

Миколаївський державний коледж економіки та харчових технологій

**Семінар-тренінг
«Обробка даних в Excel
з використанням надбудови Power Query»**

Викладач: Кореновський Г.П.

Миколаїв
2019

Обробка даних в Excel з використанням надбудови Power Query

Мета: вдосконалювати знання викладачів при створенні запитів для підключення та імпорту даних в Excel з різних зовнішніх джерел (файлів різних форматів, баз даних, Інтернету і т. д.) і трансформації отриманих таблиць з подальшим їх аналізом.

Забезпечення: інструкційно-методичний матеріал; електронні таблиці Microsoft Excel з інтегрованим компонентом Power Query; програма керування навчальним процесом NetOp School.

Хід семінару-тренінгу

Завдання

Припустимо, що кожен день ми вивантажуємо з якої-небудь корпоративної програми або бази даних текстовий файл зі звітом такого виду:

```
Звіт по продажах за січень-березня 2019 р
Дата створення: 12052019
Користувач: Пупкін

Код замовлення; Дата; Місто; Бренд; Кількість; Вартість
ST5; 20-01-2019; Чернігів; Peugeot; 6; 10,849
SW6; 30-01-2019; м.Чернівці; Mitsubishi; 1; 15,542
KJ8; 25-01-2019; Полтава; Renault; 8; 88,846
OZ5; 04-01-2019; борислав; Mitsubishi; 8; 97,127
U0I; 31-01-2019; Херсон; Toyota; 9; 42,834
8ZE; 05-01-2019; Черкаси; Subaru; 6; 21,575
UR7; 09-01-2019; МАРІУПОЛЬ; Peugeot; 8; 24,843
ST5; 20-01-2019; м.Чернігів; Peugeot; 6; 1,849
РАЗОМ = 294,465

-----

LNR; 11-02-2019; ЛУЦЬК; Saab; 5; 69,134
KLH; 27-02-2019; Рівне; Fiat; 2; 58,933
J6X; 17-02-2019; м.Житомир; Honda; 8; 9,342
H0P; 03-02-2019; Вінниця; Skoda; 2; 56,566
E7I; 09-02-2019; чернівці; Mini; 1; 71,661
A0S; 19-02-2019; Умань; Toyota; 6; 82,200
0BM; 16-02-2019; Харків; Fiat; 4; 17,882
0BM; 16-02-2019; Харків; Fiat; 4; 17,882
0BM; 16-02-2019; Харків; Fiat; 4; 17,882
P2O; 11-02-2019; Херсон; Hyundai; 9; 80,457
3G8; 30-02-2019; м.Миколаїв; Suzuki; 5; 30,952
РАЗОМ = 512,891

-----

CFJ; 07-03-2019; Дніпро; Ford; 3; 17,325
L5I; 12-03-2019; Запоріжжя; Saab; 3; 88,439
```

Зверніть увагу на такі моменти:

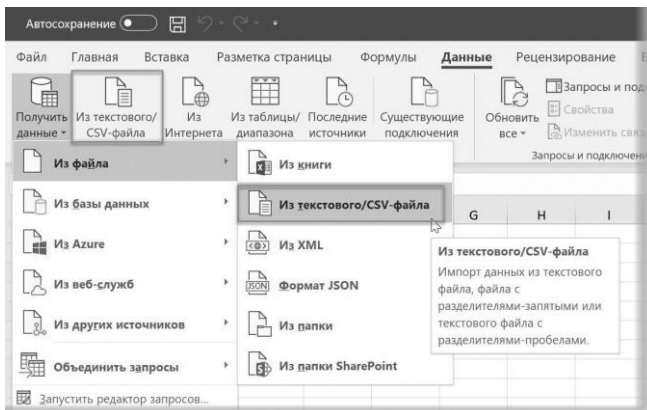
- У верхній частині є непотрібна шапка з технічними даними з вивантаження.
- Блоки даних розділені «для краси» рядком з дефісів (причому їх кількість не скрізь однакова).
- В кінці кожного блоку є підсумковий рядок із сумою (яка нам не потрібна, тому що в майбутньому ми будемо будувати зведену таблицю за цими даними).
- Дати в форматі, який не розуміє Excel (через дефіс).
- Назва міста може бути записана по-різному (з буквою «м» і без).
- У цінах використовуються неукраїнські роздільники (кома замість пробілу як тисячний роздільник).
- В даних є повтори.

Вміст цього текстового файлу потрібно завантажити в Power Query, виправити всі недоліки, а потім вивантажити назад на лист і побудувати зведену таблицю.

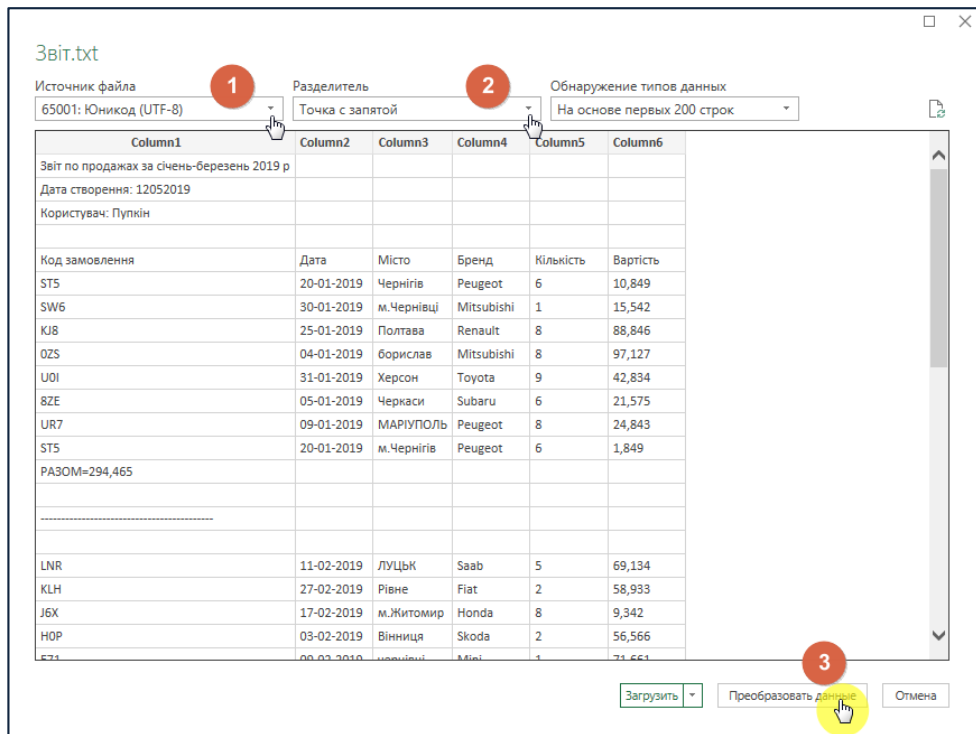
Інструкція до виконання завдання

Завантажуємо файл.

Щоб завантажити вміст цього файлу, виберемо на вкладці **Дані** команди **Отримати дані з текстового/CSV-файлу** або скористаємося однойменною кнопкою:



У вікні потрібно буде вибрати файл (D:\Temp\звіт.xlsx), потім з'явиться вікно попереднього перегляду.



Тут потрібно зробити наступне:

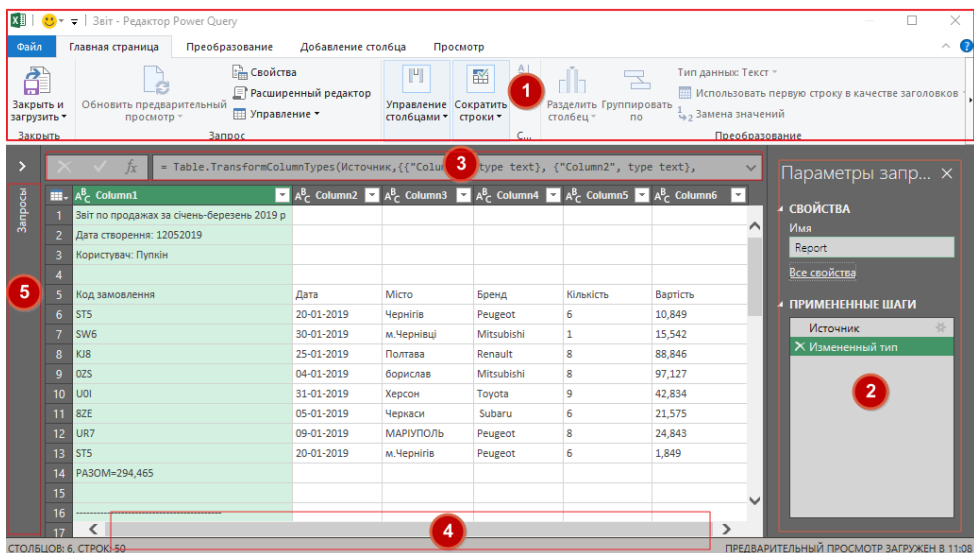
1. Перевірити правильність обраного кодування у випадяючому списку **Джерело файлу**. Зазвичай це Юнікод, але при імпорті, наприклад, файлів зі старих систем, створених ще під MS-DOS, кодування може відрізнитися. Потрібний варіант знаходиться перебором, головне, щоб український текст не перетворювався на «крязкозябли».

2. Підібрати правильний роздільник стовпчиків (в нашому випадку крапка з комою).

3. Натиснути кнопку **Перетворити** для того, щоб завантажити дані в Power Query для подальшої обробки. Сусідня кнопка **Завантажити** нам не потрібна, тому що помістить дані відразу на лист як є, а нам їх потрібно спочатку привести в порядок.

Вікно редактора запитів

Після натискання на кнопку **Змінити** ми побачимо головне вікно Power Query – редактор запитів:



1. **Стрічка**. У верхній частині вікна видно звичний інтерфейс - стрічка і вкладки **Головна**, **Перетворення**, **Додавання стовпця** і **Перегляд**. На цих вкладках знаходиться більшість потрібних нам команд для роботи з даними.

2. **Параметри** запиту. У правій частині вікна розташована панель **Параметри** запиту, де можна дати ім'я нашому запиту, а трохи нижче видно список всіх дій (кроків), які ми вже застосували до даних. Першим кроком **Джерело** був запит до файлу, а другий **Змінений тип** Power Query автоматично додав сам, намагаючись розпізнати тип даних (числа, текст, дати і тому подібне) в кожному стовпці. Будь-який крок можна відкоригувати, натиснувши на знак шестерінки праворуч від кроку, або видалити, клацнувши по хресту зліва.

3. **Рядок формул**. Всі виконані дії в Power Query записуються у вигляді послідовності команд на спеціальній вбудованій мові **M**. В рядку формул відображається команда, що відповідає поточному виділеному кроку.

4. **Панель перегляду.** Вміст будь-якої комірки можна переглянути в нижній панелі, якщо клацнути мишею в білий фон (не текст!) комірки. У Power Query, на відміну від Excel, в комірках можуть зберігатися не тільки поодинокі числа або текст, але цілі списки або навіть таблиці.

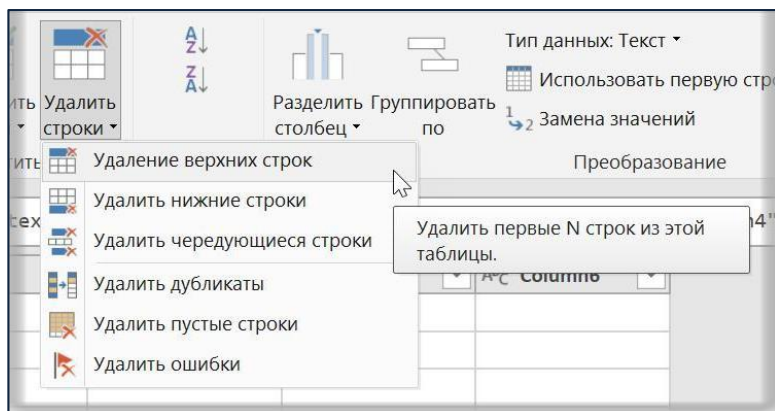
5. **Панель запитів.** В одній книзі Excel можна створити і використовувати декілька запитів. Наприклад, один запит може брати дані з листа, інший - з Інтернету, а третій - з'єднувати їх між собою і вивантажувати потім результати збирання на лист у вигляді зведеної таблиці. Швидко переключатися між запитами можна якраз за допомогою цієї панелі. За замовчуванням ця панель зазвичай згорнута (кнопка зі стрілкою в правому верхньому куті панелі).

Наводимо порядок у даних

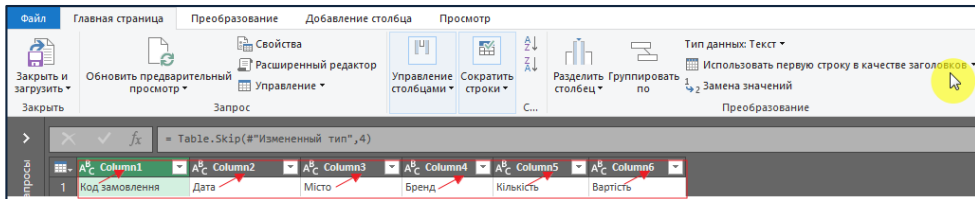
Тепер давайте зробимо всі необхідні дії для приведення наших даних в нормальний вигляд.

1. Видаліть крок **Змінений тип** в панелі **Параметрів запиту**, тому що поки ще рано призначати стовпцями типи даних. Ми зробимо це пізніше.

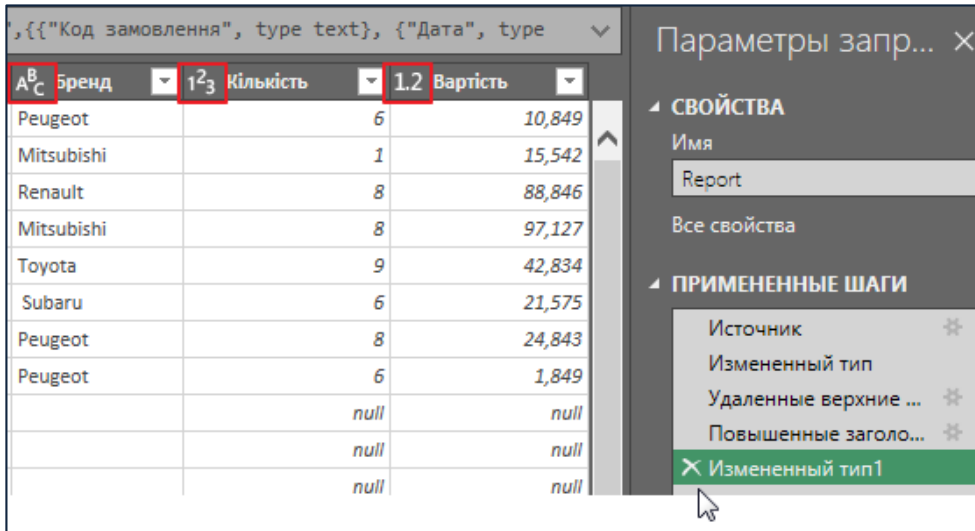
2. Прибрати чотири верхні непотрібні рядки можна на вкладці **Основна** → група **Скоротити рядки** → **Видалити рядки** → **Видалення верхніх рядків**:



3. Після видалення верхніх рядків виявиться, що перший рядок в наших даних містить, по суті, заголовки стовпців. Щоб підняти її в шапку таблиці, використовуємо на **Головній** вкладці команду **Використовувати перший рядок в якості заголовків**:



4. Зверніть увагу, що після попереднього кроку Power Query знову послужливо спробував розпізнати типи даних в стовпчиках (див. значки в шапці) і самовільно додав крок **Змінений тип**. Причому тип не скрізь визначився правильно: в останньому стовпці **Вартість** він зрозумів як роздільник цілої та дробової частини, а вона у нас насправді роздільник тисяч. Так що краще знову видалити цей крок, ми налаштуємо типи самостійно трохи пізніше:



5. Позбудемося порожніх рядків. Для цього на вкладці **Головна** виберемо **Видалити рядки** → **Видалити порожні рядки**.

6. Трохи складніше можна прибрати рядки-роздільники з дефісами і підсумки. Для цього в Power Query можна використовувати фільтр. У звичайному Microsoft Excel фільтр всього лише приховує якусь інформацію. У Power Query **фільтрація призводить до видалення непотрібних рядків**. Для цього виберемо в фільтрі по стовпцю **Код замовлення** опції **Текстові фільтри** → **Не містить**:

1	ST5	Сортировка по возрастанию	6	1,849
2	SW6	Сортировка по убыванию	1	15,542
3	KJ8	Отменить сортировку	8	88,846
4	OZ5	Очистить фильтр	8	97,127
5	U01	Удалить пустые	9	42,834
6	8ZE	Текстовые фильтры	6	21,575
7	UR7	Поиск		
8	ST5	<input checked="" type="checkbox"/> (Выбрать все)		
9	ИТОГО=294,465	<input checked="" type="checkbox"/> -----		
10	-----	<input checked="" type="checkbox"/> -----		
11	LNR	<input checked="" type="checkbox"/> OBM		
12	KLH	<input checked="" type="checkbox"/> OZ5		
13	J6X	<input checked="" type="checkbox"/> 30H		
14	HOP	<input checked="" type="checkbox"/> 3G8		
15	E71	<input checked="" type="checkbox"/> 5B5		
16	A05			
17	OBM			

У вікні можна ввести одне або декілька умов відбору:

Фильтрация строк

Применить одно или несколько условий фильтра к строкам в этой таблице.

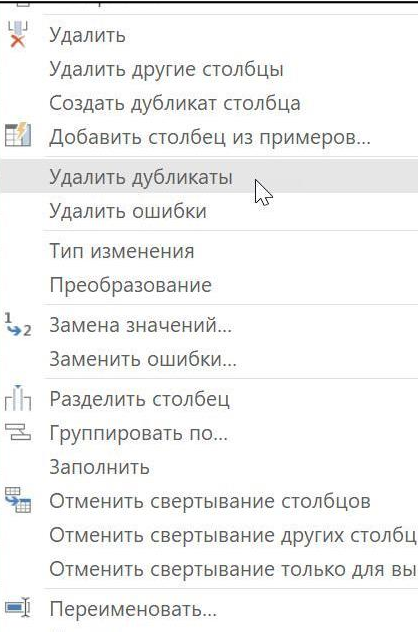
Базовый Подробнее

Сохранять строки, в которых "Код замовлення"

И Или

7. Зауважте, що в першому стовпці у нас трапляються дублікати, давайте від них позбавимося. Для цього можна використовувати контекстне меню, яке в Power Query вміє дуже багато і вкрай зручно.

Клацніть правою кнопкою миші по заголовку стовпця **Код замовлення**, виберіть відповідну команду - і справа зроблено:

1	ST5	
2	SW6	
3	KJ8	
4	OZS	
5	U0I	
6	8ZE	
7	UR7	
8	ST5	
9	LNR	
10	KLH	
11	J6X	
12	H0P	
13	E71	
14	A0S	
15	0BM	
16	0BM	
17	0BM	

8. Тепер, коли всі основні трансформації виконані, можна, нарешті, призначити кожному стовпцю відповідні типи даних. Найпростіше це зробити клацанням по значку в лівому верхньому кутку кожного стовпця, поряд з трикутником фільтра. Для стовпців **Код замовлення**, **Місто** і **Бренд** можна залишити текстовий тип даних, для стовпця **Кількість** встановити **Ціле число**, для стовпця з датою - формат **Дата**.

Найцікавішою буде ситуація з останньої колонкою **Вартість**, де у нас числа в нестандартному (для «українського» Excel) форматі. Для їх коректного розпізнавання потрібно вибрати не опцію **Десятькове число**, а останню в контекстному меню команду **Використовуючи локаль**. Під *локаллю* в даному випадку розуміється набір регіональних (локальних) налаштувань запису чисел, дат і грошових сум для різних країн. Вікно дозволить вибрати будь-який потрібний

тип даних і стандарт (в нашому випадку **Англійська (США)**), і після натискання на **ОК** числа будуть правильно інтерпретовані:

№	Місто	Бренд	Кількість	1.2	Вартість
019	Чернігів	Peugeot	6		10,849
019	м.Чернівці	Mitsubishi	1		15,542
019	Полтава	Renault	8		88,846

СВОЙСТВА

Имя

Report

Изменение типа по локали

Измените тип данных и выберите локаль источника.

Тип данных

Десятичное число

Языковой стандарт

Английский (США)

Примеры входных значений:

2,100.50

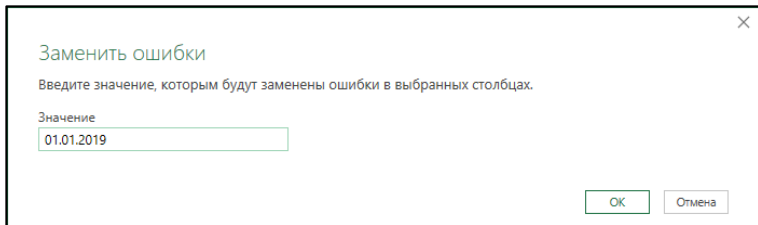
-1.50

OK Отмена

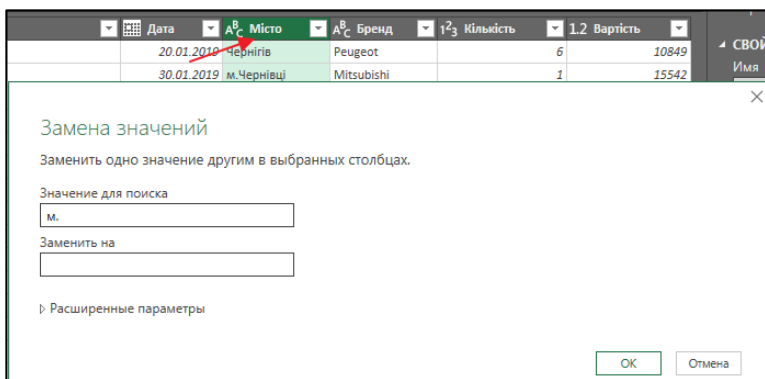
9. Зверніть увагу, що після перетворення в дату на одній з комірок з'явилася помилка. Щоб зрозуміти її причину, можна клацнути мишею в фон комірки з помилкою (але не в слово **Error**) і побачити опис помилки в нижній частині вікна.

Причина банальна: людський фактор. Питання: що робити з отриманою помилкою? Варіантів три.

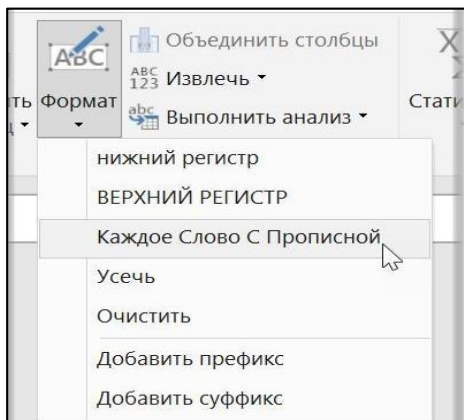
- **Залишити все як є.** Під час вивантаження на лист комірка з **Error** буде порожньою, але в процесі поновлення ми отримаємо повідомлення про знайдені помилки і подробиці по ним.
- **Видалити рядки з помилками**, виділивши стовпець з датою і вибравши на вкладці **Головна** → **Вилучити рядки** → **Видалити помилки**.
- **Замінити помилки** на будь-яке значення, наприклад на 1 січня 2019 року або будь-яку іншу відповідну дату. Для цього можна натиснути правою кнопкою миші по заголовку стовпця дати і вибрати команду **Замінити помилки** з контекстного меню:



10. Схожим чином можна позбутися, наприклад, і від букв «м» в стовпці з містами. Клацніть правою кнопкою миші по заголовку стовпця **Місто** і в контекстному меню будуть всі можливі дії для цієї колонки, в тому числі і потрібна нам команда **Заміна значень**:

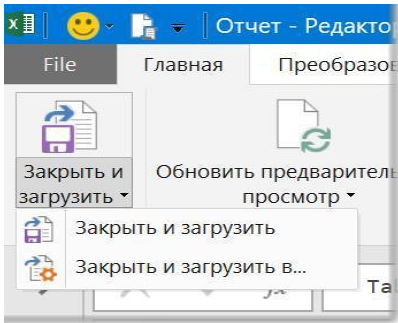


11. Тепер можна додатково виправити реєстр в стовпці з містами, вибравши на вкладці **Перетворення** → **Формат** → **Кожне слово з великої**:



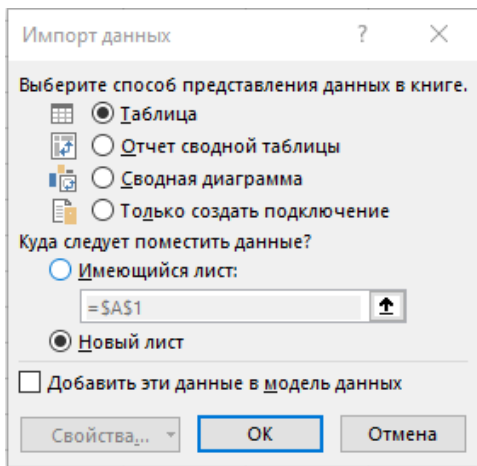
Вивантаження результатів з Power Query назад в Excel

Тепер, коли всі операції по трансформації даних виконані, отримані результати треба вивантажити з Power Query назад в Excel. Для цього служить команда **Закрити і завантажити** на вкладці **Головна**.



Зверніть увагу, що в списку присутні дві команди зі схожими назвами, але різним змістом:

- **Закрити і завантажити** - закриває вікно Power Query і вивантажує всі результати у вигляді «розумної» таблиці на доданий в поточну книгу новий лист.
- **Закрити і завантажити в ...** - виводить діалогове вікно, де можна вибрати, куди і в якому вигляді ми хочемо вивантажити результати.



Таблиця - дані вивантажуються в вигляді «Розумної» таблиці (на окремий новий або один з наявних листів).

Тільки створити підключення - дані залишаються в пам'яті і нікуди явно не вивантажуються. На підставі цих даних можна пізніше побудувати зведену таблицю або скористатися ними в іншому запиті. Цей варіант економить місце на аркуші і розмір файлу.

Звіт зведеної таблиці - теж саме, що і попередній варіант, але з побудовою зведеної таблиці по створеному підключенню.

Зведена діаграма - теж саме, що і попередній варіант, але поруч зі зведеною таблицею буде побудована ще й зведена діаграма.

Прапорець **Додати ці дані в модель даних** завантажує результати в надбудову Power Pivot для подальшого аналізу і використання.

Виберемо для початку просто вивантаження у вигляді таблиці на новий лист. Після натискання на **ОК** на порожньому доданому листі з'явиться наша «розумна» таблиця з результатами обробки з Power Query:

	A	B	C	D	E	F
1	Код замовлення	Дата	Місто	Бренд	Кількість	Вартість
2	ST5	20.01.2019	Чернівів	Peugeot	6	10849
3	SW6	30.01.2019	Чернівці	Mitsubishi	1	15542
4	KJ8	25.01.2019	Полтава	Renault	8	88846
5	OZS	04.01.2019	Борислав	Mitsubishi	8	97127
6	U0I	31.01.2019	Херсон	Toyota	9	42834
7	8ZE	05.01.2019	Черкаси	Subaru	6	21575
8	UR7	09.01.2019	Маріуполь	Peugeot	8	24843
9	LNR	11.02.2019	Луцьк	Saab	5	69134
10	KLH	27.02.2019	Рівне	Fiat	2	58933
11	J6X	17.02.2019	Житомир	Honda	8	9342
12	HOP	03.02.2019	Вінниця	Skoda	2	56566
13	E71	09.02.2019	Чернівці	Mini	1	71661
14	A0S	19.02.2019	Умань	Toyota	6	82200
15	0BM	16.02.2019	Харків	Fiat	4	17882
16	P2O	11.02.2019	Херсон	Hyundai	9	80457
17	3G8	01.01.2019	Миколаїв	Suzuki	5	30952
18	CFJ	07.03.2019	Дніпро	Ford	3	17325
19	L51	12.03.2019	Запоріжжя	Saab	3	88439
20	С6K	13.03.2019	Суми	Saab	6	61377
21	P0O	14.03.2019	Миргород	Rover	9	59120
22	5B5	26.03.2019	Тернопіль	Renault	5	30129
23	KIU	19.03.2019	Луцьк	Mitsubishi	7	18253

Запити і ...

Запити | Підключення

1 запит

Report

Заружено строк: 29.

В правій частині вікна повинна з'явитися панель **Запити та підключення**, де повинен відобразитися наш запит і кількість завантажених рядків. Якщо натиснути правою кнопкою миші на наш запит *Report* в цій панелі, то можна побачити набір всіх доступних нам команд:

- **Оновити** - повторює весь процес звернення до файлу і наступних трансформацій заново. Якщо завтра ви отримаєте новий файл з іншими даними, то досить буде підмінити його у вихідній папці (зі збереженням імені файлу, само-собою) і оновити запит.
- **Завантажити в ...** - дозволяє ще раз вивести діалогове вікно з варіантами вибору місця і типу вивантаження (таблиця, зведена, тільки підключення і т. д.)
- **Перейменувати** - якщо ви забули дати запиту зрозуміле ім'я в редакторі, то це можна легко зробити тут.
- **Видалити** - при видаленні запиту видаляються всі записані кроки, але вивантажені на лист результати залишаються на місці (тільки їх не можна буде більше оновити).

Висновок. За допомогою надбудови Power Query можна створювати прості або складні запити відповідно до власних потреб. Оскільки надбудова Power Query записує та виконує ці кроки за допомогою мови **M**, запити можна створювати з нуля (або налаштувати вручну), використовуючи потужні та гнучкі можливості сценаріїв даних безпосередньо в надбудові Power Query.

За допомогою Power Query ви можете шукати джерела даних, створювати підключення, а потім формувати дані (наприклад, видалити стовпець, змінити тип даних або об'єднати таблиці) відповідно до своїх потреб. Коли дані сформовані, ви можете поділитися своїми висновками або використовувати запит для створення звітів.

