

# **Профессии специалистов по данным**

**Воркшоп по разделам, инструментам и задачам**

Четин Константин, [https://github.com/hakeydotom/data\\_webinar\\_2022](https://github.com/hakeydotom/data_webinar_2022)

# О себе

- Выпускник математического факультета (специалист по прикладной математике)
- Начал карьеру в data-аналитике в 2018 году
- Работал на позициях Data Analyst и Data Scientist в банке, e-commerce-проекте, в рекламных технологиях (MyTarget, холдинг VK)
- С 2021 года руководжу командой из 6 аналитиков в Альфа-банке. Решаем задачи B2B.



# Data Science: какие бывают специалисты по данным?

- Data Analyst
- Data Scientist
- Data Engineer (\*ML-Engineer)



**Statistician**



**Data Scientist**





# Data Analyst - аналитик данных

## Задачи и инструменты

- Сфокусирован больше на описательном анализе данных (изучает факты)
- Погружается в предметную область, понимает как устроен бизнес и на какие показатели ориентируется (бизнес-метрики)
- Общается с бизнес-заказчиками, помогает сформулировать запрос «к данным» (гипотезу)
- Делает выводы по данным, отвечает на запросы бизнеса и презентует результаты (презентации и дашборды)
- Python, SQL, BI-инструменты, PowerPoint и всё такое

# Data Scientist - «ученый по данным»

## Задачи и инструменты

- Использует данные для предсказательного моделирования
- Меньше софт-, больше хардскиловых задач, больше кода на SQL и Python
- Использует Machine Learning - инструментарий и методы продвинутой аналитики
- Часто решает задачи в предметных областях, связанных с обработкой текстов, изображений, неструктурированных данных. Задачи как правило требуют больших hard skills



# DATA LAKE

Data scientist



Data engineer



# Data Engineer - инженер данных

## Задачи и инструменты

- Строит системы хранения и обработки массивов данных
- Отвечает за извлечение, обработку и поставку данных различным потребителям в том числе для нужд DA, DS
- Много кода на SQL, Python, Scala, Java, etc
- Больше похож на классического разработчика
- Ключевые слова: Data Warehouse, ETL

# Открытые данные для экспериментов (datasets)

- <https://github.com/awesomedata/awesome-public-datasets>
- <https://data.gov.ru/>
- <https://console.cloud.google.com/marketplace>
- <https://www.kaggle.com/>





# Что такое Jupyter project?

- Jupyter project - семейство инструментов разработки программного кода в интерактивном режиме. Чаще всего - в веб интерфейсе.
- Наиболее популярный инструмент для манипуляции данными с помощью Python для специалистов в Data Science.

