Data Scientist

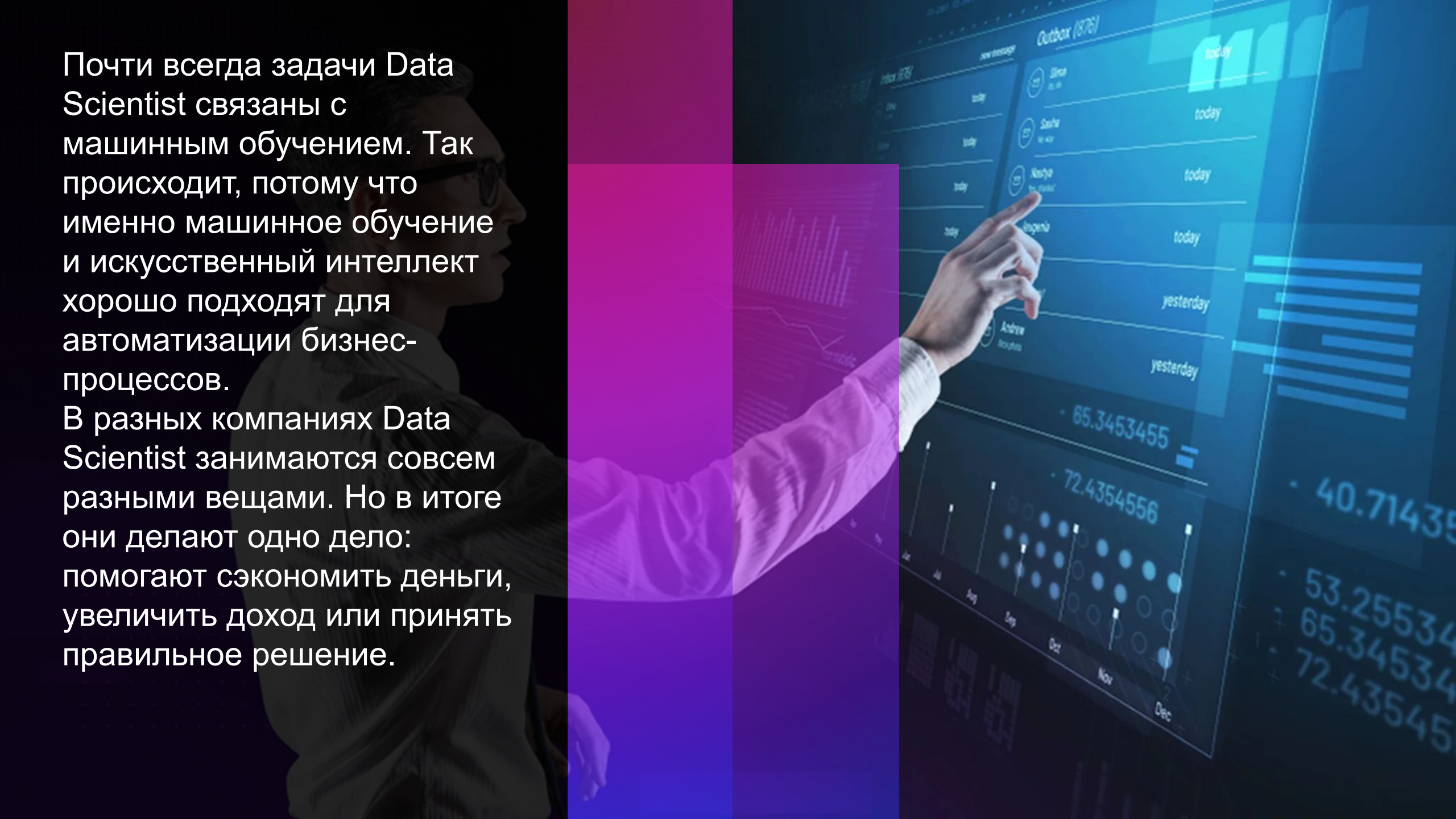
Data Scientist — кто это и что делает?

- Дословно Data Scientist переводится как «ученый данных». Но деятельность у такого специалиста не научная, а практическая: он работает с данными компании, анализирует их, ищет зависимости, делает выводы на их основе и при необходимости строит визуализации. Для этого Data Scientist использует разные математические алгоритмы, специальное программное обеспечение и инструменты разработки.
- Данные, с которыми работает Data Scientist, могут быть любыми: звук, текст, фото, видео, таблицы, документы. Если у вас есть любые данные и нужно их проанализировать — это работа для Data Scientist.

Чем занимается?

- Часто в бизнесе есть задачи, которые решаются вручную. Например, менеджер делает простенькие расчеты в Excel или руководитель магазина по своему опыту предсказывает спрос на товары. Такие ручные решения занимают много времени и часто необъективны.
- Data Scientist автоматизирует принятие таких решений и делает их более точными, основанными на данных. Он разбирается в задаче, смотрит, какие данные нужны для ее решения. Потом разрабатывает программу, которая будет автоматически считать и анализировать данные. Такая программа может либо принимать простые решения самостоятельно, либо давать более точную и полезную информацию менеджерам.

Почти всегда задачи Data Scientist связаны с машинным обучением. Так происходит, потому что именно машинное обучение и искусственный интеллект хорошо подходят для автоматизации бизнес-процессов.
В разных компаниях Data Scientist занимаются совсем разными вещами. Но в итоге они делают одно дело: помогают сэкономить деньги, увеличить доход или принять правильное решение.

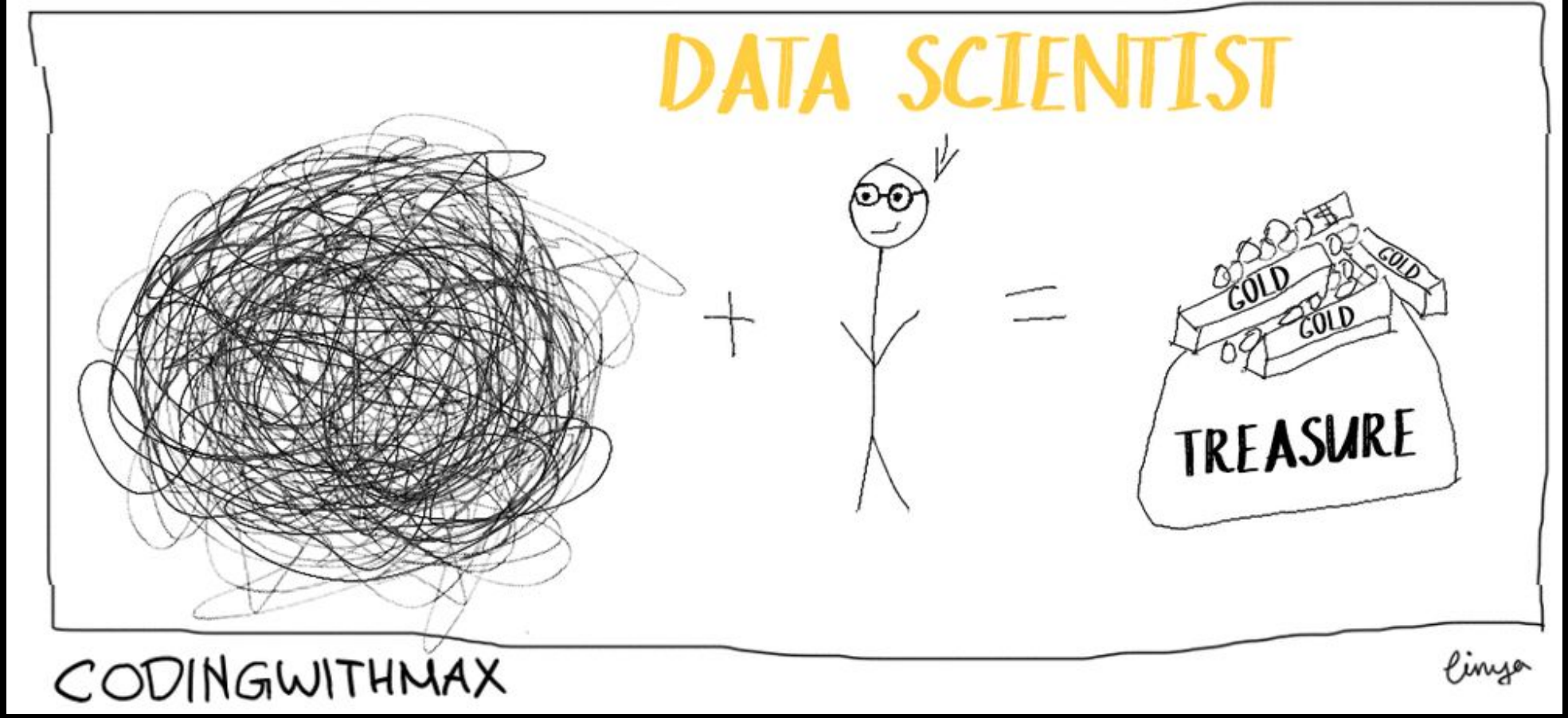


Что нужно Data Scientist для работы?

- Главное, с чем работает Data Scientist — это данные. Компания должна уже собирать, обрабатывать и хранить данные, построить для этого соответствующую инфраструктуру.
- Data Scientist обрабатывает данные, для чего часто требуются большие вычислительные мощности и специальные инструменты.
- Еще для работы Data Scientist нуждается в команде помощников. Чаще всего он работает в связке с Data Engineer и командой разработчиков. Первые обеспечивают его данными, вторые превращают разработанные модели в конкретные программы и сервисы, которыми могут пользоваться другие люди.

Что нужно знать Data Scientist?

- Data Science — дисциплина, которая лежит на стыке математики, статистики и компьютерных наук. Поэтому обычно Data Scientist должен:
- Знать математику и статистику
- Писать код, обычно на языках R и Python
- Работать с базами данных и знать язык SQL
- Владеть инструментами для работы с большими данными
- Отлаживать код и готовить к выкатке в продакшен
- Работать с программами для визуализации и презентации результатов работы: PowerPoint, Shiny/Dash, Power BI, Qlik
- Знать технологии машинного и глубокого обучения



Спасибо за внимание