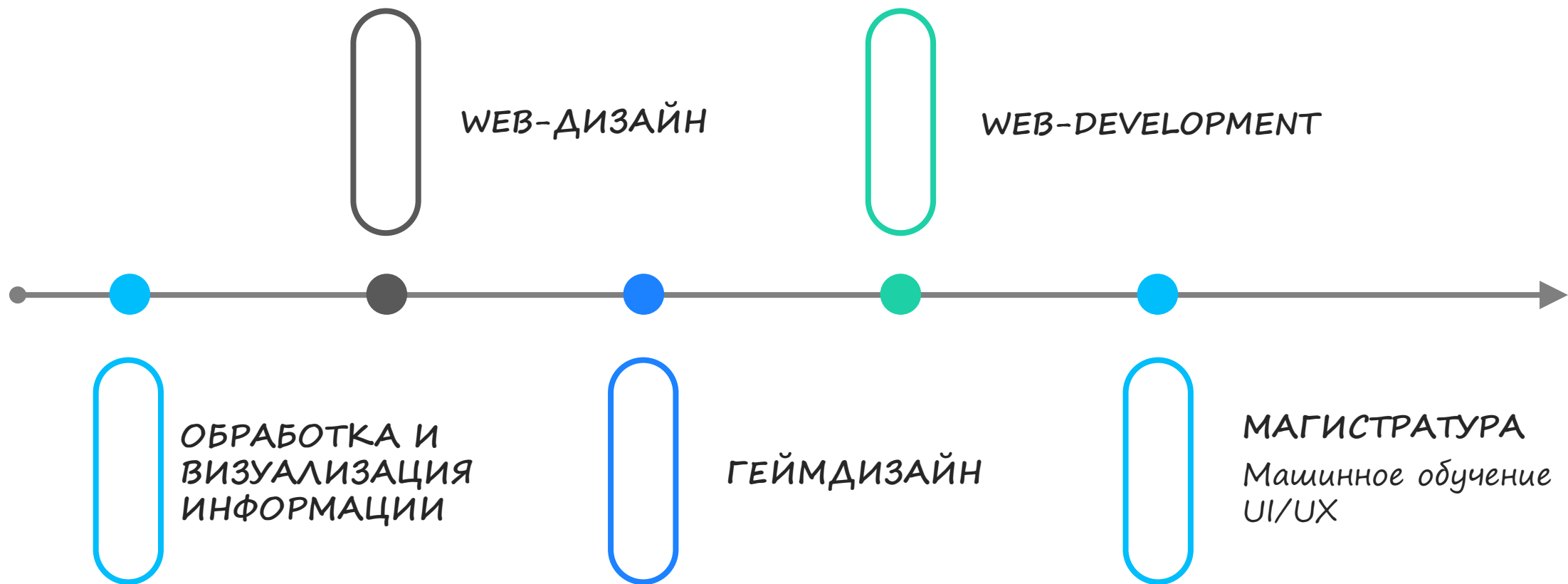


The background is a deep blue with several horizontal, wavy bands of glowing white and light blue particles, creating a sense of motion and data flow.

ОБРАБОТКА И ВИЗУАЛИЗАЦИЯ ИНФОРМАЦИИ

Наше взаимодействие



Что изучается в дисциплине?

0

Основы Python

- Среды разработки
- Строки, числа, вычисления
- Функции, циклы, условная логика

1
0

Очистка и подготовка данных

- Отсутствующие данные
- Преобразование
- Переформатирование

2

0

Специализированные библиотеки

- NumPy
- pandas

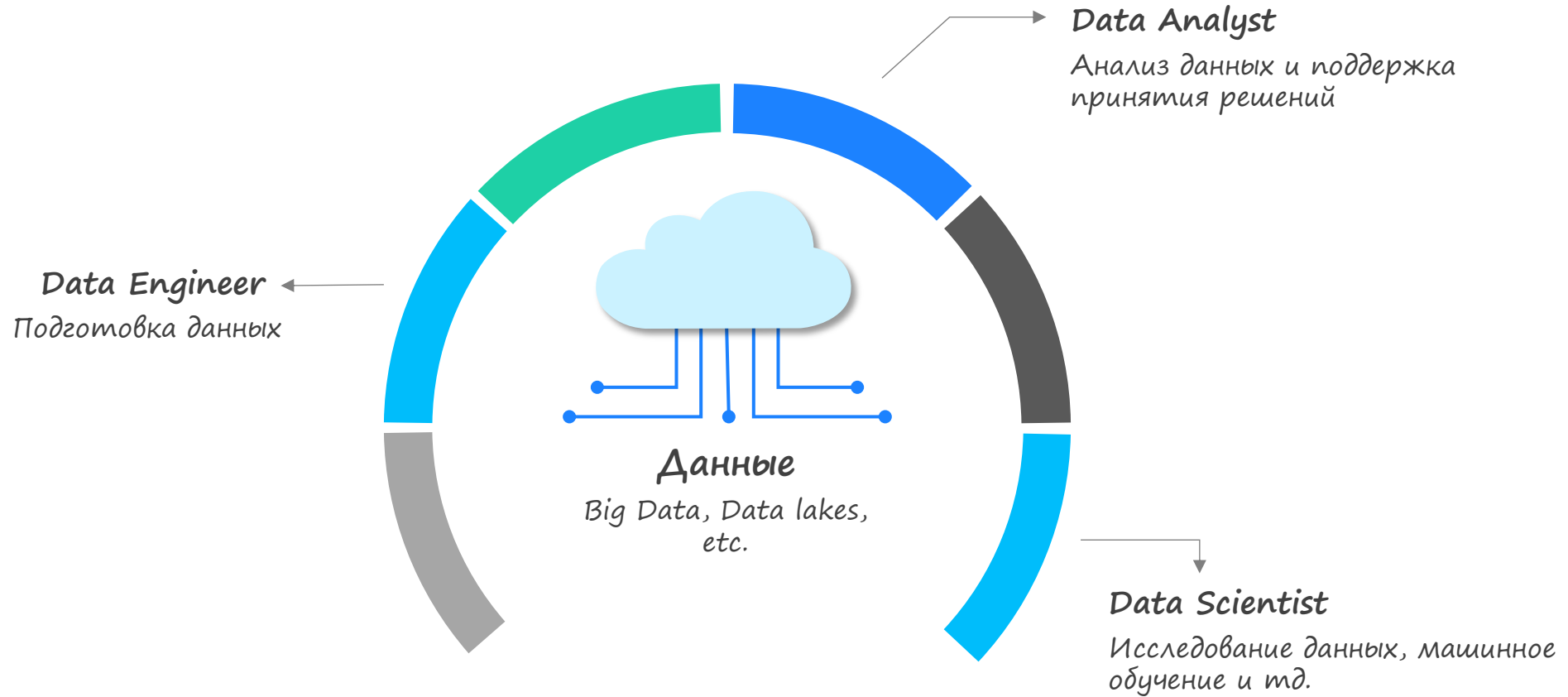
3
0

Графики и визуализация

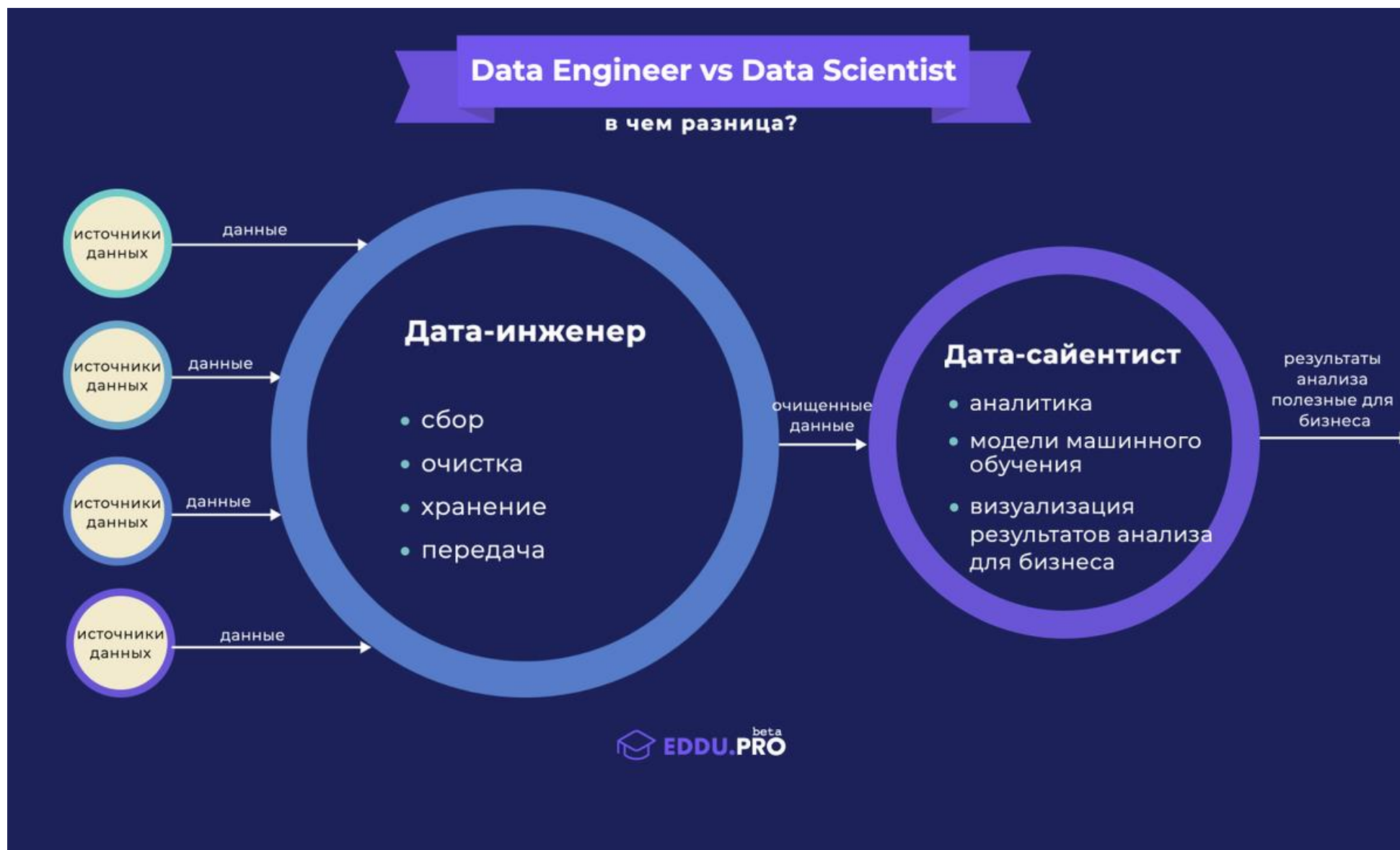
- matplotlib
- seaborn

4

Зачем изучается?



Data Engineer – что выполняет?



Инженерия данных в той или иной форме существует с тех пор, как компании начали проводить работу с данными, — например, предиктивный анализ, описательную аналитику и отчёты — и оказалась в центре внимания вместе с развитием науки о данных в 2010-х годах. Для целей этой книги очень важно определить, что означают термины «инженерия данных» и «инженер данных» (*data engineer*).

“

Инженерия данных — это набор операций, направленных на создание интерфейсов и механизмов для потока и доступа к информации. Требуются преданные своему делу специалисты — инженеры по обработке данных — для поддержания данных в доступном и пригодном для использования другими виде. Если кратко — инженеры по обработке данных настраивают и эксплуатируют инфраструктуру данных организации, подготавливая её для дальнейшего анализа аналитиками данных и учёными.

— из «Data Engineering and Its Main Concepts», AlexSoft

“

По сравнению с ранее существовавшими должностями область инженерии данных можно рассматривать как слияние бизнес-аналитики и работы с хранилищами данных, которое заимствует много функций из разработки программного обеспечения. Эта дисциплина также объединяет специализацию по работе с так называемыми распределёнными системами «больших данных», а также концепции расширенной экосистемы Hadoop, потоковой обработки и масштабных вычислений.

— Maxime Beauchemin

“

Первый тип инженерии данных ориентирован на SQL. Работа и основное хранение данных происходит в реляционных базах данных. Вся обработка данных выполняется с помощью SQL или языка на основе SQL. Иногда эта обработка данных выполняется с помощью инструмента ETL2. Второй тип инженерии данных ориентирован на большие данные. Работа и основное хранение данных осуществляются в технологиях больших данных, таких как Hadoop, Cassandra и HBase. Вся обработка данных выполняется в средах больших данных, таких как MapReduce, Spark и Flink. Хотя используется SQL, основная обработка выполняется с помощью таких языков программирования, как Java, Scala и Python.

— Jesse Anderson

“

Инженерия данных — это перемещение, манипулирование и управление данными.

— Lewis Gavin

Инженерия данных получает данные, сохраняет их и подготавливает их для использования учёными, аналитиками и другими специалистами.

Инженерия данных

разработка, внедрение и обслуживание систем и процессов, которые принимают необработанные данные и производят высококачественную, согласованную информацию, которая поддерживает последующие варианты использования, такие как анализ и машинное обучение

совокупность безопасности, управления данными, DataOps, архитектуры данных, оркестрации и разработки программного обеспечения

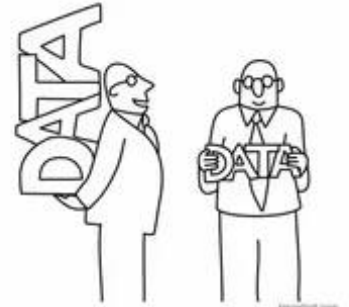
Инженер по обработке данных управляет жизненным циклом инженерии данных, начиная с получения данных из исходных систем и заканчивая предоставлением данных для сценариев использования, таких как анализ или машинное обучение

Data analyst — это специалист, который работает с данными: собирает их, обрабатывает, изучает и интерпретирует.

- ❖ Его выводы помогают принимать решения в бизнесе, в научных исследованиях, в менеджменте и в других областях.
- ❖ Некоторые задачи аналитика данных:
- ❖ Сбор данных. Аналитик аккумулирует информацию из разнообразных источников, таких как базы данных, веб-сайты, CRM-системы и другие платформы.
- ❖ Очистка и подготовка данных. После сбора информации специалист проводит её очистку, устраняя ошибки, дубликаты и пропуски, а также преобразует данные в удобный для анализа формат.
- ❖ Анализ данных. Используя статистические методы и специализированные инструменты, аналитик выявляет закономерности, тренды и аномалии, которые могут быть полезны для бизнеса.
- ❖ Визуализация и представление результатов. Создание наглядных отчётов, графиков и дашбордов позволяет представить результаты анализа в понятной форме для руководства и других заинтересованных сторон.
- ❖ Разработка рекомендаций. На основе полученных инсайтов аналитик формулирует предложения по оптимизации бизнес-процессов, повышению эффективности маркетинговых кампаний или улучшению клиентского опыта.

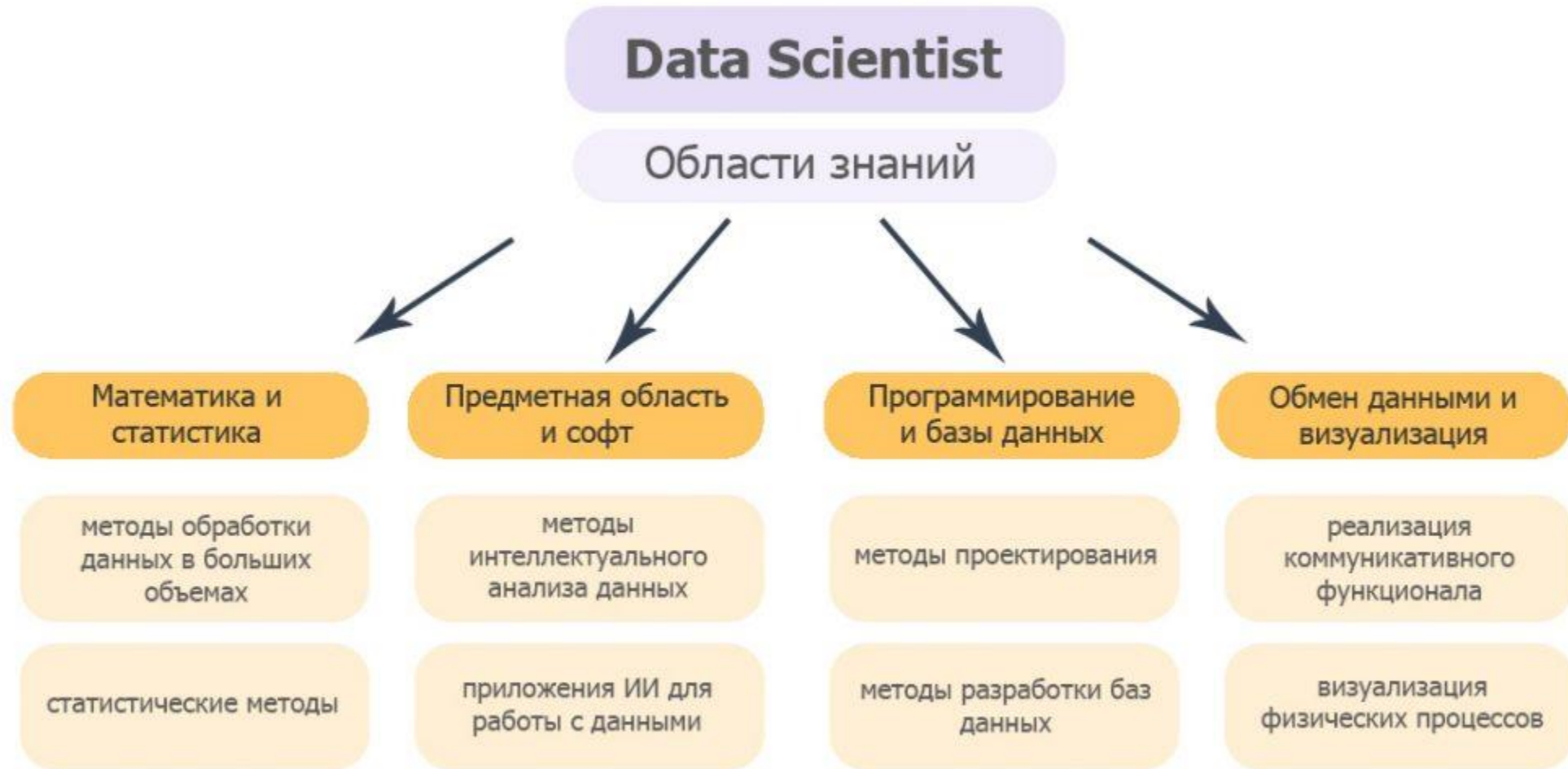
Аналитик – кто это?

Специалист, занимающийся проведением аналитических исследований и обобщений в определенной сфере деятельности. На основе результатов своей работы он способен прогнозировать процессы и разрабатывать перспективные программы развития.



"I think you'll find that mine is bigger..."

Data Scientist— это специалист по работе с данными для решения задач бизнеса. Он работает на стыке программирования, машинного обучения и математики



Что изучается в дисциплине?

0

Основы Python

- Среда разработки
- Строки, числа, вычисления
- Функции, циклы, условная логика

1
0

Очистка и подготовка данных

- Отсутствующие данные
- Преобразование
- Переформатирование

2

0

Специализированные библиотеки

- NumPy
- pandas

3
0

Графики и визуализация

- matplotlib
- seaborn

4

РЕКОМЕНДУЕМ ЫЕ ИСТОЧНИКИ

- ❖ Дэн Бейдер и др. – знакомство с Python
- ❖ Уэс Маккинни – Python и анализ данных
- ❖ "Поколение Python": курс для начинающих на платформе stepik.org

ВСЕ МАТЕРИАЛЫ ВЗЯТЫ ИЗ
ОТКРЫТЫХ ИСТОЧНИКОВ И
ПРЕДСТАВЛЕНЫ ИСКЛЮЧИТЕЛЬНО В
ОЗНАКОМИТЕЛЬНЫХ ЦЕЛЯХ



БЛАГОДАРЮ
ЗА ВНИМАНИЕ