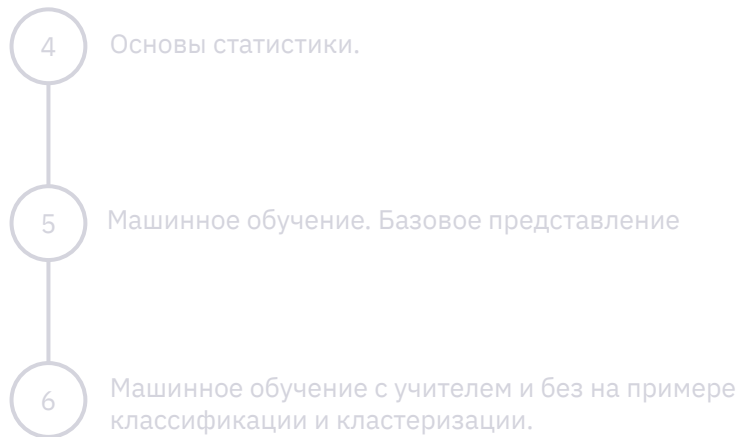
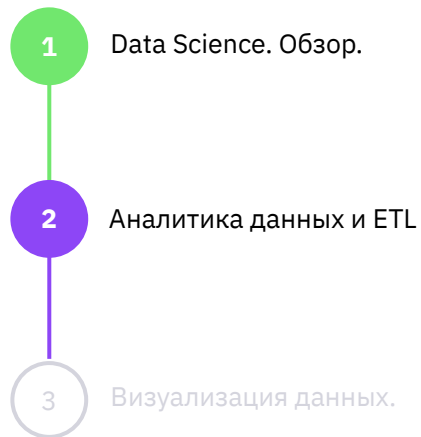


Аналитика данных и ETL

Урок 2

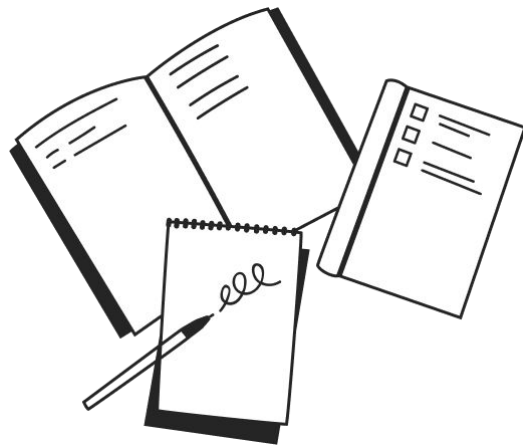


План курса



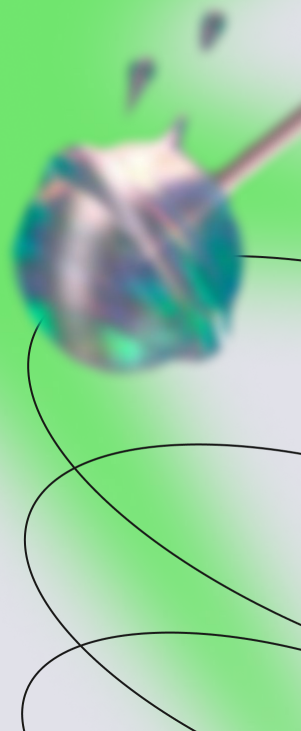
Что будет на уроке сегодня

- 📌 Что такое Data Science
- 📌 Для чего нужна наука о данных?
- 📌 Data Scientist и бизнес-аналитика
- 📌 Каковы преимущества науки о данных для бизнеса
- 📌 Методы науки о данных
- 📌 Кто такой Data Scientist
- 📌 Инструменты Data Scientista
- 📌 Облачные технологии в Data Science
- 📌 Сферы применения Data Science
- 📌 Примеры проектов Data Science





Бизнес аналитика и анализ данных





Что такое бизнес-аналитика

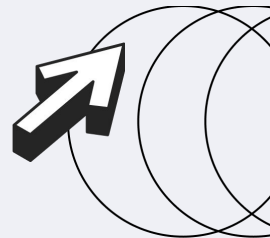
Бизнес-аналитика (BI – Business intelligence) – это набор процессов, архитектур и технологий, которые преобразуют необработанные данные в значимую информацию, стимулирующую прибыльные бизнес-действия.

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?

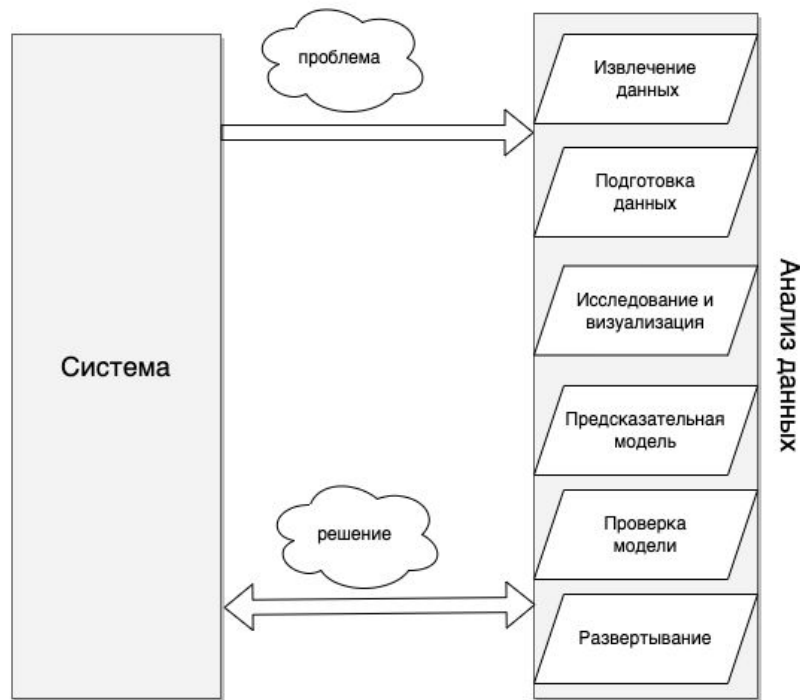




Анализ данных

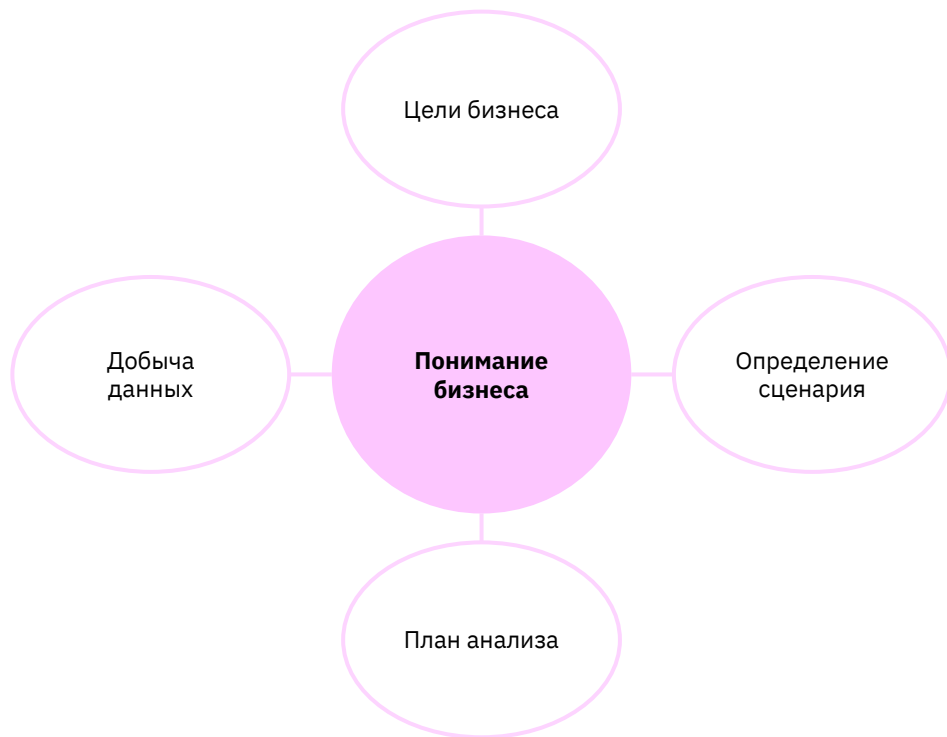
Анализ данных — это всего лишь последовательность шагов, каждый из которых играет ключевую роль для последующих. Этот процесс похож на цепь последовательных, связанных между собой этапов:

- 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.



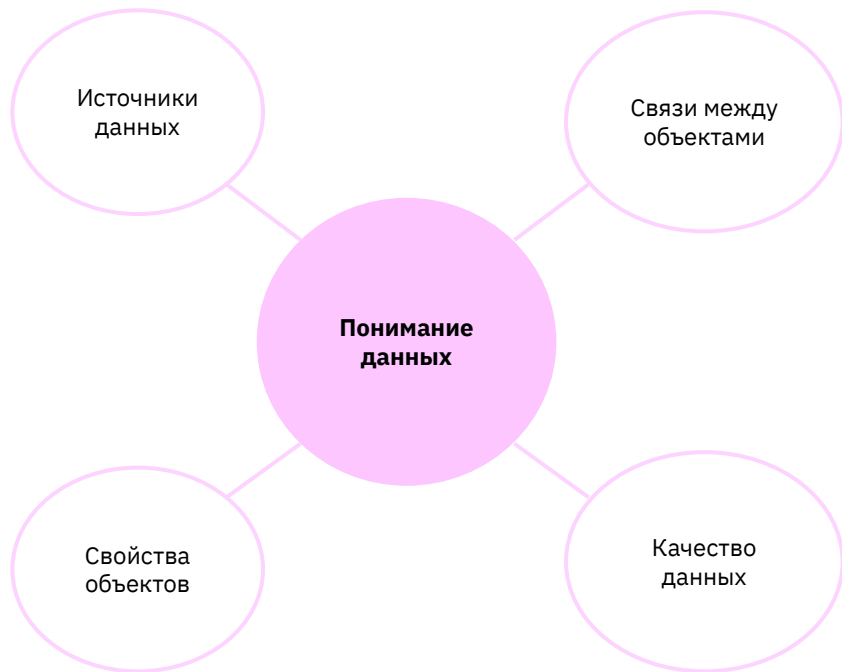


Понимание бизнеса





Понимание данных



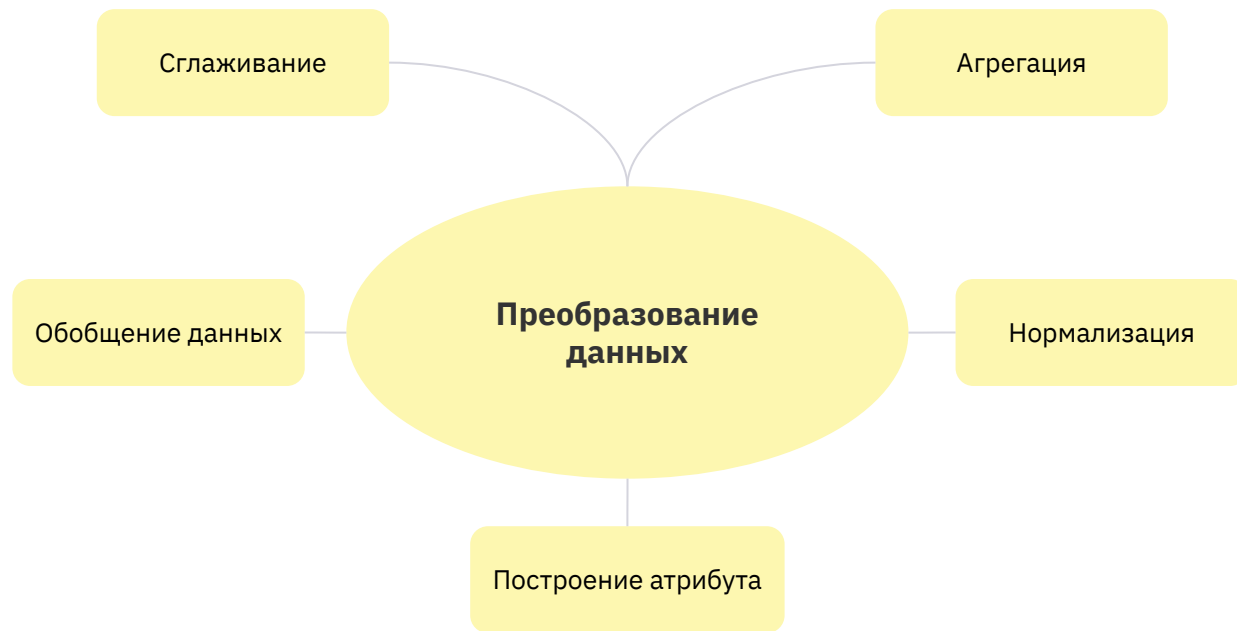


Подготовка данных





Преобразование данных

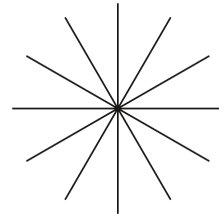
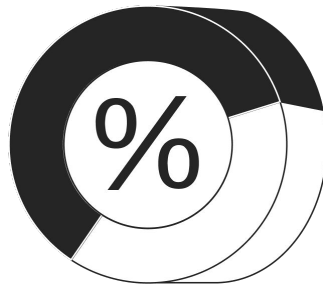




Исследование и визуализация данных

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.

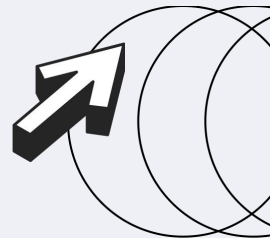


Python для базовых задач обработки данных



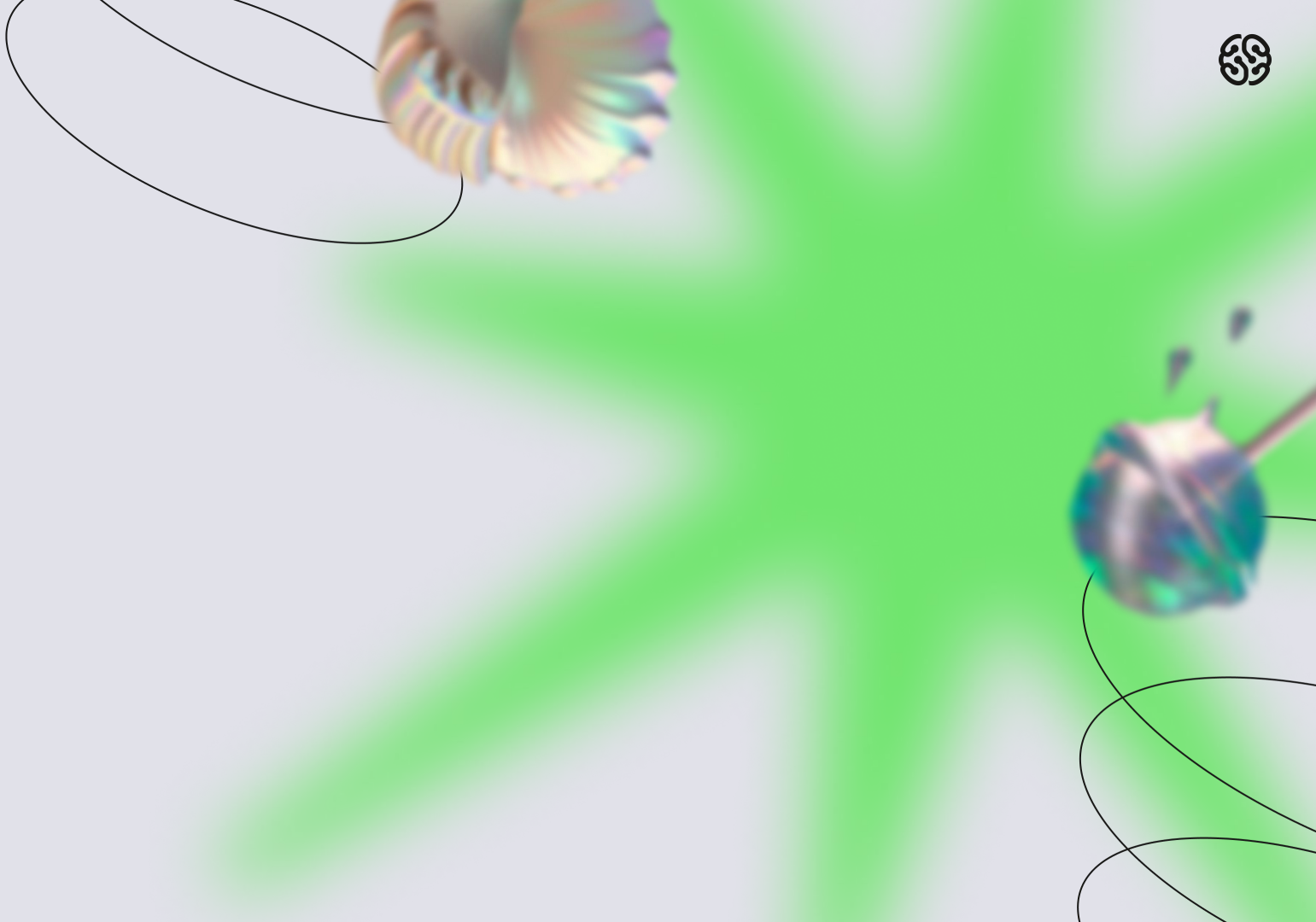


Давайте делать это в коде!





ETL





Что такое ETL

ETL – аббревиатура от Extract, Transform, Load. Это общий термин для процессов, которые происходят, когда данные переносят из нескольких систем в одно хранилище.

ETL – один из центральных процессов в системах хранения данных.

Он включает в себя:

- **извлечение** данных из различных источников
- **трансформация** и очистка данных для приведения их к единообразию или в соответствие с бизнес-задачами.
- **загрузка** в хранилище данных



Основные функции ETL

1

Извлечение

Во время извлечения данных необработанные данные копируются или экспортируются из исходных местоположений в промежуточную область

2

Трансформация

Данные преобразуются и консолидируются для предполагаемого аналитического использования

3

Загрузка

На этом последнем шаге преобразованные данные перемещаются из промежуточной области в целевое хранилище данных



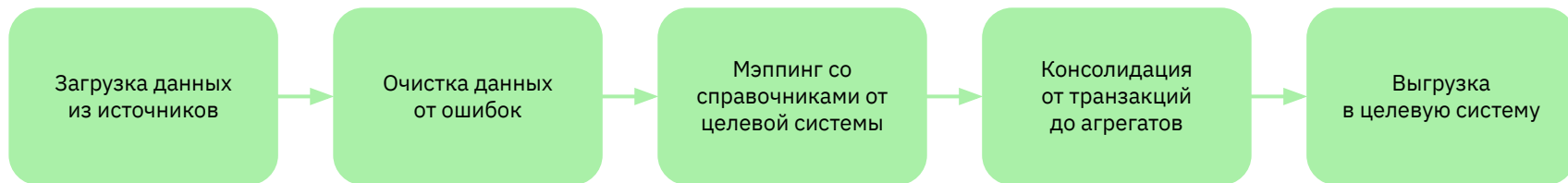
Вопрос

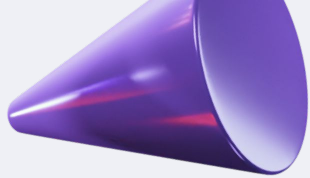
Зачем нужна ETL-система?





Как работает ETL-система





Спасибо за внимание

