

# Business Intelligence (BI)

## На этом уроке

1. О направлении BI как «фронтенде» результатов работы дата-инженера (популярные инструменты, примеры задач).
2. Как работа Data Engineer связана с BI.

## Теория

1. Что такое направление Business Intelligence.
2. Какую пользу бизнесу приносит BI.
3. Примеры BI инструментов.
4. На какие вопросы можно ответить с помощью аналитики по данным в BI инструментах (примеры).
5. Роль Data Engineer в BI: о связи ETL-процессов с подготовкой данных в эффективном и удобном для аналитики формате.

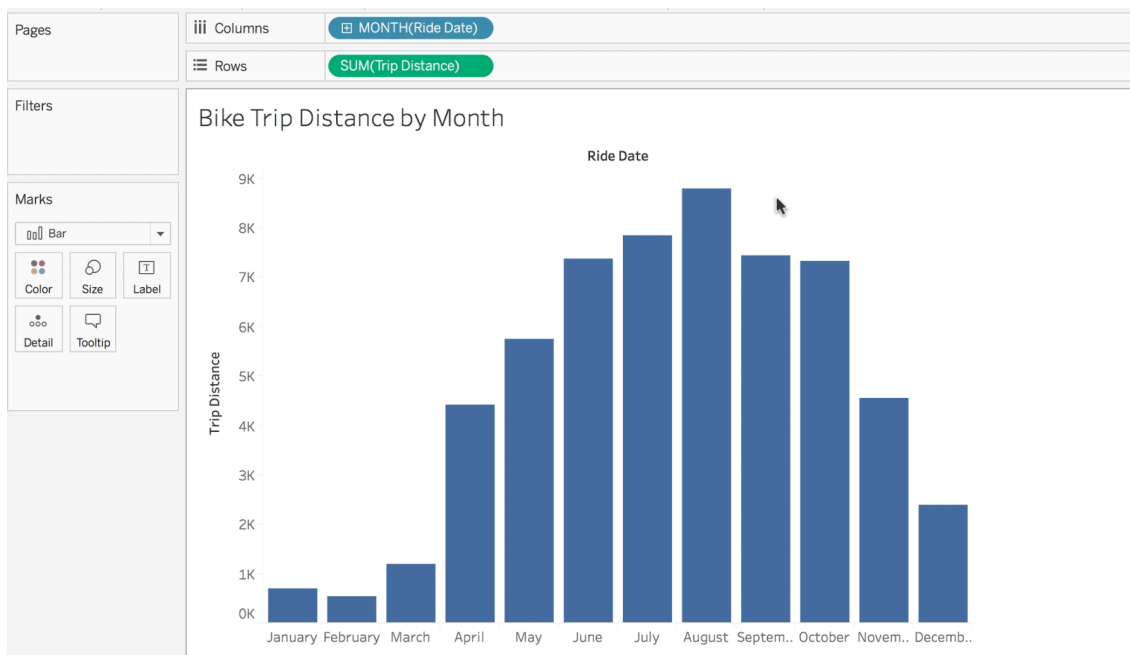
## Что такое Business Intelligence?

Для начала вот определение термина простыми словами: **Business Intelligence — это про извлечение пользы из данных**. BI помогает бизнесу принимать решения на основе данных (англ. data-driven decisions).

Если же говорить более научным языком, то **Business Intelligence — это преобразование необработанных данных в полезную информацию**. Business Intelligence состоит из:

1. Сбора данных (ETL - extract, transform, load).
2. Хранения данных в удобном для анализа виде (Многомерная модель, OLAP)
3. Отображения данных (Системы построения и предоставления отчетности)

Чтобы было нагляднее, вот как выглядит результат работы в BI инструменте:



Это график или набор графиков (т.н. *дашборд*), построенных по данным, описывающим бизнес. Например, на графике выше приведено число поездок на велосипедах в разбивке по дате поездки с группировкой по месяцам: каждый столбец отражает число поездок в указанном месяце. Такой график может быть полезен менеджеру сервиса аренды велосипедов, чтобы принимать решение о запуске маркетинговых кампаний в зависимости от сезонности.

## Исторический контекст

Термин BI не отражает своего фактического значения, потому как имеет очень долгую историю по меркам IT индустрии, а его значение адаптировалось «под нужды» бизнеса.

Направлению BI уже более 60 лет — свое начало он берет в 1960-ых годах. Тогда термин обозначал комплекс мер по распространению информации внутри организаций и за ее пределами. **Бизнес учился быть «умнее» за счет накопленных знаний:** экспертизы, данных и т.п.

К 1980-ым, то есть спустя 20 лет после появления, направление BI формализовалось и этим термином стали обозначать **подход к принятию решений, основанных на данных**. Оформилось понятие вывода, или *инсайта* (англ. insight): чтобы данные приносили пользу, из них **необходимо извлечь инсайты и представить в удобочитаемом виде** лицу, принимающему решение — например, топ-менеджеру бизнеса для выбора стратегии компании или региональному менеджеру для принятия решения об изменении объема поставок в региональное подразделение сети офисов продаж.

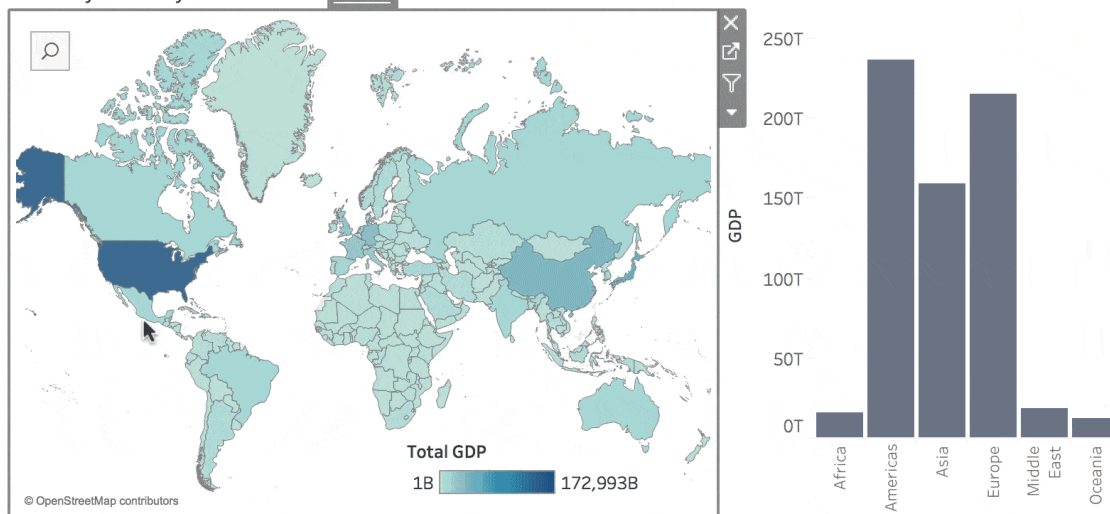
Сейчас BI — это внутренний продукт компании IT природы. В широком смысле BI как подход объединяет в себе аналитику, data mining, визуализацию, и даже инфраструктуры и инструменты для работы с данными, которые реализуют лучшие практики индустрии данных.

Таким образом, сейчас BI — это **комплекс мер и практик, направленных на принятие решения на основе данных**.

## Примеры Business Intelligence: инструменты и задачи

Обратите внимание на скринкаст из первого раздела лекции: важно, что **визуализация данных интерактивна**. Человек, который использует этот график для принятия решений, должен иметь возможность «покрутить» данные, конфигурируя способ отображения данных по своему усмотрению. Например, ограничивая выборку диапазоном дат, конкретным регионом, или даже меняя способ визуализации со столбчатой диаграммы на линейчатую или карту:

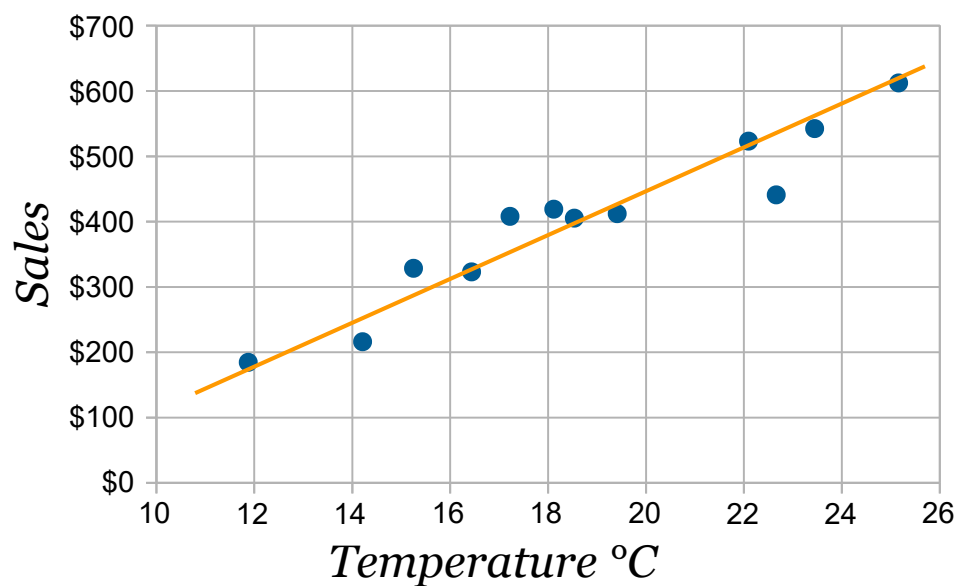
GDP by Country



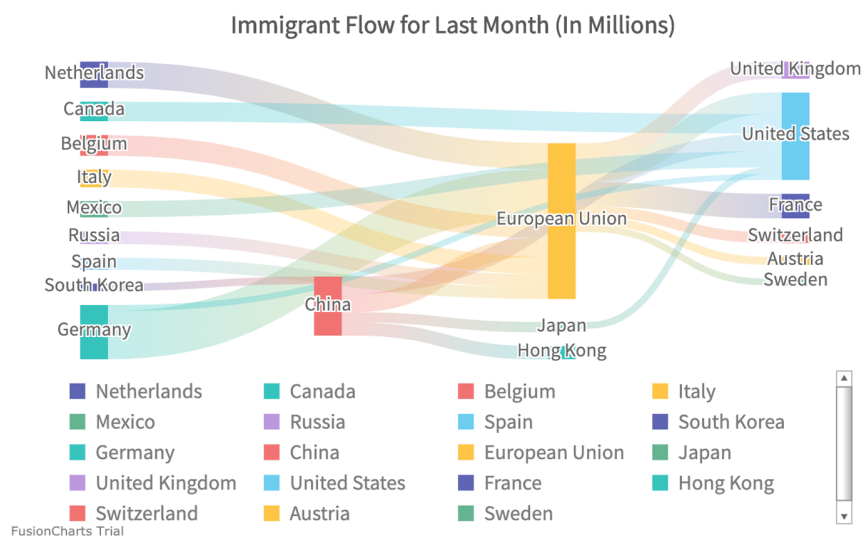
## Задачи

Хронологически BI инструменты начинают внедрять в компаниях **для построения отчетов и мониторинга бизнеса**. Стейкхолдерам (например, руководству) предоставляются графики с ежемесячной выручкой и другими показателями бизнеса: например, в случае такси агрегатора числом уникальных пользователей, средним количеством поездок на пользователя и т.п. с фильтрами по регионам, месяцам, тарифам и прочим атрибутам.

Далее на базе данных **строятся инсайты: опорные точки для принятия решений**. Например, BI инструмент может отвечать на простые вопросы о том, какой тренд преобладает в изменении выручки: нисходящий или восходящий.



А также более сложные вопросы: например, отразить зависимости в миграционных потоках.



Перечисленные задачи упрощаются, если решать их графически, поэтому все BI инструменты направлены именно на построение разнообразных визуализаций.

### Инструменты: классификация и примеры

BI инструменты - в широком смысле, это средства извлечения полезной информации из какого либо массива данных. Причём данные могут быть представлены в любом произвольном виде.

По отношению к источникам данных BI инструменты бывают:

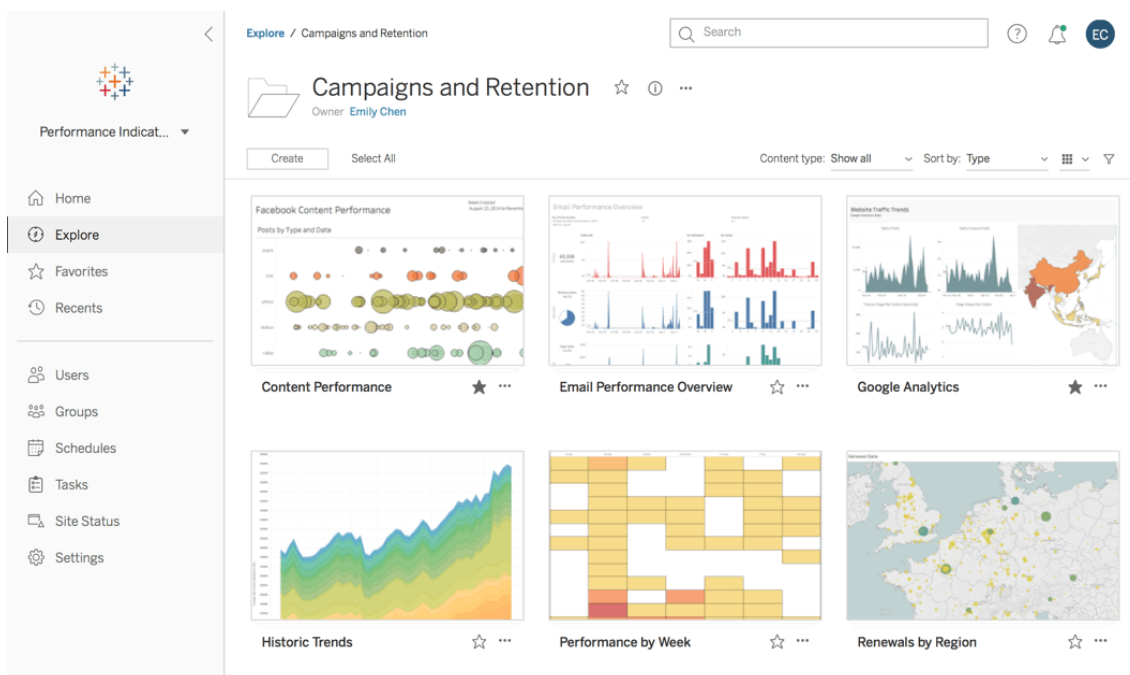
- Тонкими: используются только для отрисовки, а вычисления выполняются на стороне источника;
- Толстыми: кроме отрисовки могут хранить и производить вычисления над данными;

Так же бывают либо Web-приложениями, либо Desktop-приложениями.

- Web-приложения, это отдельный сервер в сети, к которому можно обратиться и он вернет необходимую информацию;
- Desktop-приложения, это приложения, которые устанавливаются непосредственно на рабочую станцию и работают в рамках этой рабочей станции.

Далее перечислим примеры известных BI инструментов.

#### Tableau



Один из лидеров индустрии. Приведенные выше скриншоты и скринкасты — это Tableau.

Основные преимущества:

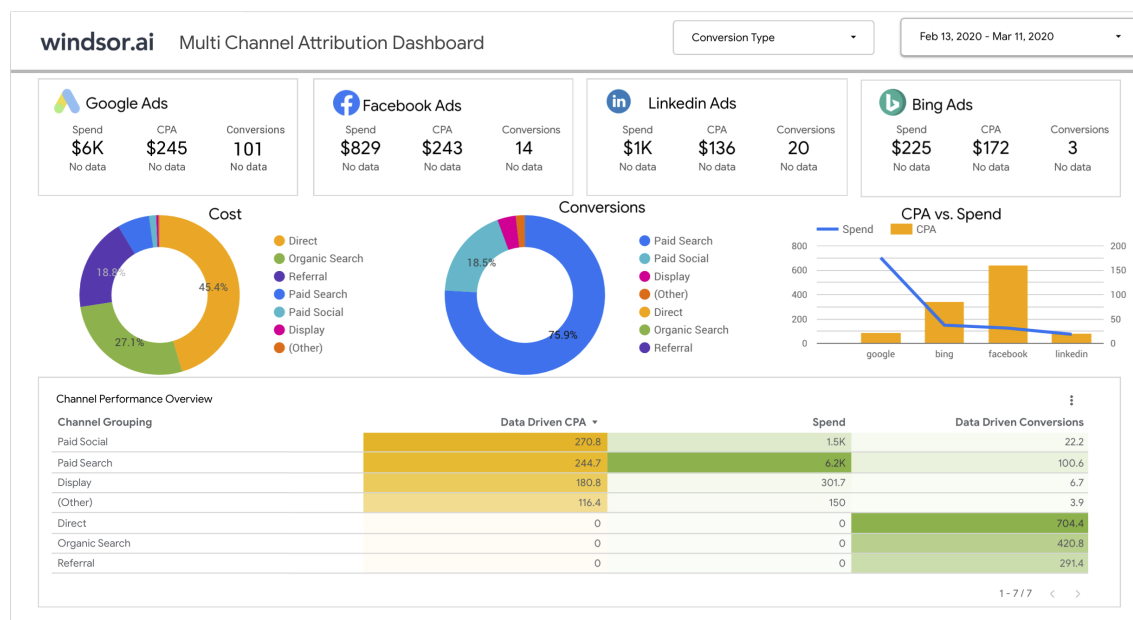
1. Большое разнообразие способов визуализации.
2. Гибкость в настройке: данные можно разбивать на категории вручную, по некоторому порогу, по размеру группы и т.п.
3. Есть возможность «кешировать» результаты вычислений на базе данных и компактно хранить в специальном формате Tableau Extract'ов ближе к клиенту: на централизованном Tableau сервере или в локально на машине пользователя.
4. Гибкая модель лицензирования. Есть десктопная и веб-версии, read-only лицензии, серверная лицензия — в общем, подходящий инструмент для большой компании. Увы, подразумевает и крупные бюджеты в тысячи долларов в год.

## Microsoft Power BI



Второй лидер в индустрии. Имеет схожие с Tableau характеристики и даже похожую модель лицензирования. Особенно удобен, если компания использует офисный стек Microsoft.

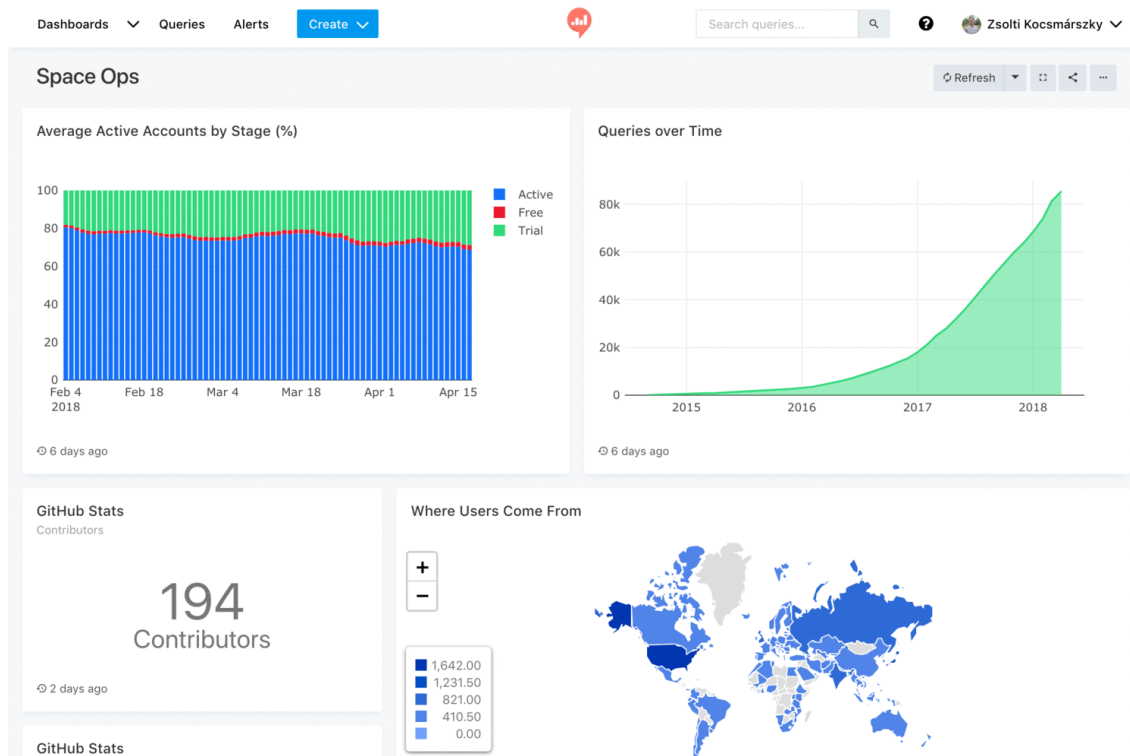
## Облачные решения



Google Cloud Data Studio



## Open Source решения



Наиболее популярны Apache Superset, Redash и Metabase. Open source решения богаты на коннекторы, потому как поддерживаются и расширяются целым сообществом, но имеют традиционный недостаток — отсутствие поддержки.

Есть ещё Grafana, которая, в большей степени заточена под создание дашбордов на основе временных метрик.

### Microsoft Excel и Google Sheets

Инструменты для работы с электронными таблицами тоже хорошо решают задачи BI и часто являются отправной точкой.





Полноценный дашборд в Excel

## О ролях в Business Intelligence

Когда мы говорим о BI в широком смысле, имея в виду работу с данными в целом и извлечение из них пользы, то мы обозначаем область ответственности роли Data Engineer. Именно **Data Engineer'ы внедряют этот подход в бизнесе, создавая платформу для работы с данными.**

Дадим пояснения об отличиях Data Engineer, Data Analyst и Data Scientist.

Расскажем об этом через хронологию развития команды. Когда компания задумывается об извлечении пользы из данных, сначала нанимают аналитика (Data Analyst) или датасаентиста (Data Scientist). Эти специалисты могут самостоятельно сконвертировать данные в инсайты (в некую осмысленную и понятную информацию, которая даёт понимание происходящих процессов). Обычно работа этих специалистов business-driven: они **выполняют прямые заказы бизнеса**, не тратя время на отстраивание практики переиспользования результатов работы — это дает **гибкость и повышает скорость доставки** результатов.

DA/DS как ремесленники: каждый из них талантлив в ручном труде, но **чтобы поставить производство на поток, нужен инженер**. Когда у данных появляется много стейкхолдеров (читайте — пользователей), приходит время наращивать обороты «бизнеса данных» внутри компании. Тогда и появляется Data Engineer, который генерализует наработки аналитиков и датасаентистов, организуя платформу для работы с данными.

**DA/DS + DE — это основной набор ролей в команде данных.** Чем больше растёт нужда в генерализации подхода, тем больше специалистов появляется (здесь речь идёт о классическом подходе): тимлид (Team Lead) для управления командой дата-инженеров как командой разработки, проектный менеджер (Project Manager) для организации процессов, архитектор хранилища данных для принятия решений о модели данных и технической платформе, системные аналитики (System Analyst) для проектирования сущностей и поддержки знаний о домене, и, иногда, выделенные специалисты по различным инструментам Business Intelligence, а также разного рода аналитики с углубленной экспертизой в каждой отдельной области бизнеса.

Основными пользователями платформы, поддерживаемой Data Engineer'ами, выступают аналитики, менеджмент и другие сотрудники компании, связанные с работой с данными.

Полезно различать области ответственности и подходы аналитиков и Data Scientist'ов. **Data Scientist ищет неявные закономерности в данных**, строит прогнозные модели. **Аналитик же ищет возможности для**

**улучшения бизнес-показателей**, видит закономерности и понятно доносит результаты анализа до лиц, принимающих решения.

Важно понимать, что перечисленные выше «должности» — это роли. Каждый сотрудник может брать на себя несколько ролей полностью или частично, что несколько размывает между ними границу. **Наиболее плотно с BI инструментами работают аналитики**: они занимаются извлечением инсайтов из данных для лиц, принимающих решения — менеджмента.

В общем виде, понимание того, кто за что ответственен несут в себе значения таких терминов как:

- ЛПР — лицо, принимающее решение,
- СППР — система поддержки принятия решений,

Почему важно знать о перечисленных выше ролях: когда вы как Data Engineer возьметесь за реализацию платформы, вам потребуется **понимать, чьи потребности вы реализуете** в первую очередь, чтобы собрать требования с нужных людей.

## Как Data Engineer помогает компании строить BI

Задачи Data Engineer'a сводятся к организации размещения данных, доступа к ним и навигации по ним удобным пользователям способом. **В идеально спроектированной платформе данных конечный пользователь может самостоятельно найти необходимые ему данные и извлечь из них пользу.**

Также Data Engineer должен понимать архитектурные особенности и узкие места используемых технологий. Подробнее о доступном наборе технических решений, их достоинствах и недостатках мы расскажем в дальнейшем.

## Домашнее задание

Как результат обучения на курсе вам потребуется **построить аналитику по своим расходам**. Чтобы подготовиться к его выполнению, найдите возможность выгрузить историю своих операций из вашего онлайн-банка. ~~Если ваш банк такое не умеет, воспользуйтесь моей реферальной ссылкой...~~ :) Желательно чтобы она была в табличном формате, лучше всего в CSV. В крайнем случае приведите ее к такому формату любым удобным вам способом.

Захватите достаточно продолжительный период, чтобы иметь возможность проанализировать ваши траты в месячном срезе.

Если вы не хотите раскрывать свои финансовые привычки преподавателю, проверяющему домашнее задание, вы можете сгенерировать случайный набор данных, приближенный к реальности (походы в продуктовые магазины случаются чаще, чем покупки спорткаров), или построить визуализацию по альтернативному датасету — встроенному в Tableau [Superstore](#).

## Источники

1. [What is business intelligence?](#)
2. [В чём разница между аналитикой данных и data science](#)
3. [Self-Service BI: обзор задач и рынка](#)