

SZKOŁA GŁÓWNA HANDLOWA W WARSZAWIE

Kolegium Analiz Ekonomicznych

**Studia Podyplomowe „Akademia Analityka z SAS, R & Python*”***

Analiza rankingu szczęśliwości dla lat 2015-2019 z wykorzystaniem programu Python.

Imię nazwisko Augustyn Olejniczak

Nr albumu 107132

Praca napisana pod kierunkiem dr inż. Adam Karwan

Warszawa 2021

Table of Contents

[Rozdział 1. Wstęp 3](#_Toc72335877)

[1.1. Charakterystyka World Happiness Report 3](#_Toc72335878)

[Rozdział 2. Analiza World Happiness Report 2015-2019 7](#_Toc72335879)

[2.1. World Happiness Report w ujęciu globalnym 7](#_Toc72335880)

[2.2. Charakterystyka czynników szczęśliwości w latach 2015-2019 16](#_Toc72335881)

[2.2. Model regresji logistycznej w oparciu o wskaźnik Produktu Krajowego Brutto 20](#_Toc72335882)

[Rozdział 3. Wnioski i podsumowanie 22](#_Toc72335883)

[Kod Python 24](#_Toc72335884)

[Bibliografia 34](#_Toc72335885)

[Spis wykresów 35](#_Toc72335886)

# Rozdział 1. Wstęp

## Charakterystyka World Happiness Report

Inicjatywa ONZ Sustainable Development Solutions Network (SDSN) została założona w 2012 roku. Nadzorowana jest przez Organizację Narodów Zjednoczonych. Uchwalona została przez Sekretarza Generalnego ONZ Ban Ki-moona, aby zmobilizować globalną wiedzę naukową i technologiczną w celu promowania praktycznego rozwiązywania problemów dla zrównoważonego rozwoju i realizacji celów zrównoważonego rozwoju (SDG). Po ich przyjęciu SDSN zobowiązuje się obecnie do wspierania realizacji celów zrównoważonego rozwoju na poziomie krajowym i międzynarodowym. SDSN ma na celu przyspieszenie wspólnego uczenia się i pomoc w przezwyciężeniu barier rozdziału prac technicznych i politycznych poprzez promowanie zintegrowanego podejścia do wzajemnie powiązanych wyzwań gospodarczych, społecznych i środowiskowych, przed którymi stoi świat. SDSN ściśle współpracuje z agencjami ONZ, wielostronnymi instytucjami finansowymi, sektorem prywatnym oraz społeczeństwem. SDSN umożliwia liderom ze wszystkich regionów uczestniczenie w rozwoju sieci, zapewniając jednocześnie skuteczne struktury podejmowania decyzji i odpowiedzialności [1].

Według statutu organizacji celem jest, aby dostarczać wysokiej jakości kontent dotyczący zrównoważonego rozwoju [2]. Głównymi adresatami są zarówno osoby uczące się, jak i praktycy. Publikowanie raportu ma na celu zwiększanie świadomości konieczności dbania o zrównoważony rozwój. Poszczególne kraje na świecie rozwijają się w różnym tempie. Samo pojęcie szczęśliwości jest subiektywne, jednak nie ulega wątpliwości, że niektóre kraje rozwijają się w sposób bardziej odpowiedzialny, ale też tworzą lepsze warunki do życia dla swoich mieszkańców. Dlatego też nie zawsze kraj, który ma lepszą ekonomię automatycznie ma wyższą pozycję w rankingu. Twórcy raportu wyodrębnili sześć wskaźników mających wpływ na poziom szczęśliwości. Są nimi:

* PKB (Produkt Krajowy Brutto) liczony na głowę jednego mieszkańca danego kraju.
* Poziom opieki społecznej.
* Średnia długość zdrowego życia.
* Poziom wolności i swobód obywatelski.
* Poziom dobroczynności.
* Poziom korupcji.

Produkt Krajowy Brutto charakteryzuje przede wszystkim stan bogactwa w danym kraju. Im wyższy ten współczynnik, tym większy jest dobrobyt w danym kraju (w wydźwięku ekonomicznym). Może to oznaczać na przykład, że godzina pracy fryzjera w Niemczech będzie zdecydowanie lepiej opłacana przez pracodawcę, niż na przykład na Ukrainie.

Poziom opieki społecznej odnosi się do wsparcia państwa dla swoich obywateli. Są to różnego rodzaju benefity i programy wsparcia. Warto zauważyć, że coraz więcej programów i form pomocy nie ma kryterium dochodowego. Jest to na przykład dodatek przysługujący dla rodziców, ale również (co jeszcze nie jest wprowadzone na szeroką skalę) dochód bezwarunkowy.

Średnia długość życia to czynnik o tyle istotny, że im wyższy poziom życia, wyższa jego jakość, tym ludzie prowadzą dłuższe, spokojniejsze życie. Kraje wysokorozwinięte cechuje wysoka średnia długość życia mieszkańców.

Poziom wolności i swobód obywatelskich to zerojedynkowa odpowiedź na pytanie „czy jest Pan\Pani zadowolony\zadowolona z poziomem swobody w podejmowaniu decyzji?”.

Następny czynnik, poziom dobroczynności, to chęć i aktywne partycypowanie w akcjach charytatywnych. Wszelkiego rodzaju zbiórki, jak i akcje dobroczynne, ale też istnienie różnego rodzaju fundacji mających na celu wspieranie osób w potrzebie to cecha szczególnie widoczna w krajach wysoko rozwiniętych.

Współczynnik korupcji oznacza subiektywny pogląd mieszkańców danego kraju względem zaufania do swoich polityków, a mianowicie do tego, czy według nich osoby decyzyjne w kraju są skorumpowane czy nie.

Oczywiste jest, że to, czy w danym państwie żyje się lepiej lub gorzej ma przede wszystkim wydźwięk subiektywny. Dla jednej osoby o wiele ważniejsze będzie to, że państwo posiada bogaty pakiet socjalny, który jest dla niego dostępny w trudnej sytuacji życiowej, a dla kogoś innego liczyć się będzie dobroczynność i dobre relacje z sąsiedztwem.

Preferencje co do wskaźników szczęśliwości zalezą również od perspektywy, z jakiej patrzy się na poziom szczęścia. Kraje skandynawskie, które od lat plasują się w czołówce rankingu charakteryzują się dość chłodnym klimatem. I choć według rankingu może żyć się tam najlepiej, to osoba, która preferuje wysoką temperaturę na zewnątrz mogłaby wcale nie czuć się szczęśliwa w tego typu kraju.

Kulturowość danego kraju może również mieć znaczny wpływ na postrzeganie ideałów szczęśliwości. To, że w danym kraju swoboda obywatelska może być znacznie ograniczona może nie mieć tak dużego wpływu na ogólne nastroje jego mieszkańców. Może to wynikać na przykład z religijności danego obywatela, który jest szczęśliwy, bez względu na to, w jakim stopniu ma tą swobodę ograniczoną.

W analizowanym okresie od 2015 do 2019 roku ranking Polski ulegał stałej poprawie, co jest przedstawione w następującej tabeli 1.

|  |  |  |
| --- | --- | --- |
| **Country** | **Ranking** | **Year** |
| Poland | 60 | 2015 |
| Poland | 57 | 2016 |
| Poland | 46 | 2017 |
| Poland | 42 | 2018 |
| Poland | 40 | 2019 |

Tabela 1 Ranking Polski w latach 2015-2019

Źródło: opracowanie własne na podstawie https://www.kaggle.com/unsdsn/world-happiness

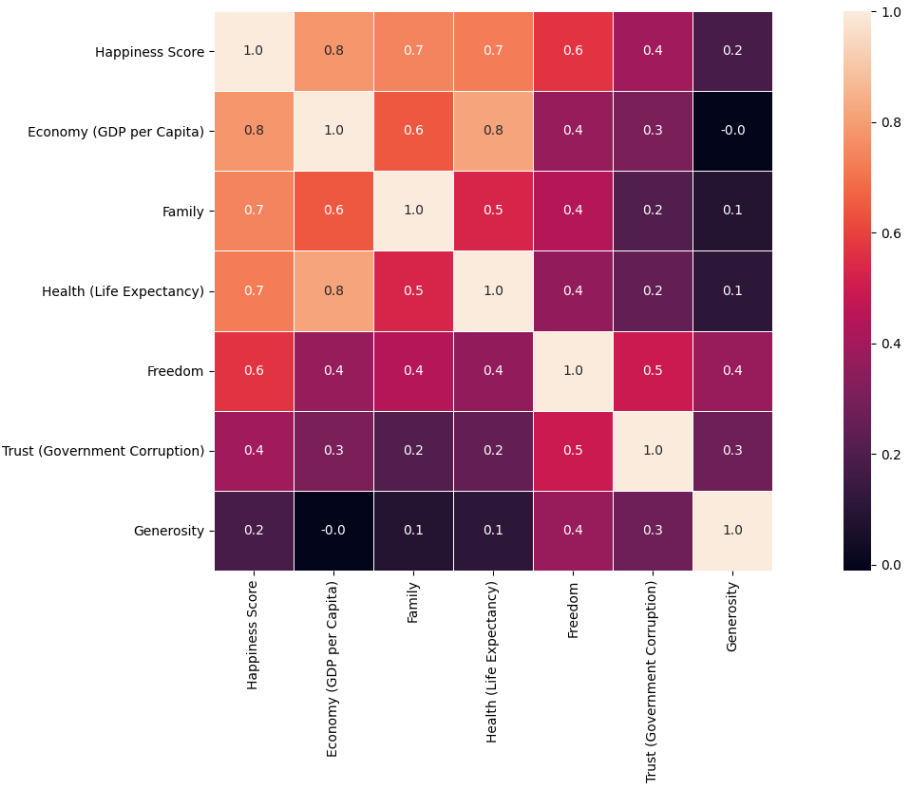
Zaczynając od roku 2020 trend dla Polski się odwrócił. Zaliczając spadek na miejsce 43, żeby w roku następnym, 2021, spaść o kolejną pozycję, plasując się na miejscu 44. Zauważalna jest analogia pomiędzy bogaceniem się kraju, a jednak zaliczeniem spadku w rankingu szczęśliwości. Pomimo, że sytuacja polskiej gospodarki uległa stosunkowo małemu spadkowi (względem innych państw na świecie podczas pandemii korona wirusa), to spadek zaufania do rządu oraz szereg innych czynników miały wpływ na to, że Polska odwróciła trend względem analizowanego w tej pracy okresu od 2015 do 2019 roku.

Ranking prezentowany przez Sustainable Development Solutions Network ma charakter poglądowy. To, że dany kraj jest na niskim lub wysokim miejscu nie musi mieć przełożenia na stan rzeczywisty. W kraju uplasowanym na wyższym miejscu faktyczny poziom szczęśliwości może być mniejszy niż w kraju, który został oceniony jako dużo niższy. Niemniej, ranking ma na celu zwiększyć świadomość odpowiedzialnego rozwoju, czyli takiego, który bierze pod uwagę czynniki ekonomiczne, ale też nie pomija dbania o dobro swoich obywateli, środowisko naturalne czy wielu innych czynników. Ma również wesprzeć środowisko akademickie dostępem do użytecznych informacji, ale też zwiększyć świadomość ludzi na problemy, z jakimi zmagają się kraje, które chcą rozwijać się w sposób odpowiedzialny[3].

# Rozdział 2. Analiza World Happiness Report 2015-2019

## 2.1. World Happiness Report w ujęciu globalnym

Na podstawie danych z lat 2015-2019 sporządzone zostały mapy pokazujące korelację pomiędzy poszczególnymi czynnikami wpływającymi na pozycję w rankingu. Rysunki 1 przedstawiają zależności pomiędzy nimi w roku 2015.



Wykres 1 Mapa korelacji 2015

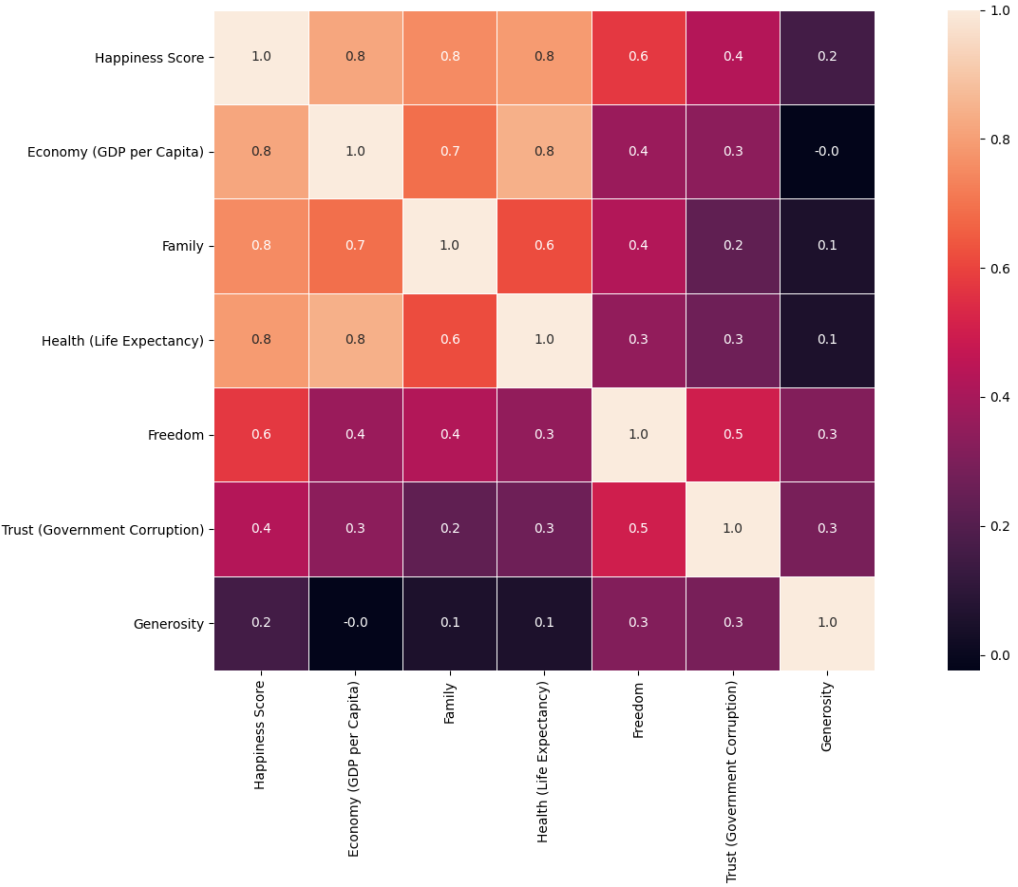
Źródło: opracowanie własne na podstawie https://www.kaggle.com/dhanyajothimani/basic-visualization-and-clustering-in-python

Największą korelacje pomiędzy wynikiem w rankingu a poszczególnymi czynnikami ma PKB na jednego mieszkańca. Wysoka korelacja jest zauważalna również pomiędzy średnią długością życia oraz Produktem Krajowym Brutto. Podobne wyniki otrzymano analizując mapy korelacji w latach 2016-2019, które są przedstawione na wykresach 2-5.



Wykres 2 Mapa korelacji 2016

Źródło: opracowanie własne na podstawie https://www.kaggle.com/dhanyajothimani/basic-visualization-and-clustering-in-python



Wykres 3 Mapa korelacji 2017

Źródło: opracowanie własne na podstawie https://www.kaggle.com/dhanyajothimani/basic-visualization-and-clustering-in-python



Wykres 4 Mapa korelacji 2018

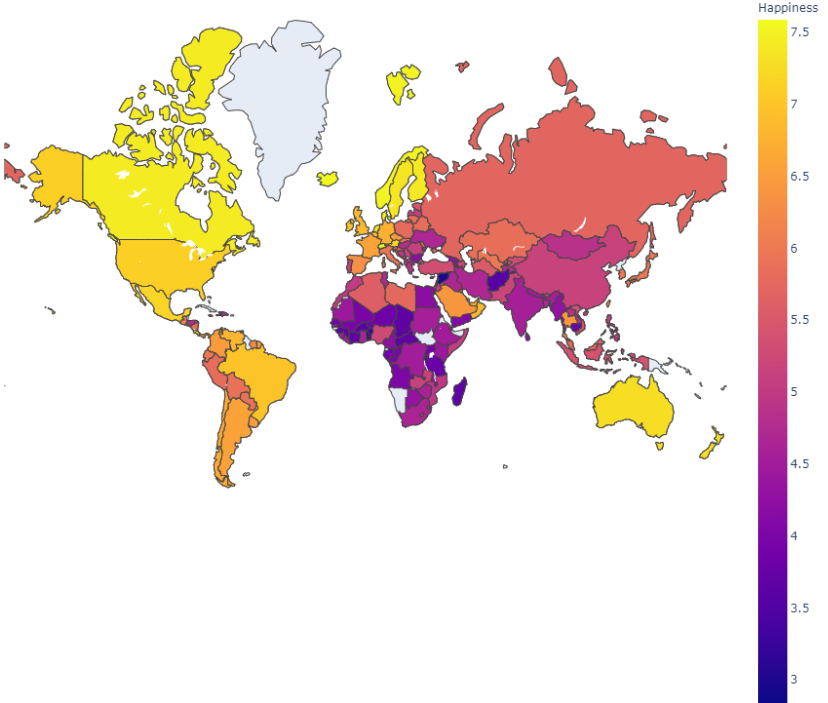
Źródło: opracowanie własne na podstawie https://www.kaggle.com/dhanyajothimani/basic-visualization-and-clustering-in-python



Wykres 5 Mapa korelacji 2019

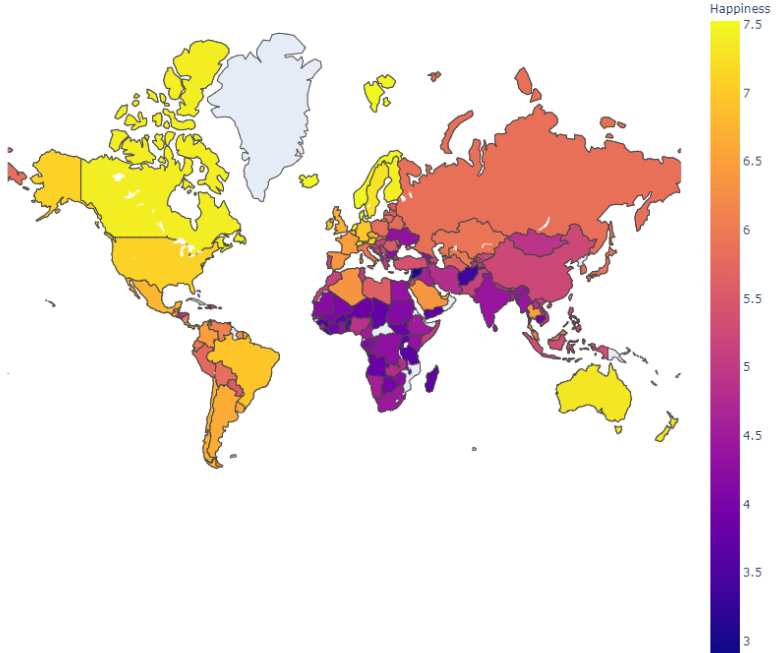
Źródło: opracowanie własne na podstawie https://www.kaggle.com/dhanyajothimani/basic-visualization-and-clustering-in-python

Kraje wysokorozwinięte, te z najwyższym wskaźnikiem PKB zajmują najwyższe pozycje w rankingu szczęśliwości. Z analizy map korelacji wynika, że im większe jest bogactwo danego kraju, tym stosunkowo wyższą pozycję otrzyma ono w rankingu. Przedstawiona zależność potwierdza również analiza poszczególnych krajów zobrazowana na mapie świata. Przedstawiają to wykresy 6-10.



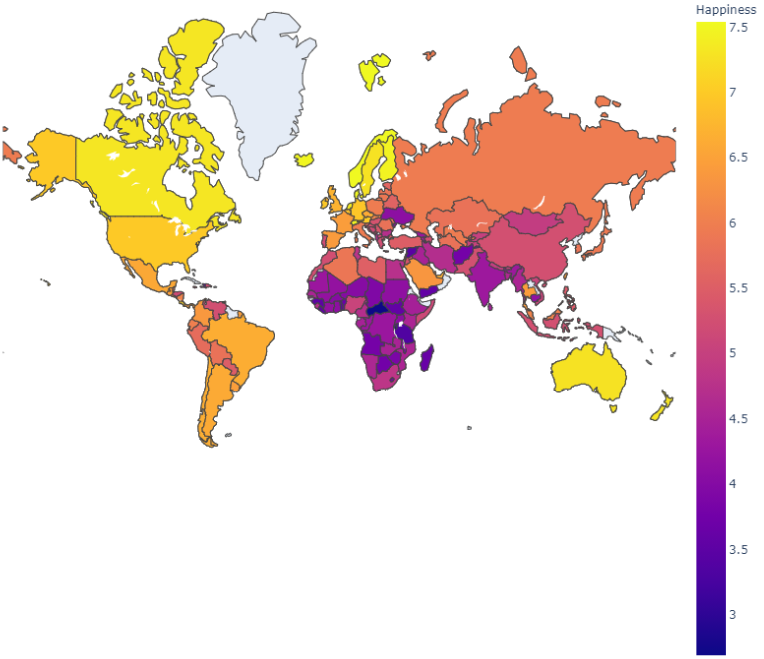
Wykres 6 Ranking poszczególnych krajów przedstawiony na mapie świata dla roku 2015

Źródło: opracowanie własne na podstawie https://www.kaggle.com/dhanyajothimani/basic-visualization-and-clustering-in-python



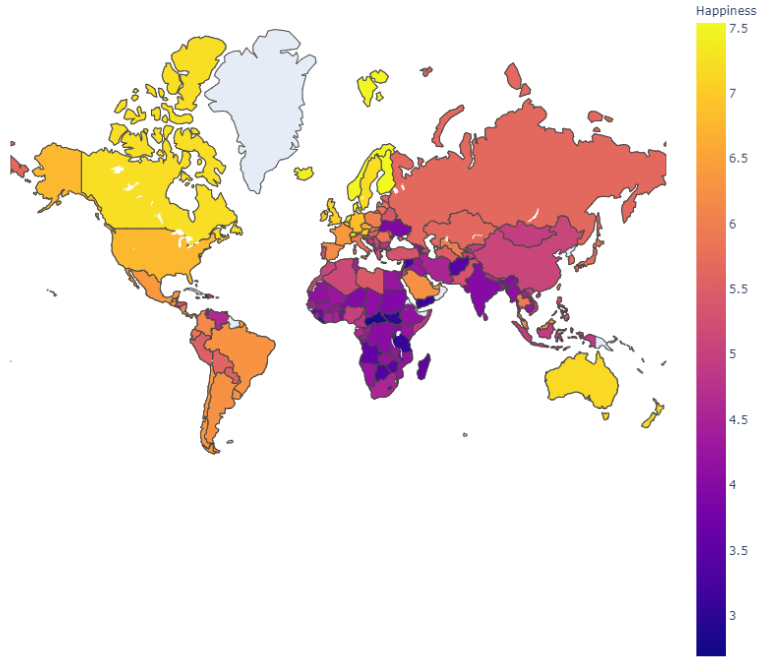
Wykres 7 Ranking poszczególnych krajów przedstawiony na mapie świata dla roku 2016

Źródło: opracowanie własne na podstawie https://www.kaggle.com/dhanyajothimani/basic-visualization-and-clustering-in-python



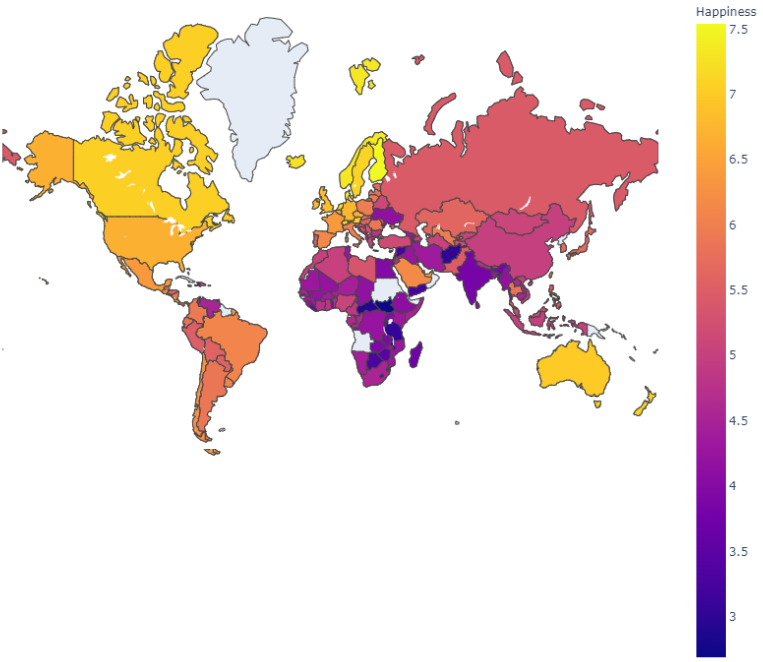
Wykres 8 Ranking poszczególnych krajów przedstawiony na mapie świata dla roku 2017

Źródło: opracowanie własne na podstawie https://www.kaggle.com/dhanyajothimani/basic-visualization-and-clustering-in-python



Wykres 9 Ranking poszczególnych krajów przedstawiony na mapie świata dla roku 2018

Źródło: opracowanie własne na podstawie https://www.kaggle.com/dhanyajothimani/basic-visualization-and-clustering-in-python

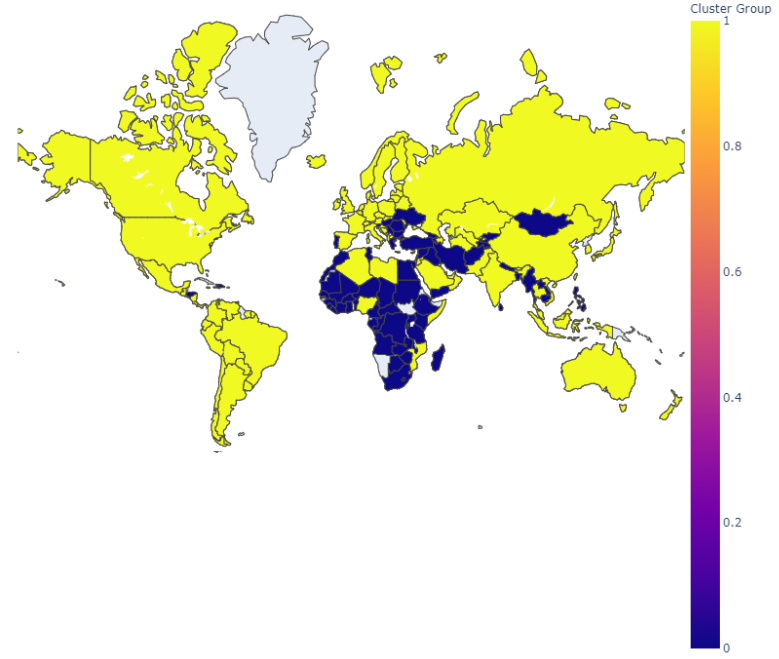


Wykres 10 Ranking poszczególnych krajów przedstawiony na mapie świata dla roku 2019

Źródło: opracowanie własne na podstawie https://www.kaggle.com/dhanyajothimani/basic-visualization-and-clustering-in-python

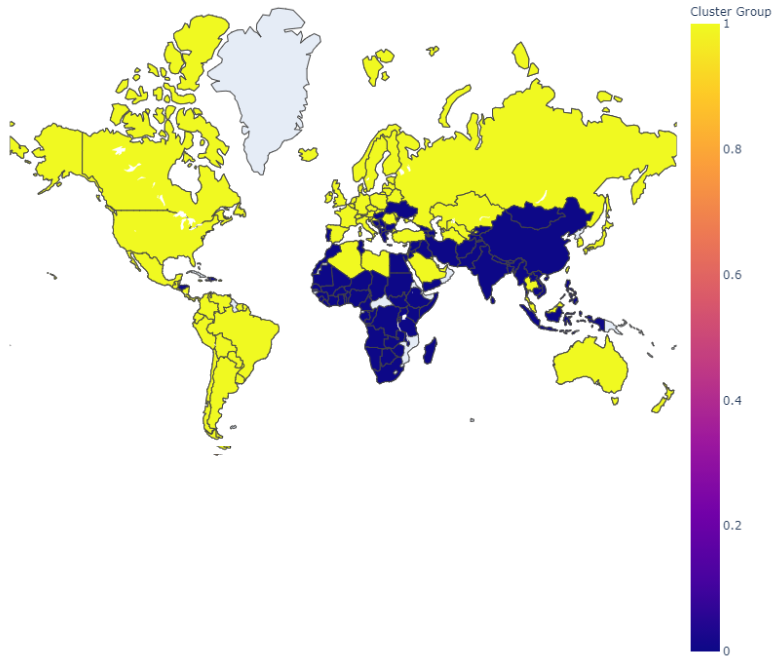
Kraje wysokorozwinięte, przede wszystkim te z Europy Zachodniej, kraje skandynawskie, Australia, Stany Zjednoczone, Kanada to niezmienna czołówka rankingu. Przedstawiona wcześniej zależność wpływu PKB na pozycje w rankingu potwierdzona jest poprzez analizę wykresów 6-10.

Dokładniejsze zobrazowanie podziału mapy świata uzyskano za pomocą algorytmu k-średnich. Kraje zajmujące wyższe miejsce w rankingu oznaczone są kolorem żółtym, zaś te zajmujące niższą pozycję w rankingu oznaczone są kolorem ciemnoniebieskim. Zobrazowany wynik przedstawiono na wykresach 11-15.



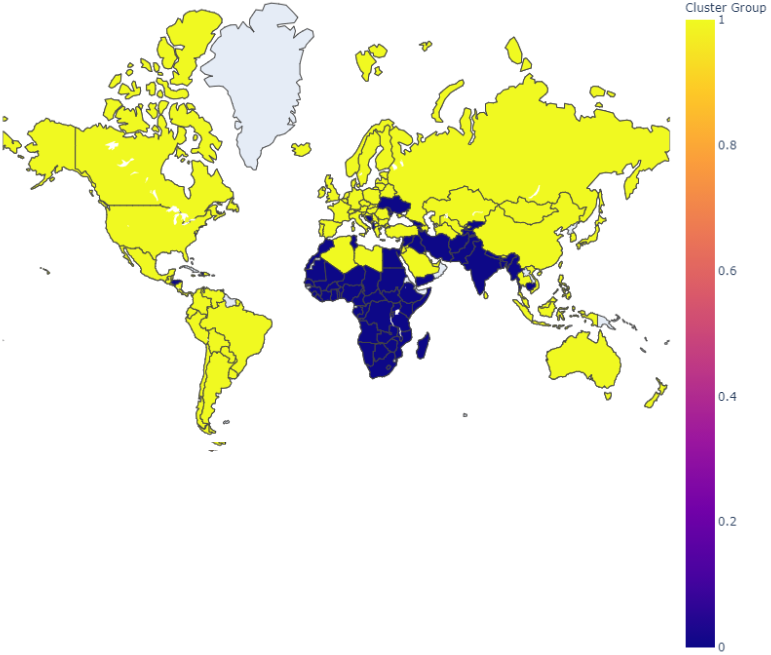
Wykres 11 Algorytm K-średnich przedstawiony na mapie świata dla danych z roku 2015

Źródło: opracowanie własne na podstawie https://www.kaggle.com/dhanyajothimani/basic-visualization-and-clustering-in-python



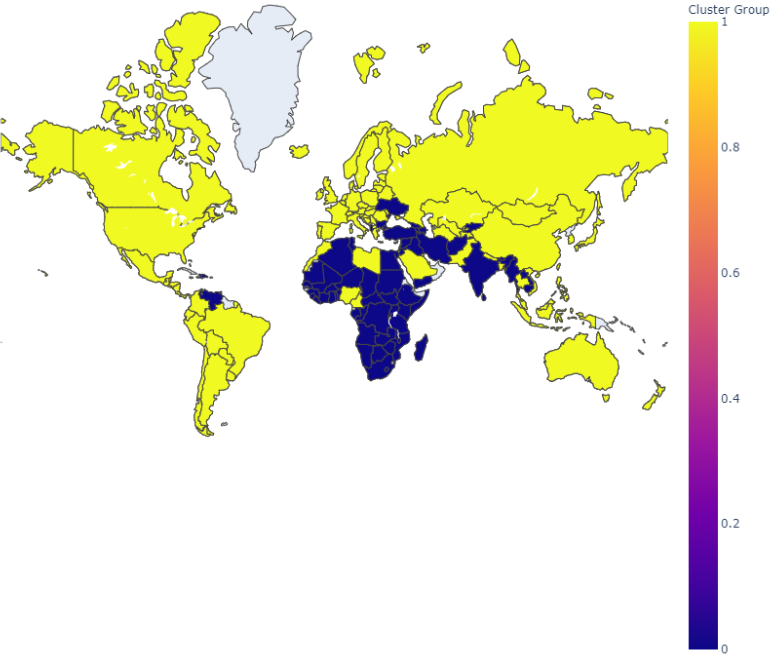
Wykres 12 Algorytm K-średnich przedstawiony na mapie świata dla danych z roku 2016

Źródło: opracowanie własne na podstawie https://www.kaggle.com/dhanyajothimani/basic-visualization-and-clustering-in-python



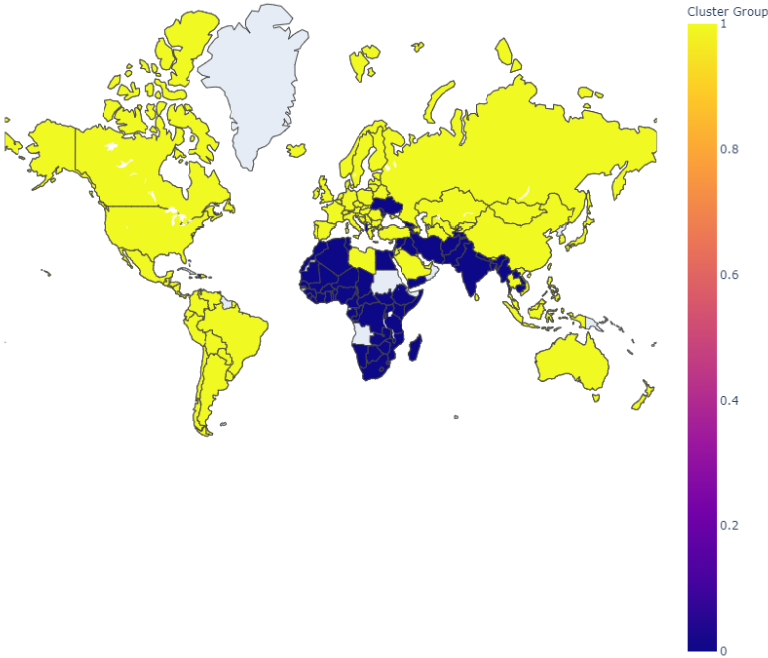
Wykres 13 Algorytm K-średnich przedstawiony na mapie świata dla danych z roku 2017

Źródło: opracowanie własne na podstawie https://www.kaggle.com/dhanyajothimani/basic-visualization-and-clustering-in-python



Wykres 14 Algorytm K-średnich przedstawiony na mapie świata dla danych z roku 2018

Źródło: opracowanie własne na podstawie https://www.kaggle.com/dhanyajothimani/basic-visualization-and-clustering-in-python



Wykres 15 Algorytm K-średnich przedstawiony na mapie świata dla danych z roku 2019

Źródło: opracowanie własne na podstawie https://www.kaggle.com/dhanyajothimani/basic-visualization-and-clustering-in-python

Analiza map po klasteryzacji, ale też obrazujących miejsce w rankingu światowym na przestrzeni lat wykazała, że nie ma większych roszad w rankingu krajów w okresie 2015 do 2019 roku. Nie oznacza to, że kraje przestały się rozwijać. Brak rozwoju skutkowałby natychmiastowym spadkiem w rankingu.

## 2.2. Charakterystyka czynników szczęśliwości w latach 2015-2019

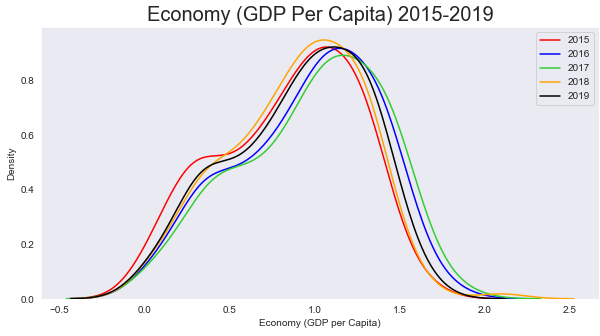
Czynniki decydujące o poziomie w rankingu charakteryzują się zmiennością na przestrzeni lat. PKB na głowę mieszkańca z roku na rok jest coraz większe dla przeważającej części krajów w rankingu. Niemniej istotność wpływu PKB na ranking jest zmienna na przestrzeni lat. Przedstawione jest to w tabeli 2.

|  |  |
| --- | --- |
| **Rok** | **Średnia wartość współczynnika PKB** |
| 2015 | 0.846137215 |
| 2016 | 0.953879809 |
| 2017 | 0.984718202 |
| 2018 | 0.891448718 |
| 2019 | 0.905147436 |

Tabela 2 Średnie wagi współczynnika PKB w latach 2015-2019

Źródło: opracowanie własne na podstawie <https://www.kaggle.com/unsdsn/world-happiness>

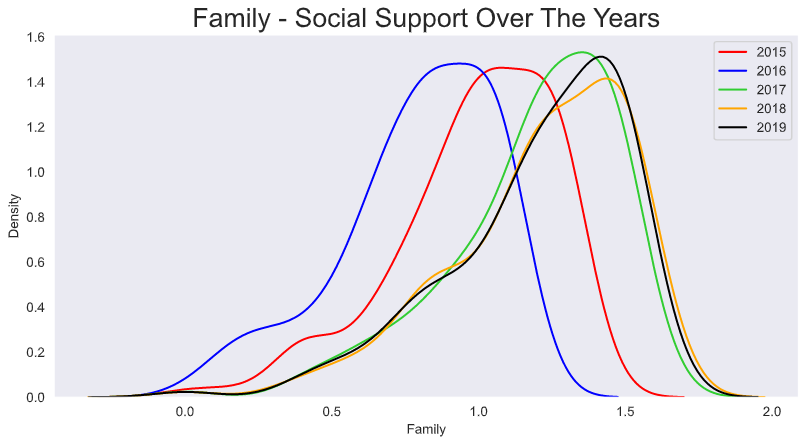
Największy wpływ Produktu Krajowego Brutto na ranking zaobserwowano w roku 2017-tym. Może to wynikać z faktu, że samo zwiększanie się bogactwa danego kraju nie jest najbardziej istotnym czynnikiem wpływającym na poziom szczęśliwości. Powyższa zależność potwierdza analiza współczynnika PKB przedstawiona na wykresie 16.



Wykres 16 Współczynnik PKB w latach 2015-2019

Źródło: opracowanie własne na podstawie https://www.kaggle.com/dhanyajothimani/basic-visualization-and-clustering-in-python

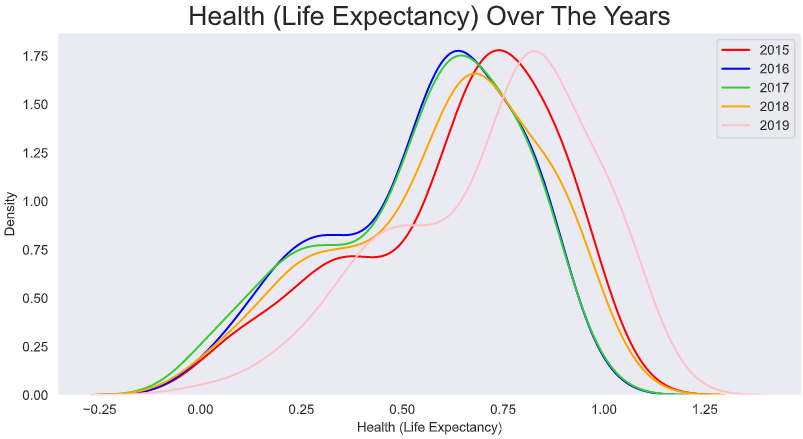
Pomimo faktu, iż w roku 2019-tym zaobserwowano największą jak dotąd liczbę krajów, dla których wartość współczynnika PKB zbliżała się do 1.5, to zmalała – względem lat poprzednich – średnia wartość współczynnika znajdującego się w zakresie 1.5 do 2.0. Może to wynikać z faktu, że zwiększyła się istotność pozostałych czynników w rankingu. Dokładniejsza analiza pozostałych czynników wykaże prawdziwość postawionej tezy.



Wykres 17 Współczynnik wsparcia dla rodzin w latach 2015-2019

Źródło: opracowanie własne na podstawie https://www.kaggle.com/dhanyajothimani/basic-visualization-and-clustering-in-python

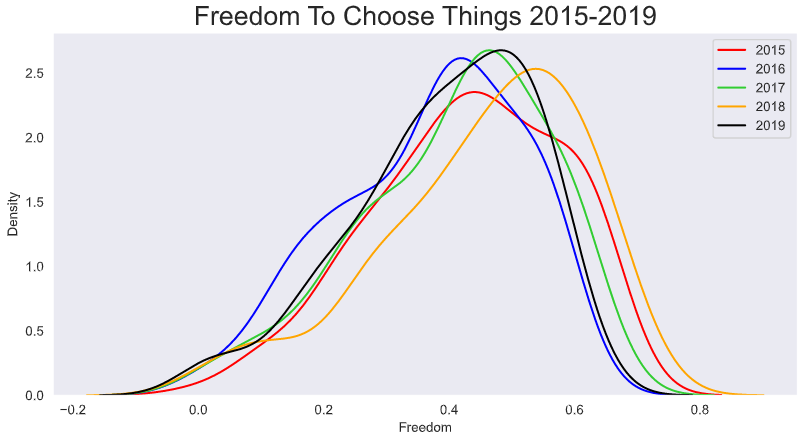
Pomimo, że poziom wsparcia dla rodzin jest ściśle skorelowany z poziomem bogactwa danego kraju (im większymi środkami kraj dysponuje, tym stosunkowo więcej pieniędzy może być przeznaczone na wsparcie socjalne), to zauważalny jest znaczny wzrost znaczenia tego czynnika na miejsce danego kraju w rankingu. W roku 2016-tym szczyt tego współczynnika plasował się na poziomie 0.8 – 1.2, zaś w roku 2019 urósł aż do poziomu 1.5 – 1.6.



Wykres 18 Współczynnik średniej długości życia w latach 2015-2019

Źródło: opracowanie własne na podstawie https://www.kaggle.com/dhanyajothimani/basic-visualization-and-clustering-in-python

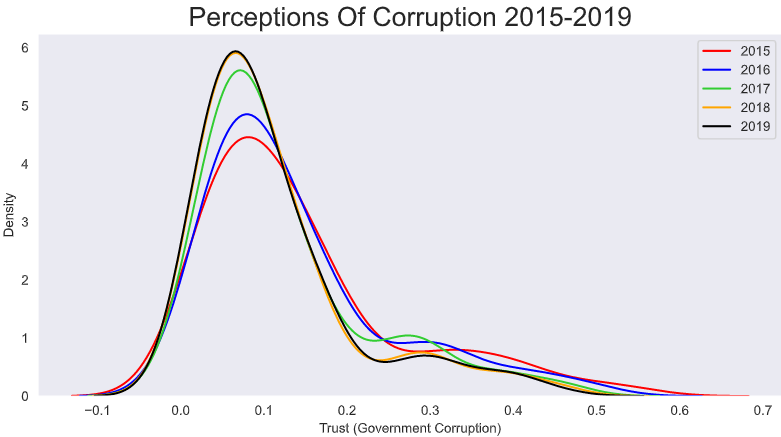
Współczynnik średniej długości życia miał najistotniejsze znaczenie w roku 2015-tym oraz 2019-tym. Im bardziej kraj jest rozwinięty, ma lepszą służbę zdrowia, tym długość życia jest dłuższa. Mniejsza waga w latach 2016-2018 mogła być skutkiem wzrostu istotność pozostałych czynników na ogólny wynik rankingu.



Wykres 19 Współczynnik swobody i wolności obywatelskiej w latach 2015-2019

Źródło: opracowanie własne na podstawie https://www.kaggle.com/dhanyajothimani/basic-visualization-and-clustering-in-python

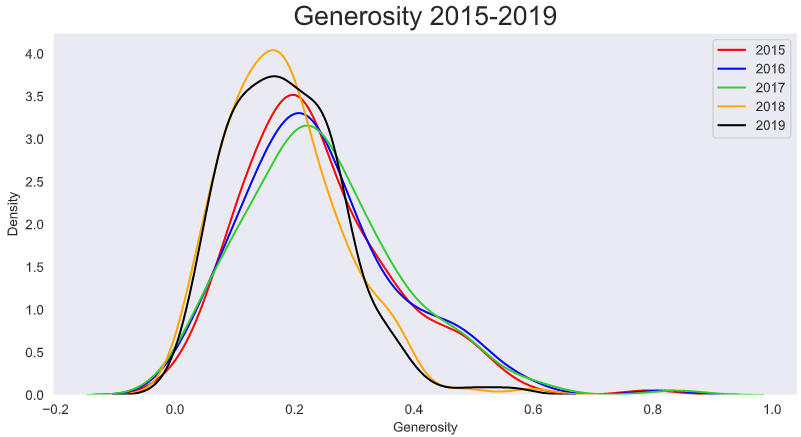
Od roku 2015 wzwyż można zauważyć znaczny wzrost istotności współczynnika wolności i swobód obywatelskich w rankingu. Biorąc pod uwagę spadek wpływu Produktu Krajowego Brutto na wysokość miejsca w rankingu istnieje zależność większego wpływu wolności w podejmowaniu decyzji na poziom szczęścia w danym kraju.



Wykres 20 Współczynnik korupcji w latach 2015-2019

Źródło: opracowanie własne na podstawie https://www.kaggle.com/dhanyajothimani/basic-visualization-and-clustering-in-python

Średni współczynnik korupcji zmalał w okresie 2015-2019. Niemniej urosło jego występowanie jako istotny czynnik wpływający na ranking. Może to oznaczać, że pomimo iż znaczenie tego współczynnika spadło, to wzrosło w krajach, w których w poprzednich latach nie miał on większego znaczenia.



Wykres 21 Współczynnik dobroczynności w latach 2015-2019

Źródło: opracowanie własne na podstawie https://www.kaggle.com/dhanyajothimani/basic-visualization-and-clustering-in-python

Istotność dobroczynności (chęci udziału w akcjach charytatywnych) spadła w perspektywie od 2015-tego do 2019-tego roku. Ten sam trend co dla współczynnika korupcji – pojawienie się go w większej ilości państw – jest zauważalny również w przypadku powyższego wskaźnika.

## 2.2. Model regresji logistycznej w oparciu o wskaźnik Produktu Krajowego Brutto

Produkt Krajowy Brutto w przeliczeniu na jednego mieszkańca jest najlepiej skorelowanym wskaźnikiem z uzyskanym wynikiem w rankingu. W oparciu o te wskaźniki zbudowany został model regresji logistycznej. Pod uwagę został wzięty nowo utworzony wskaźnik zerojedynkowy. Dla wyniku rankingu powyżej średniej przyjął wartość „1”, natomiast dla wartości poniżej średniej „0”. Dane które były modelowane to wskaźnik PKB. Dla uzyskania lepszego wyniku modelu posłużono się skalowaniem danych przy użyciu biblioteki Python sklearn, funkcja StandardScaler [4].

Uzyskany model charakteryzuje się następującym punktem przecięcia oraz nachyleniem:

model.intercept

array([-2.09695287])\_

model.coef\_

array([[3.9246237]])

Ocena zgodności modelu wynosi 0.76, co oznacza, że z 76-procentową dokładnością jest w stanie poprawnie przewidzieć wpływ Produktu Krajowego Brutto na miejsce w rankingu w 76 na 100 przypadków. Przedstawiona to poniższy zwrot:

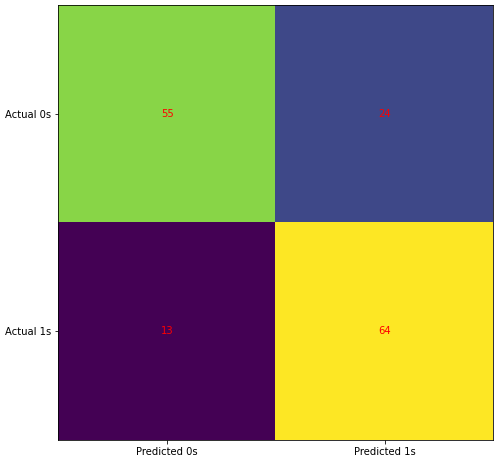
model.predict(x)

array([1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 0, 1, 1, 1, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 0, 1, 0, 1, 0, 1, 1, 1, 0, 1, 0, 1, 0, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 1, 1, 0, 1, 1, 1, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0], dtype=int64)

model.score(x, y)

0.7628205128205128

Dokładny rozkład błędnych oraz poprawnych klasyfikacji przedstawia macierz pomyłek. Prawdziwie pozytywne przypisania w modelu uzyskano dla 55 obserwacji, prawdziwie negatywne dla 64. Fałszywie negatywne przypisanie dotyczyło 13 obserwacji, a 24 zostały sklasyfikowane jako fałszywie pozytywne. Przedstawia to wykres 22.



Wykres 22 Macierz pomyłek dla modelu regresji logistycznej

Źródło: opracowanie własne na podstawie https://realpython.com/logistic-regression-python/ oraz https://www.kaggle.com/aysuncag/happiness-report-logistic-regression

Produkt Krajowy Brutto na jednego mieszkańca, pomimo, że posiada najwyższą korelację z wynikiem w rankingu, nie jest jedynym istotnym czynnikiem wpływającym na jego wynik. Potwierdzeniem tego jest fakt, że wiele krajów pomimo niższego wyniku wskaźnika PKB mają wyższy ranking, niż kraje od nich bogatsze. Raport szczęśliwości bierze pod uwagę wiele wskaźników, które mają zróżnicowany wpływ na jego strukturę.

# Rozdział 3. Wnioski i podsumowanie

Raport szczęśliwości jest próbą oceny zadowolenia z życia w danym kraju na świecie. Pod uwagę branych jest wiele czynników, których wpływ na ogólny wynik jest zróżnicowany. Przede wszystkim największa korelacja dla wpływu na wynik w rankingu przypada wskaźnikowi Produktu Krajowego Brutto na jednego mieszkańca. Wysokim poziomem skorelowania cechuje się również jego zależność względem w wsparcia dla rodzin. Nie zmienia to faktu, że każdy z analizowanych czynników ma wpływ na finalny wynik danego kraju. Wysoki wpływ PKB na wynik może świadczyć o tym, że ludzie przede wszystkim preferują godne życie. Niemniej z analizy pozostałych czynników dla lat 2015-2019 wynika, że u coraz większej liczby krajów zwiększyła się częstotliwość ich kluczowości. Istotność czynników różnych od wskazującego na poziom bogactwa danego kraju jest wysoka. Powodem tego, że zostały włączone do rankingu jest to, że znacząco wpływają one na poziom szczęśliwości obywateli danego kraju. Im wyższy poziom rozwinięcia danego kraju tym wpływ czynników innych od bezpośrednio związanych z poziomem bogactwa jest większy. Może to wynikać z faktu, że gdy zaspokojone są podstawowe, kluczowe potrzeby, to dopiero wtedy zwiększa się istotność innych czynników. Oczywiście nie jest tak, że póki podstawowe potrzeby nie są zaspokojone to inne czynniki nie mają jakiegokolwiek wpływu na poziom szczęśliwości. Niemniej, biorąc pod uwagę sytuację, w której dany obywatel jest biedny, znajduje się np. poniżej progu ubóstwa, to bez wątpliwości jest to, że ważniejsze będzie dla niego polepszenie sytuacji finansowej, otrzymanie wsparcia od państwa, niż to, że będzie mieszkał w kraju, w którym stosunkowo duży odsetek ludzi wspiera akcje charytatywne.

Struktura rankingu nie ulega znaczącym zmianom z roku na rok. Faktem jest, że wynik poszczególnego kraju może różnić się w mniejszym lub znaczniejszym stopniu, niemniej, wynik regionów pozostaje praktycznie bez zmian. Zobrazowane jest to na wykresach mapowych, gdzie zachodnia Europa, północna część Ameryki, Australia i kilka innych regionów niezmiennie plasuje się w górnej części rankingu. Szacuje się, że dysproporcja pomiędzy krajami bogatymi a biednymi będzie wzrastać. Na podstawie analizy przeprowadzonej w pracy zakłada się, że istotność czynników innych od tych bezpośrednio związanych z bogactwem danego kraju będzie wzrastać. Dla krajów wysokorozwiniętych, aspirujących na zdobycie szczytu tabeli rankingu szczęśliwości kluczowe będzie, aby wszystkie czynniki w rankingu ulegały poprawie. Przykładowo, nie jest możliwe, aby szczęśliwość w danym kraju cechowała się jedynie wysokim poziomem bogactwa albo niskim poziomem korupcji.

Pomimo, że ranking Polski w rankingu w analizowanym okresie ulegał stałej poprawie, to w najnowszym rankingu spadł aż o pięć pozycji (względem roku 2019-tego). Przypadek tego kraju obrazuje istotność czynników nie związanych z bogactwem danego kraju. Polska gospodarka w czasie pandemii – w porównaniu do innych krajów europejskich – zaliczyła niski spadek, co, gdyby pod uwagę były brane jedynie czynniki ekonomiczne, powinno skutkować awansem o kilka pozycji, a już na pewno nie spadkiem w rankingu.

Pandemia COVID-19 może być katalizatorem dla wzrostu istotności dbania o planetę. Postęp technologiczny, poprawa stanu ekonomicznego w danym kraju jest ważna, niemniej szacuje się, że ostatni rok, podczas którego ludzkość zmagała się z wirusem SARS-COV-2 przyczyni się do większego zainteresowania tematami ekologii, minimalizowania negatywnego wpływu działalności człowieka oraz większego stopnia solidarności na arenie międzynarodowej.

# Kod Python

from chart\_studio.plotly import plot, iplot as py

import plotly.graph\_objects as go

import seaborn as sns

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

from plotly.offline import download\_plotlyjs, init\_notebook\_mode, plot, iplot

import time

import warnings

from sklearn.preprocessing import StandardScaler

from sklearn.cluster import KMeans, AgglomerativeClustering, AffinityPropagation

from sklearn.mixture import GaussianMixture

import os

import sys

from plotly.offline import iplot

import os

sns.set\_style('dark')

os.getcwd()

os.chdir(r"C:\Users\augus\OneDrive\Desktop\SGH Warszawa\Praca zaliczeniowa 2\Data preparation and graphs")

HappyReport2015 = pd.read\_csv("2015.csv")

HappyReport2016 = pd.read\_csv("2016.csv")

HappyReport2017 = pd.read\_csv("2017.csv")

HappyReport2018 = pd.read\_csv("2018.csv")

HappyReport2019 = pd.read\_csv("2019.csv")

PolandShort = happy1519[happy1519["Country"].str.contains("Poland")]

Poland = PolandShort[["Country", "Happiness Rank", "Year"]]

Poland

#Import Data Sets

os.getcwd()

os.chdir(r"C:\Users\augus\OneDrive\Desktop\SGH Warszawa\Praca zaliczeniowa 2\World Happines Report Dataset")

print("Current Working Directory " , os.getcwd())

WH\_2015 = pd.read\_csv(r"C:\Users\augus\OneDrive\Desktop\SGH Warszawa\Praca zaliczeniowa 2\World Happines Report Dataset\2015.csv", lineterminator='\r')

WH\_2016 = pd.read\_csv(r"C:\Users\augus\OneDrive\Desktop\SGH Warszawa\Praca zaliczeniowa 2\World Happines Report Dataset\2016.csv", lineterminator='\r')

WH\_2017 = pd.read\_csv(r"C:\Users\augus\OneDrive\Desktop\SGH Warszawa\Praca zaliczeniowa 2\World Happines Report Dataset\2017.csv", lineterminator='\r')

WH\_2018 = pd.read\_csv(r"C:\Users\augus\OneDrive\Desktop\SGH Warszawa\Praca zaliczeniowa 2\World Happines Report Dataset\2018.csv", lineterminator='\r')

WH\_2019 = pd.read\_csv(r"C:\Users\augus\OneDrive\Desktop\SGH Warszawa\Praca zaliczeniowa 2\World Happines Report Dataset\2019.csv", lineterminator='\r')

WH\_2015.describe()

WH\_2016.describe()

WH\_2017.describe()

WH\_2017.describe()

WH\_2018.describe()

WH\_2019.describe()

#Happiness Rank in 2017 data = object. The following is to convert it to float64

WH\_2017["Happiness Rank"] = pd.to\_numeric(WH\_2017["Happiness Rank"], errors='coerce')

print("Dimension of dataset: wh.shape")

WH\_2015.dtypes

WH\_2016.dtypes

WH\_2017.dtypes

WH\_2018.dtypes

WH\_2019.dtypes

WH\_2015.head(0)

WH\_2016.head(0)

WH\_2017.head(1)

WH\_2018.head(0)

WH\_2019.head(0)

#Generating single Heat Maps

HeatMap2015 = WH\_2015[['Happiness Score', 'Economy (GDP per Capita)', 'Family', 'Health (Life Expectancy)', 'Freedom', 'Trust (Government Corruption)', 'Generosity']] #Subsetting the data

cor2015 = HeatMap2015.corr() #Calculate the correlation of the above variables

sns.heatmap(cor2015, square = True, linewidths=.5) #Plot the correlation as heat map

HM2015 = sns.heatmap(cor2015, square = True, linewidths= 0.5, annot = True, cbar = False, fmt=".1f")

plt.show()

HeatMap2016 = WH\_2016[['Happiness Score', 'Economy (GDP per Capita)', 'Family', 'Health (Life Expectancy)', 'Freedom', 'Trust (Government Corruption)', 'Generosity']] #Subsetting the data

cor2016 = HeatMap2016.corr() #Calculate the correlation of the above variables

sns.heatmap(cor2016, square = True, linewidths=.5) #Plot the correlation as heat map

HM2016 = sns.heatmap(cor2016, square = True, linewidths= 0.5, annot = True, cbar = False, fmt=".1f")

plt.show()

HeatMap2017 = WH\_2017[['Happiness Score', 'Economy (GDP per Capita)', 'Family', 'Health (Life Expectancy)', 'Freedom', 'Trust (Government Corruption)', 'Generosity']] #Subsetting the data

cor2017 = HeatMap2017.corr() #Calculate the correlation of the above variables

sns.heatmap(cor2017, square = True, linewidths=.5) #Plot the correlation as heat map

HM2017 = sns.heatmap(cor2017, square = True, linewidths= 0.5, annot = True, cbar = False, fmt=".1f")

plt.show()

HeatMap2018 = WH\_2018[['Happiness Score', 'Economy (GDP per Capita)', 'Family', 'Health (Life Expectancy)', 'Freedom', 'Trust (Government Corruption)', 'Generosity']] #Subsetting the data

cor2018 = HeatMap2018.corr() #Calculate the correlation of the above variables

sns.heatmap(cor2018, square = True, linewidths=.5) #Plot the correlation as heat map

HM2018 = sns.heatmap(cor2018, square = True, linewidths= 0.5, annot = True, cbar = False, fmt=".1f")

plt.show()

HeatMap2019 = WH\_2019[['Happiness Score', 'Economy (GDP per Capita)', 'Family', 'Health (Life Expectancy)', 'Freedom', 'Trust (Government Corruption)', 'Generosity']] #Subsetting the data

cor2019 = HeatMap2019.corr() #Calculate the correlation of the above variables

sns.heatmap(cor2019, square = True, linewidths=.5) #Plot the correlation as heat map

HM2019 = sns.heatmap(cor2019, square = True, linewidths= 0.5, annot = True, cbar = False, fmt=".1f")

plt.show()

#Creating World Maps

data2015 = dict(type = 'choropleth',

           locations = WH\_2015['Country'],

           locationmode = 'country names',

           z = WH\_2015['Happiness Score'],

           text = WH\_2015['Country'],

           colorbar = {'title':'Happiness'})

layout2015 = dict(title = 'Happiness Index 2015',

             geo = dict(showframe = False,

                       projection = {'type': 'mercator'}))

choromap2015 = go.Figure(data = [data2015], layout=layout2015)

plot(choromap2015)

data2016 = dict(type = 'choropleth',

           locations = WH\_2016['Country'],

           locationmode = 'country names',

           z = WH\_2016['Happiness Score'],

           text = WH\_2016['Country'],

           colorbar = {'title':'Happiness'})

layout2016 = dict(title = 'Happiness Index 2016',

             geo = dict(showframe = False,

                       projection = {'type': 'mercator'}))

choromap2016 = go.Figure(data = [data2016], layout=layout2016)

plot(choromap2016)

data2017 = dict(type = 'choropleth',

           locations = WH\_2017['Country'],

           locationmode = 'country names',

           z = WH\_2017['Happiness Score'],

           text = WH\_2017['Country'],

           colorbar = {'title':'Happiness'})

layout2017 = dict(title = 'Happiness Index 2017',

             geo = dict(showframe = False,

                       projection = {'type': 'mercator'}))

choromap2017 = go.Figure(data = [data2017], layout=layout2017)

plot(choromap2017)

data2018 = dict(type = 'choropleth',

           locations = WH\_2018['Country'],

           locationmode = 'country names',

           z = WH\_2018['Happiness Score'],

           text = WH\_2018['Country'],

           colorbar = {'title':'Happiness'})

layout2018 = dict(title = 'Happiness Index 2018',

             geo = dict(showframe = False,

                       projection = {'type': 'mercator'}))

choromap2018 = go.Figure(data = [data2018], layout=layout2018)

plot(choromap2018)

data2019 = dict(type = 'choropleth',

           locations = WH\_2019['Country'],

           locationmode = 'country names',

           z = WH\_2019['Happiness Score'],

           text = WH\_2019['Country'],

           colorbar = {'title':'Happiness'})

layout2019 = dict(title = 'Happiness Index 2019',

             geo = dict(showframe = False,

                       projection = {'type': 'mercator'}))

choromap2019 = go.Figure(data = [data2019], layout=layout2019)

plot(choromap2019)

#Clustering

ss.fit\_transform(HeatMap2015)

ss = StandardScaler()

C\_2015 = ss.fit\_transform(HeatMap2015)

C\_2016 = ss.fit\_transform(HeatMap2016)

C\_2017 = ss.fit\_transform(HeatMap2017)

C\_2018 = ss.fit\_transform(HeatMap2018)

C\_2019 = ss.fit\_transform(HeatMap2019)

x = pd.DataFrame(C\_2015)

HeatMap2015.dropna()

#Plot the clusters obtained using k means

fig = plt.figure()

ax = fig.add\_subplot(111)

scatter = ax.scatter(HeatMap2015['Economy (GDP per Capita)'],HeatMap2015['Trust (Government Corruption)'],

                     c=kmeans2015[0],s=50)

ax.set\_title('K-Means Clustering')

ax.set\_xlabel('GDP per Capita')

ax.set\_ylabel('Corruption')

plt.colorbar(scatter)

plt.show()

fig = plt.figure()

ax = fig.add\_subplot(111)

scatter = ax.scatter(ClusterDropNA2016['Economy (GDP per Capita)'],ClusterDropNA2016['Trust (Government Corruption)'],

                     c=kmeans[0],s=50)

ax.set\_title('K-Means Clustering 2016')

ax.set\_xlabel('GDP per Capita')

ax.set\_ylabel('Corruption')

plt.colorbar(scatter)

plt.show()

fig = plt.figure()

ax = fig.add\_subplot(111)

scatter = ax.scatter(ClusterDropNA2017['Economy (GDP per Capita)'],ClusterDropNA2017['Trust (Government Corruption)'],

                     c=kmeans[0],s=50)

ax.set\_title('K-Means Clustering 2017')

ax.set\_xlabel('GDP per Capita')

ax.set\_ylabel('Corruption')

plt.colorbar(scatter)

plt.show()

fig = plt.figure()

ax = fig.add\_subplot(111)

scatter = ax.scatter(ClusterDropNA2018['Economy (GDP per Capita)'],ClusterDropNA2018['Trust (Government Corruption)'],

                     c=kmeans[0],s=50)

ax.set\_title('K-Means Clustering 2018')

ax.set\_xlabel('GDP per Capita')

ax.set\_ylabel('Corruption')

plt.colorbar(scatter)

plt.show()

fig = plt.figure()

ax = fig.add\_subplot(111)

scatter = ax.scatter(ClusterDropNA2019['Economy (GDP per Capita)'],ClusterDropNA2019['Trust (Government Corruption)'],

                     c=kmeans[0],s=50)

ax.set\_title('K-Means Clustering 2019')

ax.set\_xlabel('GDP per Capita')

ax.set\_ylabel('Corruption')

plt.colorbar(scatter)

plt.show()

print("Średnia wartość PKB w roku 2015:", np.mean(HappyReport2015["Economy (GDP per Capita)"]))

print("Średnia wartość PKB w roku 2016:", np.mean(HappyReport2016["Economy (GDP per Capita)"]))

print("Średnia wartość PKB w roku 2017:", np.mean(HappyReport2017["Economy (GDP per Capita)"]))

print("Średnia wartość PKB w roku 2018:", np.mean(HappyReport2018["Economy (GDP per Capita)"]))

print("Średnia wartość PKB w roku 2019:", np.mean(HappyReport2019["Economy (GDP per Capita)"]))

HappyReport2015

HappyReport2016

HappyReport2017 = HappyReport2017.rename(columns=({'Happiness.Rank':'Happiness Rank',

'Happiness.Score':'Happiness Score','Economy..GDP.per.Capita.':'Economy (GDP per Capita)',

'Health..Life.Expectancy.':'Health (Life Expectancy)',

'Trust..Government.Corruption.':'Trust (Government Corruption)',

'Dystopia.Residual':'Dystopia Residual','Whisker.high':'Whisker high',

'Whisker.low':'Whisker low'}))

HappyReport2018 = HappyReport2018.rename(columns=({'Overall rank':'Happiness Rank','Score':'Happiness Score',

 'Country or region':'Country', 'GDP per capita':'Economy (GDP per Capita)',

 'Healthy life expectancy':'Health (Life Expectancy)','Freedom to make life choices':'Freedom',

 'Perceptions of corruption':'Trust (Government Corruption)',"Social support":'Family'}))

HappyReport2019 = HappyReport2019.rename(columns=({'Overall rank':'Happiness Rank','Score':'Happiness Score',

 'Country or region':'Country', 'GDP per capita':'Economy (GDP per Capita)',

 'Healthy life expectancy':'Health (Life Expectancy)','Freedom to make life choices':'Freedom',

 'Perceptions of corruption':'Trust (Government Corruption)',"Social support":"Family"}))

HappyReport2018.info()

#2018 has null value in column = "Perceptions of corruption"

#Happines

plt.figure(figsize =(10,5))

sns.kdeplot(HappyReport2015['Happiness Score'], color='red', label='2015')

sns.kdeplot(HappyReport2016['Happiness Score'], color='blue', label='2016')

sns.kdeplot(HappyReport2017['Happiness Score'], color='limegreen', label='2017')

sns.kdeplot(HappyReport2018['Happiness Score'], color='orange', label='2018')

sns.kdeplot(HappyReport2019['Happiness Score'], color='pink', label='2019')

plt.title('Happiness 2015-2019', size=25)

plt.legend()

plt.show()

plt.figure(figsize=(10,5))

sns.kdeplot(HappyReport2015['Economy (GDP per Capita)'],color='red', label="2015")

sns.kdeplot(HappyReport2016['Economy (GDP per Capita)'],color='blue', label="2016")

sns.kdeplot(HappyReport2017['Economy (GDP per Capita)'],color='limegreen', label="2017")

sns.kdeplot(HappyReport2018['Economy (GDP per Capita)'],color='orange', label="2018")

sns.kdeplot(HappyReport2019['Economy (GDP per Capita)'],color='black', label="2019")

plt.title("Economy (GDP Per Capita) 2015-2019", size=20)

plt.legend()

plt.show()

plt.figure(figsize=(10,5))

sns.kdeplot(HappyReport2015['Family'],color='red', label='2015')

sns.kdeplot(HappyReport2016['Family'],color='blue', label='2016')

sns.kdeplot(HappyReport2017['Family'],color='limegreen', label='2017')

sns.kdeplot(HappyReport2018['Family'],color='orange', label='2018')

sns.kdeplot(HappyReport2019['Family'],color='black', label='2019')

plt.title('Family - Social Support Over The Years',size=20)

plt.legend()

plt.show()

plt.figure(figsize=(10,5))

sns.kdeplot(HappyReport2015['Health (Life Expectancy)'],color='red', label='2015')

sns.kdeplot(HappyReport2016['Health (Life Expectancy)'],color='blue', label='2016')

sns.kdeplot(HappyReport2017['Health (Life Expectancy)'],color='limegreen', label='2017')

sns.kdeplot(HappyReport2018['Health (Life Expectancy)'],color='orange', label='2018')

sns.kdeplot(HappyReport2019['Health (Life Expectancy)'],color='pink', label='2019')

plt.title('Health (Life Expectancy) Over The Years',size=20)

plt.legend()

plt.show()

plt.figure(figsize=(10,5))

sns.kdeplot(HappyReport2015['Freedom'],color='red', label='2015')

sns.kdeplot(HappyReport2016['Freedom'],color='blue', label='2016')

sns.kdeplot(HappyReport2017['Freedom'],color='limegreen', label='2017')

sns.kdeplot(HappyReport2018['Freedom'],color='orange', label='2018')

sns.kdeplot(HappyReport2019['Freedom'],color='black', label='2019')

plt.title('Freedom To Choose Things 2015-2019',size=20)

plt.legend()

plt.show()

plt.figure(figsize=(10,5))

sns.kdeplot(HappyReport2015['Trust (Government Corruption)'],color='red', label='2015')

sns.kdeplot(HappyReport2016['Trust (Government Corruption)'],color='blue', label='2016')

sns.kdeplot(HappyReport2017['Trust (Government Corruption)'],color='limegreen', label='2017')

sns.kdeplot(HappyReport2018['Trust (Government Corruption)'],color='orange', label='2018')

sns.kdeplot(HappyReport2019['Trust (Government Corruption)'],color='black', label='2019')

plt.title('Perceptions Of Corruption 2015-2019',size=20)

plt.legend()

plt.show()

plt.figure(figsize=(10,5))

sns.kdeplot(HappyReport2015['Generosity'],color='red', label='2015')

sns.kdeplot(HappyReport2016['Generosity'],color='blue', label='2016')

sns.kdeplot(HappyReport2017['Generosity'],color='limegreen', label='2017')

sns.kdeplot(HappyReport2018['Generosity'],color='orange', label='2018')

sns.kdeplot(HappyReport2019['Generosity'],color='black', label='2019')

plt.title('Generosity 2015-2019',size=20)

plt.legend()

plt.show()

ScoreList = list(HappyReport2019.Score)

for i in range(0,5):

    print(ScoreList[i])

dataAbove=[]

dataBelow=[]

for i in range(len(ScoreList)):

    if ScoreList[i]>np.mean(ScoreList):

        dataAbove.append(ScoreList[i])

    else:

        dataBelow.append(ScoreList[i])

print(len(dataAbove))

print(len(dataBelow))

HappyReport2019.drop(["Overall rank", "Country or region"], axis=1, inplace=True)

HappyReport2019.Score=[1 if each>np.mean(ScoreList) else 0 for each in HappyReport2019.Score]

y=HappyReport2019.Score.values

x = HappyReport2019["GDP per capita"]

x\_data=HappyReport2019.drop(["Score", "Social support", "Healthy life expectancy", "Freedom to make life choices","Generosity","Perceptions of corruption"],axis=1)

x=(x\_data - np.min(x\_data))/(np.max(x\_data) - np.min(x\_data)).values

scaler = StandardScaler()

x\_train = scaler.fit\_transform(x\_train)

model = LogisticRegression(solver='liblinear', random\_state=0).fit(x, y)

model.classes\_

model.intercept\_

model.coef\_

model.predict\_proba(x)

model.predict(x)

model.score(x, y)

confusion\_matrix(y, model.predict(x))

cm = confusion\_matrix(y, model.predict(x))

fig, ax = plt.subplots(figsize=(8, 8))

ax.imshow(cm)

ax.grid(False)

ax.xaxis.set(ticks=(0, 1), ticklabels=('Predicted 0s', 'Predicted 1s'))

ax.yaxis.set(ticks=(0, 1), ticklabels=('Actual 0s', 'Actual 1s'))

ax.set\_ylim(1.5, -0.5)

for i in range(2):

    for j in range(2):

        ax.text(j, i, cm[i, j], ha='center', va='center', color='red')

plt.show()

# Bibliografia

[1] <https://sdsnyouth.org/about>

[2] <https://www.unsdsn.org/about-us>

[3] <https://worldhappiness.report/faq/>

[4] <https://realpython.com/logistic-regression-python/>

# Spis wykresów

[Wykres 1 Mapa korelacji 2015 7](#_Toc72054263)

[Wykres 2 Mapa korelacji 2016 8](#_Toc72054264)

[Wykres 3 Mapa korelacji 2017 8](#_Toc72054265)

[Wykres 4 Mapa korelacji 2018 9](#_Toc72054266)

[Wykres 5 Mapa korelacji 2019 9](#_Toc72054267)

[Wykres 6 Ranking poszczególnych krajów przedstawiony na mapie świata dla roku 2015 10](#_Toc72054268)

[Wykres 7 Ranking poszczególnych krajów przedstawiony na mapie świata dla roku 2016 11](#_Toc72054269)

[Wykres 8 Ranking poszczególnych krajów przedstawiony na mapie świata dla roku 2017 11](#_Toc72054270)

[Wykres 9 Ranking poszczególnych krajów przedstawiony na mapie świata dla roku 2018 12](#_Toc72054271)

[Wykres 10 Ranking poszczególnych krajów przedstawiony na mapie świata dla roku 2019 12](#_Toc72054272)

[Wykres 11 Algorytm K-średnich przedstawiony na mapie świata dla danych z roku 2015 13](#_Toc72054273)

[Wykres 12 Algorytm K-średnich przedstawiony na mapie świata dla danych z roku 2016 14](#_Toc72054274)

[Wykres 13 Algorytm K-średnich przedstawiony na mapie świata dla danych z roku 2017 14](#_Toc72054275)

[Wykres 14 Algorytm K-średnich przedstawiony na mapie świata dla danych z roku 2018 15](#_Toc72054276)

[Wykres 15 Algorytm K-średnich przedstawiony na mapie świata dla danych z roku 2019 15](#_Toc72054277)

[Wykres 16 Współczynnik PKB w latach 2015-2019 16](#_Toc72054278)

[Wykres 17 Współczynnik wsparcia dla rodzin w latach 2015-2019 17](#_Toc72054279)

[Wykres 18 Współczynnik średniej długości życia w latach 2015-2019 18](#_Toc72054280)

[Wykres 19 Współczynnik swobody i wolności obywatelskiej w latach 2015-2019 18](#_Toc72054281)

[Wykres 20 Współczynnik korupcji w latach 2015-2019 19](#_Toc72054282)

[Wykres 21 Współczynnik dobroczynności w latach 2015-2019 19](#_Toc72054283)

[Wykres 22 Macierz pomyłek dla modelu regresji logistycznej 21](#_Toc72054284)