

Czy profil użytkownika wpływa na aktywność w aplikacji?

Aleksandra Zachajska, Konstanty Andrzejczak

Wprowadzenie

Rozwój technologii mobilnych i powszechna dostępność urządzeń elektronicznych znacząco wpłynęły na sposób, w jaki ludzie zdobywają wiedzę i rozwijają swoje umiejętności. Aplikacje mobilne, takie jak Duolingo, stanowią doskonały przykład tej rewolucji edukacyjnej, umożliwiając użytkownikom naukę języków obcych w sposób elastyczny, dostosowany do ich indywidualnych potrzeb i harmonogramu. Od momentu wprowadzenia na rynek w 2011 roku, Duolingo zdobyło ogromną popularność na całym świecie, przyciągając miliony użytkowników, którzy regularnie korzystają z aplikacji, aby doskonalić swoje umiejętności językowe.

Za sukcesem aplikacji stoi nie tylko jej innowacyjna metoda nauki, oparta na gamifikacji, ale także zdolność do dostosowania się do różnorodnych potrzeb użytkowników. Aplikacja stała się szczególnie popularna w Stanach Zjednoczonych, Europie i Azji, a jej użytkownicy reprezentują szeroką gamę krajów, kultur oraz poziomów zaawansowania. Zjawisko to skłania do refleksji nad tym, w jaki sposób profil użytkownika – obejmujący takie czynniki jak wiek, kraj pochodzenia, płeć czy wcześniejsze doświadczenia z językiem – może wpływać na jego zaangażowanie oraz efektywność nauki. Celem niniejszej pracy jest zbadanie, w jaki sposób różne cechy użytkowników aplikacji Duolingo kształtują ich aktywność, motywację oraz postępy w nauce, a także jakie czynniki mogą decydować o sukcesie w tym procesie edukacyjnym.

Dane

Aby przeprowadzić tę analizę, posłużymy się danymi pochodzącymi z publicznego repozytorium MatthewWolff/DuolingoAnalysis na GitHubie, które zawierają szczegółowe informacje na temat użytkowników, ich aktywności w aplikacji, liczby rozpoczętych i zakończonych lekcji, a także poziomu ich zaangażowania w naukę.

<https://github.com/MatthewWolff/DuolingoAnalysis.git>

Autor nie zamieścił informacji o licencji na jakiej udostępniono dane jednak dane zostaną użyte wyłącznie w celach edukacyjnych do wykonania niniejszej analizy funkcjonalnej. Dane

zostały zebrane za pomocą ankiety przeprowadzonej przez Duolingo w okresie od 1 sierpnia do 5 listopada 2018 roku. Ankieta obejmowała pytania dotyczące aktywności użytkowników, korzystania z aplikacji oraz ich profilu demograficznego, takiego jak wiek, płeć czy kraj zamieszkania.

Opis zmiennych

Dane w analizie zawierają 29 zmiennych opisujących profile użytkowników Duolingo oraz ich aktywność w aplikacji. Wśród kluczowych miar aktywności użytkowników znajdują się: liczba dni aktywnych, liczba rozpoczętych i zakończonych lekcji oraz najdłuższa seria korzystania z aplikacji. Dodatkowo dane demograficzne, takie jak wiek, kraj pochodzenia i status zatrudnienia, umożliwiają zrozumienie różnorodności użytkowników aplikacji.

Table 1: Pełny opis zmiennych w zbiorze danych Duolingo.

Zmienna	Opis	Jednostki/Wartości
user_id	Unikalny identyfikator użytkownika	Tekst
age	Grupa wiekowa użytkownika	Kategorie: 'Under 18', '18-34', '35-54', '55-74', '75 or older'
annual_income	Roczny dochód użytkownika	Kategorie: '\$0 - \$25,000', '\$26,000 - \$75,000', '\$76,000 - \$150,000', '> \$150,000'
country	Kraj zamieszkania użytkownika	Kody krajów, np. 'US', 'JP', 'DE'
duolingo_platform	Platforma używana przez użytkownika	Kategorie: 'Android', 'iOS', 'Web'
duolingo_subscriber	Status subskrypcji Duolingo Plus	Kategorie: 'Yes', 'No', 'Previously Paid'
duolingo_usage	Częstotliwość korzystania z aplikacji Duolingo	Kategorie: 'Daily', 'Weekly', 'Monthly'
employment_status	Status zatrudnienia użytkownika	Kategorie: 'Employed full-time', 'Part-time', 'Retired', 'Unemployed'
future_contact	Czy użytkownik zgodził się na kontakt w przyszłości	Prawda/Falsz
gender	Płeć użytkownika	Kategorie: 'Male', 'Female', 'Other'
other_resources	Liczba innych zasobów używanych do nauki języków obcych	Liczba całkowita
primary_language_difficulty	Doświadczenia zaangażowania w naukę języka głównego	Kategorie: 'Low', 'Medium', 'High'
primary_language_proficiency	Przebieg postępów w nauce języka głównego	Kategorie: 'Poor', 'Average', 'Good'

Zmienna	Opis	Jednostki/Wartości
primary_language_motivation	Motywacja użytkownika do nauki języka głównego	Tekst (deklaracje motywacji)
primary_language_additional_info	Dodatkowe informacje o motywacji do nauki głównego języka	Tekst (dodatkowe informacje o motywacji)
primary_language_proficiency	Samoocena oceniony poziom znajomości języka głównego	Kategorie: 'Beginner', 'Intermediate', 'Advanced'
student	Czy użytkownik jest studentem	Prawda/Falsz
survey_complete	Czy użytkownik ukończył ankietę	Prawda/Falsz
time_spent_seconds	Czas spędzony na wypełnianiu ankiety (w sekundach)	Liczba całkowita (sekundy)
duolingo_start_date	Data rozpoczęcia korzystania z aplikacji	Data (YYYY-MM-DD)
highest_course_progress	Najwyższy osiągnięty postęp w kursie (w procentach)	Procent (0-100)
took_placement_test	Czy użytkownik wykonał test poziomujący	Prawda/Falsz
purchased_subscription	Czy użytkownik wykupił subskrypcję Duolingo Plus	Prawda/Falsz
highest_crown_count	Najwyższa liczba zdobytych koron	Liczba całkowita
n_active_days	Liczba dni, w których użytkownik był aktywny	Liczba całkowita
n_lessons_started	Liczba lekcji rozpoczętych przez użytkownika	Liczba całkowita
n_lessons_completed	Liczba lekcji ukończonych przez użytkownika	Liczba całkowita
longest_streak	Najdłuższa seria dni korzystania z aplikacji	Liczba całkowita
n_days_on_platform	Łączna liczba dni od rejestracji użytkownika w aplikacji	Liczba całkowita

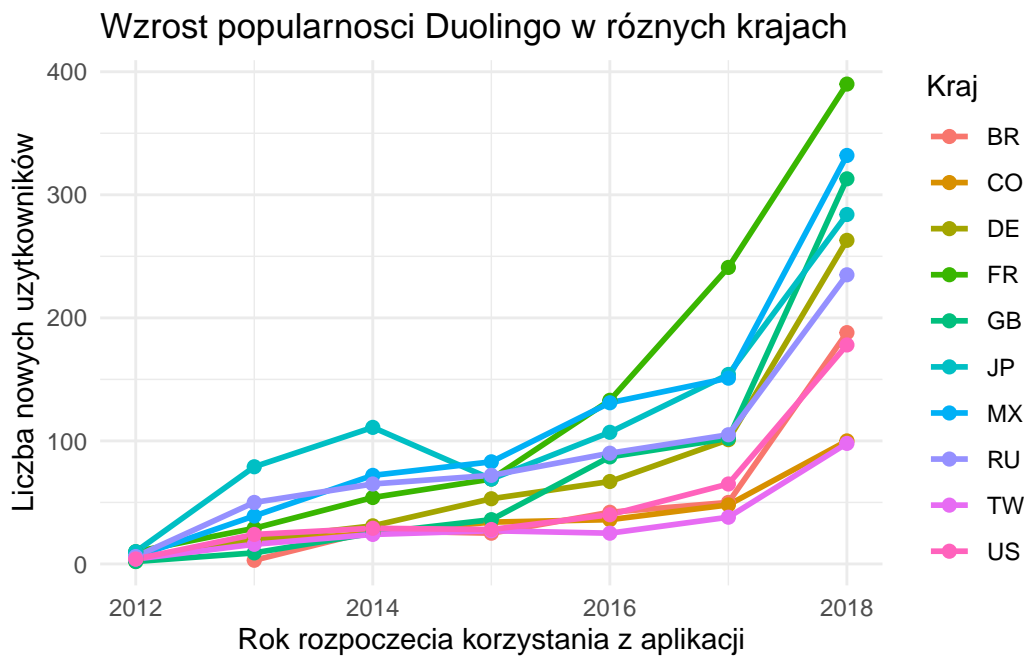
Statystyki opisowe zmiennych liczbowych

```
# A tibble: 1 x 72
```

```

other_resources_count time_spent_seconds_count highest_course_progress_count
                        <int>                        <int>                        <int>
1                      5488                      5488                      5488
# i 69 more variables: highest_crown_count_count <int>,
#   n_active_days_count <int>, n_lessons_started_count <int>,
#   n_lessons_completed_count <int>, longest_streak_count <int>,
#   n_days_on_platform_count <int>, other_resources_mean <dbl>,
#   time_spent_seconds_mean <dbl>, highest_course_progress_mean <dbl>,
#   highest_crown_count_mean <dbl>, n_active_days_mean <dbl>,
#   n_lessons_started_mean <dbl>, n_lessons_completed_mean <dbl>, ...

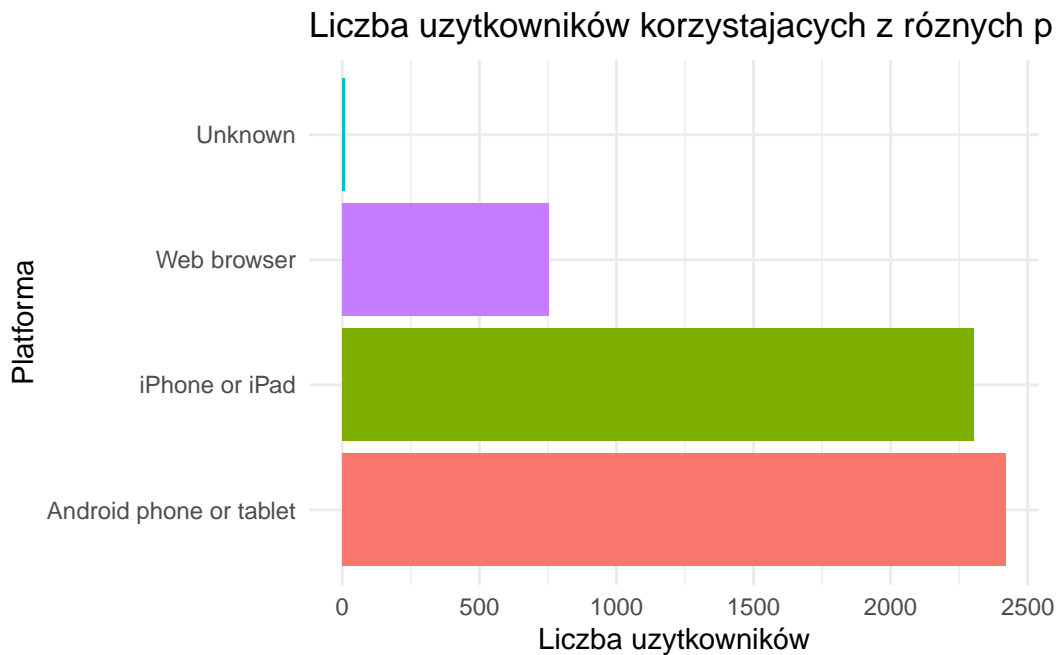
```



```

<ScaleContinuousPosition>
Range:
Limits: 0 -- 1

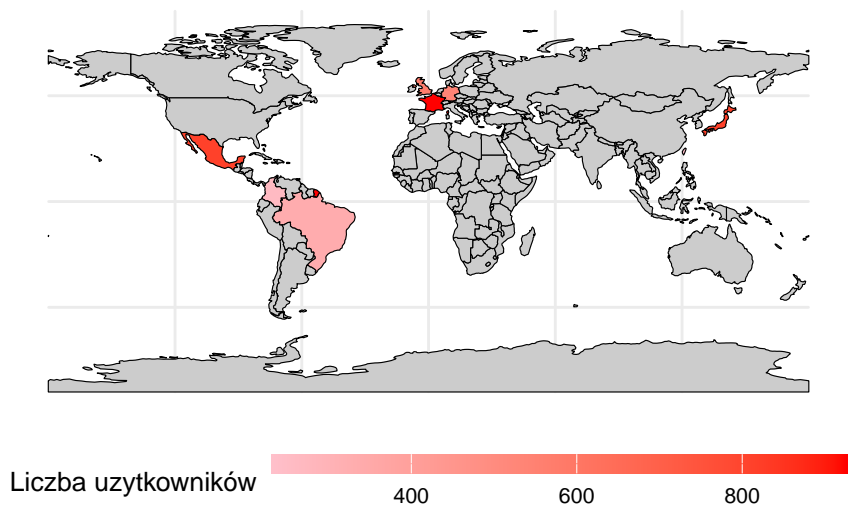
```



Kto sięga po aplikację?

Rozkład użytkowników Duolingo według krajów

Kolor przedstawia liczbę użytkowników w danym kraju



Zródło: Dane Duolingo

```
# A tibble: 10 x 1
  country
  <chr>
1 Japan
2 Russian Federation
3 Colombia
4 Taiwan
5 United Kingdom
6 Brazil
7 Mexico
8 Germany
9 France
10 United States
```

Rozkład użytkowników Duolingo według krajów

Kolor przedstawia liczbę użytkowników w danym kraju

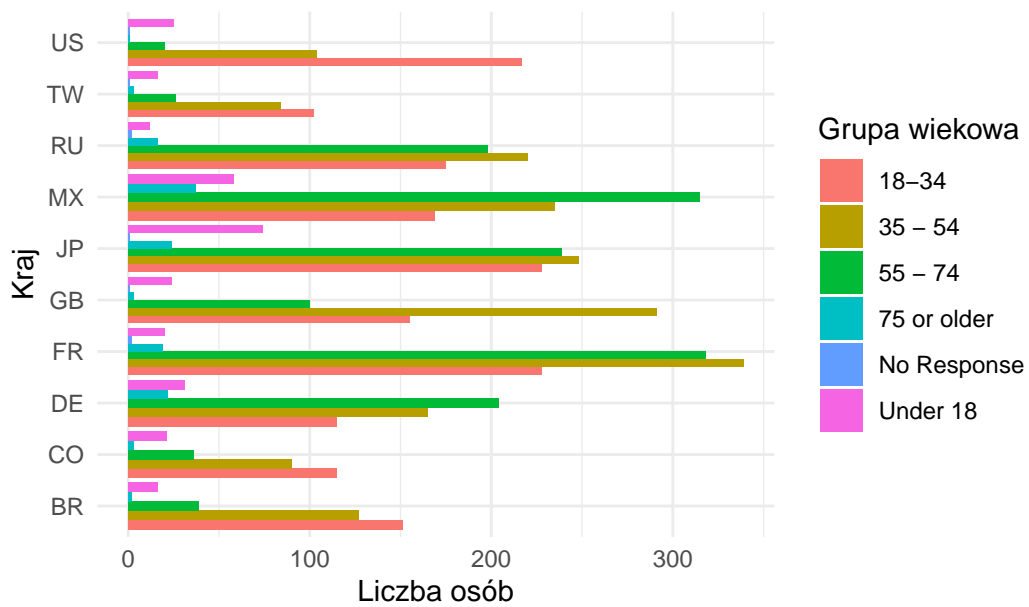


Liczba użytkowników

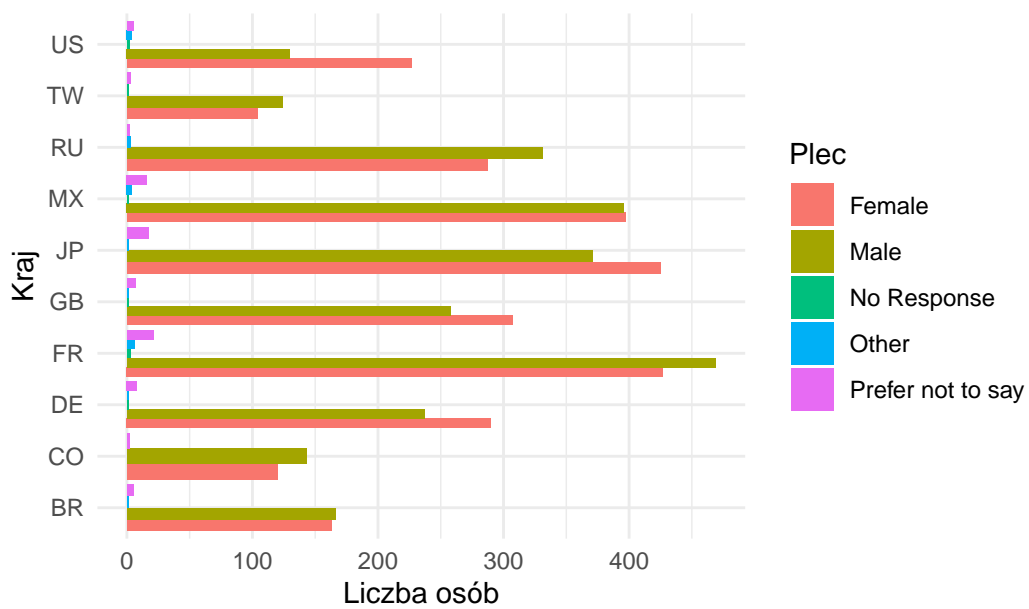
150 200 250 300

Zródło: Dane Duolingo

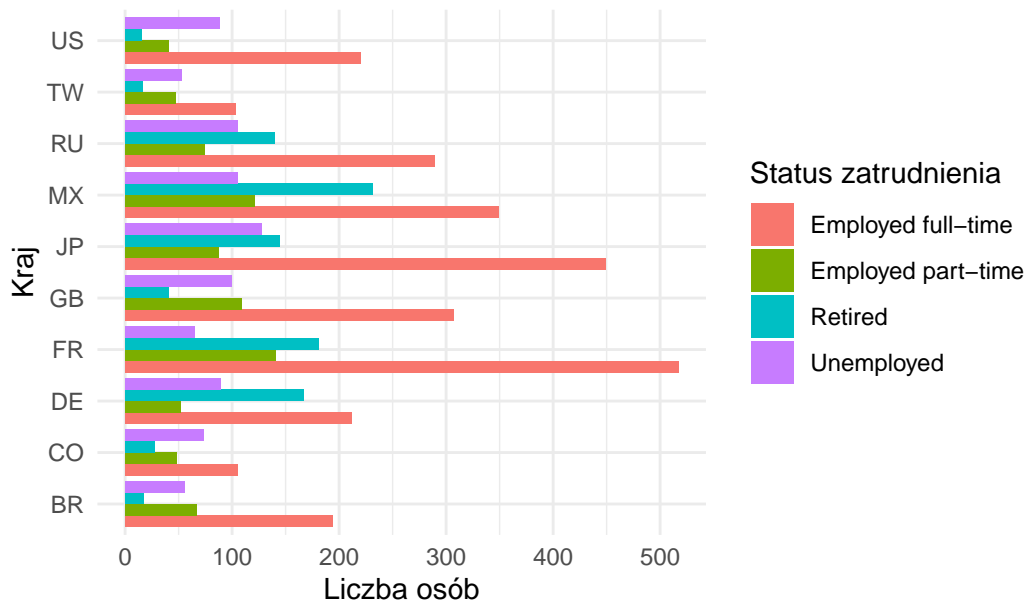
Liczba osób w różnych grupach wiekowych w różnych krajach



Liczba osób według płci w różnych krajach



Liczba osób według statusu zatrudnienia w różnych krajach



Śmietnik

Wprowadzenie

We wprowadzeniu napisać z jakich dat są zebrane dane i zaznaczyć daty widoczne w zestawieniu kiedy uczestnicy podlegający spisowi rozpoczęli korzystanie z aplikacji odnieść się do daty wprowadzenia aplikacji na rynek i mniej więcej kiedy aplikacja stała się popularna można dołączyć dane dotyczące dynamiki wejścia na rynki w różnych krajach w szczególności tych uwzględnionych w zestawieniu

[1] 4

Kto sięga po aplikację (ogólna charakterystyka użytkowników)?

Analiza jakie kraje są brane pod uwagę w każdym z nich możemy zestawić ile jest osób zatrudnionych w jakim wymiarze i ilu w jakim wieku jakiej płci (ogólnie info)

Profil użytkownika

Analiza profilu użytkowników dokładna pod kątem używania aplikacji tu zestawić: -> zaangażowanie -> motywację -> czy miał już styczność z językiem -> poziom deklarowany vs wyznaczony w aplikacji

Dynamika rozwoju przy użyciu aplikacji -> Zestawienie liczby rozpoczętych lekcji w zestawieniu z zakończonymi -> liczba dni na platformie vs liczba aktywnych dni -> najdłuższa seria