

Raport

Analiza danych użytkowników platformy Netflix

Wykonały:

Anna Vezdenetska nr albumu: 107938

Aleksandra Kot nr albumu: 104061

Spis treści

Wprowadzenie.....	2
Przegląd danych.....	3
Wyświetlane informacje w menu i ich opis.....	5
1. Wykres bazy użytkowników serwisu NETFLIX z podziałem na płeć oraz grupę wiekową.....	5
2. Wykres zależności, między wiekiem użytkowników serwisu NETFLIX, a urządzeniem, z którego korzystają.....	6
3. Wykres zależności, między krajem zamieszkania użytkowników, a wykorzystywanym planem subskrypcji użytkowników serwisu NETFLIX.....	7
4. Wykres zależności, między krajem zamieszkania użytkowników, a urządzeniem, z którego korzystają użytkownicy serwisu NETFLIX.....	8
5. Wykres, przedstawiający histogram kolumny Age.....	9
6. Wykres przedstawiający rozkład normalny.....	10
7. Wykres skośności.....	11
8. Wykres kurtozy.....	12
9. Wykres słupkowy wyników testu statystycznego chi-kwadrat.....	13
10. Wykres typu heatmap wyników testu chi-kwadrat.....	14
11. Wykres typu heatmap wyników testu chi-kwadrat.....	14
12. Zakończ przegląd wykresów.....	14
13. Autorki projektu.....	14

Wprowadzenie

Zbiór danych użyty w analizie to baza użytkowników serwisu Netflix, która zawiera informacje o subskrypcjach, dochodach miesięcznych, datach dołączenia i ostatnich płatnościach, kraju zamieszkania, wieku, płci, urządzeniach i długości subskrypcji użytkowników serwisu Netflix. W analizie użyto różnych narzędzi i wykresów, aby zrozumieć charakterystykę zbioru danych i zależności między zmiennymi.

Analizowane dane pochodzą z pliku, zawierającego zbiór baz danych użytkowników serwisu NETFLIX. Zbiór pochodzi ze strony Kaggle i jest publicznie dostępny do użytku. Także podgląd do danego [pliku](#) jest załadowany razem z plikiem projektu na platformie github.

Sam projekt można odnaleźć przechodząc za poniższym linkiem do platformy github, gdzie znajdują się [projekt](#).

Oprócz tego, [kopia projektu](#), dostosowana do uruchomienia online, znajduje się w folderze projektu w postaci Jupyter Notebook. Można ją uruchomić za pomocą takiego narzędzia jak np. Google Colab.

Przegląd danych

Przeprowadzono analizę danych z wykorzystaniem bibliotek takich jak numpy, pandas, seaborn, matplotlib, statsmodels, scipy.stats, i sklearn. Aktualnie używana wersja biblioteki sklearn to sklearn.__version__.

Dane zostały wczytane z pliku CSV dostępnego pod adresem [link](#). Zbiór danych zawiera informacje o subskrypcjach użytkowników, przychodach, datach dołączenia, płatnościach, kraju zamieszkania, wieku, płci, urządzeniach i długości planu subskrypcji.

Poniżej przedstawiono początkowe rekordy zbioru danych:

LP	Subscription Type	Monthly Revenue	Join Date	Last Payment Date	Country	Age	Gender	Device	Plan Duration
1	Basic	10	15-01-22	10-06-23	United States	28	Male	Smartphone	30
2	Premium	15	05-09-21	22-06-23	Canada	35	Female	Tablet	30
3	Standard	12	28-02-23	27-06-23	United Kingdom	42	Male	Smart TV	30
4	Standard	12	10-07-22	26-06-23	Australia	51	Female	Laptop	30
5	Basic	10	01-05-23	28-06-23	Germany	33	Male	Smartphone	30

Tabela 1. Tabela przedstawia 5 początkowych rekordów analizowanego zbioru danych. Zbiór danych zawiera 2500 rekordów i 9 kolumn.

Przeprowadzono selekcję cech i wyświetlone nazwy kolumn:

- Subscription Type
- Monthly Revenue
- Join Date
- Last Payment Date
- Country
- Age
- Gender
- Device

- Plan Duration

Wygenerowano także podstawowe statystyki zbioru danych, takie jak średnie, odchylenia standardowe, minimum, maksimum, percentyle, itp. Poniżej przedstawiono wyniki dla kolumn "Monthly Revenue", "Age" i "Plan Duration":

- Średni przychód miesięczny wynosi około 12,51.
- Średni wiek użytkowników wynosi około 38,80.
- Długość planu subskrypcji wynosi zawsze 30 dni.

Przedstawiono podstawowe informacje o danych, takie jak typy danych w każdej kolumnie, liczba wartości różnych od wartości "null", oraz wykorzystanie pamięci.

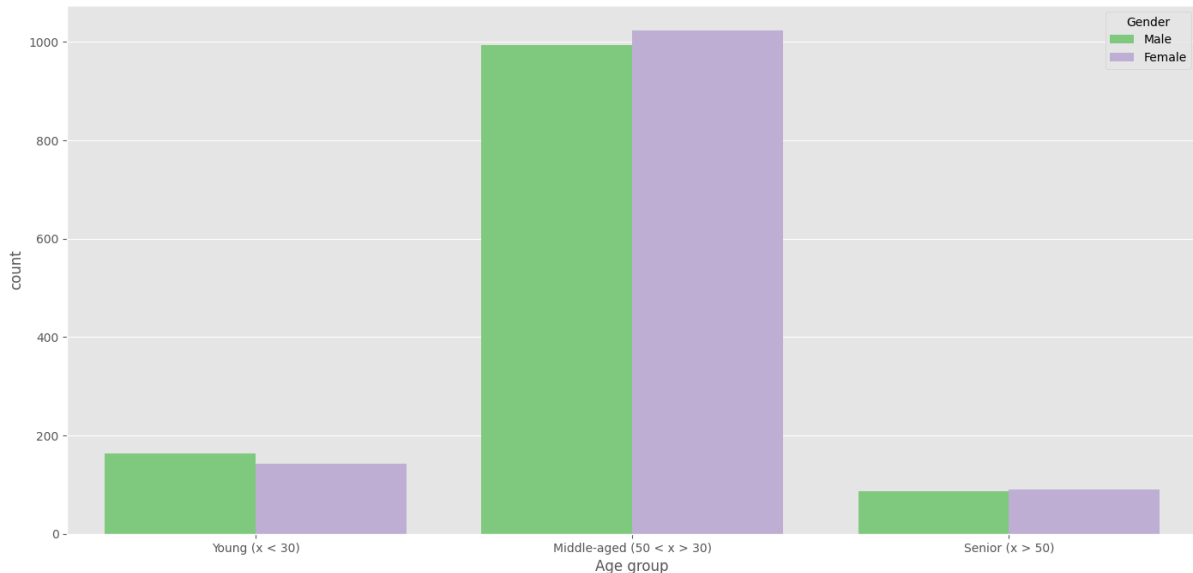
W zbiorze danych nie występują brakujące wartości, co oznacza, że dane są kompletne.

Dokonano zliczenia unikalnych wartości w poszczególnych kolumnach, takich jak "Subscription Type", "Monthly Revenue", "Join Date", "Last Payment Date", "Country", "Age", "Gender", "Device", "Plan Duration".

Wyświetlane informacje w menu i ich opis

Menu posiada 13 pozycji:

1. Wykres bazy użytkowników serwisu NETFLIX z podziałem na płeć oraz grupę wiekową

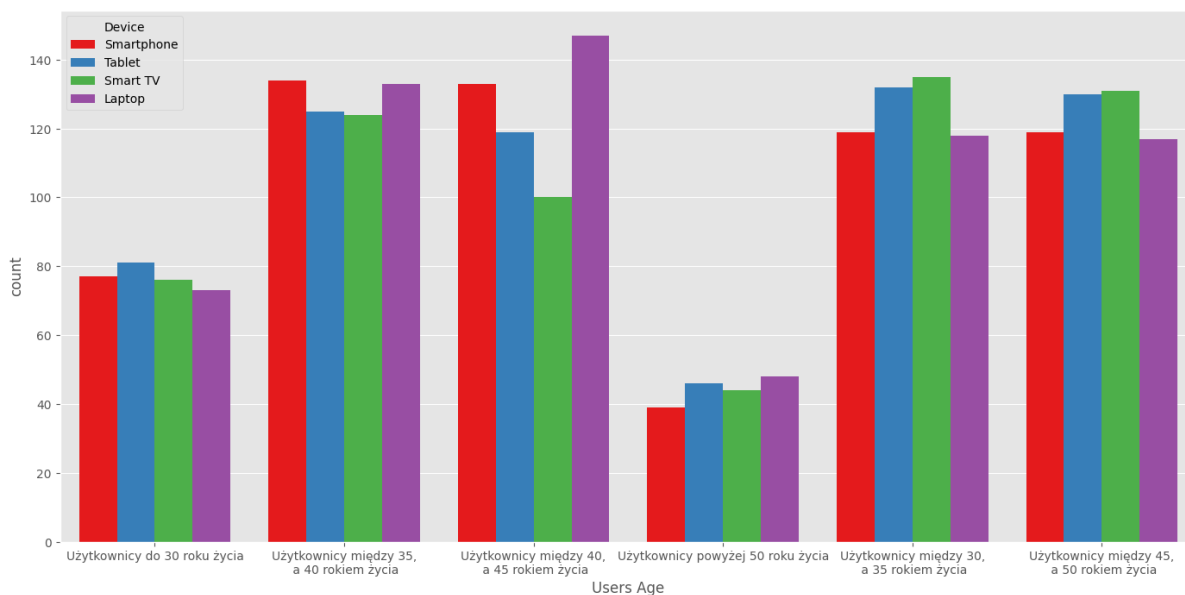


Rycina 1. Przedstawiony wykres słupkowy uwzględnia płeć (mężczyźni-zielony i kobiety-fioletowy) oraz podział na grupy wiekowe.

Powyższy wykres wykorzystuje słupki, aby przedstawić proporcje użytkowników w różnych kategoriach, takich jak płeć i grupy wiekowe. Każda kategoria (na przykład mężczyźni i kobiety, oraz różne grupy wiekowe) ma swoje osobne słupki, które są porównywane ze sobą, co ułatwia wizualizację różnic i zależności między tymi kategoriami.

Ten wykres przedstawia demograficzny podział użytkowników serwisu NETFLIX. Dzięki niemu możemy zobaczyć, ile kobiet i mężczyzn korzysta z platformy oraz jaki jest zakres wiekowy najbardziej aktywnych użytkowników. Może to pomóc w dostosowaniu treści dostępnych na platformie do preferencji demograficznych.

2. Wykres zależności, między wiekiem użytkowników serwisu NETFLIX, a urządzeniem, z którego korzystają



Rycina 2. Przedstawiony wykres słupkowy uwzględnia podział na grupy wiekowe oraz urządzenia z których najczęściej korzystają

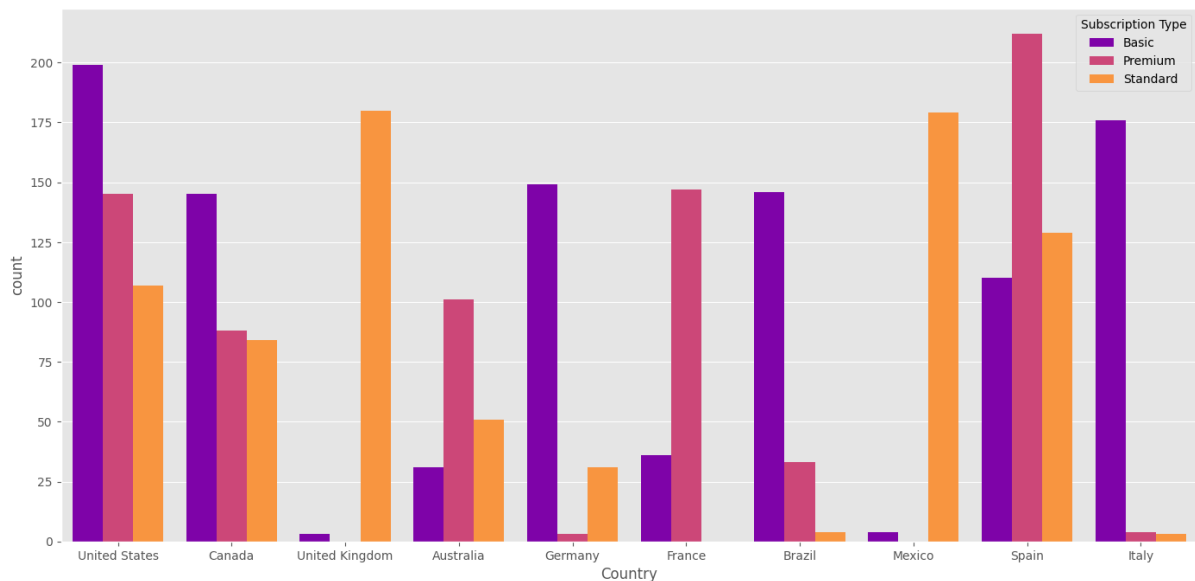
Dane do tego wykresu obejmują wiek użytkowników serwisu NETFLIX oraz informacje o rodzaju urządzenia, z którego korzystają (np. smartfony, telewizory, komputery). To pozwala zrozumieć, czy wiek użytkowników ma wpływ na to, z jakiego rodzaju urządzeń korzystają.

Analiza danych przedstawionych na wykresie nie wskazuje jednak jakichś większych preferencji, jednak warto zaznaczyć że grupa wiekowa 40-45 oddaje preferencje laptopom w przeciwieństwie do Smart TV. Dana grupa wiekowa może preferować laptopy ze względu na ich uniwersalność i mobilność. Laptopy pozwalają użytkownikom na oglądanie treści zarówno w domu, jak i poza nim, to może być szczególnie ważne dla osób pracujących zdalnie lub podróżujących. Może to także wynikać z faktu, że grupa wiekowa 40-45 lat ma łatwiejszy dostęp do laptopów niż do nowoczesnych smart TV.

Także warto zaznaczyć że użytkownicy w grupie wiekowej 35-40 preferują smartfony i laptopy, z czego można również wywnioskować że dla tej grupy wiekowej ważna jest mobilność.

Natomiast grupa wiekowa 45-50 oraz powyżej 50 używają coraz rzadziej smartfony i częściej korzystają z tabletów lub Smart TV. Może to być związane z sytuacją finansową lub preferencjami oglądania treści na większych ekranach.

3. Wykres zależności, między krajem zamieszkania użytkowników, a wykorzystywanym planem subskrypcji użytkowników serwisu NETFLIX



Rycina 3. Przedstawiony wykres słupkowy uwzględnia kraj zamieszkania użytkowników i rodzaj wybranego planu subskrypcji(podstawowy - fioletowy, standardowy - pomarańczowy, premium - różowy)

Wykresy dla Meksyku i Zjednoczonego Królestwa wykazują podobne preferencje subskrybentów. W obu tych krajach plan subskrypcji "Premium" nie cieszy się dużą popularnością, a dominującym wyborem jest "Standard". To może sugerować, że subskrybenci w tych krajach są bardziej skłonni do wybierania tańszych opcji subskrypcji, być może ze względu na aspekty finansowe.

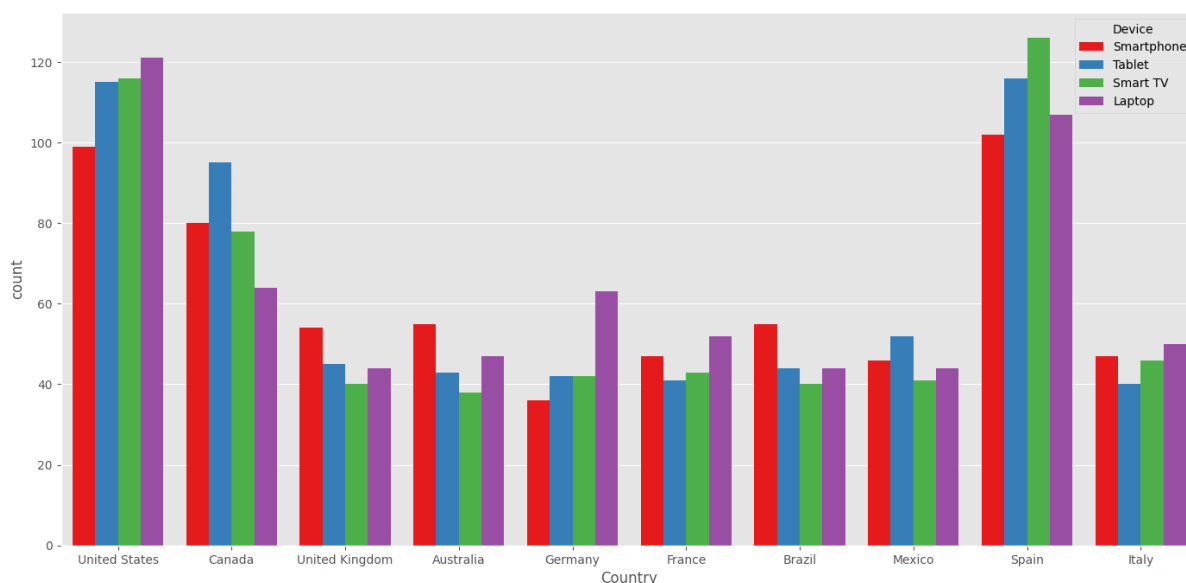
W Hiszpanii, Francji i Australii plan subskrypcji "Premium" zdaje się być bardzo popularny. To może wynikać z większego zainteresowania użytkowników oglądaniem treści w jakości HD lub Ultra HD oraz z większymi możliwościami jednoczesnego oglądania na wielu urządzeniach, co oferuje plan "Premium".

W Niemczech i Włoszech wykresy wskazują, że plan "Bazowy" jest najpopularniejszym wyborem subskrybentów. To może być związane z dążeniem do oszczędności i wybieraniem podstawowej subskrypcji bez dodatkowych funkcji.

W Stanach Zjednoczonych i Kanadzie różnice występują głównie w ilości subskrybentów, ale ogólnie widoczne jest podobieństwo preferencji. "Bazowy" plan jest najpopularniejszy, ale "Premium" zajmuje drugie miejsce. W obu tych krajach subskrybenci mogą być bardziej skłonni do wydawania więcej na subskrypcję, wybierając plan "Premium".

Preferencje dotyczące planów subskrypcji mogą być związane zarówno z aspektami finansowymi, jak i kulturowymi w poszczególnych krajach.

4. Wykres zależności, między krajem zamieszkania użytkowników, a urządzeniem, z którego korzystają użytkownicy serwisu NETFLIX



Rycina 4. Przedstawiony wykres słupkowy uwzględnia kraj zamieszkania użytkowników i urządzenie, z którego korzystają (Smartphone-czerwony; Tablet-niebieski; Smart TV-zielony; Laptop-fioletowy)

Wykres wskazuje, że w tych krajach laptopy są preferowanym rodzajem urządzenia do korzystania z usługi Netflix. To może wynikać z tego, że laptopy są wszechstronne, mobilne i pozwalają na oglądanie treści zarówno w domu, jak i poza nim. Ponadto, te kraje są znane z wysokiego poziomu dostępu do laptopów.

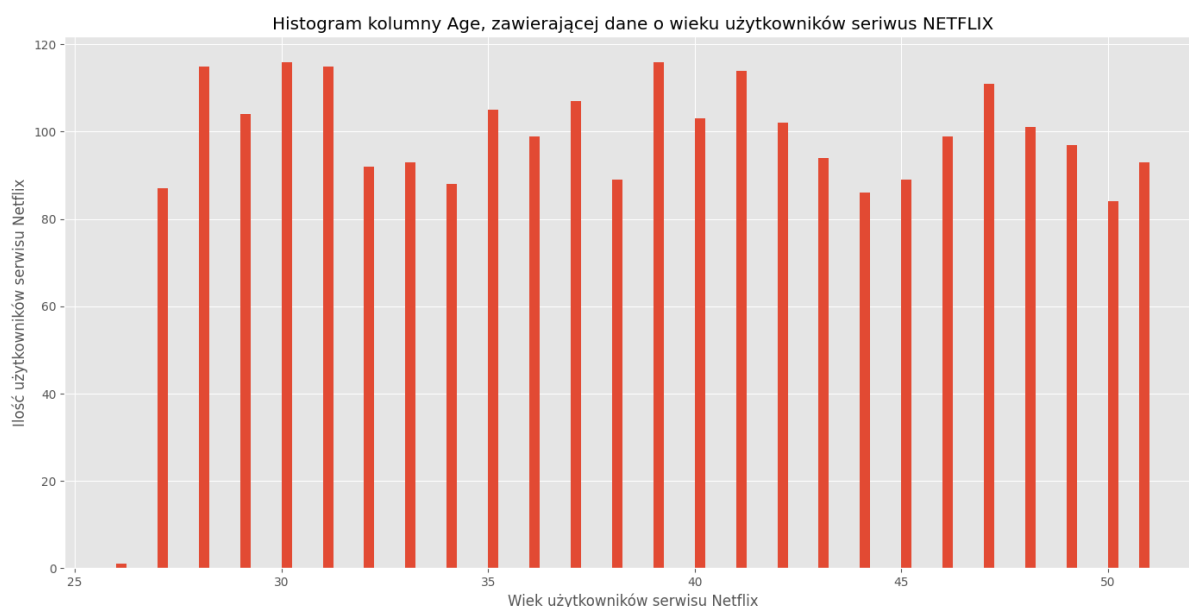
Kanada i Meksyk wykazują preferencję w kierunku tabletów jako głównego urządzenia do oglądania treści na Netflix. Może to sugerować, że mieszkańcy tych krajów lubią większy ekran niż smartfony, ale niekoniecznie potrzebują pełnowymiarowego laptopa do oglądania.

Zjednoczone Królestwo, Australia i Brazylia zdają się preferować smartfony jako główne urządzenie do korzystania z Netflix. To może być związane z tym, że smartfony są wygodne i przenośne, co pozwala użytkownikom na oglądanie treści w dowolnym miejscu i czasie.

Hiszpania wyróżnia się na wykresie jako kraj, w którym preferowane są Smart TV. To może być związane z tym, że telewizory inteligentne oferują duży ekran i wygodę oglądania treści na większym ekranie w domowym otoczeniu.

Dana analiza pokazuje, że preferencje urządzeń do oglądania treści na Netflix różnią się w zależności od kraju i mogą być wynikiem różnych czynników, takich jak dostępność urządzeń, preferencje osobiste użytkowników oraz różnice kulturowe.

5. Wykres, przedstawiający histogram kolumny Age

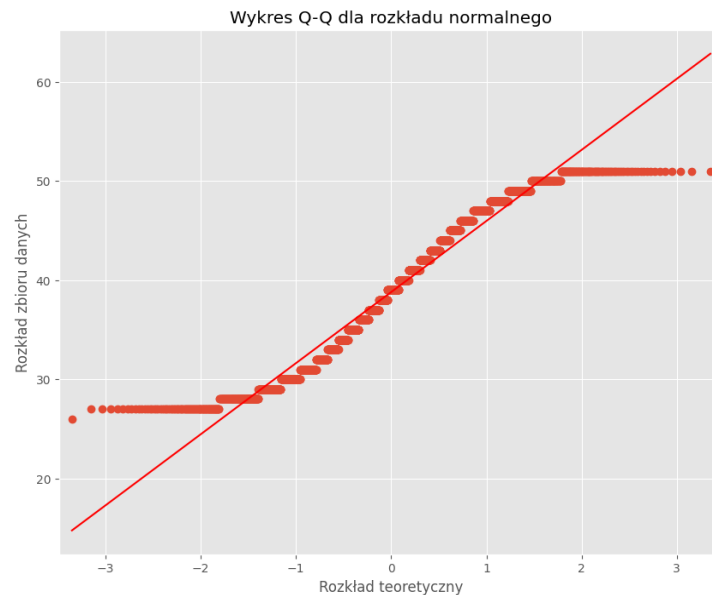


Rycina 5. Histogram słupkowy przedstawia rozkład wieku subskrybentów serwisu Netflix. Na osi poziomej znajdują się zakresy wiekowe, a na osi pionowej liczba subskrybentów w każdym z tych zakresów wiekowych.

Histogram wykazuje, że wiele osób koncentruje się w określonych przedziałach wiekowych, co może być wynikiem preferencji treści dostępnych na Netflix dla tych grup wiekowych. Analiza histogramu pozwala zaobserwować, że najwięcej użytkowników jest w przedziałach 28-31 i 39-41 lat, co sugeruje, że Netflix jest popularny wśród osób w wieku 28-31 lat i 39-41 lat. To może wynikać z różnych preferencji oglądania treści, potrzeb i stylu życia w tych grupach wiekowych. Także z obserwacji wynika, że w wieku 25-26 lat jest mniej użytkowników w porównaniu do innych grup wiekowych. To może być związane z tym, że osoby w tym wieku mogą być bardziej zajęte nauką lub pracą, lub mogą nie mieć stabilności finansowej, co może wpływać na ich czas na oglądanie treści na platformie. Histogram pozwala również zauważyć, że istnieją inne grupy wiekowe, które mają swoje własne preferencje dotyczące subskrypcji Netflix.

Histogram wieku użytkowników serwisu Netflix jest narzędziem, które pozwala na bardziej szczegółową analizę preferencji wiekowych grup użytkowników i wnioskowanie na ich temat. Dana analiza wskazuje na różnice w preferencjach oglądania treści w zależności od wieku i sugerują, że wiek może wpływać na dostępność czasu oraz stabilność finansową użytkowników, co jest ważnym aspektem przy planowaniu treści i strategii biznesowej platformy. Dzięki tym wnioskom można lepiej dostosować ofertę treści do różnych grup wiekowych, co może pomóc w przyciąganiu i zatrzymywaniu klientów.

6. Wykres przedstawiający rozkład normalny



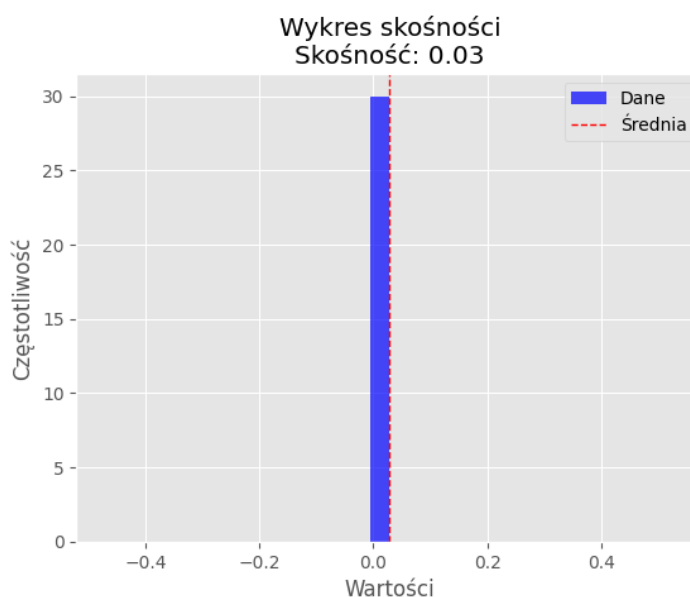
Rycina 6. Wykres Q-Q dla rozkładu normalnego. Na osi poziomej rozkład teoretyczny, a na osi pionowej rozkład zbioru danych.

Wykres kwantyl-quantyl (q-q plot) jest przydatnym narzędziem do wizualnej oceny, czy dane rzeczywiście pochodzą z rozkładu normalnego. Punkty na wykresie q-q plot układają się wzdłuż linii przekątnej ($y = x$), a nawet nakładają się na nią w środkowej części wykresu. To wskazuje na to, że dane z kolumny "Age" są zbliżone do rozkładu normalnego.

Otrzymane wartości kwantyli (mediany, pierwszego i trzeciego kwartyła) również pomagają zrozumieć rozkład danych. Mediana wynosi 39.0, co oznacza, że połowa użytkowników Netflix ma mniej niż 39 lat, a połowa więcej niż 39 lat. Pierwszy kwantyl (25th percentile) wynosi 32.0, co oznacza, że 25% użytkowników ma mniej niż 32 lata, a 75% więcej niż 32 lata. Trzeci kwantyl (75th percentile) wynosi 45.0, co oznacza, że 75% użytkowników ma mniej niż 45 lat, a 25% więcej niż 45 lat. To pozwala na określenie, jak wiek użytkowników jest rozkładany w badanej populacji.

Podsumowując, wykres q-q plot i obliczone kwantyle potwierdzają, że dane z kolumny "Age" są zbliżone do rozkładu normalnego i mogą być traktowane jako dane zmiennych losowych o rozkładzie zbliżonym do normalnego.

7. Wykres skośności

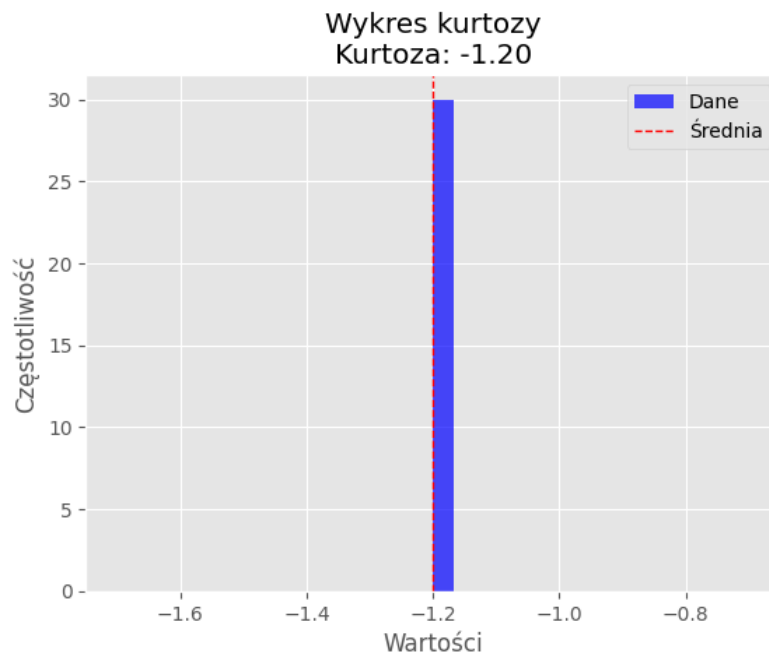


Rycina 7. Wykres skośności. Na osi poziomej wartość, a na osi pionowej częstotliwość.

Wykres skośności jest ważnym narzędziem w analizie danych i pomaga zrozumieć, jak dane są rozkładane wokół swojej średniej wartości. Skośność mierzy asymetrię rozkładu danych. Wartość skośności równa 0 wskazywałaby na symetryczny rozkład wokół średniej, natomiast wartości dodatnie oznaczają, że rozkład jest przesunięty bardziej w prawo względem średniej. Analiza danych wskazuje na to, że skośność wynosi 0.03, co jest bliskie zeru, co oznacza, że rozkład danych jest tylko nieznacznie prawostronnie skośny. To wskazuje, że dane są dość symetryczne wokół średniej wartości wieku.

Wartości dodatnie skośności sugerują, że ogólna tendencja w danych jest przesunięta w kierunku wyższych wartości wieku, ale jest to przesunięcie dość niewielkie. Dlatego można uznać, że rozkład wieku użytkowników jest dość równomierny i nie wykazuje znaczącej asymetrii.

8. Wykres kurtozy



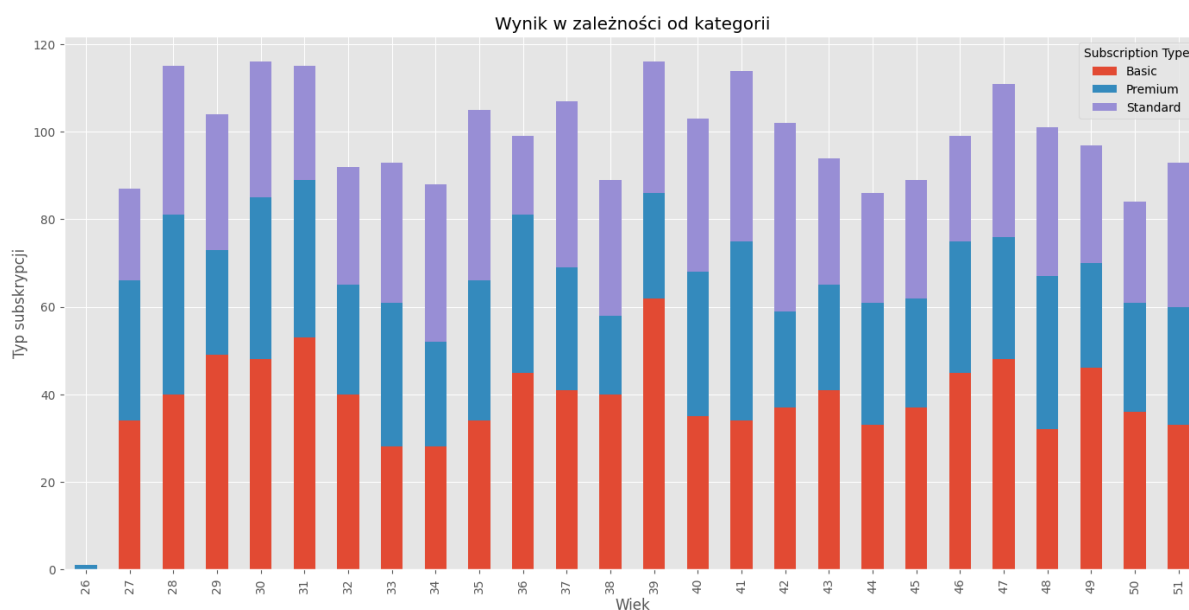
Rycina 8. Wykres kurtozy. Na osi poziomej wartość, a na osi pionowej częstotliwość.

Wartość kurtozy kolumny "Age" wynosząca około -1.2 sugeruje, że rozkład danych jest bardziej "płaski" niż rozkład normalny. Oznacza to, że dane w kolumnie "Age" mają ogony, które są mniej ekstremalne niż w rozkładzie normalnym.

Jeśli kurtoza wynosiłaby 0, to rozkład danych byłby dokładnie zgodny z rozkładem normalnym. Wartości kurtozy mniejsze niż 0 wskazują na bardziej "płaski" rozkład danych, co oznacza, że dane mają mniej ekstremalnych wartości odstających w porównaniu do rozkładu normalnego.

Podsumowując, wartość kurtozy sugeruje, że rozkład wieku użytkowników serwisu Netflix jest bardziej "płaski" niż rozkład normalny, co może oznaczać, że dane te mają mniej wartości odstających na skrajnych końcach rozkładu wieku.

9. Wykres słupkowy wyników testu statystycznego chi-kwadrat



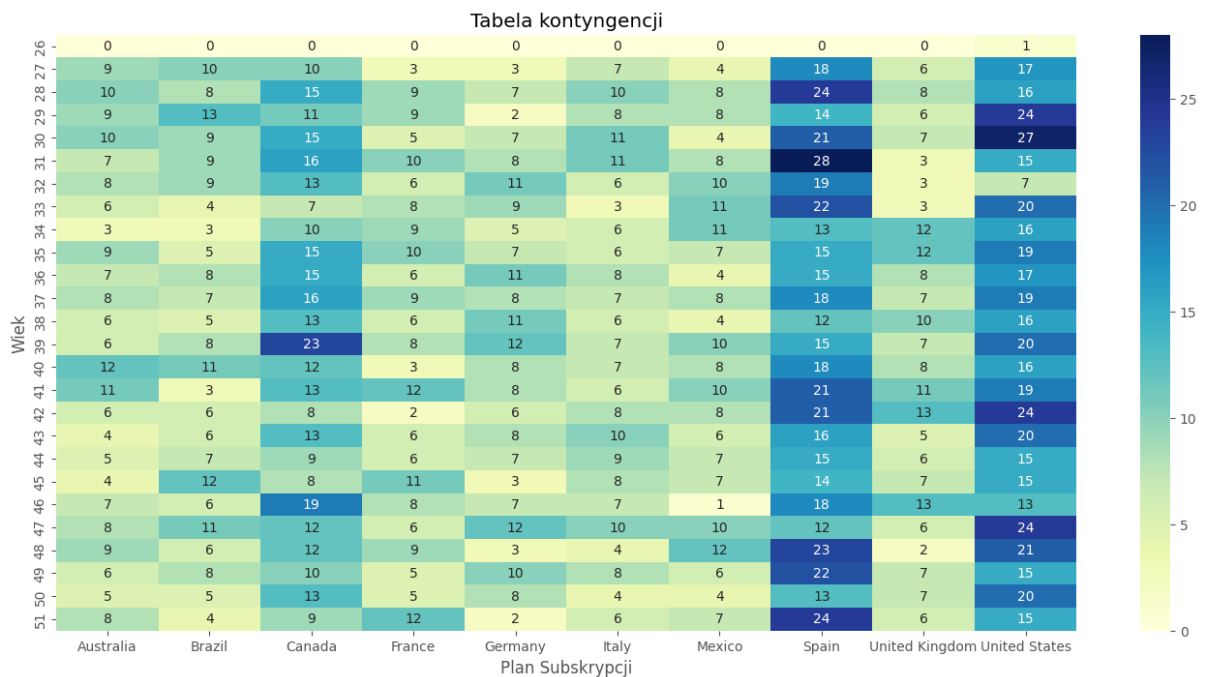
Rycina 9. Wykres słupkowy przedstawiający zależność między wiekiem a wybieranym typem subskrypcji. Na osi poziomej jest wartość wiekowa, a na osi pionowej Typ subskrypcji (Podstawowy-czerwony; Premium-niebieski; Standard-fioletowy)

Powyższy wykres słupkowy przedstawia wyniki testu statystycznego chi-kwadrat, który analizuje zależność między wiekiem użytkowników a wybieranym typem subskrypcji na platformie Netflix. Ten rodzaj wykresu jest używany w analizie statystycznej do oceny, czy istnieje zależność między dwiema zmiennymi kategorialnymi.

Wynik testu chi-kwadrat wynosi 73.13. Jest to miara, która ocenia, jak bardzo obserwowane dane różnią się od wartości oczekiwanych w przypadku braku zależności między zmiennymi. Wyższa wartość statystyki chi-kwadrat sugeruje większą odmienność między obserwowanymi a oczekiwanymi danymi. P-wartość wynosi 0.0181. Jest to poziom istotności testu, czyli prawdopodobieństwo, że zależność między wiekiem a typem subskrypcji jest statystycznie istotna. P-wartość poniżej 0.05 wskazuje na istotność statystyczną, co sugeruje, że istnieje zależność między tymi zmiennymi. Test chi-kwadrat zawsze ma określoną liczbę stopni swobody, która wynosi 50 w tym przypadku. Stopnie swobody odzwierciedlają liczbę niezależnych zmiennych w analizie.

Wyniki tego testu sugerują, że istnieje statystycznie istotna zależność między wiekiem użytkowników a wyborem typu subskrypcji na platformie Netflix. Innymi słowy, wiek użytkowników może wpływać na to, jaki rodzaj subskrypcji wybierają.

10. Wykres typu heatmap wyników testu chi-kwadrat



Rycina 10. Tabela kontyngencji. Na osi poziomej znajdują się różne rodzaje planów subskrypcji (Podstawowy, Standardowy, Premium) oraz kraje, w których mieszkają użytkownicy. Na osi pionowej znajdują się grupy wiekowe. Wartości w komórkach tabeli reprezentują ilość użytkowników, którzy należą do danej kombinacji planu subskrypcji “Premium”, kraju i grupy wiekowej

Powyższa tabela kontyngencji ilustruje zależność między planem subskrypcji, krajem zamieszkania użytkowników, a grupą wiekową. Tabela ta jest używana do analizy statystycznej i pozwala zrozumieć, jak te zmienne są ze sobą powiązane.

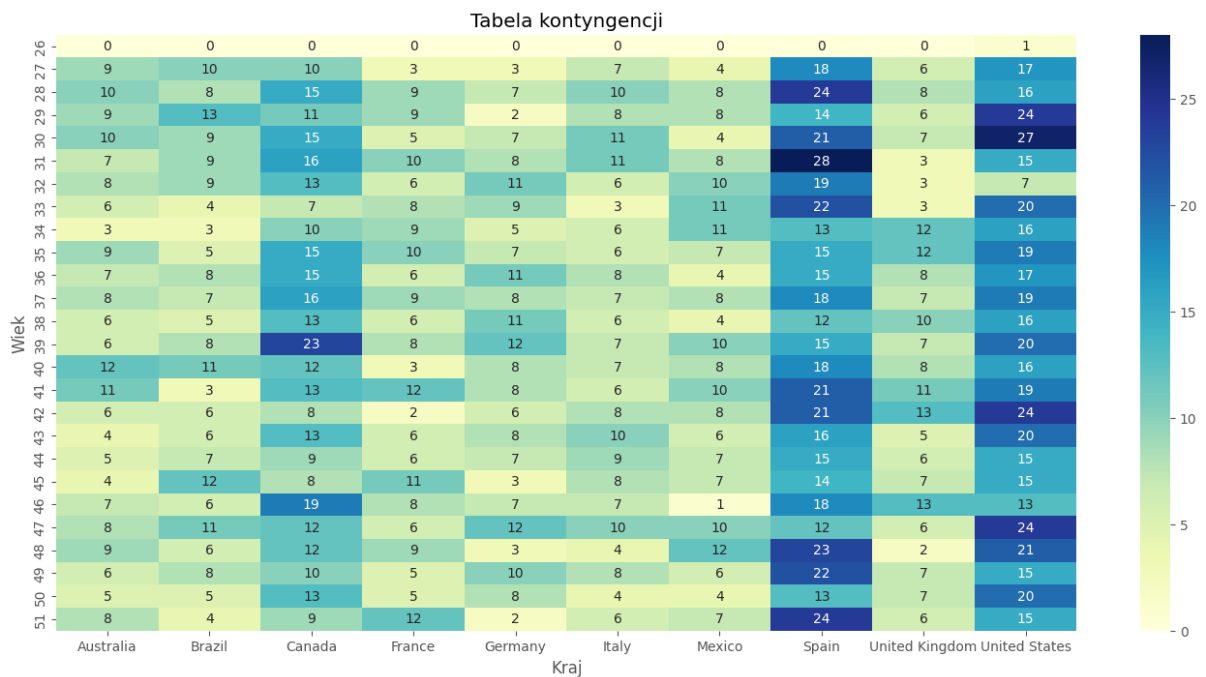
Każdy element w tabeli zawiera liczbę użytkowników, którzy należą do konkretnej kombinacji wieku, planu subskrypcji i kraju. Liczby te są wynikiem analizy danych i mogą pomóc zrozumieć, jakie tendencje występują wśród użytkowników Netflix w różnych grupach wiekowych i krajach. Tabela kontyngencji zawiera oczekiwane wartości (oczekiwane liczby użytkowników w każdej komórce), które byłyby uzyskane, gdyby nie istniała żadna zależność między planem subskrypcji, wiekiem i krajem zamieszkania użytkowników. Porównuje się te oczekiwane wartości z rzeczywistymi danymi, aby ocenić, czy istnieją różnice istotne statystycznie.

Wyniki testu chi-kwadrat (przedstawione również w rycinie 9) wskazują na to, że istnieje statystycznie istotna zależność między tymi zmiennymi. Heatmapa jest używana do wizualizacji tych wyników. Im bardziej komórki tabeli są wypełnione kolorem (w tym przypadku odcieniem niebieskiego), tym bardziej różnią się od oczekiwanych wartości, co może wskazywać na istnienie zależności.

Ogólnie rzecz biorąc, tabela kontyngencji i heatmapa pomagają w analizie danych i identyfikacji wzorców oraz zależności między zmiennymi kategorialnymi. W tym przypadku pokazuje ona, że wiek użytkowników i plan subskrypcji nie są niezależne od siebie i różnią się w zależności od kraju zamieszkania użytkowników Netflix.

Różnice w preferencjach subskrypcyjnych mogą wynikać z różnic kulturowych między krajami, jak również z ogólnych warunków ekonomicznych i sytuacji finansowej w danym regionie. Na przykład, w krajach o wyższym poziomie życia i stabilnej sytuacji ekonomicznej, użytkownicy mogą być bardziej skłonni do wyboru planów subskrypcji Premium, podczas gdy w krajach o niższym poziomie życia i mniejszych dochodach mogą bardziej preferować plany Podstawowe. Różnice w dostępie do technologii i zwyczajach kulturalnych mogą również wpływać na te preferencje.

11. Wykres typu heatmap wyników testu chi-kwadrat



Rycina 11. Tabela kontyngencji. Na osi poziomej znajdują się kraje, w których mieszkają użytkownicy. Na osi pionowej znajdują się grupy wiekowe. Wartości w poszczególnych komórkach tabeli przedstawiają liczbę użytkowników, którzy należą do konkretnej kombinacji kraju i grupy wiekowej.

Tabela kontyngencji oraz wyniki testu chi-kwadrat wskazują, że nie istnieje statystycznie istotna zależność między wiekiem użytkowników a krajem zamieszkania użytkowników Netflix. Wynik testu chi-kwadrat jest stosunkowo wysoki (Statystyka testu chi-kwadrat: 220.031), a p-wartość jest duża (P-wartość: 0.581), co oznacza brak istotności statystycznej. Stopnie swobody wynoszą 225.

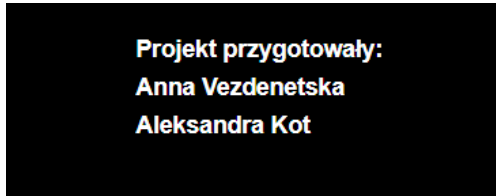
Wartości w komórkach tabeli kontyngencji i oczekiwane wartości są zbliżone, co wskazuje na brak znaczących odchyłeń od oczekiwanych wartości w wyniku analizy. To sugeruje, że wybór subskrypcji na Netflix nie jest znacząco związany z wiekiem użytkowników w badanych krajach.

Wnioskiem z tej analizy jest to, że preferencje subskrypcyjne użytkowników Netflix w tych krajach nie różnią się znacząco w zależności od wieku. Może to oznaczać, że oferta i dostępność planów subskrypcji są podobne dla różnych grup wiekowych w tych krajach, lub że inne czynniki, takie jak zainteresowania czy sytuacja finansowa, wpływają na wybór planu subskrypcji bardziej niż wiek.

12. Zakończ przegląd wykresów.

Ta opcja pozwala na zakończenie przeglądu wykresów i wyjście z programu. Jeśli już obejrzałeś wszystkie dostępne wykresy i analizy, możesz wybrać tę opcję, aby zakończyć działanie programu.

13. Autorki projektu.



Rycina 12. Pokazuje przykładowe okno, uzyskane po uruchomieniu [programu](#) za pośrednictwem Google Colab.

Ta opcja wyświetla okno z informacjami o autorach projektu. Możesz poznać imiona autorów oraz dowiedzieć się, kto pracował nad tym projektem. Jest to przydatne, jeśli chcesz dowiedzieć się więcej o osobach, które przygotowały ten program.

Podsumowanie raportu:

Raport rozpoczyna się od wprowadzenia, które przedstawia analizowany zbiór danych, zawierający informacje o użytkownikach serwisu Netflix, takie jak subskrypcje, dochody, wiek, płeć, urządzenia i wiele innych. Zbiór danych składa się z 2500 rekordów i 9 kolumn.

Następnie raport przegląda różne analizy danych, używając narzędzi takich jak wykresy, statystyki opisowe i testy statystyczne. Oto główne wnioski z analizy:

Demograficzny podział użytkowników: Wykres przedstawiający demograficzny podział użytkowników serwisu Netflix pokazuje, ile kobiet i mężczyzn korzysta z platformy oraz jakie grupy wiekowe są najbardziej aktywne. To może być przydatne do dostosowania treści na platformie do preferencji demograficznych.

Preferowane urządzenia: Analiza wieku użytkowników i rodzaju urządzenia, z którego korzystają, pokazuje, że laptopy są preferowanym urządzeniem, ale preferencje różnią się w zależności od grup wiekowych i kraju zamieszkania.

Zależność między krajem a planem subskrypcji: Wykresy ukazują różnice w preferencjach subskrypcyjnych między krajami. To może wynikać z różnic kulturowych i ekonomicznych między regionami.

Rozkład wieku użytkowników: Histogram wieku użytkowników pozwala zobaczyć, że wiele osób koncentruje się w określonych przedziałach wiekowych, co może pomóc w dostosowaniu treści do różnych grup wiekowych.

Analiza rozkładu normalnego: Wykresy Q-Q plot i wartości kwantyli sugerują, że rozkład wieku użytkowników jest zbliżony do rozkładu normalnego.

Skośność i kurtoza: Analiza skośności i kurtozy wskazuje, że rozkład wieku jest dość symetryczny i ma mniej wartości odstających niż rozkład normalny.

Zależność między wiekiem a typem subskrypcji: Test chi-kwadrat wykazuje istotną statystycznie zależność między wiekiem a wyborem typu subskrypcji. Preferencje subskrypcyjne różnią się w zależności od wieku.

Zależność między krajem a typem subskrypcji: Analiza testem chi-kwadrat i tabela kontyngencji pokazują, że istnieje zależność między krajem a typem subskrypcji użytkowników Netflix.

Brak zależności między wiekiem a krajem: Test chi-kwadrat i tabela kontyngencji nie wykazują statystycznie istotnej zależności między wiekiem a krajem zamieszkania użytkowników.

Podsumowując, raport dostarcza obszernej analizy danych użytkowników serwisu Netflix, uwzględniając wiek, kraj zamieszkania i preferencje subskrypcyjne. Analiza ta może być użyteczna dla firmy Netflix do lepszego dostosowania oferty treści i subskrypcji do różnych grup użytkowników.

Wnioski:

Na podstawie przeprowadzonej analizy danych użytkowników serwisu NETFLIX można wyciągnąć następujące wnioski:

- ➔ Preferencje subskrypcyjne różnią się między krajami, a także w zależności od wieku użytkowników.
- ➔ Wiek użytkowników może wpływać na wybór typu subskrypcji na platformie Netflix.
- ➔ Dane dotyczące wieku użytkowników Netflix są zbliżone do rozkładu normalnego, ale mają bardziej "płaski" rozkład niż rozkład normalny.
- ➔ Nie ma statystycznie istotnej zależności między wiekiem użytkowników a krajem zamieszkania użytkowników Netflix.
- ➔ Analiza danych może pomóc w dostosowaniu oferty treści i strategii biznesowej platformy do różnych grup wiekowych i kulturowych.