

# **Разработка системы анализа медицинских изображений для эпидемиологического мониторинга COVID-19**

Яровикова Анастасия  
студентка группы M24-525

# Архитектура решения

- Предобработка данных (заполнены пропуски, удалены дубликаты, выявлены и обработаны аномальные значения)
- SQL-аналитика
- фильтрация и категоризация данных, обработка в PySpark, реализация UDF
- Визуализация

Используемые технологии: PySpark 3.5.1, Python 3.12.12, Pandas, Matplotlib

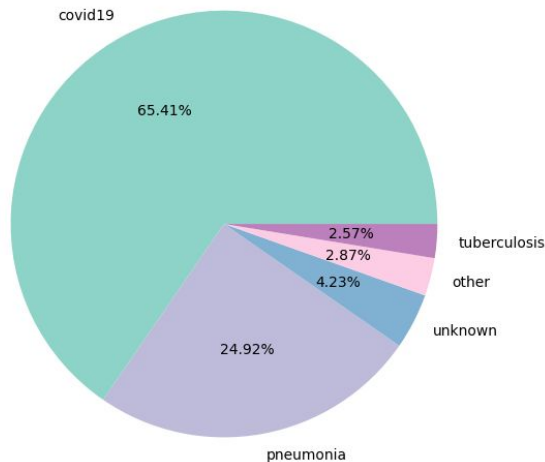
# Ключевые статистики

- 433 записи covid19 в наборе из 950 строк
- средний возраст заболевших covid19 - 55 лет, минимальный - 20, максимальный - 94
- Чаще обнаружен у мужчин: Male - 284, Female - 149
- Наибольшее количество случаев заболеваемости в 2020 году (~72% выборки)
- Основные проекции: AP, PA
- Категория covid19 после унификации занимает ~65%

# Визуализации и выводы

Проект показал, что среди пациентов с COVID-19 преобладают взрослые люди 20-60 лет, большая часть - лица старше 50 лет. Анализ временных трендов выявил колебания количества исследований в годы обострения эпидемии - 2019 и 2020.

Распределение диагнозов



Распределение по возрастным группам

