

COVID19



Разработка системы анализа медицинских
изображений для эпидемиологического
мониторинга COVID 19

Итоговый проект - Инфраструктура Больших
Данных / Попова Виктория



АНАЛИТИЧЕСКАЯ СИСТЕМА МОНИТОРИНГА COVID-19 НА ОСНОВЕ РЕНТГЕНОВСКИХ ДАННЫХ

Источник данных: *COVID-19 CHEST X-RAY DATASET (METADATA.CSV)*

ПРЕДОБРАБОТКА ДАННЫХ В **PYSPARK**:

- ОЧИСТКА И ЗАПОЛНЕНИЕ ПРОПУСКОВ
- УНИФИКАЦИЯ ДИАГНОЗОВ И ПРОЕКЦИЙ
- ОБРАБОТКА ДАТ И ДУБЛИКАТОВ

АНАЛИТИКА В **SPARK SQL**:

- АГРЕГАТЫ
- ОКОННЫЕ ФУНКЦИИ
- ВРЕМЕННЫЕ ТРЕНДЫ

ХРАНЕНИЕ РЕЗУЛЬТАТОВ В ФОРМАТЕ **PARQUET**

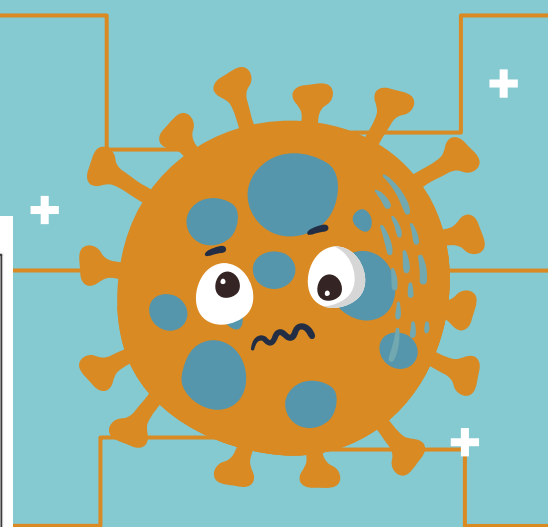
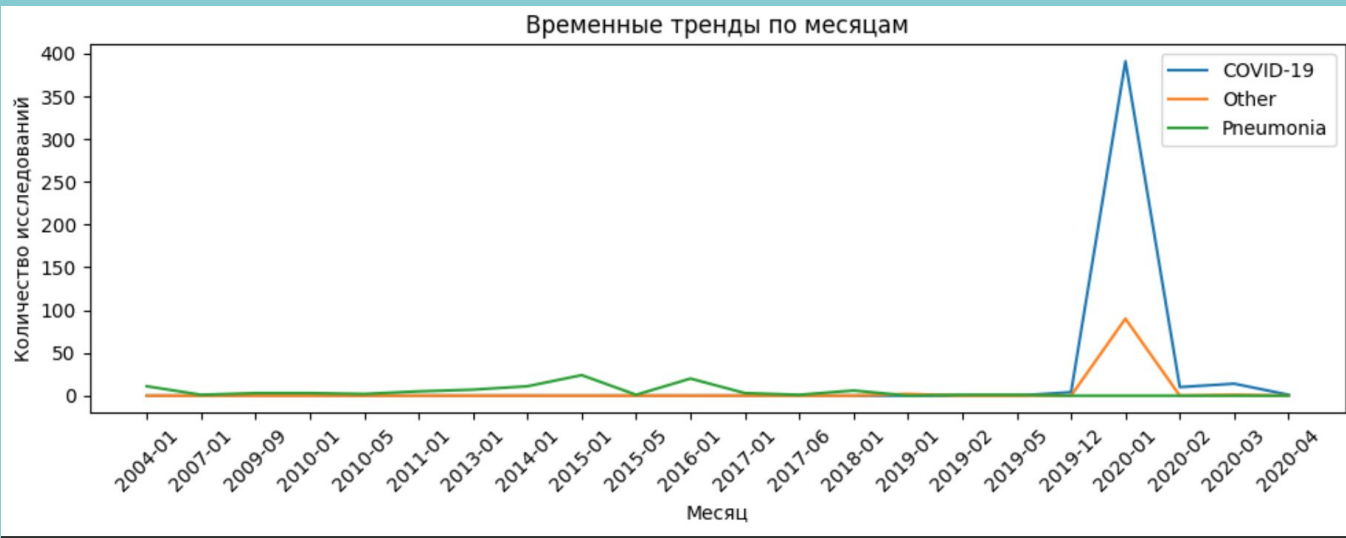
ВИЗУАЛИЗАЦИЯ С ИСПОЛЬЗОВАНИЕМ **PANDAS + MATPLOTLIB**



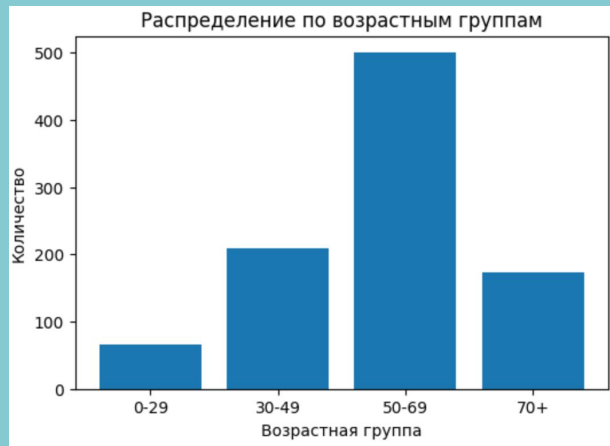
ОСНОВНЫЕ ХАРАКТЕРИСТИКИ ДАТАСЕТА

COVID-19 является доминирующим диагнозом, преимущественно у мужчин старше 50 лет; чаще всего используются проекции RA и AP.

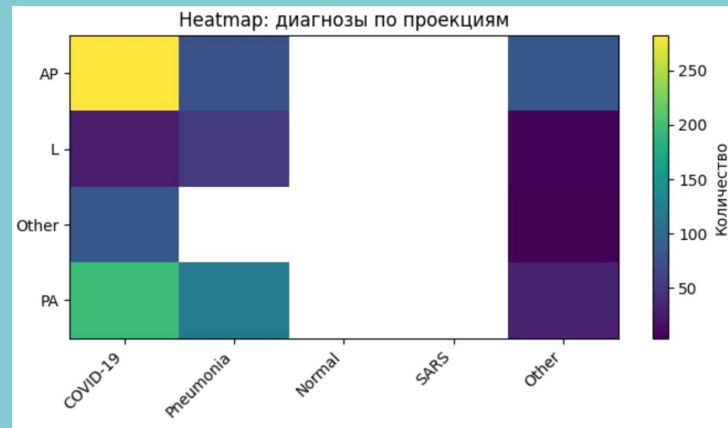
Анализ временных трендов показал чёткую связь между ростом количества исследований и началом пандемии COVID-19, подтверждая применимость Spark-аналитики для задач эпидемиологического мониторинга.



RESEARCH AND PUBLICATIONS



Анализ взаимосвязи диагнозов и рентгеновских проекций выявил, что для пациентов с COVID-19 наиболее часто используются проекции PA и AP. Это согласуется с клинической практикой и подтверждает репрезентативность используемого датасета.



Анализ распределения пациентов по возрастным группам показал, что наибольшее количество случаев COVID-19 приходится на возрастную группу 50–69 лет. Это подтверждает повышенную уязвимость старших возрастных групп и подчёркивает важность их приоритетного эпидемиологического мониторинга.