



Выводы что было сделано



Самостоятельная настройка и запуск Apache Hadoop и Apache Spark на базе WSL. Загрузка исходных методанных в HDFS, их комплексная очистка, унификация и обогащение с использованием PySpark. Применение партиционирования и бакетирования для оптимизации хранения и доступа.

Выполнение разнообразных и сложных SQL-запросов, включая оконные функции, самосоединения и аналитические подзапросы, для выявления закономерностей в данных и идентификации нестандартных случаев.

Подготовка данных и построение моделей классификации (Random Forest Classifier) для предсказания COVID-19, с высокой точностью и анализом важности признаков.

Создание информативных графиков, отображающих распределение диагнозов, демографические характеристики пациентов и временную динамику.