[Цели работы](#с9)

[Теоретическая часть](#с10)

[С чем имеем дело](#_Творения_Дашульки)

[Немного о tableau](#с1)

[Плюсы](#с2)

[Минусы](#с3)

[Кто использует](#с4)

[Как работает](#с5)

[Рубрика «Звучит как тост»](#с6)

[А чего добился ты?](#с7)

[Интересности](#с8)

[Эксперты рекомендуют   
Ход лабораторной работы   
Творческое задание  
Данные  
Содержание отчета  
Контрольные вопросы](#с16)

**Цели нашей лабораторной**

* Знакомство с ПО Tableau desktop
* Освоение базовых инструментов программы
* Научиться импортировать данные из разных источников и работать с ними
* Делать визуальные выборки, строить диаграммы, отчеты
* Протестировать работу с картами
* Наслаждаться красивой цветовой гаммой

**Теоретическая часть**

# **С чем имеем дело**

*Tableau* – система интерактивной аналитики, позволяющая в кратчайшие сроки проводить глубокий и разносторонний анализ больших объемов данных. Чем она хороша? Не нужно быть гением, набирать 300 баллов по егэ или учиться в МФТИ, чтобы понять, как работает программа, разобраться с ее инструментами и начать строить свои красивые диаграммы.

***Что умеет tableau?***

* Радовать нас дружелюбным и приятным интерфейсом (вы скоро сами в этом убедитесь)
* Черпать данные различных форматов (БД sql server, текстовые файлы, карты(!) и прочие)
* Анализировать данные (вау, неожиданно)
* Создавать крутые диаграммы, отчеты, таблицы

***Что не умеет tablau?***

* Разговаривать на русском языке
* Давать лицензию студентам

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**Немного о tableau**

Миссия Tableau — помочь людям увидеть и понять данные.

* В основу своих продуктов Tableau встраивает лучшие визуальные методы, искусственный интеллект и возможности машинного обучения, чтобы сделать мощную аналитику проще для всех. Это включает в себя умные рекомендации по объединениям базы данных и подготовительным этапам очистки, расширенную аналитику с перетаскиванием, такую ​​как кластеризация и прогнозирование, и взаимодействие на естественном языке, которое понимает намерения пользователя и контекст данных (сейчас здесь было немного умного текста, но сделайте вид, что вы все поняли)

**Плюсы**

**+** Комбинирование данных из разных БД и других источников (проверено)

**+** Tableau позволяет нескольким пользователям одновременно работать над отчетом в реальном времени

**+** Богатая библиотека платформы включает в себя облака слов, пузырьковые и древовидные диаграммы, которые позволяют достигнуть более высокого уровня понимания своих данных и их контекста (тоже что-то умное из алгоритмов)

**+** Интуитивно понятный интерфейс

**+** Не требуется специфичных технических знаний, например, SQL (ну блиииин, зря ходил(а) на БД)

**+** Повышается самостоятельность пользователей в работе с данными (условно, бабулька тоже может пилить тут диаграммки)

**+** Интерактивность отчетности

**+** Сокращается время подготовки отчетности

**+** Никаких настроек

**+** Возможность поделиться результатами работы

**+** Доступ из любой точки мира

**+** Импорт дэшбордов на веб-страницы

**+** Любые ОС (mac, IOS, Android, ну и всеми любимая Win)

**+** Визуализация данных (вроде этот термин мы уже где-то слышали)

**+** Группы и наборы. Пользователь легко может сделать выборку

**+** КАРТЫ!!! Являются одной из сильных сторон. Tableau предлагает отличную функциональность, которая включает постоянно обновляемые карты и дополнительную информацию (по какой-то лицензии третьих сторон), например, население (без каких –либо дополнительных затрат для клиента)

**+** Только Tableau может использовать SSAS Multidimensional и PowerPivot "local" Cubes как источники данных и строить к ним визуализацию

**+** Наличие готовых отраслевых решений

**+** Высокая скорость получения результата

**+** Постоянно совершенствуется, регулярно выпуская обновления

**Минусы**

**-** Параметры. Невозможно использовать динамические параметры, формулы для расчета значений, динамические названия формул, изменять значение одного параметра при смене другого и тд. Это лишает возможности создания дэшбордов, сравнения YTD без выбора периода

**-** Нельзя сравнивать агрегированные и не агрегированные данные. Например, запись вида [OrderDate]<=max([OrderDate]) неверна. Также, в IF условии должны быть либо агрегированные, либо нет условия и выражение

**-** Не обрабатывает иерархические связи (за исключением дат)

**-** Нет глобального поиска (грустненько)

**-** Оптимизация производительности является сложной задачей

**-** Дополнительные средства разработки. Tableau делает ставку на визуализацию данных, забывая о синтаксисе выражений и тп. Таким образом, простая задача показа объекта в зависимости от параметра является легкой задачей в QlikView, но необходимо примененять для ее реализации обходные пути в Tableau.

**-** Нет возможности комментирования (тоже грустно, хотя не уверена, что здесь это настолько необходимо)

**-** Необходимость предварительной обработки данных (должны быть структурированы)

**-** Даёт возможность низкоуровневой защиты на уровне данных, однако реализация выглядит ненадёжно

**Кто использует**

В основе Tableau – две инновационные технологии: VizQL, благодаря которой кликом мыши можно создавать сложные визуализации, и Data Engine, обеспечивающая анализ миллионов строк в считанные секунды. Tableau уже используют 65000 пользователей в 100 странах мира, в том числе – сотрудники Amazon, BP, Apple, Walmart, Heineken, Starbucks, Pfizer, Coca-Cola, Bank of America и Barclays.

Масштабные внедрения Tableau в СНГ:

* Wargaming (разработчик – World of Tanks) – компания активно развивает много-терабайтное хранилище данных с использованием технологий Big Data (hadoop, kafka, oracle и др.), а для аналитики используется Tableau.
* YOTA NETWORKS – ещё одно крупное хранилище на HP Vertica и аналитика на Tableau. Короткий комментарий от одного из участников проекта.

**Как работает**

Алгоритм работы с данными в Tableau очень простой и состоит из трех основных этапов:

* выбор источника данных
* визуализация данных
* публикация результата



**Рубрика «Звучит как тост»**

*Данные могут рассказать важную историю. Они полагаются на то, что вы дадите им громкий и убедительный голос.*

// Стивен Фью

**А чего добился ты?**

* Шесть лет подряд Tableau остается в квадранте лидеров среди ключевых игроков рынка бизнес-аналитики
* Tableau стал стандартным языком аналитики для бизнес-пользователей и специалистов по аналитике, признанным Forbes как один из самых быстрорастущих технических навыков и один из наиболее ценных для получения более высокой компетенции в профессиях в области наук о данных

**Интересности**

* Самое активное сообщество пользователей, которые создают тысячи обучающих видео, блогов и форумов. У них есть крутецкий сайт с различными статьями и марафонами. Лови, только не закапай слюнями клавиатуру от восторга <https://www.datayoga.ru/>
* Сообщество также является силой, которая делает добро с данными. Более 12 000 новых визуализаций представили миру некоммерческие организации, журналисты, блоггеры, политические энтузиасты и многие другие через Tableau Public, которая в прошлом году насчитывала более 1,5 миллиардов просмотров. Благодаря партнерству с PATH, Tableau и ее сообщество помогли снизить смертность от малярии на юге Замбии на 90%. А Tableau Foundation обязался предоставить 100 миллионов долларов на программное обеспечение, обучение и гранты в период до 2025 года для поддержки некоммерческих организаций, использующих данные для продвижения по сложным вопросам - от исследований заболеваний и ухода за пациентами до образования и государственных расходов.

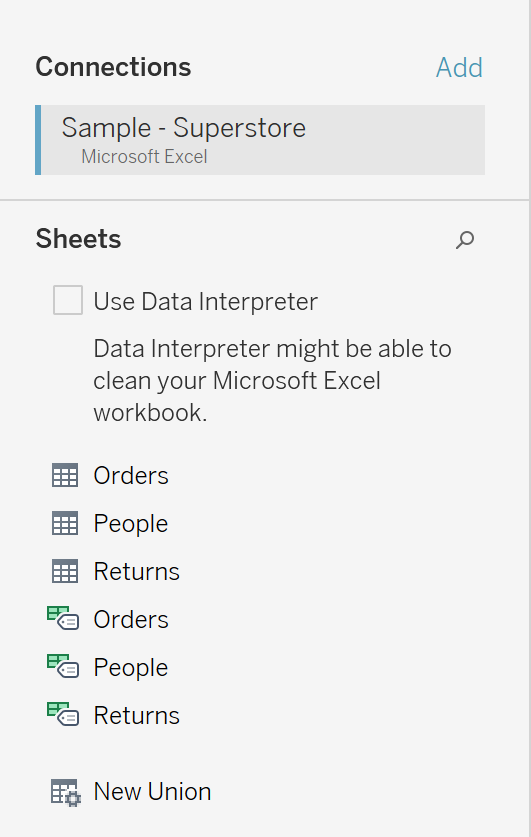
**Эксперты рекомендуют**

Использовать Tableau, если есть:

* Хранилище данных (DWH/DataWareHouse)
* Big Data
* Панели ключевых показателей (dashboards) с доступом в существующее хранилище данных (DWH/DataWareHouse) или в SSAS Multi-Dimensional Cubes
* Возможность создавать dashboard и аналитику силами самих бизнес-пользователей и в качестве конечных решений использовать MS Excel/Tableau с данными из SSAS Cubes/Microsoft / SQL Server data
* Легкость в визуализации существующих Excel PowerPivot баз данных
* Web interface для панели ключевых показателей и аналитических приложений
* Рассылка представлений для аналитики другим пользователям по запросу

**Ход лабораторной работы**

1. Открываем программу Tableau и сталкиваемся с нуждой выбора источника данных. Для начала мы будем работать с файлом Excel «Sample – superstore», который предварительно у вас всех уже есть. Соответственно, выбираем вкладку Microsoft Excel под заголовком to a text file.
2. Выбрав нужные данные, мы увидим следующую картину слева:

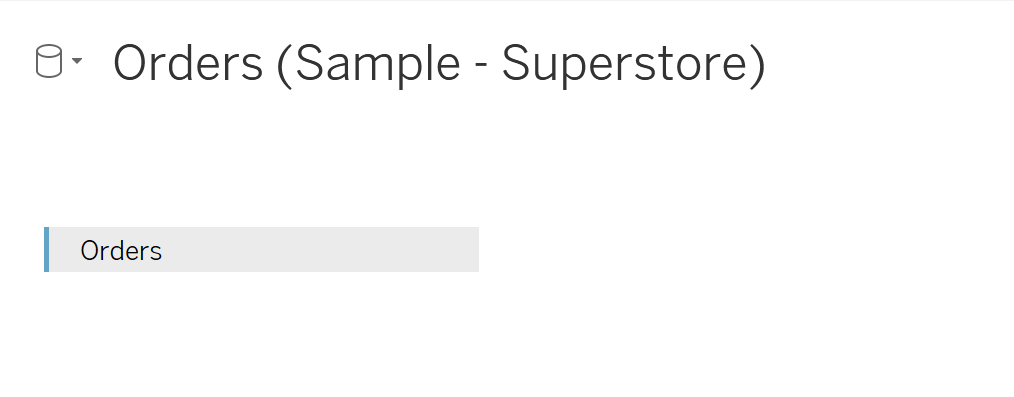
  
Connections – наши соединения с данными.

Add – добавить новые данные.

Use data Interpreter – инструмент, который помогает привести к приличному виду экселевский документ, распознав шапку и название столбцов.

Ниже перечислены доступные для использования таблицы.

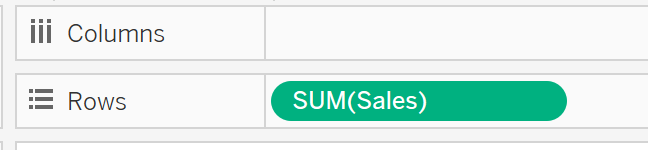
Мы можем подключить еще данные, причем из других источников (текстовые, например). Внутри Tableau они будут распознаваться как две БД с таблицами и данными.

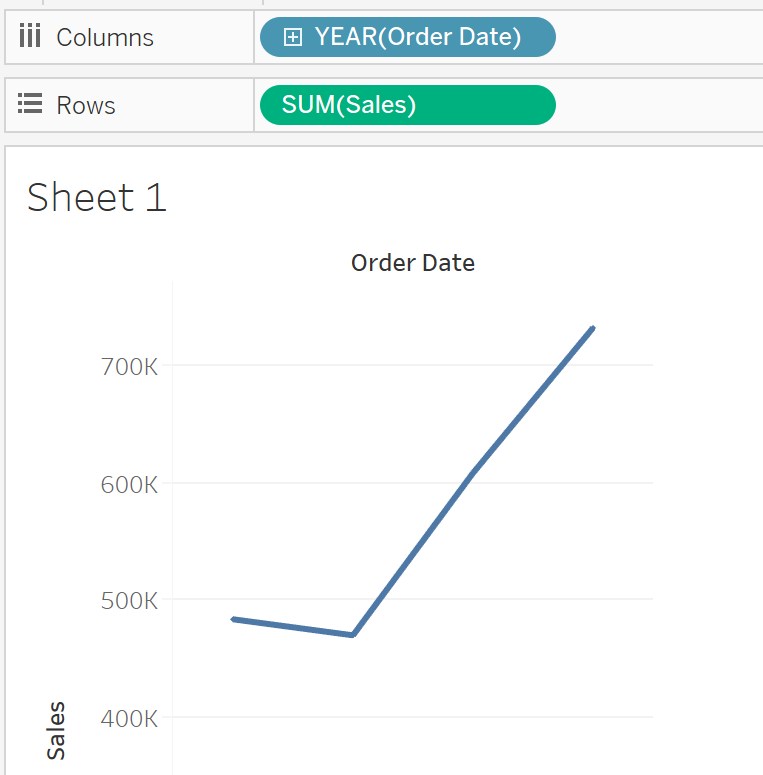
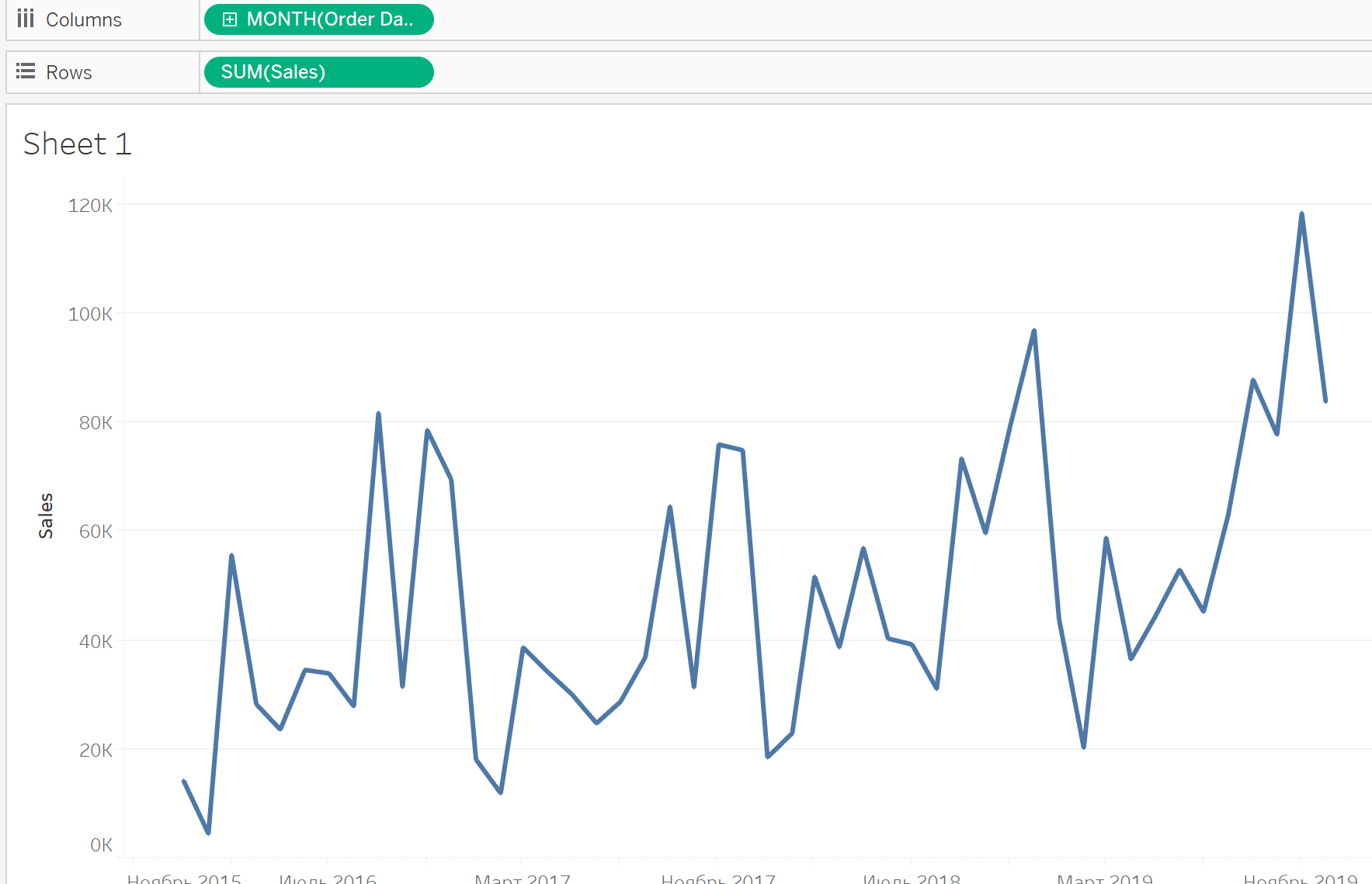
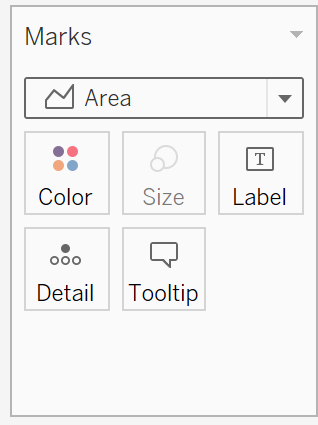
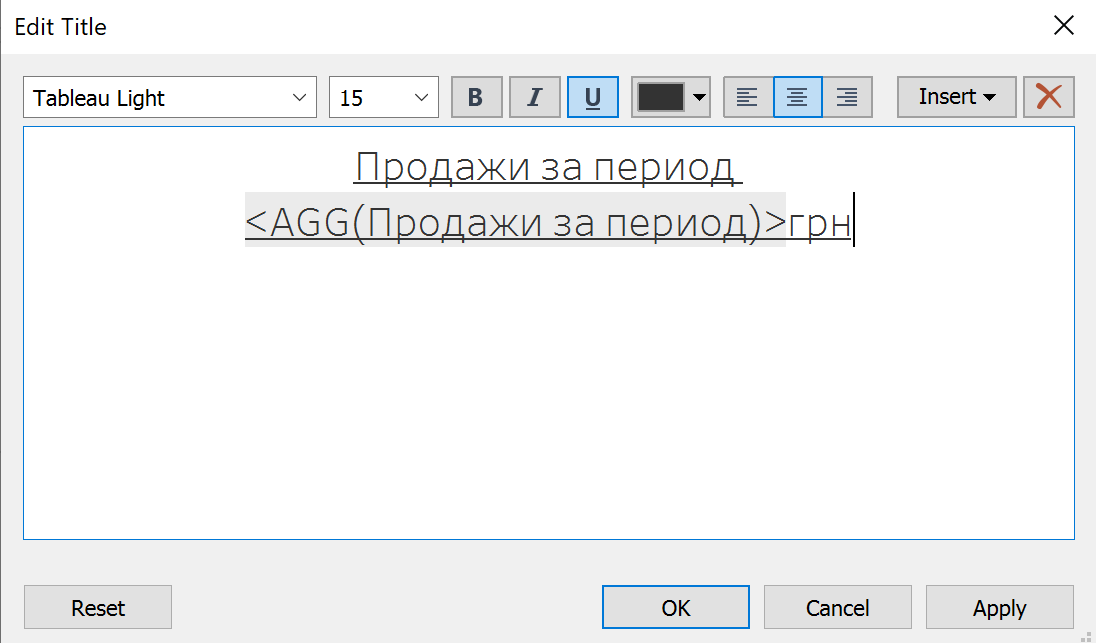
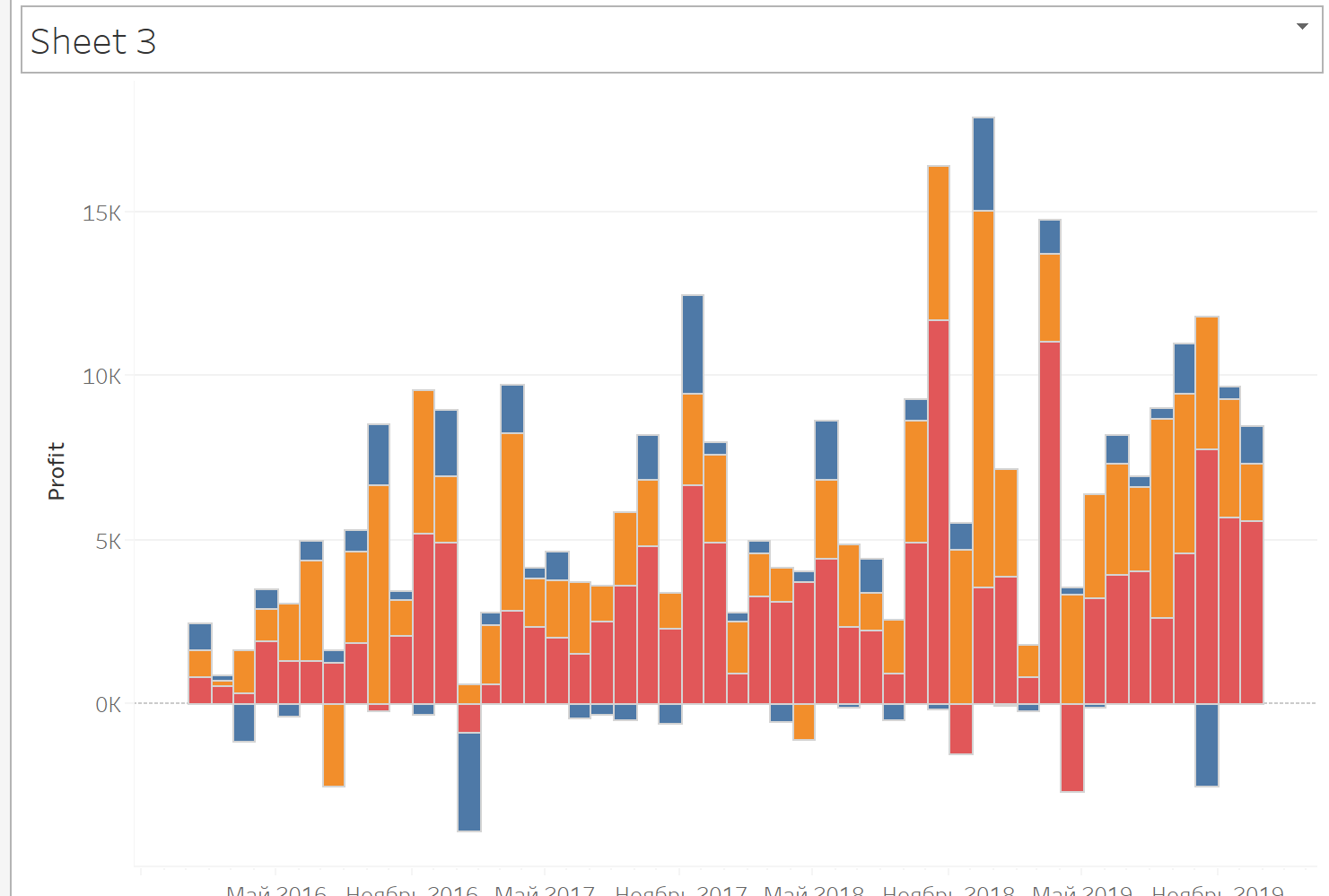
1. Сегодня мы будем работать с таблицей Orders (а почему бы и нет?). Возьмем ее и перетащим сюда  
   
2. Live – живое подключение к источнику данных

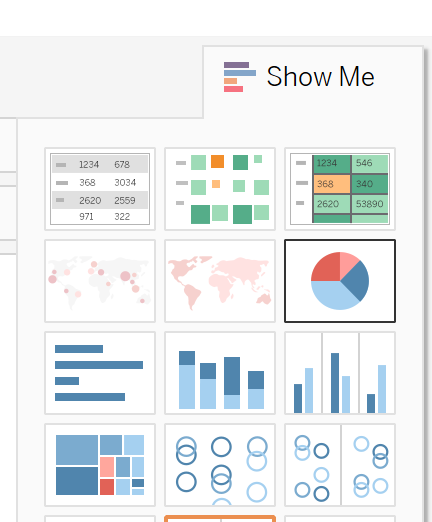
Extract – выжимка из источника данных, то есть мы обращаемся к ним каждый раз, когда формируем какие-либо виды отчетов.

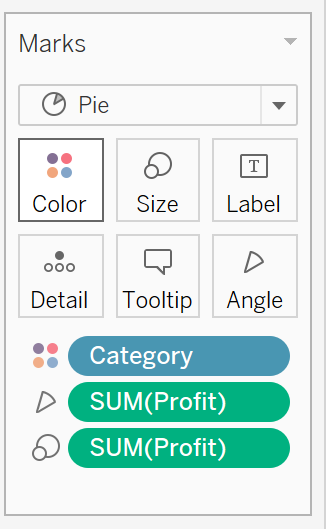
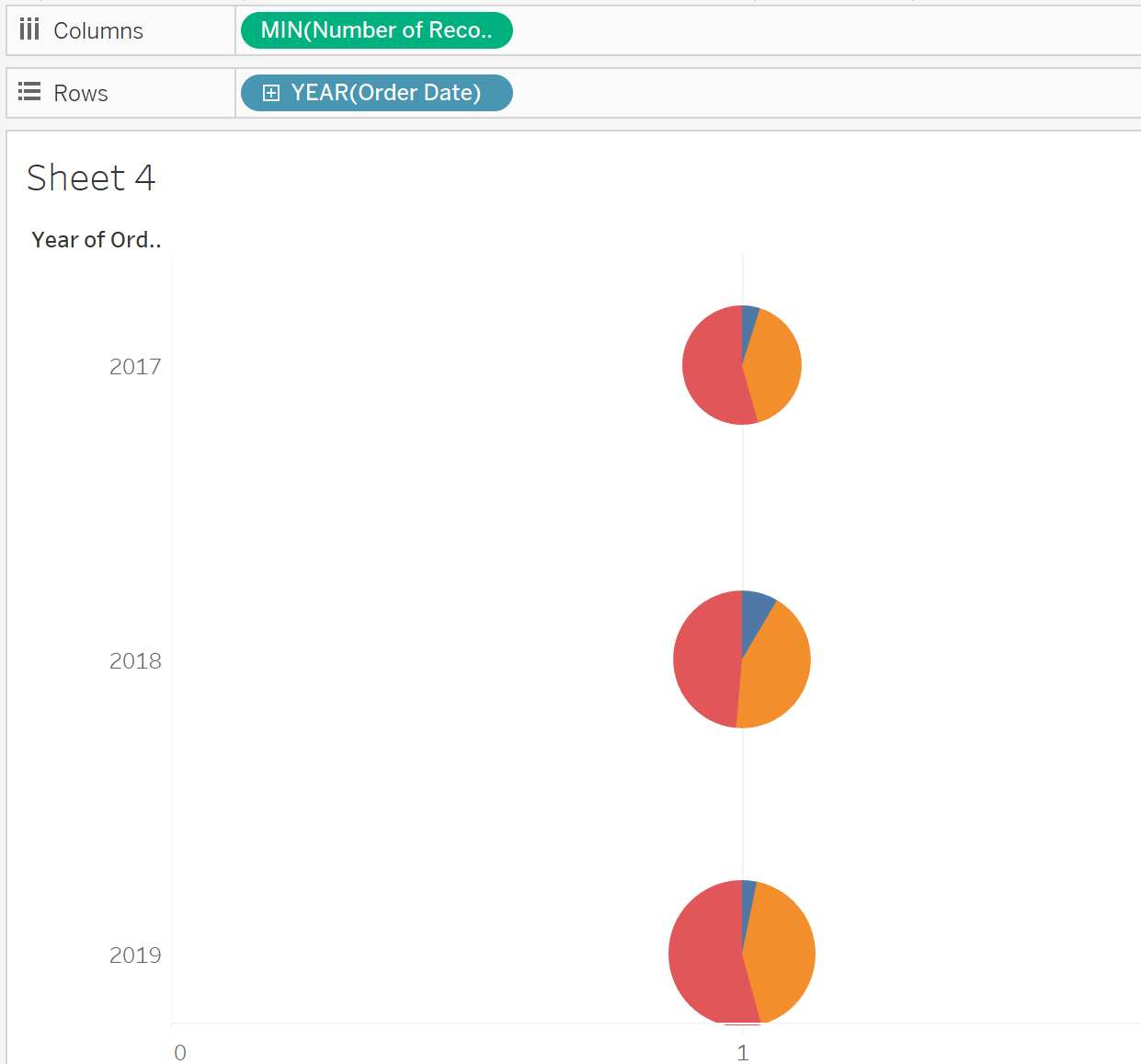
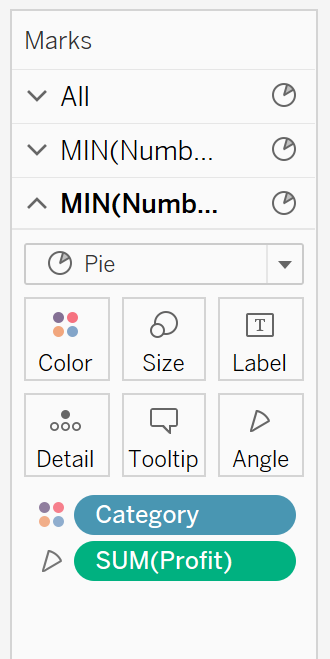
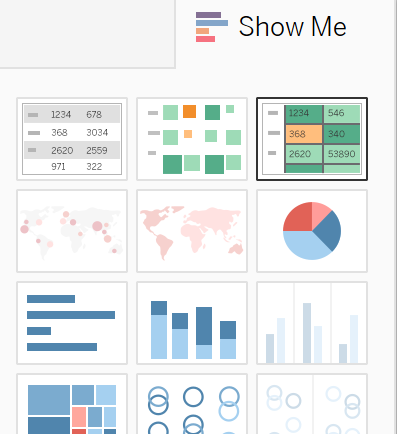
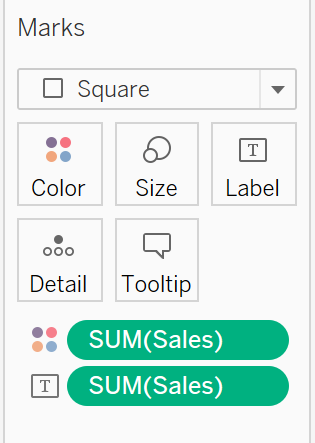
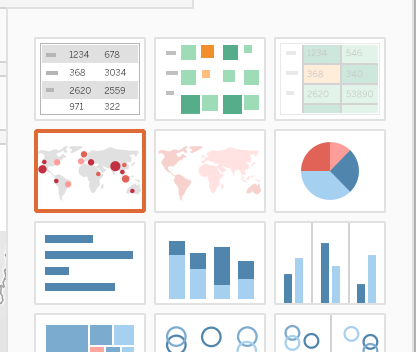
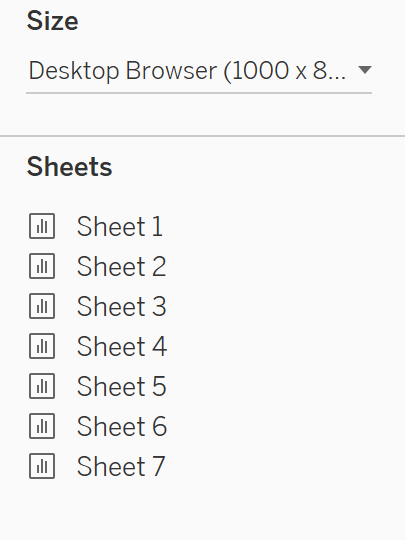
В нашей лабораторной будем использовать Live.

1. В самом низу страницы слева переходим во вкладку sheet1
2. Мы будем строить 6 различных диаграмм, для каждой из которых нам понадобится отдельный лист (sheet). В конце все листы мы прикрепим на одну интерактивную доску (dashboard).
3. Для начала построим график продаж за определенный период. Слева у нас два столбца – dimension and measures – какие-то факты, то есть данные, и измерения (что можно посчитать). Нам понадобятся «продажи» и «даты заказов». Кликаем два раза мышкой по «sales» из measures и видим, что он попадает сюда



1. Далее кликаем два раза на «Order date» и получаем это  
   
2. Мы хотим, чтобы продажи у нас суммировались не по годам, а по месяцам, поэтому выбираем второй month в выпадающем списке при двойном нажатии на YEAR(order date). После этого состояние данного параметра перейдет автоматически из Discrete в Continuous. Должно все выглядеть так   
   
3. Далее играемся с видом нашего графика. В поле *marks* найдем выпадающий список и посмотрим различные варианты отображения данных. Сейчас нас интересует *Area.* Выберем ее, а далее нажмем на *Сolor* и зададим цвет на свой вкус. Вуаля   
   
4. Создадим новое вычисляемое поле. Зачем? Для облегчения жизни. Щелкаем правой кнопкой мыши по “sales” в measures -> create -> calculated field. В открывшемся окошке надо прописать следюущее: WINDOW\_SUM(SUM([Sales])), а назвать это безобразие – продажи за период. Теперь нажимаем ОК и видим в measures новое поле. Перетаскиваем его в marks для того, чтобы оно отображалось у нас в качестве функции.
5. Теперь изменим название нашего графика, щелкнув по нему левой кнопкой мыши дважды. В названии можно креативить, но лучше называть график по его содержимому. Должно у вас получится следующее:  
     
   Название прописываем ручками, а функцию выбираем из выпадающего списка insert. С дизайном текста можно поиграться, чтобы было все по красоте.
6. В поле marks можно нажать на tooltip и изменить параметры осей графика, текста на них и тд. Делать это не обязательно, но надо попробовать, чтобы понять, как.
7. Далее в самом низу нажимаем левой кнопкой мыши два раза на sheet1 и задаем ему новое название (желательно такое же, как и графика).
8. Переходим к построению второго графика. Он будет очень похож на первый. Для этого нам нужен новый лист, который мы можем создать, нажав на самый левый плюсик   
   
9. Теперь будем находить прибыль за определенный период. Два раза нажимаем левой кнопкой мыши на profit, затем на Order date. Order date также разбиваем по месяцам, как и в пункте выше.
10. Поменяем цвет нашего графика на какой-нибудь. Во вкладке цвет выберем пунктик, чтобы график у нас был с точками. Изменим ширину линии во вкладке marks -> size.
11. Далее создадим еще одно вычисляемое поле как в пункте выше, только теперь не по продажам, а по прибыли. Делаем все то же самое.
12. Теперь сменим название нашего графика и нашего листа как в примере выше.
13. Снова можем поиграться с осями, но это не обязательно.
14. Создаем новый (уже третий, если кто пропустил) лист, на котором будем считать прибыль по различным категориям товаров.
15. Жмем два раза левой кнопкой мыши по profit, затем по order date, и снова в выпадающем списке выбираем MONTH.
16. В поле marks в типе графика выбираем bar.
17. Из dimensions берем category и накладываем на marks -> color. Должно получиться что-то похожее.   
    
18. Во вкладке color выберем приятные для глаз цвета.
19. Изменим название графика, название листа и поиграемся с осями. ГОТОВО
20. Создаем новый, четвертый, лист. Будем делать диаграмму «Прибыль по годам и категориям».
21. Удерживаем ctrl и выбираем category, order date, profit. Затем выбираем круговую диаграмму



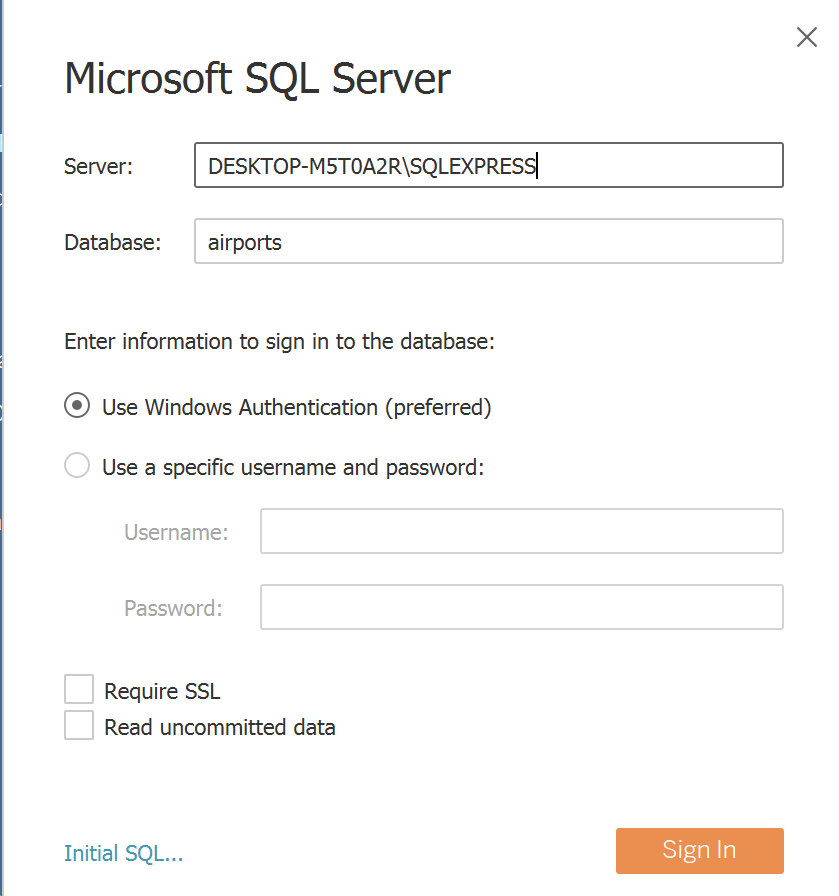
1. Удаляем здесь последний пункт  
   
2. Перетаскиваем number of records из measures в columns. Обращаемся к выпадающему списку этого поля, выбираем measures -> minimum. Должно получиться так   
   
3. Проделываем предыдущий пункт второй раз (должно получиться два столбика диаграмм)
4. Выбираем вот это вот поле и удаляем для него параметр category  
   
5. Диаграммы второго столбика стали серыми. Уменьшаем размер кругов в них (size), а затем выбираем цвет белый (color).
6. Теперь открываем выпадающее меню в правом пунктике MIN(number of records) -> dual axis (совмещаем две диаграммы в одну).
7. Меняем имя диаграммы, листа, играемся с осями и их названием.
8. Создаем пятый лист. В нем будут у нас «Тренды по дням недели»
9. Левой кнопкой мыши дважды кликаем на sales, затем на order date. У order date обращаемся к выпадающему списку, в первой верхней группе “year, month…” выбираем more -> weekday.
10. Выбираем выделенное отображение  
    
11. Из marks удаляем последнюю строчку, из-за чего на нашей диаграмме исчезнут цифры (в них нет нужды, так как при наведении на нужный прямоугольник нам видны все данные, а также переход цветов от светлого к темному означает возрастание значений)  
    
12. В rows перетаскиваем order date, в выпадающем списке в первой группе «year, month…» выбираем month. Теперь наша таблица показывает сумму продаж по месяцам в каждый из дней недели. Интересная штука.
13. Можно изменить цветовую гамму диаграммы, можно изменить период подсчета суммы (выбрать кварталы, дни и тд). Экспериментируйте на ваше усмотрение и не забудьте поменять имя диаграммы, имя листа и убрать лишние параметры с осей, если они вам мешают.
14. Создаем лист 6 и наконец-то приступаем к работе с картой. Что нам для этого нужно? Вместе с ctrl выбирает city и country из dimension и sales из measures. Далее выбираем карту  
    
15. У вас должна появиться и прогрузиться типичная карта. Проблема в том, что все города у нас лежат в пределах United States, поэтому местность не будет столь разнообразной. С помощью size мы регулируем размер наших точек на карте.
16. Перетаскиваем sales на color и меняем цветовую гамму (желательно выбрать какую-то яркую, так как города с минимальной суммой продаж будут априори бледными, а хочется их видеть).
17. Даем имя нашей карте и нашему листу. Не забываем сохраняться.
18. Далее нажимаем на средний плюсик и создаем наш первый Dashboard  
    
19. Слева во вкладке sheets отображаются все наши листы, которые мы можем спокойненько перетащить на интерактивную доску и расставить, как заблагорассудится.   
    
20. Также можно удалить название листов с интерактивной доски, если вам они кажутся неуместными.
21. Сохраняем проект и показываем Даше, Ивану или Наташе)

**Творческое задание**

Это задание обязательное, так как оно позволит вам самим лучше познакомиться с инструментами Tableau. ИМХО, программа действительно стоящая, интересная, а главное, - забавная.

Теперь мы будем экспортировать данные из sql server. Если у вас на устройстве нет sql server и sql management studio, то подсядьте к товарищу и наслаждайтесь жизнью.

Надо зайти в sql management studio, подключиться к серверу, создать новую БД под названием Airports, а в ней – новый запрос. Далее переходим по третьей ссылке из данных и натыкаемся на код запроса. Копируем его, вставляем в наш запрос, удаляем первые 3 строчки и последние 2 строчки и выполняем.

Создаем в Tableau новый файл. Выбираем источник данных (вверху на панели инструментов data -> new data source). Теперь выбираем в качестве типа Microsoft SQL Server. Вылезет такое окошко:  


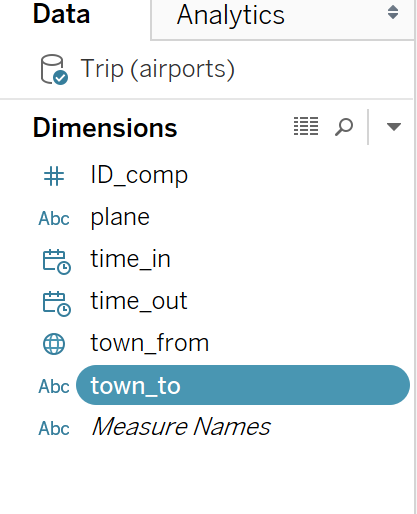
В верхнюю строчку из sql management studio копируем имя нашего сервера, в нижней строчке прописываем имя нашей БД (Airports). Теперь мы присоединили данные. Далее алгоритм такой же, как и в случае с файлом Excel. Мы выбираем нужную таблицу, перетаскиваем ее, смотрим, какие в ней данные и работаем с ними.

Прежде чем браться за творчество внимательно изучите все таблицы этой БД, посмотрите на их связи, атрибуты.

Далее выбираете любую(ые) понравившееся таблицы и делаете свою уникальную Dashboards из 6 различных листов. Обязательным условием является наличие минимум одной карты.

По выполнении творческого задания можете позвать Дашу, Ивана или Наташу и рассказать, что вы сделали, освоили ли вы еще какие-либо новые инструменты ну и вот это все.

**Примечание**

Прежде чем работать с картой в новой БД, надо убедиться в правильном формате географических данных.   


Выбираем, например, город. Если напротив него стоит значок глобуса, значит данные у нас в географическом формате, и все хорошо. В противном случае мы левой кнопкой мыши щелкаем по ТИПУ данных, внизу видим пункт Geographic role -> city (country/region).

**Данные**

1. <https://www.tableau.com/products/desktop/download> - скачивание пробной 14дневной версии Tableau desktop
2. <https://drive.google.com/open?id=1liuvJsX8bgokN-iJRRpYqcMTbx3MeasM> – ссылка на Excel таблицу с данными для основного задания
3. <http://www.sql-ex.ru/download/aero_script.sql> - ссылка на скрипт для творческого задания

**Содержание отчета**

1. Титульный лист (куда ж без него)
2. Цель работы
3. Ход выполнения со скриншотами итоговых листов и досок, с пояснением
4. Выводы

**Контрольные вопросы**

1. Для чего используется Tableau?
2. Назовите основные достоинства и недостатки Tableau
3. Чем sheet отличается от Dashboard?
4. Какие инструменты можно найти в Tableau?
5. Работу с данными каких типов поддерживает Tableau?