

Занятие 1

Знакомство с компонентами Power BI. Создание первых запросов.



pavek@emea-box.com


**Павел
Козлов**

Специалист по технологиям Microsoft

**Перед тем как двигаться
дальше**

Перед тем как двигаться дальше

Проверьте установку Power BI Desktop

В списке установленных программ должен быть Power BI Desktop. Вот его ярлык:  Power BI Desktop

Если Power BI Desktop не установлен, пройдите по [ссылке](#). Лучше устанавливать 64-битную версию (имя файла установки оканчивается на 64).

Power BI Desktop пока существует только для Windows.

Цели занятия

- Узнать, что такое Power BI
- Познакомиться с концепцией эффективной работы с данными
- Познакомиться с Power Query

**О чём поговорим, что
сделаем**

План занятия

- Какие приложения и сервисы входят в Power BI
- Какие есть инструменты Power BI
- Какие бизнес-задачи решаются на этапе загрузки и преобразования данных
- Какие бывают источники и типы данных
- Познакомимся с некоторыми преобразованиями данных в запросе

Приложения и сервисы, входящие в Power BI

	Создание отчетов	Просмотр отчетов
PowerBI.com - облачный Power BI Report Server - локальный	без подготовки данных	☑
Приложение Power BI Desktop	☑	можно, но неестественно
Приложение Power BI	✗	☑
Excel 2013/2016/2019/365 для Windows	☑	✗

Концепция эффективной работы с данными

Что такое анализ данных?

Анализ данных – это поиск ответов на вопросы.

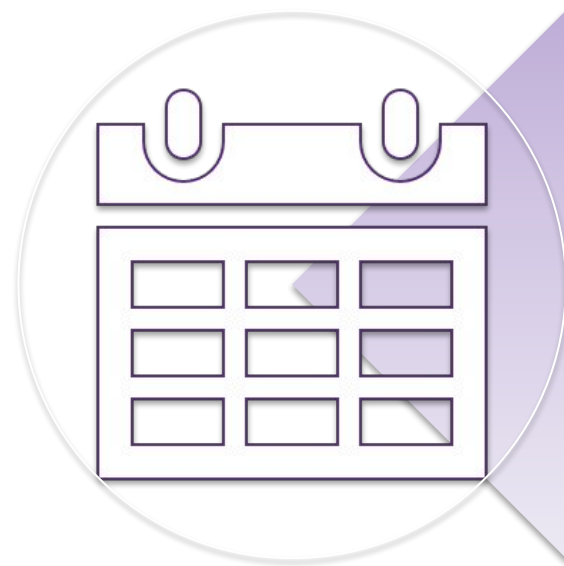
Вопросы ставит нам наша деятельность.

Ответы – предположительно скрыты в
имеющихся данных.

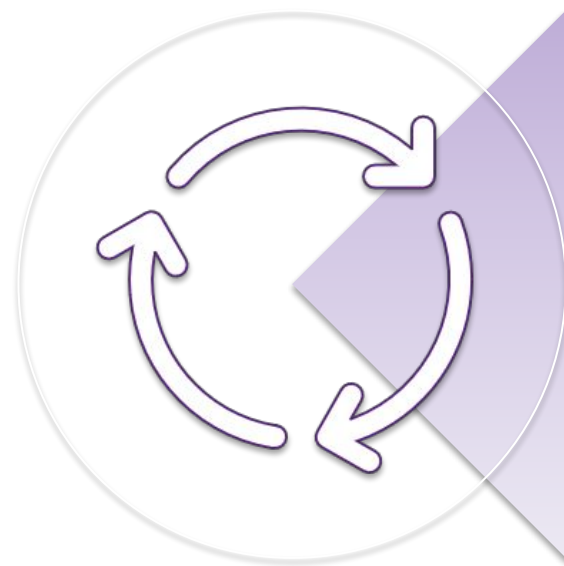
Важно! Сами по себе вопросы в данных не содержатся. Поэтому бессмысленно пытаться "анализировать данные" в надежде найти "что-нибудь интересненькое".



Принцип четырех шагов



Принцип табличности



Принцип универсальности

Решение любой задачи по анализу данных в Power BI похоже на приготовление еды. И там, и там есть 4 этапа.

	Еда	Данные	Применяемый инструмент
1	Еду нужно принести на кухню и подготовить к приготовлению	Данные нужно загрузить и подготовить к анализу. Это – создание набора данных	Power Query
2	Еду нужно приготовить	Данные нужно проанализировать – найти ответы на стоящие вопросы. Это – моделирование данных	Power Pivot
3	Еду нужно превратить в красивые блюда	Результаты анализа нужно визуализировать. Это – построение отчета	Power View
4	Готовые блюда нужно подать на стол	Подготовленный отчет нужно опубликовать. Это – публикация отчета	Сервис Power BI

Power BI Desktop

Что такое «табличный вид»

Столбцы
содержат
однородные,
т.е.
сравнимые,
показатели

Филиал	Дата	Выручка
Казань	17.02.2017	14 467 399 ₽
Ижевск	10.04.2017	4 756 509 ₽
Самара	11.10.2017	25 125 576 ₽
Уфа	15.10.2017	27 552 787 ₽
Ижевск	26.02.2017	1 130 714 ₽
Нижний Новгород	12.09.2017	1 321 378 ₽
Уфа	05.11.2017	843 071 ₽
Уфа	04.12.2017	3 288 184 ₽
Нижний Новгород	21.12.2017	12 694 405 ₽
Ижевск	04.08.2017	6 988 957 ₽
Нижний Новгород	09.06.2017	8 136 830 ₽
Уфа	09.12.2017	6 058 300 ₽
Самара	23.05.2017	10 803 119 ₽
Ижевск	06.03.2017	3 137 428 ₽

Каждая строка -
запись об одном и
только одном
событии

Нет объединенных и нет пустых ячеек.
Исключение: нулевые, либо
отсутствующие значения могут быть
пустыми ячейками

Задача 1

Таблица или нет?

Задача 1: таблица или не таблица? (1)

Филиал	Тип	Выручка	Количество
Ижевск	Легкий	218606016	241
Ижевск	Средний	238373698	315
Ижевск	Тяжелый	232728701	222
Итого Ижевск	Все типы	689708415	778
Казань	Легкий	378668372	373
Казань	Средний	206221305	201
Казань	Тяжелый	303427543	351
Итого Казань	Все типы	888317220	925
Новгород	Легкий	239427690	239
Новгород	Средний	179920899	213
Новгород	Тяжелый	156373418	179
Итого Новгород	Все типы	575722007	631
Самара	Легкий	240771550	322
Самара	Средний	287859827	312
Самара	Тяжелый	174695811	232
Итого Самара	Все типы	703327188	866

Задача 1: таблица или не таблица? (1)

Филиал	Тип	Выручка	Количество
Ижевск	Легкий	218606016	241
Ижевск	Средний	238373698	315
Ижевск	Тяжелый	232728701	222
Итого Ижевск	Все типы	689708415	778
Казань	Легкий	378668372	373
Казань	Средний	206221305	201
Казань	Тяжелый	303427543	351
Итого Казань	Все типы	888317220	925
Новгород	Легкий	239427690	239
Новгород	Средний	179920899	213
Новгород	Тяжелый	156373418	179
Итого Новгород	Все типы	575722007	631
Самара	Легкий	240771550	322
Самара	Средний	287859827	312
Самара	Тяжелый	174695811	232
Итого Самара	Все типы	703327188	866

Нет. Есть строки с промежуточными итогами

Задача 1: таблица или не таблица? (2)

Филиал	Тип	2017		2018	
		Выручка	Количество	Выручка	Количество
Ижевск	Легкий	55008560	44	58842953	52
Ижевск	Средний	55404792	83	74386120	82
Ижевск	Тяжелый	81792128	71	56260631	59
Казань	Легкий	131685018	134	109811775	103
Казань	Средний	6422235	10	77060682	66
Казань	Тяжелый	37463965	60	74292753	93
Новгород	Легкий	30925954	46	44807505	61
Новгород	Средний	47769169	54	45952916	73
Новгород	Тяжелый	54494944	47	20540415	29
Самара	Легкий	39639705	70	52031150	52
Самара	Средний	129731626	130	52111674	57
Самара	Тяжелый	72185842	96	10354957	17

Задача 1: таблица или не таблица? (2)

Филиал	Тип	2017		2018	
		Выручка	Количество	Выручка	Количество
Ижевск	Легкий	55008560	44	58842953	52
Ижевск	Средний	55404792	83	74386120	82
Ижевск	Тяжелый	81792128	71	56260631	59
Казань	Легкий	131685018	134	109811775	103
Казань	Средний	6422235	10	77060682	66
Казань	Тяжелый	37463965	60	74292753	93
Новгород	Легкий	30925954	46	44807505	61
Новгород	Средний	47769169	54	45952916	73
Новгород	Тяжелый	54494944	47	20540415	29
Самара	Легкий	39639705	70	52031150	52
Самара	Средний	129731626	130	52111674	57
Самара	Тяжелый	72185842	96	10354957	17

Нет. Есть
объединенные ячейки
с заголовками годов.
Однородные
значения (выручка и
количество)
разнесены по двум
столбцам

Задача 1: таблица или не таблица? (3)

Продавец	Квартал1	Квартал2	Квартал3	Квартал4
Гоголь	77344814	36020950	92945563	97079704
Горький	68526537	88693054	38606045	78470629
Дорофеев	79762016	66990500	71329109	63912904
Лермонтов	127196858	82727695	59833589	104332427
Некрасов	52710398	67249634	111525281	63250850
Островский	70567776	67401782	72263380	72751931
Пелевин	47663654	29091273	48675088	52191780
Пушкин	120220006	45936212	72614865	79815211
Толстой	45322581	36313956	58011850	196065987
Тургенев	68925794	51972902	108429837	63946057
Чехов	169240066	235025296	140985660	127739819

Задача 1: таблица или не таблица? (3)

Продавец	Квартал1	Квартал2	Квартал3	Квартал4
Гоголь	77344814	36020950	92945563	97079704
Горький	68526537	88693054	38606045	78470629
Дорофеев	79762016	66990500	71329109	63912904
Лермонтов	127196858	82727695	59833589	104332427
Некрасов	52710398	67249634	111525281	63250850
Островский	70567776	67401782	72263380	72751931
Пелевин	47663654	29091273	48675088	52191780
Пушкин	120220006	45936212	72614865	79815211
Толстой	45322581	36313956	58011850	196065987
Тургенев	68925794	51972902	108429837	63946057
Чехов	169240066	235025296	140985660	127739819

Нет. Однородные значения для разных кварталов в разных столбцах

Задача 1: таблица или не таблица? (4)

Тип	Выручка	Количество
Легкий	218606016	241
Средний	238373698	315
Тяжелый	232728701	222
Легкий	378668372	373
Средний	206221305	201
Тяжелый	303427543	351
Легкий	239427690	239
Средний	179920899	213
Тяжелый	156373418	179
Легкий	240771550	322
Средний	287859827	312
Тяжелый	174695811	232

Задача 1: таблица или не таблица? (4)

Тип	Выручка	Количество
Легкий	218606016	241
Средний	238373698	315
Тяжелый	232728701	222
Легкий	378668372	373
Средний	206221305	201
Тяжелый	303427543	351
Легкий	239427690	239
Средний	179920899	213
Тяжелый	156373418	179
Легкий	240771550	322
Средний	287859827	312
Тяжелый	174695811	232

Да. Все в порядке

Задача 1: таблица или не таблица? (5)

Продавец	Показатель	Значение
Гоголь	Выручка	303391031
Гоголь	Кол-во	349
Горький	Выручка	274296265
Горький	Кол-во	289
Дорофеев	Выручка	281994529
Дорофеев	Кол-во	326
Лермонтов	Выручка	374090569
Лермонтов	Кол-во	388
Некрасов	Выручка	294736163
Некрасов	Кол-во	351
Островский	Выручка	282984869
Островский	Кол-во	365
Пелевин	Выручка	177621795
Пелевин	Кол-во	253
Пушкин	Выручка	318586294
Пушкин	Кол-во	356

Задача 1: таблица или не таблица? (5)

Продавец	Показатель	Значение
Гоголь	Выручка	303391031
Гоголь	Кол-во	349
Горький	Выручка	274296265
Горький	Кол-во	289
Дорофеев	Выручка	281994529
Дорофеев	Кол-во	326
Лермонтов	Выручка	374090569
Лермонтов	Кол-во	388
Некрасов	Выручка	294736163
Некрасов	Кол-во	351
Островский	Выручка	282984869
Островский	Кол-во	365
Пелевин	Выручка	177621795
Пелевин	Кол-во	253
Пушкин	Выручка	318586294
Пушкин	Кол-во	356

Нет. И в столбце «Показатель», и в столбце «Значение» разнородные показатели. Фактически, одному событию соответствуют две строки

Задача 1: таблица или не таблица? (6)

Выручка	Город	Количество
218606016	Самара	241
238373698	Казань	315
232728701	Пермь	222
378668372	Уфа	373
206221305	Москва	201
303427543	Владимир	351
239427690	Суздаль	239
179920899	Ярославль	213
156373418	Псков	179

Задача 1: таблица или не таблица? (6)

Выручка	Город	Количество
218606016	Самара	241
238373698	Казань	315
232728701	Пермь	222
378668372	Уфа	373
206221305	Москва	201
303427543	Владимир	351
239427690	Суздаль	239
179920899	Ярославль	213
156373418	Псков	179

Да. Хотя столбец с числовыми показателями стоит первым, что непривычно, тем не менее, все в порядке

Задача 1: таблица или не таблица? (7)

Месяц	Выручка
Янв	11М
Фев, мар	11М
Апр	10М
Май	15М
Июн, авг, дек	10М
Июл	10М
Сен	14М
Окт	15М
Ноя	14М

Задача 1: таблица или не таблица? (7)

Месяц	Выручка
Янв	11М
Фев, мар	11М
Апр	10М
Май	15М
Июн, авг, дек	10М
Июл	10М
Сен	14М
Окт	15М
Ноя	14М

Нет. И в столбце «Месяц» в некоторых строках содержатся записи о нескольких событиях

Задача 1: таблица или не таблица? (8)



2017	Самара
2017	Казань
2017	Пермь
2017	Уфа

Задача 1: таблица или не таблица? (8)

2017	Самара
2017	Казань
2017	Пермь
2017	Уфа

Да. Хотя отсутствуют заголовки столбцов, это – таблица

Задача 1: таблица или не таблица? (9)

Москва

Задача 1: таблица или не таблица? (9)

Москва

Да. Хотя отсутствует заголовок столбца и есть только одна строка, это – таблица. Просто очень маленькая

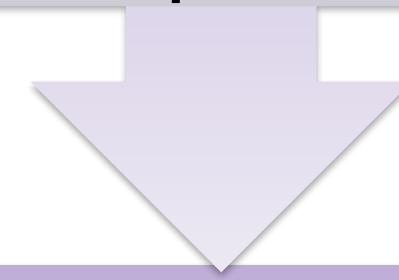
Мы должны стремиться к созданию таких решений, чтобы при допустимом изменении исходных данных не требовалось внесения изменений в решение.

**Какие задачи решаются на
этапе создания набора
данных**

Какие задачи решаются на этапе создания набора данных

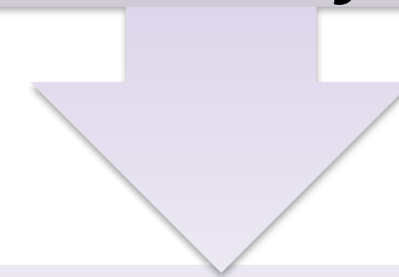
Выбор данных

Подключение к источнику данных и выбор той части данных, с которыми мы будем далее работать



Очистка данных

Удаление в выбранных исходных данных всего ненужного и приведение данных к табличному виду



Загрузка данных

Определение того, какие именно таблицы должны грузиться в модель

**Что может быть
источником данных**

Что может быть источником данных

Файлы

- Excel
- Текстовые
- CSV
- XML
- JSON
- PDF
- Папки
SharePoint

Базы данных

- SQL Server
- SQL Analysis Services
- Oracle
- MySQL
- SAP Hana
- SAP BW
- SQL Azure
- Azure Cosmos

Веб-сервисы

- Списки
SharePoint Online
- Common Data Service
- Salesforce
- Google Analytics
- Adobe Analytics
- Facebook
- GitHub
- MailChimp

Прочее

- Веб-сайты
- Active Directory
- Microsoft Exchange
- Списки
SharePoint
- Скрипт R
- Скрипт Python

Полный список источников приведен в данной [статье](#)

Задача 2

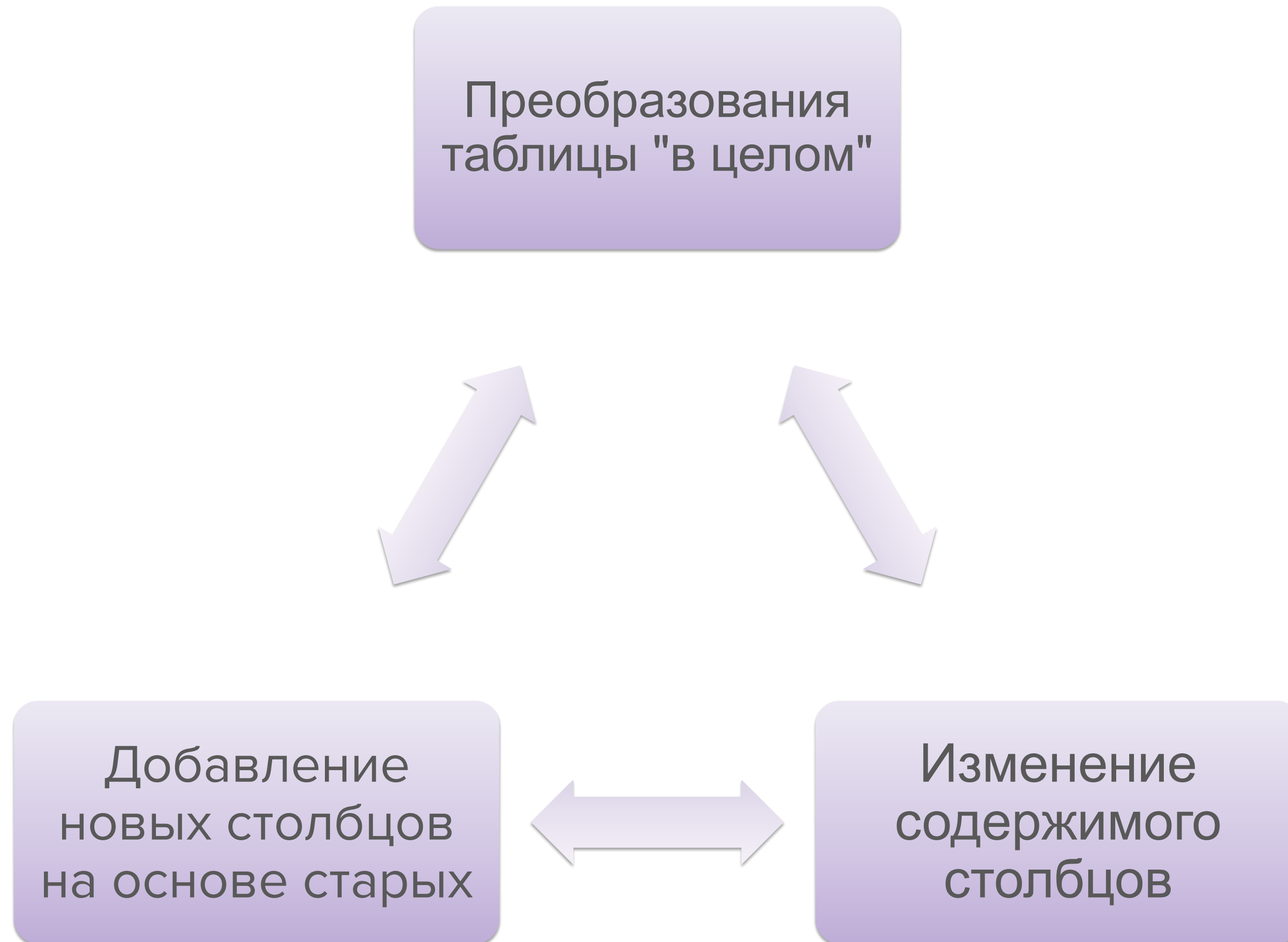
**Пробуем подключаться к
данным**

Задача 2: Пробуем подключаться к данным

1. К текстовому файлу (на примере файла «Пример.txt»)
2. К файлу Excel (на примере файла «Пример.xlsx»)
3. К веб-сайту (на примере сайта ЦБ РФ)
4. К папке с файлами (и объединим данные из файлов в одну таблицу)

**Какие преобразования
доступны в Power Query**

Какие преобразования доступны в Power Query



Задача 3

**Преобразуем таблицы «В
целом»**

Задача 3

1. Удалить столбец «Количество» в таблице из текстового файла;
2. Поменять оставшиеся столбцы местами в таблице из текстового файла;
3. Удалить строку с курсом евро в таблице с веб-сайта;
4. Удалить дубликаты в таблице из папки;
5. Поменять типы данных в таблице из папки;
6. Удалить ошибки в таблице из папки.

Задача 4

Очистить данные в файле

Задача 4

В данных, загруженных из Excel:

1. Удалить строки над заголовками столбцов;
2. Сделать строчку с «шапкой» строкой заголовков таблицы;
3. Убрать строки в которых филиал – «Казань»;
4. Убрать столбец «Количество»;
5. Сделать столбец «Дата» самым левым;
6. Переименовать столбец «Сумма» в «Выручка».

Что мы сегодня узнали

- Что такое Power BI и из чего он состоит
- Что такое таблица и как ее отличить от нетаблиц
- Откуда Power BI может брать данные
- А также, научились делать преобразования таблиц «в целом»

Полезные материалы

- [Документация по Power BI Desktop](#)
- [Сообщество \(англ.\)](#)
- [Блог Microsoft Power BI](#)

Домашнее задание

**Определить целевые
города в РФ для открытия
нового филиала**

Про домашние задания

1. Читайте, пожалуйста, внимательно задание. Очень внимательно;
2. Выполняйте, пожалуйста, все условия задания в точности так, как они сформулированы. Не срезайте углы;
3. Если что-то непонятно как делать – пересмотрите материалы занятия. Для выполнения задания не требуется ничего, чего нет в материалах занятия;
4. Домашние задания нужно делать. Это самая главная часть обучения. Можно научиться только делая задания и не просматривая материалы. Но нельзя научиться только просматривая материалы;
5. После задания почти всегда есть скриншот с требуемым результатом. Если Ваш результат отличается от скриншота – доработайте свой результат;
6. Если Вы не укладываетесь в срок сдачи, можно запросить продление срока. Но лучше укладываться в срок;
7. Срок решения – не оценивается. Качество решения – оценивается. Поэтому лучше сдать решение позже, но качественное;
8. При поиске решения помните, что решение должно быть универсальным: оно должно правильно работать и с другими исходными данными.

Ваша компания хочет выйти на новый для себя рынок городов в России с населением 200-250 тыс. человек (включительно). Для этого подбирается пилотный город для выхода. Руководство ждет от вас список подходящих городов.

В качестве источника используйте эту страницу в [Википедии](#). Данные по численности берите за 2018 год.

Требуется создать запрос, дающий таблицу из двух столбцов: «Город», с названиями городов, отсортированными от «а» до «я», и «Численность», содержащий данные по численности населения.

Файл .rbix с результатом нужно загрузить в личный кабинет.

То, что должно получиться в результате, приведено на следующем слайде.

Решать эту задачу с помощью «Таблицы из примеров» не нужно.

Домашнее задание 1, результат

A^B_C Город <input type="text"/>	1^2_3 Численность <input type="text"/>
Ангарск	226
Бийск	202
Благовещенск	225
Братск	229
Великий Новгород	223
Дзержинск	231
Комсомольск- на-Амуре	248
Королёв	223
Люберцы	203
Мытищи	212
Нальчик	239
Нижекамск	238
Орск	229
Псков	211
Старый Оскол	224
Сыктывкар	245
Таганрог	250
Шахты	234
Энгельс	226

Замечания к домашнему заданию

1. В качестве источника данных нужно использовать именно ту ссылку, которая приведена в условии. Выгружать страницу Википедии в файл и затем подключаться к этому файлу – не нужно;
2. При отсеивании каких-либо значений нужно стараться указать принцип отсеивания, а не перечислять все отсеиваемые значения. В первом случае – у нас будет соблюдаться принцип универсальности решения (см. слайд 55 «Про домашние задания», п. 8) и при новых значениях исходных данных результат будет по-прежнему правильный. Во втором случае – неуказанные явно ненужные значения не отсеются и проникнут в результат.



НЕТОЛОГИЯ
групп

Спасибо за внимание!

Павел Козлов



pavek@emea-box.com