УДК 004.77:311.2
DOI: 10.31866/2617-796X.4.1.2021.236947

Булига Костянтин,
кандидат технічних наук, доцент кафедри комп'ютерних наук,
Київський національний університет культури і мистецтв,
Київ, Україна
piton54@ukr.net
https://orcid.org/0000-0002-9537-3226

Толмач Марина,
викладач, заступник декана факультету дистанційного навчання,
Київський національний університет культури і мистецтв,
Київ, Україна
margo@knukim.edu.ua
https://orcid.org/0000-0002-7020-1348

ВІЗУАЛІЗАЦІЯ СТАТИСТИЧНИХ ДАНИХ НА ПЛАТФОРМІ POWER BI

Метою статті є використання новітніх інформаційних технологій для наочного відображення впливу різноманітних чинників на поширення пандемії COVID-19. Докладно розглянуто створення дашборду (PoltP, 2017) для візуалізації статистичних даних епідемії COVID-19 навесні 2021 року.

Метод дослідження – системний аналіз статистичних даних.

Новизною проведеного дослідження є об'єднання різнобічної статистичної інформації для попереднього якісного аналізу складної епідеміологічної ситуації.

Висновки. Викладений у статті матеріал дає наочний приклад використання різноманітних даних для візуалізації епідеміологічного стану. Показано можливість завантаження статистичних даних будь-якого формату в спільне джерело для аналізу впливу на загальний рівень поширення захворювання.

Ключові слова: пандемія COVID-19; MS Power BI; Power Query; Power Pivot; MS Excel; візуалізація; дашборд; якісний аналіз.

Вступ. Захворювання на COVID-19, яке на початку 2020 року сприймали як рідкісну екзотичну хворобу, перетворилося на пандемію. Характерною особливістю її розвитку є непередбачуваність подальшого перебігу (https://www.dw.com/uk/use-pro-koronavirus/a-529937882-7). Для оцінки перспективи росту кількості випадків захворювання використовують різноманітні методи статистичного аналізу. Наприклад, у роботі «Прогноз розвитку епідемії COVID-19 в Україні на 23 листопада – 7 грудня 2020 року ("Прогноз РГ-29")» було використано найсучаснішу модель Facebook Prophet, яка демонструє високу ефективність для моделювання часових рядів, що містять аномальні дати, різні види сезонності та лінійну чи нелінійну динаміку впливу різних складових моделі, але аналіз наведених резуль-

© Булига К. Б.
© Толмач М. С.

татів прогнозування показав, що жоден статистичний прогноз не дає хоча б приблизного кількісного результату (К. Булига та О. Булига, 2020).

У такій ситуації доцільно на початковому етапі аналізу використовувати візуалізацію будь-якої інформації, дотичної до загальної статистики. Це дає змогу якимось чином знайти складні причинно-наслідкові зв'язки.

Результати дослідження. У роботі створено дашборд для дослідження впливу різноманітних статистичних чинників, які безпосередньо або опосередковано стосуються пандемії COVID-19, на основі інформації з порталів відкритих даних. Об'єднання даних з різних інформаційних джерел і створення візуалізації виконано на платформі Microsoft Power BI (https://powerbi.microsoft.com/ru-ru/).

Для створення використано:

CSV-файл з порталу відкритих даних (https://epistat.wiv-isp.be/covid/), який містить статистичні дані про кількість випадків захворювання на COVID-19 за регіонами, віковими групами та статтю у Бельгії за період з 16.03.2020 по 16.04.2021, статистичні дані сайту (https://aqicn.org/map/belgium/ru/) з якості повітря в Бельгії.

Примітка: Бельгію обрали тому, що саме по цій країні на порталі відкритих даних було знайдено найбільше статистичної інформації, яку можна використати для ілюстрації етапів підготовки даних для візуалізації.

Загальна технологія візуалізації складається з таких етапів:

1. У середовищі MS Excel за допомогою надбудов MS Power Query та MS Power Pivot створюється допоміжний запит, який стане джерелом для створення візуалізації в MS Power BI DeskTop. Для цього із сайту (https://epistat.wiv-isp.be/covid/) скопійовано посилання на CSV-файл і за допомогою команд *Дані => Створити запит => З інших джерел => З інтернету* створено запит *Випадки* (рис. 1).

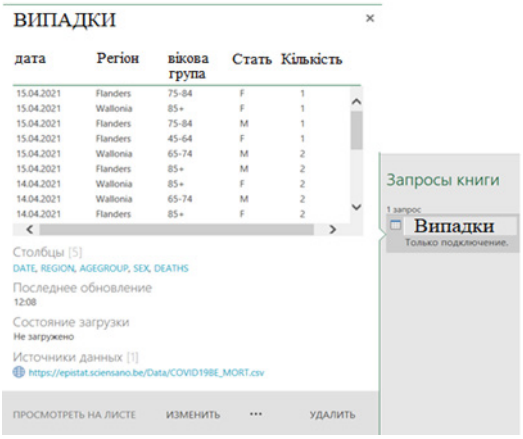

Рис. 1. Допоміжний запит «Випадки»

2. За допомогою імпорту таблиці MS Excel із сайту (https://aqicn.org/map/belgium/ru) і подальшого перетворення її на «розумну» створюється допоміжний запит *Чистота повітря* (рис. 2).

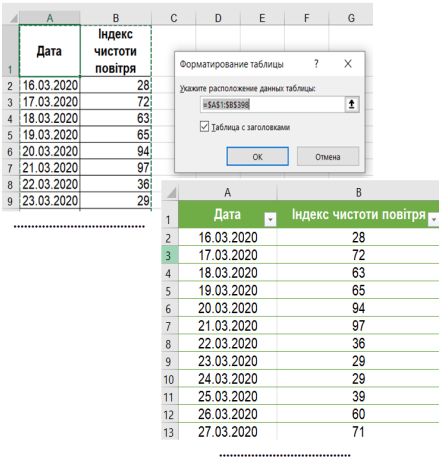

Рис. 2. Допоміжний запит «Чистота повітря»

3. Тепер маємо два запити: Випадки і Чистота повітря. На основі запиту Випадки створюємо такі запити, як *Випадки_Групування_Дати, Випадки_Групування_Вік*, *Випадки_Групування_Стать, Випадки_Групування_Регіон*, з метою групування даних таблиці Випадки за означеними параметрами. Для цього виконуємо команду *Перетворення => Групувати дані* (рис. 3, 4).

4. Далі об'єднуємо дані з двох запитів: *Випадки_Групування_Дати* і Чистота повітря. В результаті зявляється запит Злиття_Дати_Чистота. Виконується команда Дані => Створити запит => Об'єднати => Об'єднання (рис. 5).

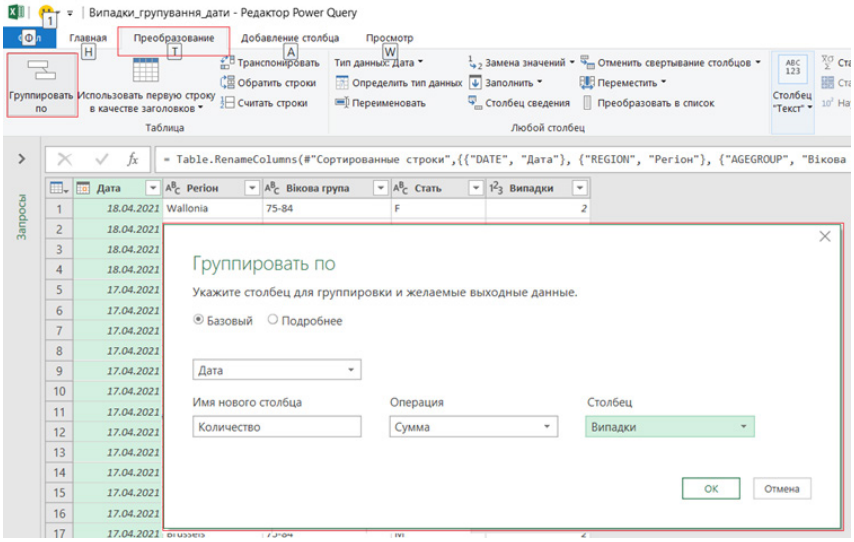

Рис. 3. Групування в запитах

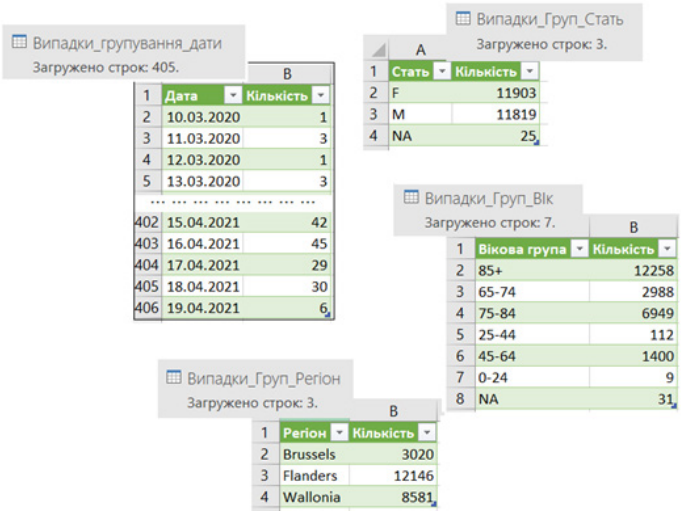

Рис. 4. Запити, що створені групуванням

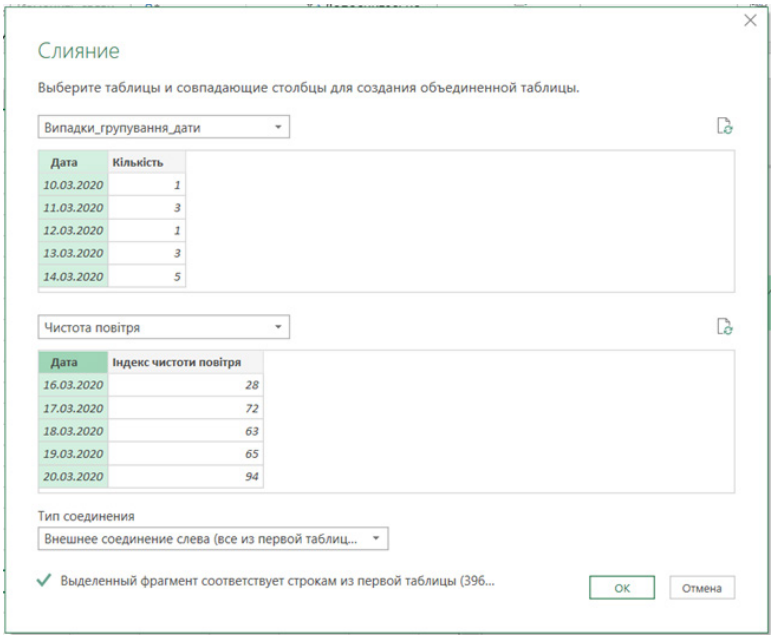

Рис. 5. Злиття запитів

Обов'язково потрібно простежити, щоб усі запити були додані до моделі даних (рис. 6).

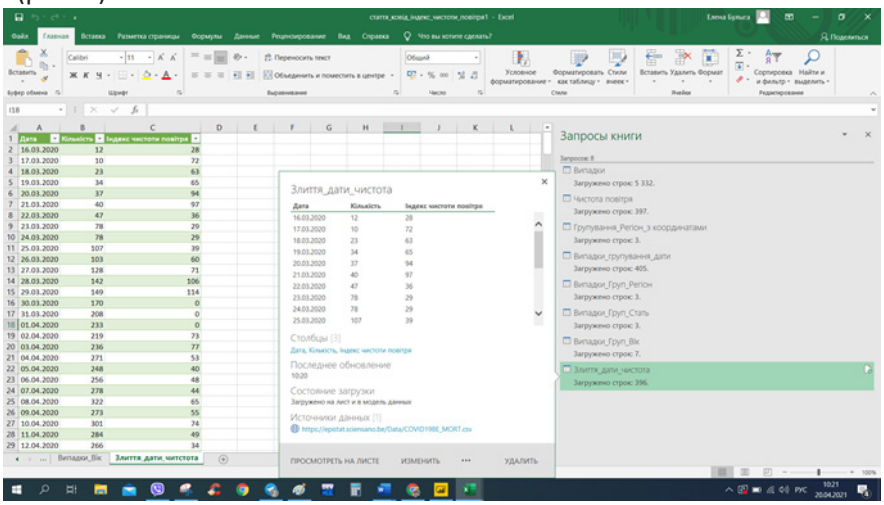

Рис. 6. Усі запити були додані до моделі даних

5. Якщо це так, то наступним кроком буде створення схеми даних у середовищі MS Power Pivot. Виконується така команда: *Вкладка Power Pivot => Управління => Подання діаграми*.

6. На основі цих запитів створюється схема даних (рис. 7).

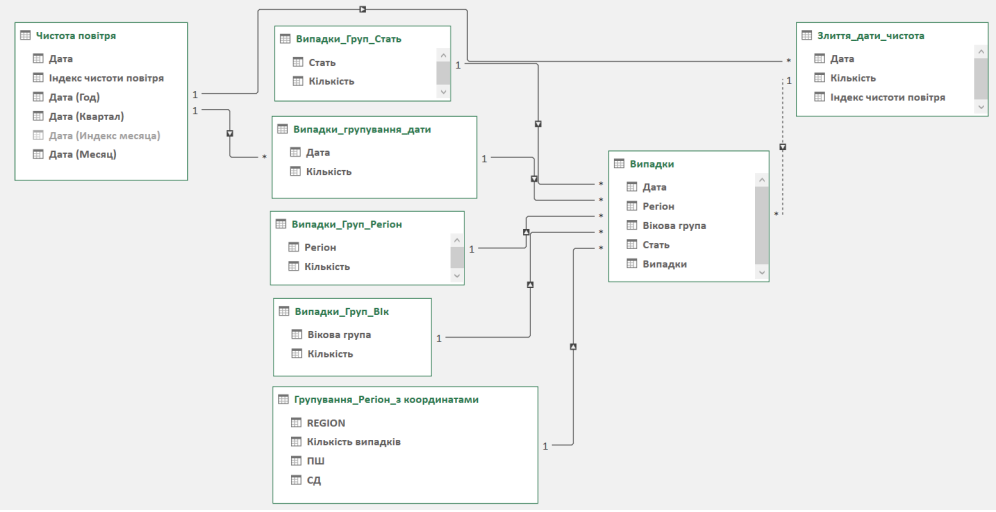

Рис. 7. Схема даних

7. На останньому етапі створюється візуалізація в середовищі MS Power BI DeskTop. У результаті маємо інтерактивний дашборд, що наведено на рис. 8.

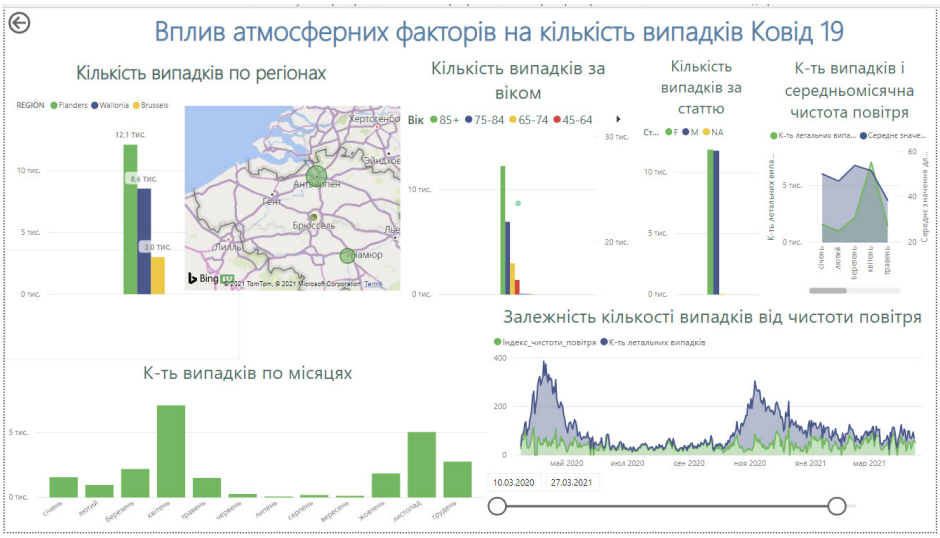

Рис. 8. Кінцевий вигляд дашборду

Інтерактивність дашборду можна побачити в онлайн-режимі, якщо перейти за посиланням (https://app.powerbi.com/view?r) на сайт Power BI. Наприклад, фрагмент дашборду на рис. 9 показує кореляцію кількості летальних випадків і чистоти повітря в листопаді 2020 року.

Рис. 9. Фрагмент інтерактивного дашборду

Висновки. Створення дашборду на платформі Power BI дає змогу в інтерактивному режимі візуалізувати вплив різноманітних чинників для якісного аналізу статистичної ситуації з подальшим виокремленням важливих умов впливу.

СПИСОК ПОСИЛАНЬ

Аналіз поточної ситуації та моделювання сценаріїв поширення захворювання COVID-19, 2020. *Міністерство охорони здоров'я України.* [online] 18 червня 2020. Доступно: <https://moz.gov.ua/article/news/analiz-potochnoi-situacii-ta-modeljuvannja-scenariiv-poshirennja-zahvorjuvannja-covid-19> [Дата звернення 21 квітня 2021].

Булига, К.Б. та Булига, О.А., 2020. Використання інформаційних технологій для аналізу епідеміологічного стану. *Цифрова платформа: інформаційні технології в соціокультурній сфері*, 3(2), с.161-169.

Вплив атмосферних факторів на кількість випадків Ковід 19. [online] Доступно: <https://app.powerbi.com/view?r=eyJrIjoiYWYyZDEyMGEtY2UzZi00MGVhLWExODQtNjE2YmQ4ODUyZmNlIiwidCI6Ijc2OGJjNWJlLTk0Y2UtNDkxMS1hYjAxLWUyMzMwNGQ1MjdlZiIsImMiOjl9> [Дата звернення 21 квітня 2021].

Епідемічна ситуація щодо covid-19 у регіонах України, 2020. *Центр громадського здоров'я Міністерства охорони здоров'я України*. [online] 09 листопада 2020. Доступно: <https://www.phc.org.ua/news/epidemichna-situaciya-schodo-covid-19-u-regionakh-ukraini> [Дата звернення 21 квітня 2021].

Загрязнение воздуха в Бельгии: качество воздуха на карте в режиме реального времени. *The World Air Quality Project*. [online] Доступно: <https://aqicn.org/map/belgium/ru/> [Дата обращения 21 апреля 2021].

Прогноз розвитку епідемії COVID-19 в Україні на 23 листопада – 7 грудня 2020 року («Прогноз РГ-29»), 2020. *Національна академія наук України*. [online] 23 листопада 2020. Доступно: <http://www.nas.gov.ua/UA/Messages/Pages/View.aspx?MessageID=7187> [Дата звернення 21 квітня 2021].

Система моніторингу поширення епідемії коронавірусу. *Апарат РНБО України.* [online] Доступно: <https://covid19.rnbo.gov.ua/> [Дата звернення 21 квітня 2021].

Усе про коронавірус (досьє). *Deutsche Welle*. [online] Доступно: <https://www.dw.com/uk/use-pro-koronavirus/a-52993788> [Дата обращения 21 апрель 2021].

COVID-19. *Sciensano*. [online] Available at: <https://epistat.wiv-isp.be/covid/ [Accessed 21 April 2021].

PoltP, 2017. Дашборд – что это и почему он будет вам полезен или современный способ сделать тайное явным. *Habr*. [online] 27 ноября 2017. Доступно: <https://habr.com/ru/company/devexpress/blog/341972> [Дата обращения 17 марта 2021].

Power BI. [online] Available at: <https://powerbi.microsoft.com/ru-ru/[Accessed 17 March 2021].

REFERENCES

Analiz potochnoi sytuatsii ta modeliuvannia stsenariiv poshyrennia zakhvoriuvannia COVID-19 [Analysis of the current situation and modeling of scenarios for the spread of the disease COVID-19], 2020. *Ministry of Health of Ukraine*. [online] 18 June 2020. Available at: <https://moz.gov.ua/article/news/analiz-potochnoi-situacii-ta-modeljuvannja-scenariiv-poshirennja-zahvorjuvannja-covid-19> [Accessed 21 April 2021].

Bulyha, K.B. and Bulyha, O.A, 2020. Vykorystannia informatsiinykh tekhnolohii dlia analizu epidemiolohichnoho stanu [The use of information technology to analyze the epidemiological situation]. *Digital Platform: Information Technologies in Sociocultural Sphere*, 3(2), pp.161-169.
COVID-19. *Sciensano*. [online] Available at: <https://epistat.wiv-isp.be/covid/ [Accessed 21 April 2021].
Epidemichna sytuatsiia shchodo covid-19 u rehionakh Ukrainy. Tsentr hromadskoho zdorovia [Epidemic situation regarding covid-19 in the regions of Ukraine], 2020. *Ministry of Health of Ukraine*. [online] 09 November 2020. Available at: <https://www.phc.org.ua/news/epidemichna-situaciya-schodo-covid-19-u-regionakh-ukraini> [Accessed 21 April 2021].
PoltP, 2017. Dashbord – chto eto i pochemu on budet vam polezen ili sovremennyi sposob sdelat tainoe iavnym [Dashboard – what it is and why it will be useful to you or a modern way to make the secret explicit]. *Habr.* [online] 27 November 2017. Available at: <https://habr.com/ru/company/devexpress/blog/341972/> [Accessed 17 March 2021].
Power BI. [online] Available at: <https://powerbi.microsoft.com/ru-ru/[Accessed 17 March 2021].
Prohnoz rozvytku epidemii COVID-19 v Ukraini na 23 lystopada – 7 hrudnia 2020 roku («Prohnoz RH-29») [Forecast of the COVID-19 epidemic in Ukraine on November 23 – December 7, 2020 («Forecast RG-29»)], 2020. *The National Academy of Sciences of Ukraine*. [online] 23 November 2020. Available at: <http://www.nas.gov.ua/UA/Messages/Pages/View.aspx?MessageID=7187> [Accessed 21 April 2021].
Systema monitorynhu rasshyrenyia эpydemyy koronavirusu [System for monitoring the spread of the coronavirus epidemic]. *National Security and Defense Council of Ukraine*. [online] Available at: <https://covid19.rnbo.gov.ua/> [Accessed 21 April 2021].
Use pro koronavirus (dosie) [All about the coronavirus (dossier)]. *Deutsche Welle*. [online] Available at: <https://www.dw.com/uk/use-pro-koronavirus/a-52993788> [Accessed 21 April 2021].
Vplyv atmosfernykh faktoriv na kilkist vypadkiv Kovid 19 [Influence of atmospheric factors on the number of cases Kovid 19]. [online] Available at: <https://app.powerbi.com/view?r=eyJrIjoiYWYyZDEyMGEtY2UzZi00MGVhLWExODQtNjE2YmQ4ODUyZmNlIiwidCI6Ijc2OGJjNWJlLTk0Y2UtNDkxMS1hYjAxLWUyMzMwNGQ1MjdlZiIsImMiOjl9> [Дата звернення 21 квітня 2021].
Zagriaznenie vozdukha v Belgii: kachestvo vozdukha na karte v rezhime realnogo vremeni [Air pollution in Belgium: real-time map air quality]. *The World Air Quality Project*. [online] Available at: <https://aqicn.org/map/belgium/ru/> [Дата обращения 21 апреля 2021].

UDC 004.77:311.2

Bulyha Kostiantyn,
PhD of Technical Sciences, Associate Professor
at the Department of Computer Science,
Kyiv National University of Culture and Arts,
Kyiv, Ukraine
piton54@ukr.net
https://orcid.org/ 0000-0002-9537-3226

Tolmach Maryna,
Lecturer, Deputy Dean of the Faculty of Distance Learning,
Kiev National University of Culture and Arts,
Kyiv, Ukraine
margo@knukim.edu.ua
https://orcid.org/0000-0002-7020-1348

STATISTICAL DATA VISUALIZATION ON THE POWER BI PLATFORM

The purpose of the article is to use the latest information technologies to clearly reflect the impact of various factors on the spread of the COVID-19 pandemic. The creation of a dashboard (PoltP, 2017) to visualize the statistics of the COVID-19 epidemic in the spring of 2021 is considered in detail.

The research methodology lies in the use of a systematic analysis of statistical data.

The novelty of the study is the combination of diverse statistical information for preliminary qualitative analysis of a complex epidemiological situation.

Conclusions. The material presented in the article gives a clear example of the use of various data to visualize the epidemiological situation. The possibility of loading statistical data of any format into a common source for the analysis of the influence on the general level of disease distribution is shown.

Keywords: COVID-19 pandemic; MS Power BI; Power Query; Power Pivot; MS Excel; visualization; dashboard; qualitative analysis.

УДК 004.77:311.2

Булыга Константин,
кандидат технических наук, доцент кафедры компьютерных наук,
Киевский национальный университет культуры и искусств,
Киев, Украина
ovch05@ukr.net
https://orcid.org/0000-0003-4255-5816

Толмач Марина,
преподаватель, заместитель декана факультета дистанционного обучения,
Киевский национальный университет культуры и искусств,
Киев, Украина
margo@knukim.edu.ua
https://orcid.org/0000-0002-7020-1348

ВИЗУАЛИЗАЦИЯ СТАТИСТИЧЕСКИХ ДАННЫХ НА ПЛАТФОРМЕ POWER BI

Целью статьи является использование новейших информационных технологий для наглядного отображения влияния различных факторов на распространение пандемии COVID-19. Подробно рассмотрено создание дашборда (PoltP, 2017) для визуализации статистических данных эпидемии COVID-19 весной 2021 года.

Метод исследования – системный анализ статистических данных.

Новизной проведенного исследования является объединение разносторонней статистической информации для предварительного качественного анализа сложной эпидемиологической ситуации.

Выводы. Изложенный в статье материал дает наглядный пример использования различных данных для визуализации эпидемиологической обстановки. Показана возможность загрузки статистических данных любого формата в общий источник для анализа влияния на общий уровень распространения заболевания.

Ключевые слова: пандемия COVID-19; MS Power BI; Power Query; Power Pivot; MS Excel; визуализация; дашборд; качественный анализ.

12.05.2021