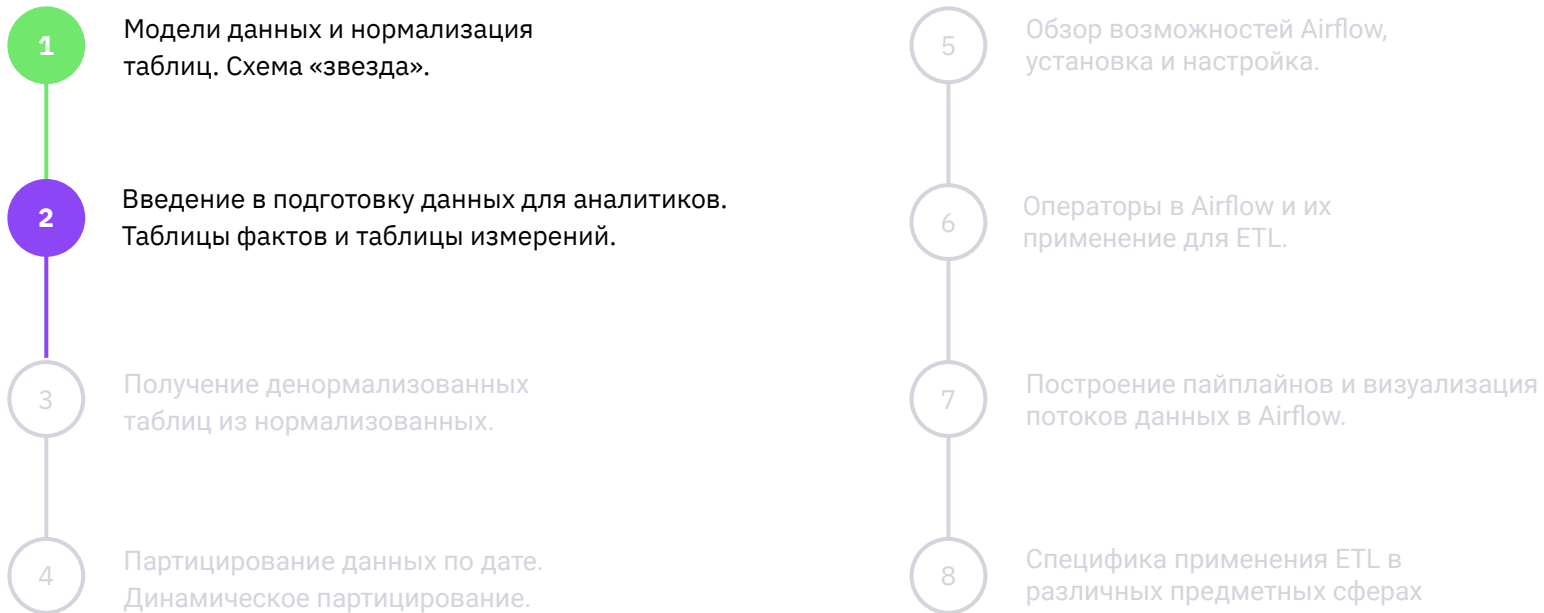


Введение в подготовку данных для аналитиков. Таблицы фактов и таблицы измерений.

Урок 2











План курса



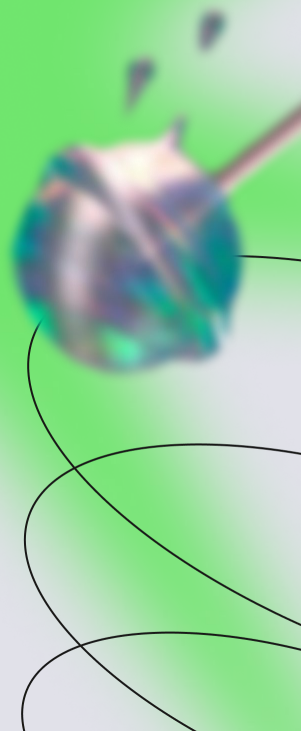


Что будет на уроке сегодня

-  Тенденции в бизнес аналитике
-  Понимание бизнеса и данных
-  Подготовка и преобразование данных
-  Исследование и визуализация данных
-  Моделирование
-  Оценка
-  Развертывание
-  Таблицы фактов и измерений



Введение в подготовку данных для аналитиков





Что такое бизнес-аналитика

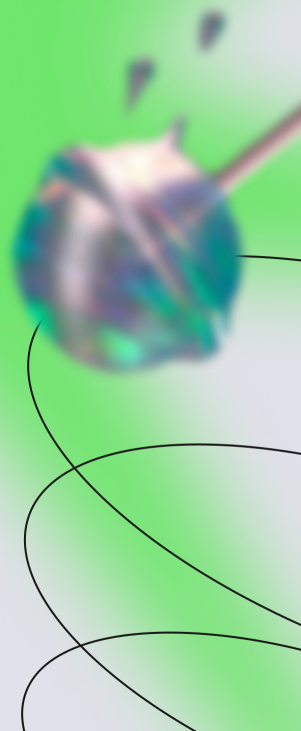
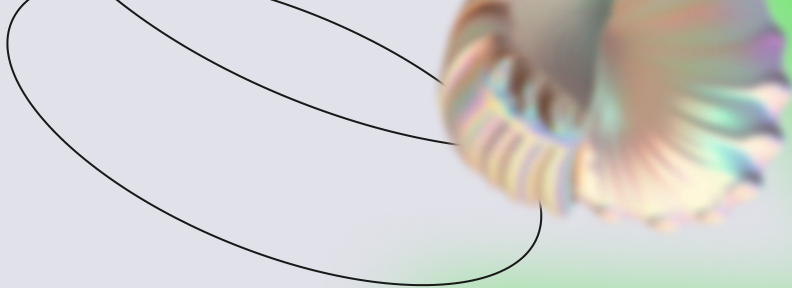
Бизнес-аналитика (BI – Business intelligence) – это набор процессов, архитектур и технологий, которые преобразуют необработанные данные в значимую информацию, стимулирующую прибыльные бизнес-действия.

Почему бизнес-аналитика имеет важное значение:

- Измерение: создание KPI (ключевых показателей эффективности) на основе исторических данных
- Определите и установите критерии для различных процессов.
- С помощью BI-систем организации могут выявлять рыночные тенденции и выявлять бизнес-проблемы, которые необходимо решить.
- BI помогает визуализировать данные, что повышает качество данных и, следовательно, качество принятия решений.
- Системы BI могут использоваться не только предприятиями, но и МСП (малые и средние предприятия)



Несколько примеров





Тенденции в бизнес-аналитике



Искусственный интеллект

отчет Gartner указывает, что ИИ и машинное обучение теперь выполняют сложные задачи, выполняемые человеческим интеллектом. Эта возможность используется для анализа данных в реальном времени и создания отчетов на приборной панели.



Collaborative BI

программное обеспечение BI в сочетании с инструментами совместной работы, в том числе социальными сетями, и другими новейшими технологиями расширяют возможности коллективной работы и обмена информацией для совместного принятия решений.



Embedded BI

позволяет интегрировать программное обеспечение BI или некоторые его функции в другое бизнес-приложение для расширения и расширения его функций отчетности.



Облачная аналитика

BI-приложения скоро будут предлагаться в облаке, и все больше компаний будут переходить на эту технологию. Согласно их прогнозам, в течение пары лет расходы на облачную аналитику будут расти в 4,5 раза быстрее



Вопрос

Какое назначение у BI? Какие плюсы и минусы вы видите у систем бизнес-аналитики?

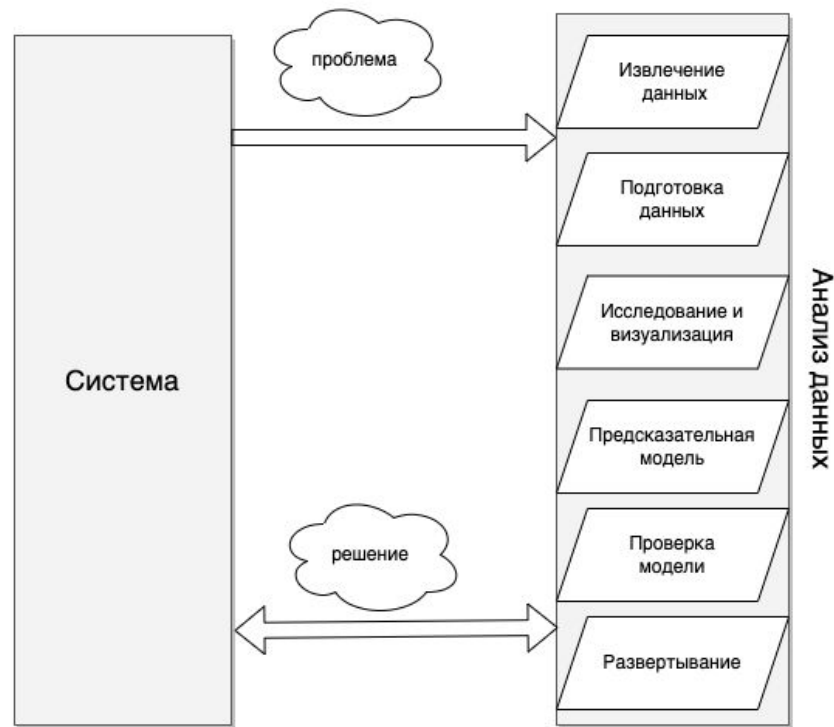




Анализ данных

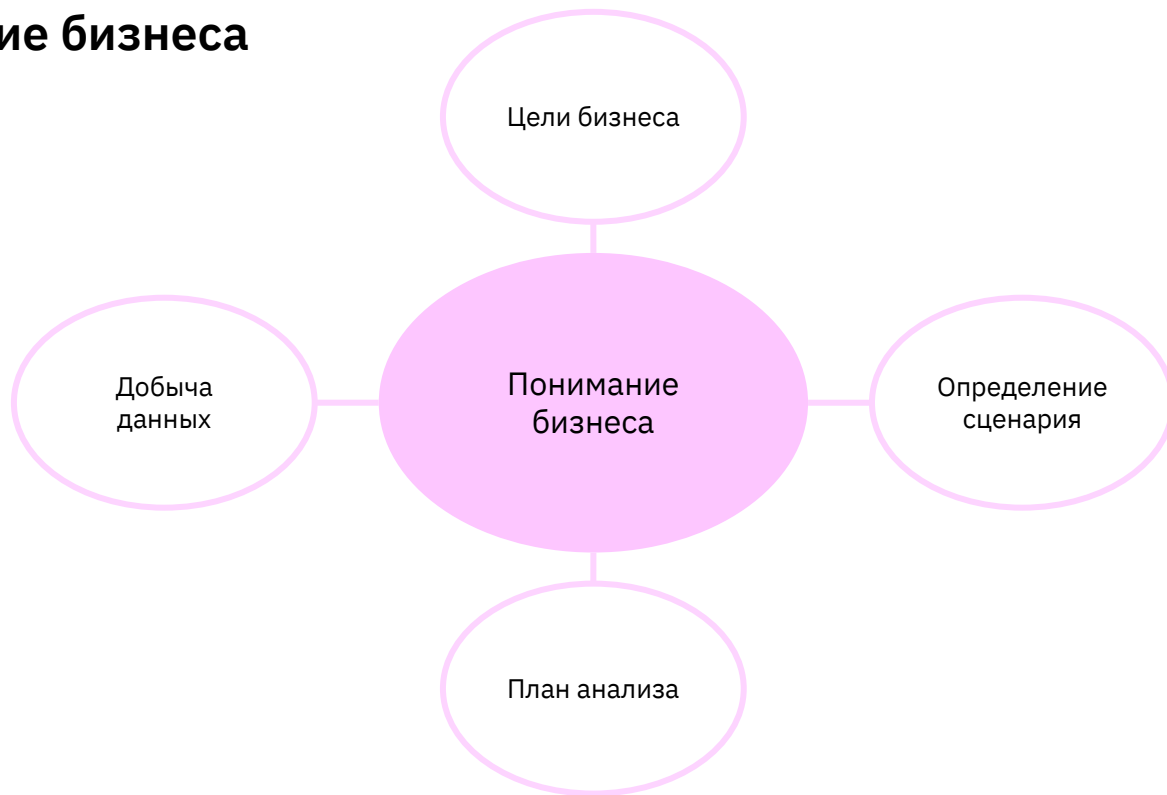
Анализ данных — это всего лишь последовательность шагов, каждый из которых играет ключевую роль для последующих. Этот процесс похож на цепь последовательных, связанных между собой этапов:

- Определение проблемы;
- Извлечение данных;
- Подготовка данных — очистка данных;
- Подготовка данных — преобразование данных;
- Исследование и визуализация данных;
- Моделирование;
- Оценка (проверка) модели;
- Развертывание — визуализация и интерпретация результатов;
- Развертывание — развертывание решения.



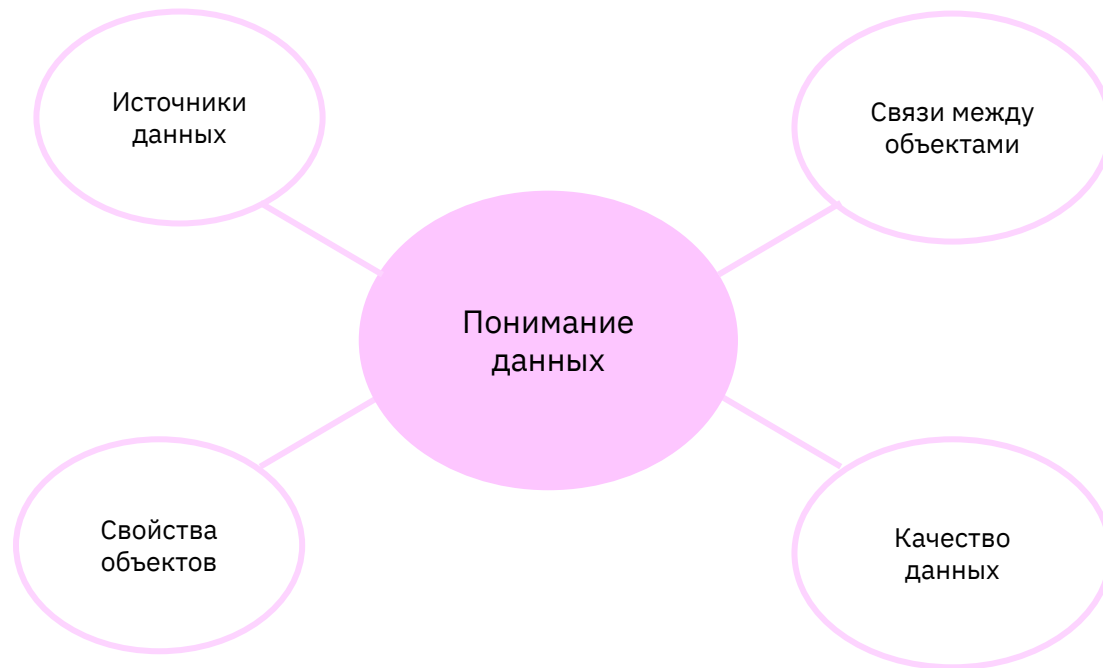


Понимание бизнеса





Понимание данных





Подготовка данных





Преобразование данных





Исследование и визуализация данных

Изучение данных — это их анализ в графической или статистической репрезентации с целью поиска моделей или взаимосвязей. **Визуализация — лучший инструмент для выделения подобных моделей.**

- Обобщение данных;
- Группировка данных;
- Исследование отношений между разными атрибутами;
- Определение моделей и тенденций;
- Построение моделей регрессионного анализа;
- Построение моделей классификации.



Моделирование

- **Модели классификации:** если полученный результат — качественная переменная.
 - **Регрессионные модели:** если полученный результат числовой.
 - **Кластерные модели:** если полученный результат описательный.
- линейная регрессия,
 - логистическая регрессия,
 - классификация,
 - дерево решений,
 - метод k-ближайших соседей.



Оценка



Результаты, полученные с помощью модели интеллектуального анализа данных, должны оцениваться в соответствии с бизнес-целями.



Получение понимания бизнеса является итеративным процессом. Фактически, при понимании, новые бизнес-требования могут быть повышены из-за интеллектуального анализа данных



Принято решение о переходе модели на этап развертывания.



Развертывание

- 💡 Знания или информация, обнаруженные в процессе извлечения данных, должны быть понятны для нетехнических заинтересованных сторон
- 💡 Создан подробный план развертывания для доставки, обслуживания и мониторинга обнаружений интеллектуального анализа данных
- 💡 Окончательный отчет по проекту создается с учетом извлеченных уроков и ключевых событий в ходе проекта. Это помогает улучшить деловую политику организации.



Таблицы фактов и таблицы измерений





Таблицы фактов

Таблица фактов — является основной таблицей хранилища данных. Как правило, она содержит сведения об объектах или событиях, совокупность которых будет в дальнейшем анализироваться. Обычно говорят о четырех наиболее часто встречающихся типах фактов. К ним относятся:

Transaction facts

основаны на отдельных событиях (типичными примерами которых являются телефонный звонок или снятие денег со счета с помощью банкомата)

Snapshot facts

Основаны на состоянии объекта (например, банковского счета) в определенные моменты времени, например на конец дня или месяца. Типичными примерами таких фактов являются объем продаж за день или дневная выручка;

Line-item facts

Основаны на том или ином документе (например, счете за товар или услуги) и содержат подробную информацию об элементах этого документа (например, количестве, цене, проценте скидки);

Event or state facts

Представляют возникновение события без подробностей о нем (например, просто факт продажи или факт отсутствия таковой без иных подробностей)



Таблица размеров

Типы измерения	Определение
Согласованные размеры	Соответствующие размеры – это тот факт, к которому они относятся. Это измерение используется более чем в одной схеме или в Datamart.
Размеры аутригера	Измерение может иметь ссылку на другую таблицу измерений. Эти вторичные размеры называются размерами аутригеров. Этот вид измерений следует использовать осторожно.
Сжатые размеры свертки	Сжатие Размеры свертки являются подразделением строк и столбцов базового измерения. Эти виды измерений полезны для разработки агрегированных таблиц фактов.
Соединения таблицы размеров	Размеры могут иметь ссылки на другие размеры. Тем не менее, эти отношения могут быть смоделированы с размерами аутригера.
Ролевые измерения	Одно физическое измерение помогает ссылаться несколько раз в таблице фактов, поскольку каждая ссылка связана с логически отдельной ролью измерения.
Размеры барахла	Это коллекция случайных транзакционных кодов, флагов или текстовых атрибутов. Он не может логически принадлежать какому-либо определенному измерению.
Вырожденные размеры	Вырожденное измерение без соответствующего измерения. Он используется в транзакции и для сбора таблиц фактов моментальных снимков. Этот вид измерения не имеет своего измерения, поскольку он получен из таблицы фактов.
Заменяемые размеры	Они используются, когда одна и та же таблица фактов связана с разными версиями одного и того же измерения.
Шаг Размеры	Последовательные процессы, такие как события веб-страниц, в большинстве случаев имеют отдельную строку в таблице фактов для каждого шага процесса. Он сообщает, где конкретный шаг следует использовать в общем сеансе.



Таблица размеров

Параметры	Таблица факторов	Таблица измерений
Определение	Измерения, метрики или факты о бизнес-процессе.	Сопоставительная таблица с таблицей фактов содержит описательные атрибуты, которые будут использоваться в качестве ограничения запроса.
Характеристика	Расположен в центре схемы звезды или снежинки и окружен размерами.	Подключен к таблице фактов и расположен по краям схемы звезды или снежинки
Дизайн	Определяется их зерном или самым атомарным уровнем.	Должен быть объемным, описательным, полным и гарантированным качеством.
Задача	Таблица фактов – это измеримое событие, для которого собираются данные таблицы измерений и используются для анализа и составления отчетов.	Сбор справочной информации о бизнесе.
Тип данных	Таблицы фактов могут содержать информацию о продажах по таким параметрам, как «Продукт» и «Дата».	Таблица измерения Evert содержит атрибуты, которые описывают детали измерения. Например, размеры продукта могут содержать идентификатор продукта, категорию продукта и т. д.
Ключ	Таблица фактических первичных ключей отображается как внешние ключи для измерений.	Таблица измерений имеет столбцы первичного ключа, которые однозначно определяют каждое измерение.
Место хранения	Помогает хранить метки отчетов и фильтровать значения доменов в таблицах измерений.	Загрузите подробные атомарные данные в размерные структуры.
Иерархия	Не содержит иерархии	Содержит иерархии. Например, Местоположение может содержать страну, пин-код, штат, город и т. д.



Ключевая разница



Таблица фактов содержит измерения, метрики и факты о бизнес-процессе, тогда как таблица измерений является дополнением к таблице фактов, которая содержит описательные атрибуты, которые будут использоваться в качестве ограничения запроса



Таблица фактов расположена в центре схемы «звезда» или «снежинка», а таблица «Измерения» расположена по краям схемы «звезда» или «снежинка»



Таблица фактов определяется их гранью или ее наиболее атомарным уровнем, тогда как таблица измерений должна быть многословной, описательной, полной и гарантированной по качеству



Таблица фактов помогает хранить метки отчетов, тогда как таблица измерений содержит подробные данные



Таблица фактов не содержит иерархии, тогда как таблица измерений содержит иерархии



Спасибо за внимание

