

Sprawozdanie końcowe — Dashboard COVID-19

Uczestnicy grupy

Imię i nazwisko	Numer albumu
Kamil Bednarski	143334
Paweł Drzyzga	143200
Łukasz Hapeta	143211

Z uwagi na zmianę datasetu (poprzedni nie nadawał się na dashboard), wstawiamy nowe pytania do zbioru. Zbiór zawiera dane ze wszystkich krajów na świecie z liczbą zachorowań, zgonów, wyzdrowień oraz aktywnych przypadków podczas epidemii COVID od stycznia 2020 do lipca 2020.

Link do zbioru: <https://www.kaggle.com/datasets/imdevskp/corona-virus-report>

Opracowane pytania

1. Jakie były kraje z najwyższą śmiertelnością?
2. Jakie były kraje z najniższą śmiertelnością?
3. Jak przebiegał rozwój pandemii na świecie od stycznia 2020 do lipca 2020?
4. Jak przebiegał rozwój pandemii w Europie od stycznia 2020 do lipca 2020?
5. Jak przebiegał rozwój pandemii w Polsce od stycznia 2020 do lipca 2020?
6. Rozwój COVID-u w Polsce na tle świata.
7. Rozwój COVID-u w Polsce na tle Europy.
8. Liczba zachorowań na milion Polaków.
9. Liczba zachorowań na milion Europejczyków.
10. Liczba zachorowań na milion osób na świecie.

Opis realizacji

Głównym celem projektu było stworzenie dashboardu, który umożliwia analizę danych dotyczących pandemii COVID-19. Początkowo głównym celem było stworzenie dashboardu z danymi dotyczącymi pandemii COVID-19 w Polsce, jednak w trakcie realizacji projektu zdecydowaliśmy się na rozszerzenie zakresu danych o informacje z całego świata oraz dostosowanie całego dashboardu do analizy danych z różnych krajów.

Aby całkowicie nie rezygnować z początkowego pomysłu, dużo początkowych ustawień jest dostosowanych do Polski, jednak w każdej chwili można zmienić kraj i analizować dane z innych państw, nie tylko pojedynczo.

Nie da się ukryć, że głównym elementem projektu jest część stworzona w R. Wynika to z tego, że R jest bardzo dobrym narzędziem do analizy danych i wizualizacji. Wykorzystaliśmy bibliotekę Shiny, która pozwala na tworzenie interaktywnych aplikacji webowych w R. Dzięki temu mogliśmy stworzyć dashboard, który umożliwia użytkownikom interakcję z danymi. Dodatkowo, praca w R daje dużo większą swobodę w dostosowywaniu dashboardu do naszych potrzeb, w przeciwieństwie do narzędzi typu Tableau, które są bardziej ograniczone.

Dashboard w R

Od ostatniego sprawozdania, w którym został przedstawiony wstępny dashboard, został on znacznie rozbudowany. Dodano wiele nowych funkcji, które umożliwiają lepszą analizę danych. Jest możliwość analizy danych z jednego bądź wielu krajów, porównywanie danego kraju z uśrednionymi wynikami z innych krajów, jak i przeglądanie danych w wersji rankingowej, z możliwością przenoszenia się do szczegółowych informacji z danego kraju.

Wszystkie te funkcje są dostępne w interaktywnym dashboardzie, który można uruchomić lokalnie na komputerze. Wystarczy zainstalować R oraz odpowiednie biblioteki, a następnie uruchomić plik `apka.R` w środowisku RStudio. Z nazwą pliku wiąże się dziwna sprawa, ponieważ kiedy plik posiadał nazwę `app.R`, to były problemy z uruchomieniem aplikacji, a konkretnie z uruchomieniem innych plików, zawartych w środowisku, więc zmieniliśmy nazwę na `apka.R` i problem zniknął.

Od czasu ostatniej prezentacji dashboardu, również została zmodyfikowana część wizualna, która jest teraz bardziej przejrzysta i estetyczna. Pozbyto się znacznej ilości pustych przestrzeni, dodatkowo nie trzeba już nic scrollować w dół, aby zobaczyć wszystkie elementy dashboardu. Wszelkiego rodzaju słupki zawsze są w kolejności malejącej, a sama kolorystyka w całym dashboardzie jest spójna i przyjemna dla oka.

Z większych zmian, została wyrzucona ostatnia zakładka "Country Details", w której mogliśmy podejrzeć szczegółowe informacje z danego kraju. Została ona zastąpiona przez interakcję z wykresem w części "Country Ranking". Wystarczy kliknąć w słupek odpowiadający danemu krajowi, aby przenieść się do szczegółowych informacji z tego kraju. Jest tam wymienione przede wszystkim "miejsce w rankingu" w konkretnej kategorii, jaka akurat była wybrana na wykresie (confirmed, deaths, recovered, active), a także wykresy, statystyki, tabela z danymi jak i możliwość pobrania tych danych w formacie CSV.

Dashboard w Tableau

Dodatkowo został stworzony dashboard w Tableau, który jest bardziej ograniczony, ale również pozwala na analizę danych dotyczących pandemii COVID-19. W naszym przypadku jest on dużo bardziej uboższy, ponieważ ogranicza się głównie do czegoś podobnego, jak pierwszy dashboard w R. Wynika to głównie z ograniczeń narzędzia Tableau, które nie pozwala na tak dużą swobodę w tworzeniu interaktywnych aplikacji webowych jak R. Stworzenie dashboardu w Tableau, który by odpowiadał naszemu pomysłowi, zajęłoby dużo więcej czasu i wymagałoby większej ilości pracy, dlatego zdecydowaliśmy się na stworzenie prostszego dashboardu, a skupiliśmy się na dashboardzie w R.

Link do dashboardu w Tableau Public <https://public.tableau.com/app/profile/pawe.drzyzga/viz/OWAD2025Dashboard/Dashboard1>

Porównanie prac

Praca w R i Tableau różni się głównie podejściem do analizy danych. R jest językiem programowania, który pozwala na dużą swobodę w tworzeniu interaktywnych aplikacji webowych, natomiast Tableau jest narzędziem do wizualizacji danych, które jest bardziej ograniczone. W R możemy stworzyć aplikację, która będzie dostosowana do naszych potrzeb, natomiast w Tableau musimy się dostosować do narzędzia. Z jednej strony, Tableau jest bardziej przyjazne dla użytkownika, ponieważ nie wymaga znajomości programowania, pozwala na szybkie tworzenie wizualizacji danych i jest bardziej intuicyjne. Z drugiej strony, R daje nam większą swobodę w tworzeniu interaktywnych aplikacji webowych, które mogą być dostosowane do naszych potrzeb. W przypadku R tak naprawdę ogranicza nas tylko wyobraźnia, a w przypadku Tableau jesteśmy ograniczeni przez narzędzie.

Używanie Tableau w naszym przypadku nie miało większego sensu, ponieważ nie było możliwości stworzenia dashboardu, który by odpowiadał naszym potrzebom. Stawianie sobie takich ograniczeń nie miało sensu, kiedy jesteśmy w stanie stworzyć coś dużo lepszego i bardziej funkcjonalnego czy kreatywniejszego w R. Stąd uważamy, że praca w R, choć na początku może wydawać się trudniejsza czy bardziej wymagająca, jest dużo bardziej opłacalna w dłuższej perspektywie czasowej.

Historia

Pierwsze uruchomienie dashboardu prowadzi nas do pierwszego panelu pod nazwą "Global Trends", z automatycznie wybranym krajem "Polska" oraz pełnym zasięgiem dat, jaki był dostępny w zbiorze danych. Możemy tu zobaczyć wykresy przedstawiające rozwój pandemii w Polsce, sprawdzić jak niektóre dedyzje np rządowe wpłynęły na rozwój pandemii. Posiadamy statystyki jak na dany (ostatni) dzień wygląda sytuacja, mamy również histogramy, jak i wykres kołowy, który dobrze obrazuje udział poszczególnych grup w całości potwierdzonych przypadków COVID-19 w Polsce. Dużą zaletą tej sekcji jest to, że nie musimy się ograniczać tylko do Polski, możemy zmienić kraj na dowolny inny i zobaczyć jak wygląda sytuacja w danym kraju. Co więcej, możemy wybrać kilka krajów, jakiś konkretny kontynent, czy nawet cały świat i zobaczyć jak wygląda sytuacja w danym regionie.

Kolejną sekcją jest "Country Comparison", która pozwala na porównanie wybranych krajów pod względem liczby zachorowań, zgonów, wyzdrowień i aktywnych przypadków.

Możemy tu zobaczyć, jak dany kraj wypada na tle innych krajów, czy to wybranych pojedynczo, czy też uśrednionych wyników z dowolnej liczby innych krajów. Jest to bardzo przydatne narzędzie, które pozwala na szybką analizę danych i porównanie sytuacji w różnych krajach. Szczególnie, że oprócz samych wykresów, mamy również tabelę z danymi, która pozwala na dokładniejsze przeanalizowanie sytuacji w danym kraju. Wykresy liniowe zostały rozdzielone na 4, aby nie było problemu z czytelnością. Tutaj szczególnie jest przydatny wykres kołowy, który pozwala na szybkie zobaczenie udziału poszczególnych krajów w całkowitej liczbie zachorowań, zgonów, wyzdrowień i aktywnych przypadków. Oczywiście tutaj tak samo jak wcześniej, możemy wybrać sobie zakres dat jaki tylko potrzebujemy.

Trzecią sekcją jest "Country Ranking", która pozwala na przeglądanie krajów w formie rankingu, według wybranej kategorii (liczba zachorowań, zgonów, wyzdrowień, aktywnych przypadków). Możemy tu zobaczyć, które kraje mają najwięcej zachorowań, zgonów, wyzdrowień i aktywnych przypadków. Możemy sobie sortować cały wykres czy to rosnące, czy to malejąco, tak jak wybrać liczbę wyświetlanych krajów z przedziału [1,25]. Dodatkowo, klikając na odpowiedni słupek (odpowiadający danemu państwu) zostaniemy przeniesieni do sekcji, która pokazuje bardziej szczegółowe informacje o danym kraju, takie jak jego statystyki, interaktywny wykres liniowy do głębszej analizy, histogram, tabela z danymi, a co ciekawsze, możliwość pobrania całej tej tabelki do pliku w formacie csv. Dodatkowo na samej górze jest wyświetlana nazwa kraju, rodzaj przypadku, który mieliśmy wybrany w Country Ranking, jak i miejsce tego kraju w tym właśnie rankingu. Możemy oczywiście wrócić do poprzedniego panelu korzystając z przeznaczonego do tego przycisku, a następnie wybrać tam kraj referencyjny. Wybranie takiego kraju sprawi, że zostanie on wyróżniony, a dodatkowo na wykresie zobaczymy kraje "sąsiadujące" z nim, jeśli chodzi o wynik w danej kategorii. Interakcja działa tutaj w identyczny sposób.

Wnioski końcowe

Projekt poświęcony analizie danych dotyczących pandemii COVID-19 zaowocował stworzeniem dwóch dashboardów — w R i Tableau — z czego główny nacisk został położony na rozwiązanie oparte na środowisku R. Dzięki wykorzystaniu bibliotek takich jak 'shiny', 'ggplot2' czy 'plotly', udało się stworzyć w pełni interaktywną aplikację webową, która pozwala użytkownikowi na dynamiczną analizę danych, porównywanie krajów, przeglądanie rankingów, a także dostęp do szczegółowych informacji o wybranym państwie.

Stworzony dashboard w R pozwala w przejrzysty i intuicyjny sposób odpowiadać na wszystkie postawione na początku pytania badawcze, takie jak porównania zachorowań i zgonów w ujęciu globalnym, kontynentalnym czy krajowym, a także umożliwia znacznie szersze analizy, nieuwjęte bezpośrednio w zestawie początkowych problemów. Elastyczność oraz możliwości dostosowywania interfejsu sprawiają, że narzędzie to może być z powodzeniem wykorzystywane także w innych kontekstach analitycznych.

Dashboard w Tableau, mimo że również zawiera podstawowe informacje o przebiegu pandemii, pełni w tym przypadku funkcję pomocniczą. Ograniczenia platformy Tableau w zakresie interaktywności i swobodnego modelowania logiki działania aplikacji sprawiły, że to środowisko R okazało się znacznie bardziej adekwatne do założonych celów projektu.