

АНАЛИЗ И ВИЗУАЛИЗАЦИЯ ДАННЫХ О COVID-19 В POWER BI

Ошмян А. В.

*Белорусский государственный университет, Минск, Беларусь,
e-mail: osmananna683@gmail.com*

Сложно представить современный мир без веб-технологий. Социальные сети, почта, поисковые системы, навигационные сервисы – все это используется большинством из нас ежедневно. Практически все интернет-ресурсы работают с базами данных. На основе баз данных с использованием языка SQL построен почти каждый сайт, социальная сеть и поисковая система. Но информация в базах данных не выглядит наглядно, сложно рассмотреть тенденцию ее изменения.

Поэтому для более наглядного представления материала по выбранной теме можно использовать комплексное программное обеспечение бизнес-анализа компании Microsoft, Power BI. Оно позволяет наглядно и интерактивно отслеживать основные тенденции, зависимости и отклонения показателей. Power BI включает множество средств визуализации: графики, линейчатые диаграммы, круговые диаграммы, точечные и пузырьковые диаграммы и многое другое. Можно выбрать тот способ, который поможет наиболее наглядно отобразить информацию.

Данная работа посвящена обработке, анализу и визуализации данных о Covid-19 в Power BI [1]. Данные были взяты из открытого источника, а именно из хранилища данных COVID-19 Центра системных наук и инженерии (CSSE) Университета Джонса Хопкинса на GitHub [2]. Этот онлайн-сервис позволяет отслеживать распространение коронавируса по всей планете. Сервис обновляется ежедневно и отображает актуальные данные, собранные Всемирной организацией здравоохранения. Данные хранятся в виде CSV-файлов. Каждый файл представляет данных за один день по всем странам мира и содержит информацию по каждой стране об общем числе подтвержденных случаев заболевания COVID 19, числе умерших от COVID 19 и выздоровевших.

Дополнительную сложность анализу этих данных придавал тот факт, что в различные периоды времени хранились данные с различающейся степенью детализации. А значит в таблицах получалось различное количество столбцов и однотипные столбцы имели разные названия в разные периоды времени. Поэтому потребовалась очистка данных и приведение их к единообразному виду. После того как очищенные данные были извлечены с GitHub, был проведен их анализ с использованием языка DAX.

Выборка делалась для данных по Беларуси, странам с которыми граничит Беларусь и странами с численностью населения сопоставимой с численностью населения Беларуси за период с начала 2020 года по 15 апреля 2021 года. Анализ данных помогает оценить масштабы заболевания и сравнить эффективность борьбы с ним в различных странах.

На рисунках 1–4 активные случаи – это общее количество случаев за вычетом общего количества выздоровевших и общего количества смертей.

На рисунках 1, 2 представлено число новых и активных случаев заболевания COVID 19 соответственно по Беларуси и по России.

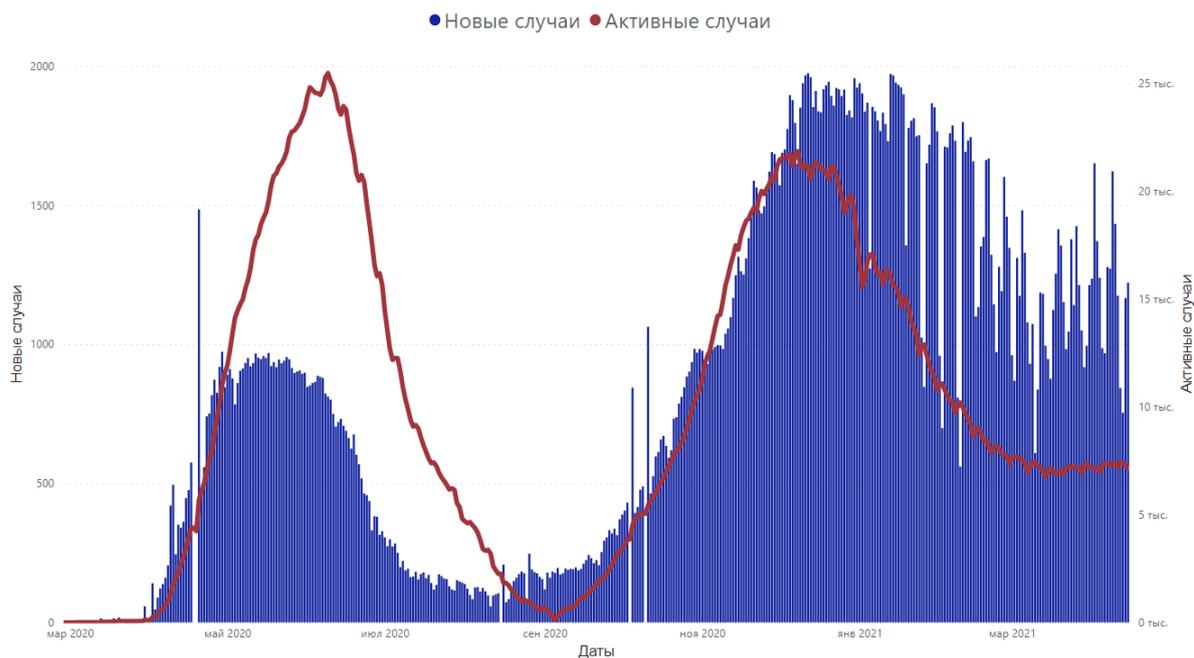


Рис. 1 Число новых и активных случаев заболевания COVID 19 по Беларуси

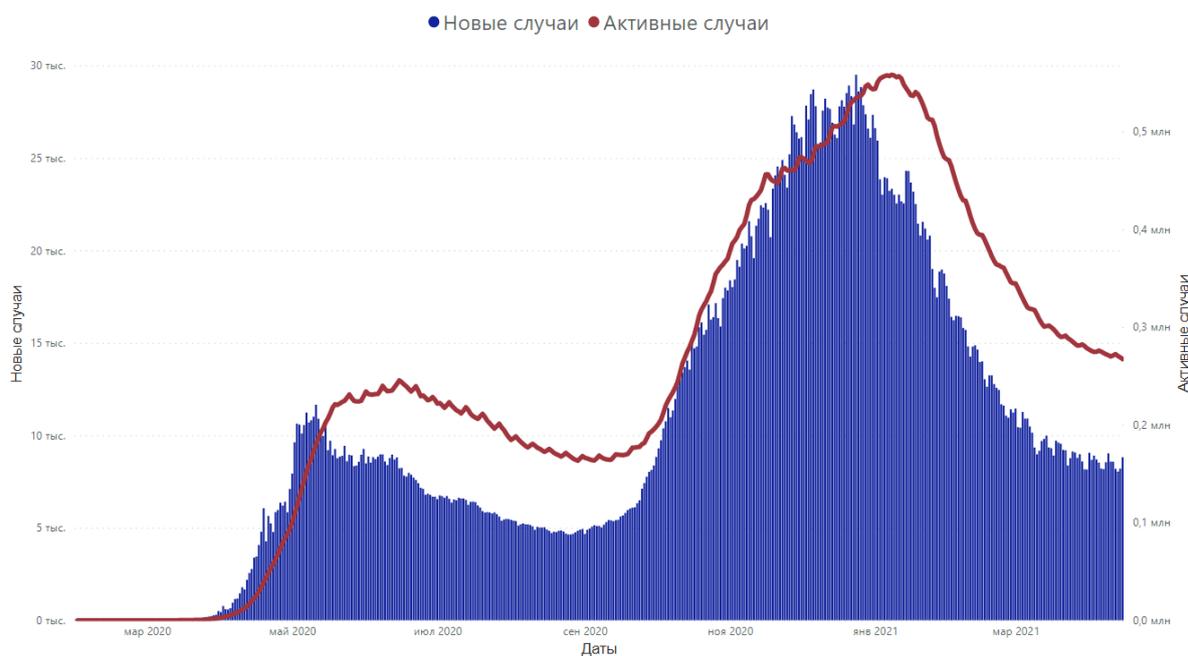


Рис. 2 Число новых и активных случаев заболевания COVID 19 по России

На рисунках 3, 4 дни лечения – это наименьшее количество дней, предшествующих рассматриваемому дню, сумма всех новых случаев за которые больше или равна числу активных случаев на рассматриваемую дату.

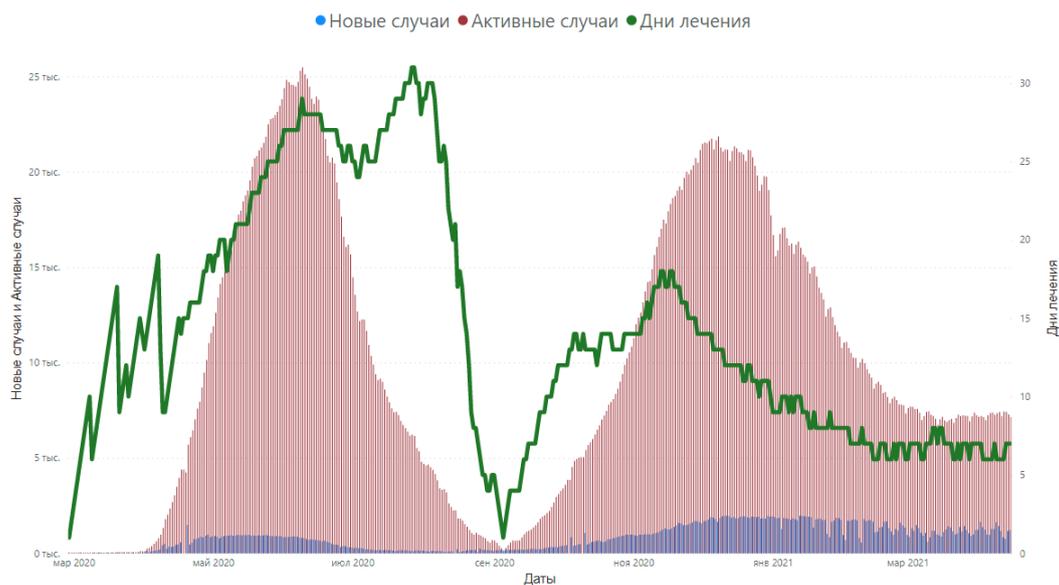


Рис. 3 Число новых и активных случаев заболевания COVID 19 и средней продолжительности лечения по Беларуси

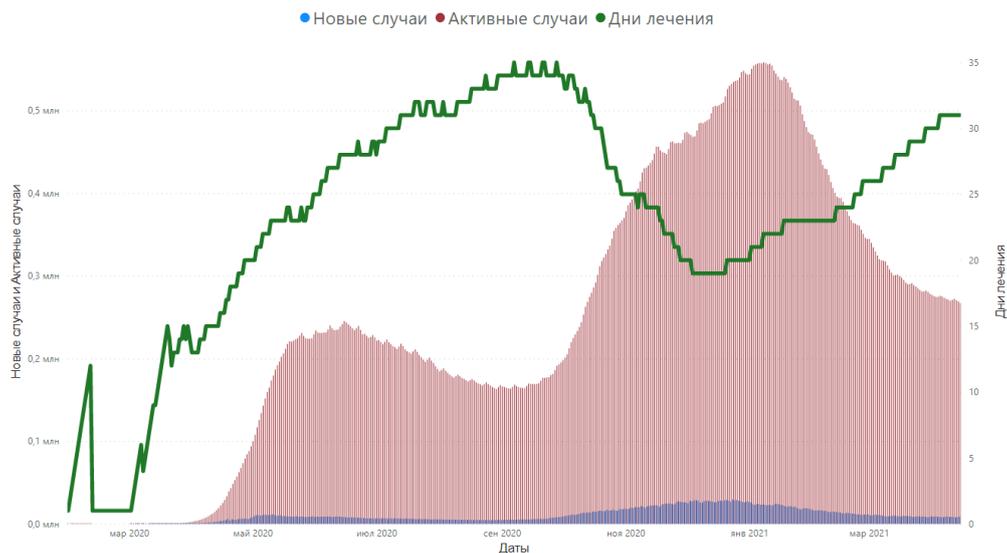


Рис. 3 Число новых и активных случаев заболевания COVID 19 и средней продолжительности лечения по России

В данной работе была показана эффективность использования различных средств визуализации для анализа и отображения информации на примере данных о Covid-19. Последний год пандемия волнует весь мир, она нанесла огромный ущерб всему миру и привела к серьезным глобальным социальным и экономическим проблемам, а также к проблемам в образовании.

Литература

1. Get started with Power BI Desktop [Электронный ресурс] – Режим доступа: <https://docs.microsoft.com/en-us/power-bi/fundamentals/desktop-getting-started>
2. CSSEGISandData Covid-19 [Электронный ресурс] – Режим доступа: https://github.com/CSSEGISandData/COVID-19/tree/master/csse_covid_19_data/csse_covid_19_daily_reports