первые два столбца, *Year* и *Month*, и пройдите *Преобразование* → *Объединить столбцы*. В окне *Объединить столбцы* выберите разделитель пробел, введите *Date* в поле *Имя нового столбца*, щелкните *Ок*. Пройдите *Главная* → *Добавить запросы* → *Добавить запросы*. В окне *Добавление* в качестве второй выберите таблицу *2018 Revenues*. Щелкните *Ок*.

Измените тип столбца *Revenue* на *Десятичное число*, а столбца *Date* – на *Дата*. Загрузите запросы: *Revenues* – *Только создать подключение* и *Добавить эти данные в модель данных*, а запрос *2018 Revenues* – *Только создать подключение*. Файл решения C14E01 - Solution.xlsx.

### Сравнение таблиц и отслеживание действий хакеров

В предыдущем упражнении вы восстановили отчет о доходах после кибератаки. Согласно исследованию группы по кибербезопасности, злоумышленники подделали значения доходов в хранилище данных. Ваша задача состоит в сравнении доходов в вашем отчете с измененными данными в файле C14E02 - Compromised.xlsx для определения, какие записи подделаны хакерами. Хотя найти различия можно с помощью сводных таблиц или мер DAX, попытайтесь решить задачу с помощью Power Query.

Вот несколько подсказок, которые помогут вам:

1. Хакеры внесли изменения в столбец *Revenue*. Можно объединить две таблицы для нахождения несоответствий. Используйте все столбцы, кроме столбца *Revenue*, для слияния, а затем разверните столбец *Revenue* из второй таблицы. Примените фильтр к двум столбцам *Revenue* для нахождения различий.
2. Чтобы найти новые строки, добавленные хакерами, можно объединить таблицу из вашего отчета с скомпрометированной таблицей. Тип соединения – анти-соединение справа. Таким образом, будут обнаружены строки, которые есть в скомпрометированной таблице и нет в вашем отчете.

Импортируйте таблицу *Revenues* из файла C14E01 - Solution.xlsx. Переименуйте запрос в *Correct*. Импортируйте таблицу *Revenues* из файла C14E02 - Compromised.xlsx. Переименуйте запрос в *Compromised*. На панели *Запросы* выберите *Correct*. Пройдите *Главная* → *Объединить запросы* → *Объединить запросы в новый*. В окне *Слияние* в таблице *Correct*, удерживая нажатой клавишу <Ctrl>, выбирайте столбцы *Date*, *Color*, *Supplier Category*, *Supplier Name*, *Custom Category* и *City*. Обязательно сохраняйте порядок следования столбцов. В качестве второй выберите таблицу *Compromised*. В таблице *Compromised*, удерживая нажатой клавишу <Ctrl>, выбирайте поля *Date*, *Color*, *Supplier Category*, *Supplier Name*, *Custom Category* и *City*.

В поле *Тип соединения* выберите *Внутреннее (только совпадающие строки)*. Щелкните *Ок*. Переименуйте новый запрос в *Compromised Rows* и разверните столбец *Compromised*. На панели *Развернуть* отмените выбор всех столбцов, кроме *Revenue*. Снимите галочку *Использовать исходное имя столбца как префикс*. Переименуйте новый столбец в *Compromised.Revenue*.

Чтобы отфильтровать строки, в которых доходы в столбцах *Revenue* и *Compromised.Revenue* не равны необходимо применить фильтр, который недоступен в пользовательском интерфейсе. Можно сгенерить код фильтра через пользовательский интерфейс, а затем изменить его в строке формул. Щелкните мышью на элементе фильтра в заголовке столбца *Revenue* и выберите *Числовые фильтры* → *Не равно* → 12345. В строке формул увидим:

```
= Table.SelectRows("#Переименованные столбцы", each [Revenue] <> 12345)
```

Измените формулу на

```
= Table.SelectRows("#Переименованные столбцы", each [Revenue] <> [Compromised.Revenue])
```

Остались только строки с различными значениями в *Revenue* и *Compromised.Revenue*.

	Date	Color	Supplier Category	Supplier Name	Custom Category	City	Revenue	Compromised.Revenue
1	01.12.2015	White	Novelty				149,5	-1000
2	01.01.2017	White	Novelty				119,6	-1000
3	01.07.2017	Blue	Clothing				276	0
4	01.07.2017	Blue	Packaging				7969,5	10000000
5	01.07.2017	Blue	Packaging				8970	0
6	01.07.2017	Blue	Packaging				3967,5	10000000
7	01.07.2017	Light Brown	Clothing				2484	10000000

Рис. 9. Строки с различными значениями в *Revenue* и *Compromised.Revenue*

### Обнаружение следов деятельности хакеров во взломанной таблице

Теперь найдем строки в скомпрометированной таблице, которые были добавлены хакерами и не найдены в исходной таблице. Продолжите предыдущее упражнение. На панели *Запросы* выберите *Correct*. Пройдите *Главная* → *Объединить запросы* → *Объединить запросы в новый*. В окне *Слияние* в таблице *Correct*, удерживая нажатой клавишу <Ctrl>, выбирайте столбцы *Date*, *Color*, *Supplier Category*, *Supplier Name*, *Custom Category* и *City*. Обязательно сохраняйте порядок следования столбцов. В качестве второй выберите таблицу *Compromised*. В ней удерживая нажатой клавишу <Ctrl>, выбирайте столбцы *Date*, *Color*, *Supplier Category*, *Supplier Name*, *Custom Category* и *City*.

В поле *Тип соединения* выберите *Анти-соединение справа (только строки во второй таблице)*. Щелкните *Ок*. Переименуйте запрос в *Hacker's Instructions*. Удалите все столбцы кроме *Compromised*. Нажмите значок развернуть рядом с заголовком *Compromised*. Извлеките все поля, но снимите галочку *Использовать исходное имя столбца как префикс*. Найдите сообщение хакера.

	Date	Color	Supplier Category	Supplier Name	Customer Category	City	1.2 Revenue
1	11.04.2018	The	hacker	was	here	today	0
2	11.04.2018	If	you want	your	data	back	0
3	11.04.2018	Call	this	number	for	instructions	1234567890

Рис. 10. Сообщение хакера

Загрузите запрос *Hacker's Instructions* в таблицу на лист Excel. Файл решения C14E02 - Solution.xlsx.