

# Klastrowanie publicznych baz związków

---

Iwona Raczkowska  
Jakub Kościukiewicz

# 1. W obrębie poszczególnych grup związków chemicznych można zaobserwować wspólne cechy korelujące ze źródłem ich danych.

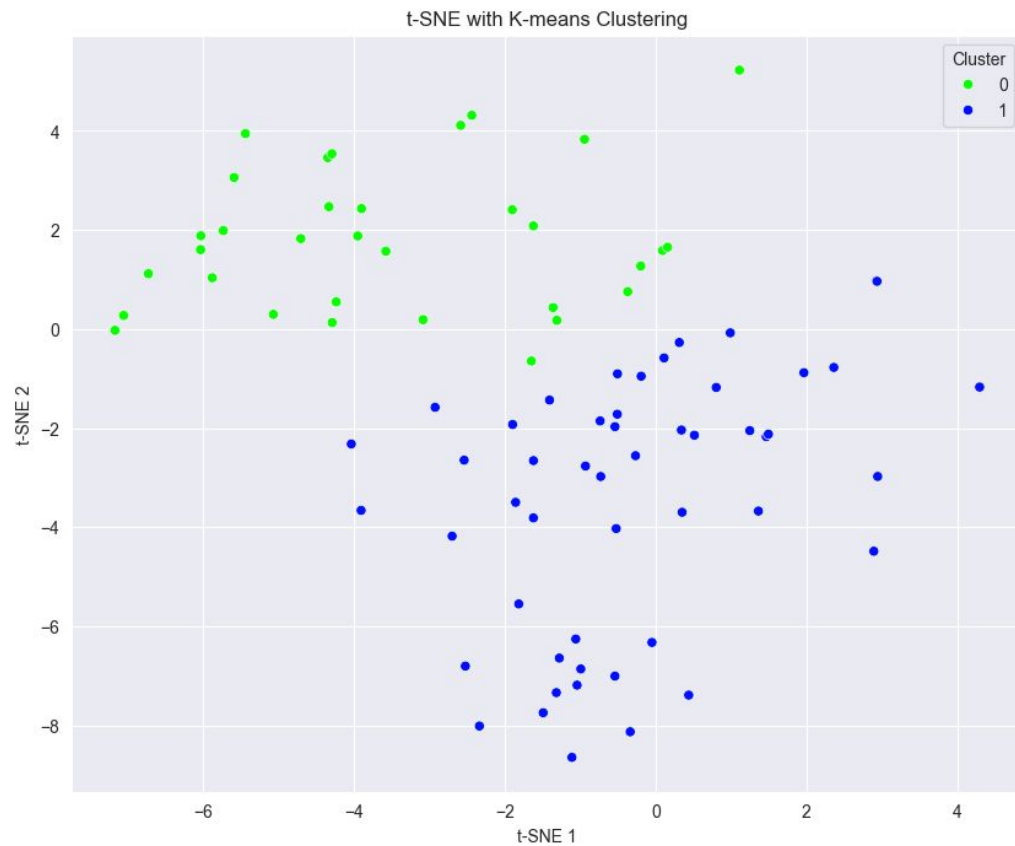
**Wybór reprezentacji związków chemicznych: alkaloidy + związki, które nie są alkaloidami**

Alkaloidy - grupa naturalnie występujących zasadowych związków organicznych (na ogół heterocyklicznych), głównie pochodzenia roślinnego, zawierających azot.

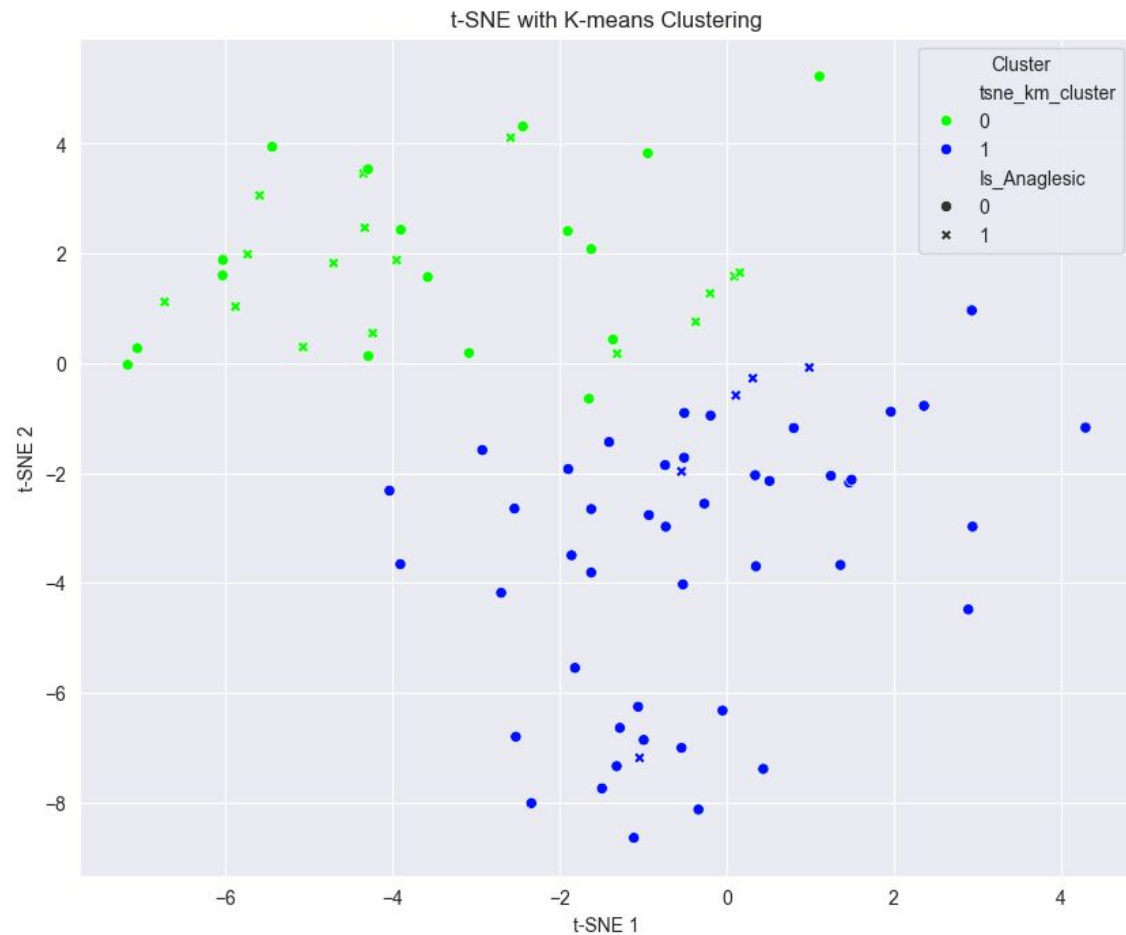
W obrębie alkaloidów do wspólnych cech można zaliczyć obecność w strukturze pierścienia z azotem, możliwość wykorzystania jako narkotyczny lek przeciwbólowy, naturalne pochodzenie, zasadowy charakter i na podstawie tych informacji przeprowadzona została klasteryzacja.



# Wyniki:



# Wyniki:



# Wyniki:

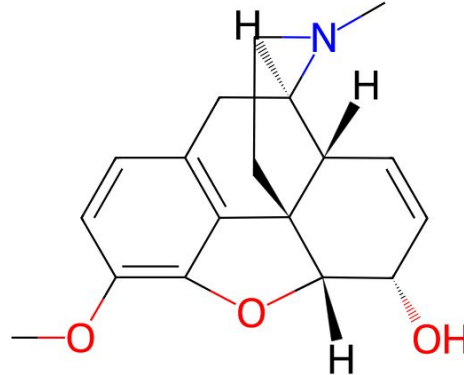


# Kodeina - zaklasyfikowana jako alkaloid



CX Acidic pKa:

13.78



**ANALGESIC**

(Kratom) Mitragynina - nie została zaklasyfikowana jako alkaloid

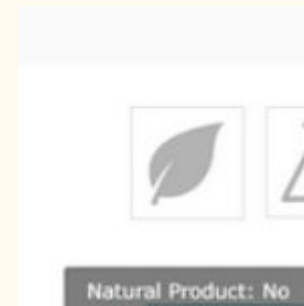
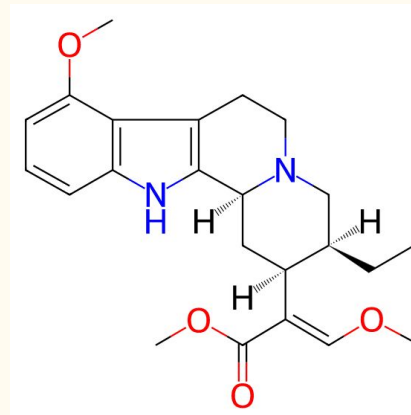


Mitragyna speciosa

Od wieków w Azji Południowo-Wschodniej spożycie mitragyniny w celach medycznych i rekreacyjnych

Rozprzestrzeniające użycie mitragyniny jako narkotyk rekreacyjny, jak i lek w Europie i obu Amerykach.

Suszone liście kratomu, które są parzone na herbatę lub mielone i umieszczane w kapsułkach.



Molecule Features

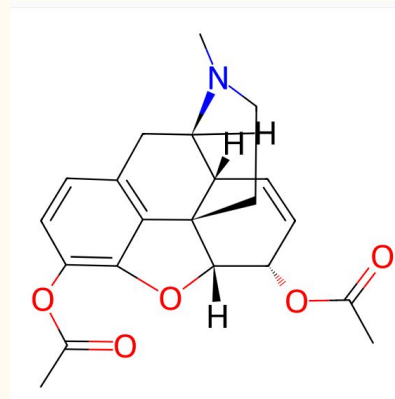
# Heroína, diacetylmorfina - zaklasyfikowana jako alkaloid



Półsyntetyczny opioid (diacetylowana pochodna morfiny)



Półsyntetyczne opioidy, choć wywodzą się z naturalnych alkaloidów nie są już klasyfikowane jako alkaloidy, ponieważ ich struktura została zmodyfikowana chemicznie.





Wnioski:

W obrębie poszczególnych grup związków chemicznych można zaobserwować wspólne cechy korelujące ze źródłem ich danych.

2. Związki chemiczne nie są równomiernie rozłożone w przestrzeni chemicznej, co można zbadać poprzez analizę ich rozmieszczenia w wybranej przestrzeni.



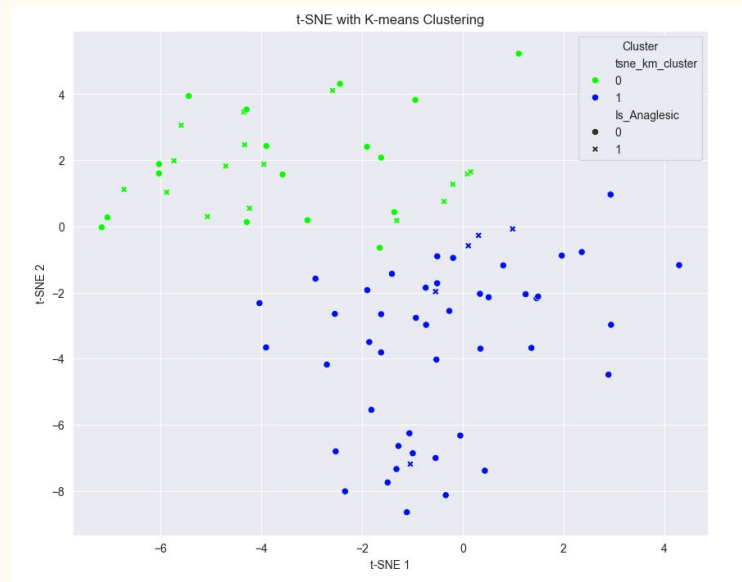
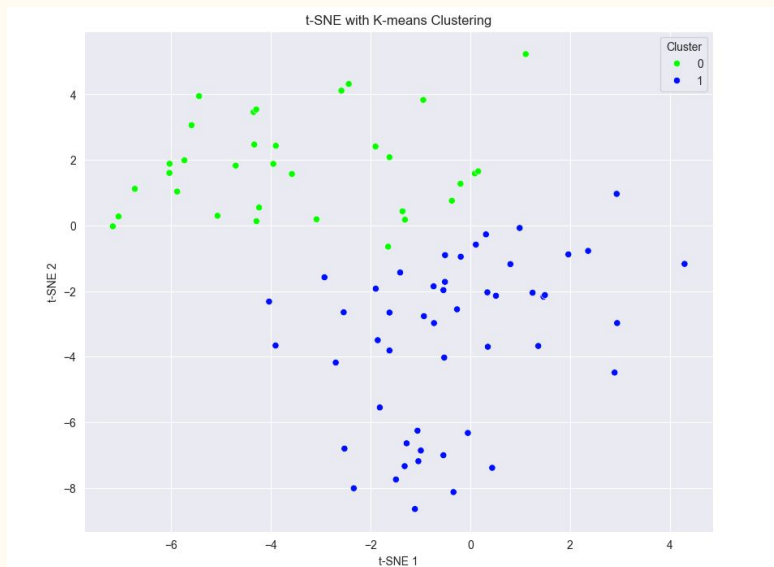
**Przestrzeń chemiczna** to zbiór wszystkich możliwych związków chemicznych, w tym wszystkich znanych cząsteczek leków oraz tych, które dopiero mają zostać odkryte. Szacuje się, że całkowita liczba związków tworzących przestrzeń chemiczną wynosi  $10^{60}$ . Do tej pory zbadano tylko niewielki ułamek tej przestrzeni. W bazie ChEMBL znajduje się 2,4 mln związków.

### **Klastry jako obszary wysokiej gęstości**

Związki o podobnych właściwościach chemicznych i biologicznych grupują się w przestrzeni chemicznej, tworząc klastry. Leki działające na podobne cele biologiczne mogą wykazywać podobieństwa w ich profilach chemicznych, co skutkuje ich zbliżonym położeniem w przestrzeni. Zatem obecność klastrów będzie w stanie potwierdzić hipotezę.

**Wybór reprezentacji związków chemicznych:** alkaloidy + związki, które nie są alkaloidami

# Wyniki:



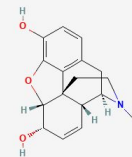
## Wnioski:

Związki chemiczne nie są równomiernie rozłożone w przestrzeni chemicznej, co można zbadać poprzez analizę ich rozmieszczenia i gęstości w wybranej przestrzeni.

### 3. Klastrowanie związków chemicznych na podstawie wartości ich aktywności odkryje niespójności pomiędzy różnymi testami biologicznymi.

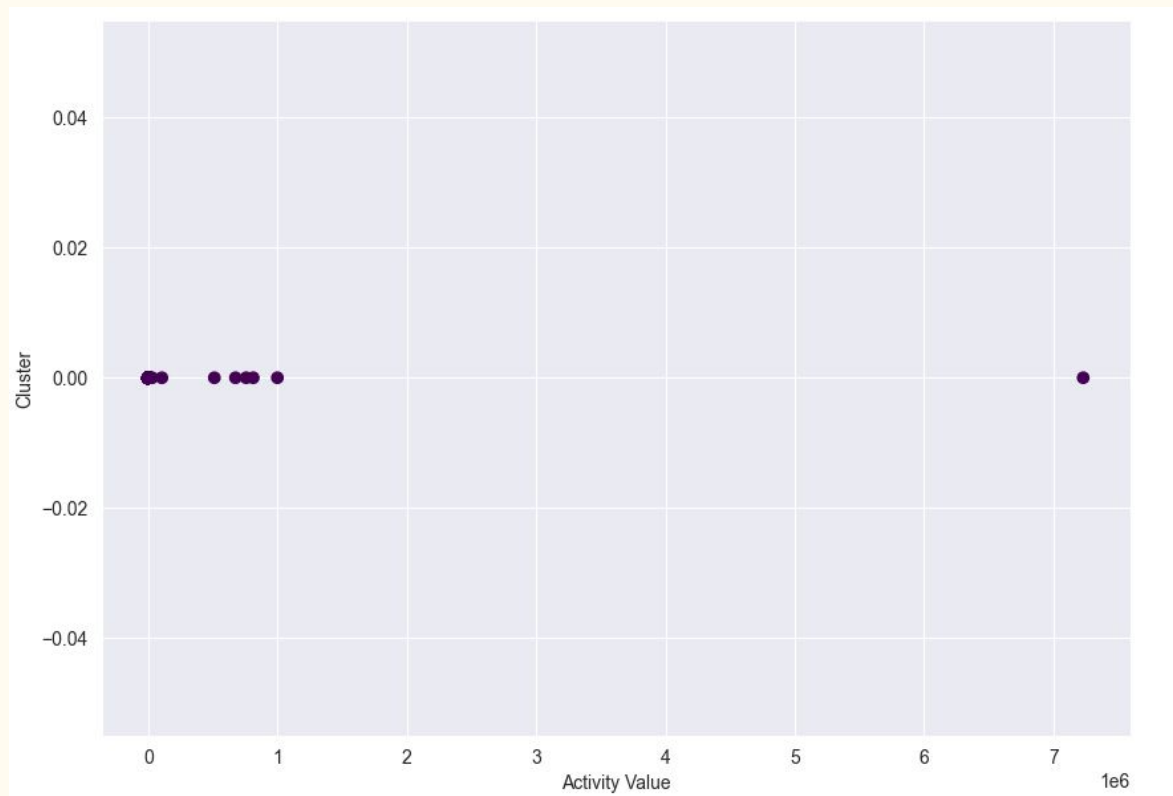
- Różne grupy badawcze mogą przeprowadzać badania na tym samym związku, ale w różnych warunkach laboratoryjnych, co może prowadzić do powstania wielu zestawów danych dotyczących tego samego związku. Różnice między sprzętem stosowanym w różnych ośrodkach naukowych mogą wpływać na niespójność danych.
- Różne parametry aktywności: Związek może być testowany pod kątem różnych typów aktywności, takich jak inhibicja, powinowactwo, toksyczność, itp. Miary aktywności mogą być określane jako  $K_i$ ,  $IC_{50}$  itp.
- Różne cele biologiczne: Aktywność związku mogła być testowana przeciwko różnym wariantom konkretnego receptora lub przeciwko innym receptorom w różnych organizmach lub systemach komórkowych.

**Wybór reprezentacji związków chemicznych: morfina**

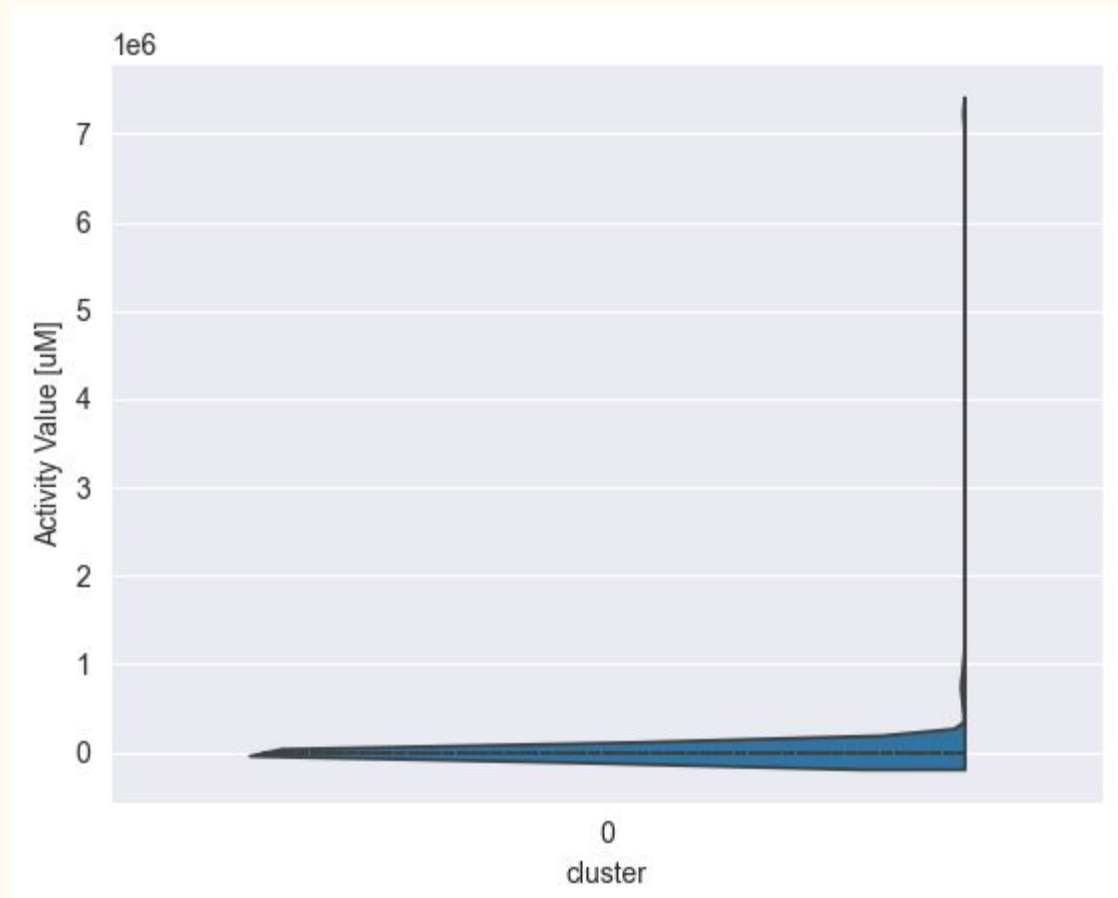


Wyniki:

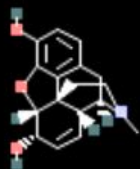
KMeans



Wyniki:



# Wyniki:



Inhibition of the twitch response of the guinea pig longitudinal muscle strip to electrical stimulation in a concentration-dependent manner

Activity Outcome: Unspecified Activity Type: pD2 Activity Value: 7230000  $\mu\text{M}$

BioAssay AID: 79676

Target Name: *Cavia porcellus* (domestic guinea pig)

Substance SID: 103169185 Compound CID: 5288826



50% Inhibition of stereospecific [3H]-naltrexone ( $10\text{e-}9\text{ M}$ ) binding towards opiate receptor in rat brain homogenate

Activity Outcome: Active Activity Type: IC50 Activity Value: 0.000027  $\mu\text{M}$

BioAssay AID: 145933

Target Name: Oprm1 - opioid receptor, mu 1 (Norway rat)

Substance SID: 103169185 Compound CID: 5288826



# Wyniki:

Pierwszy test mierzy skurcz mięśnia, co jest fizjologiczną odpowiedzią wymagającą większego stężenia morfiny do osiągnięcia efektu.

Drugi test mierzy wiązanie do receptora, co jest bardziej bezpośrednim i precyzyjnym pomiarem interakcji leku z receptorem opioidowym



"Unspecified" z powodu braku precyzyjnych danych, niejasnych wyników eksperymentu, automatycznego przetwarzania danych, niepełnych danych eksperymentalnych lub trudności w klasyfikacji wyników dla specyficznych typów testów. W przypadku badań farmakologicznych, takie sytuacje mogą się zdarzać i wymagają dalszej analizy oraz weryfikacji, aby dokładnie określić wynik aktywności danej substancji.



Wnioski:

Klastrowanie związków chemicznych na podstawie wartości ich aktywności odkrywa niespójności pomiędzy różnymi testami biologicznymi.

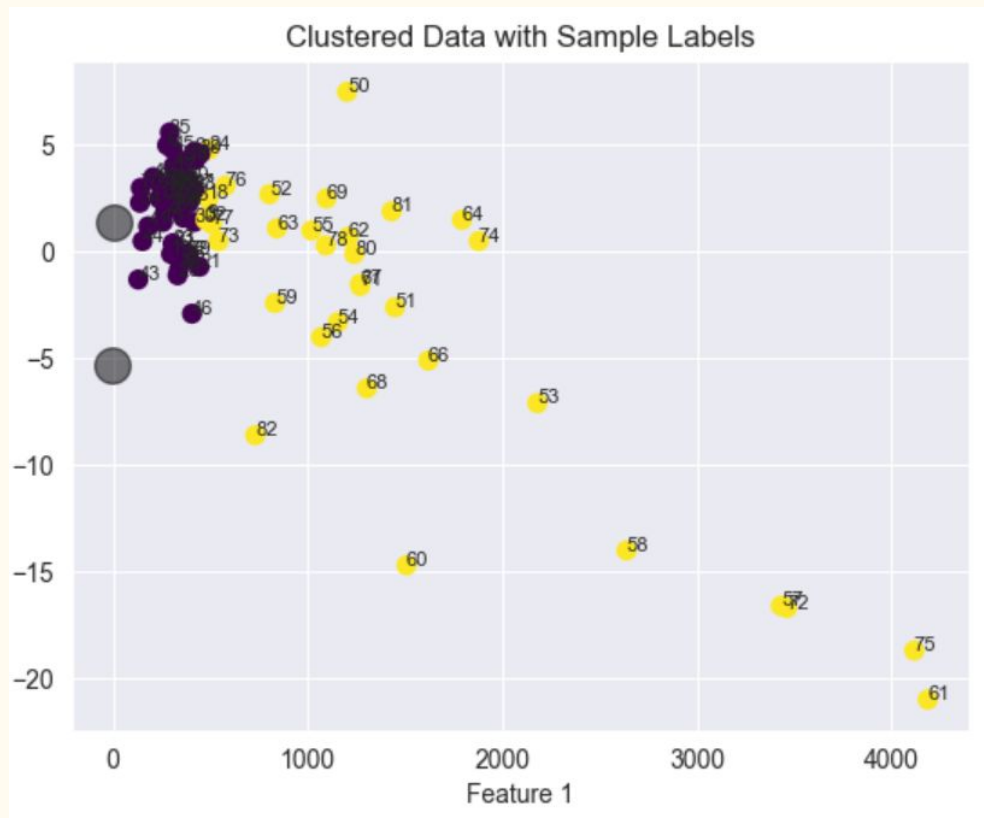
## 4. Leki spełniające cztery kryteria Lipińskiego mają szansę stać się skutecznym lekiem doustnym.

- Masa molowa  $< 500$  Da
- Współczynnik podziału oktanol: woda ( $\log P$ )  $< 5$
- Liczba donorów wiązania wodorowego  $< 5$
- Liczba akceptorów wiązania wodorowego  $< 10$



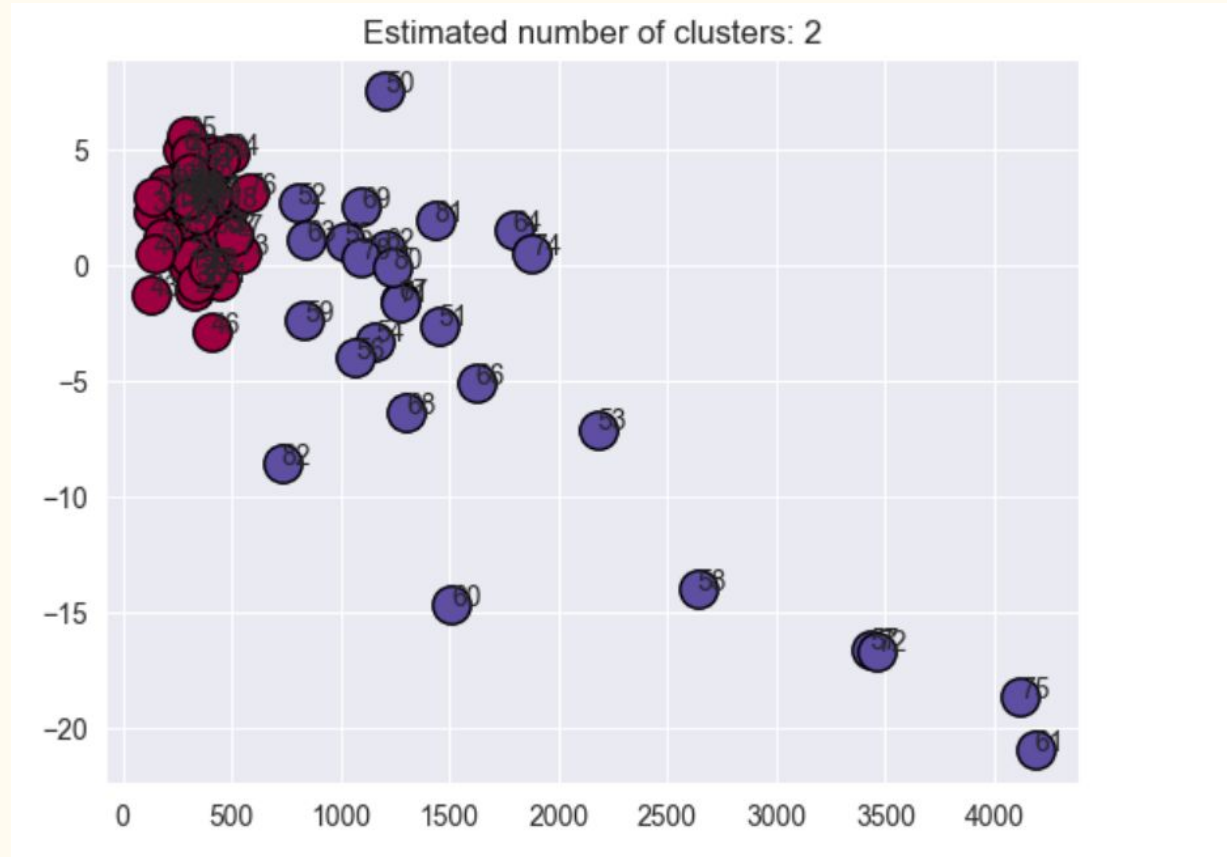
Wyniki:

KMeans



Wyniki:

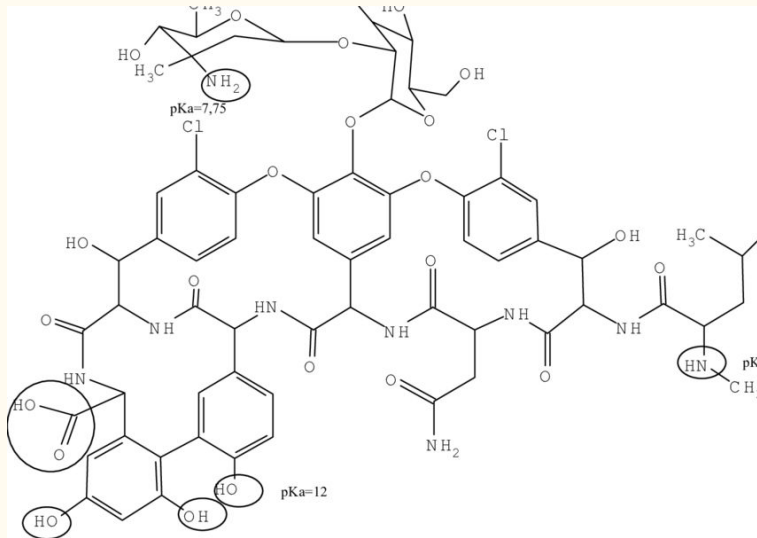
DBSCAN



# Wyniki:



# Wankomycyna



Property Name	Property Value
Molecular Weight	1449.2 g/mol
XLogP3-AA	-2.6
Hydrogen Bond Donor Count	19
Hydrogen Bond Acceptor Count	26

**Zastosowanie:** Stosowana w leczeniu poważnych zakażeń wywołanych przez bakterie Gram-dodatnie, w tym zakażeń skóry, zapalenia płuc oraz zakażeń krwi.

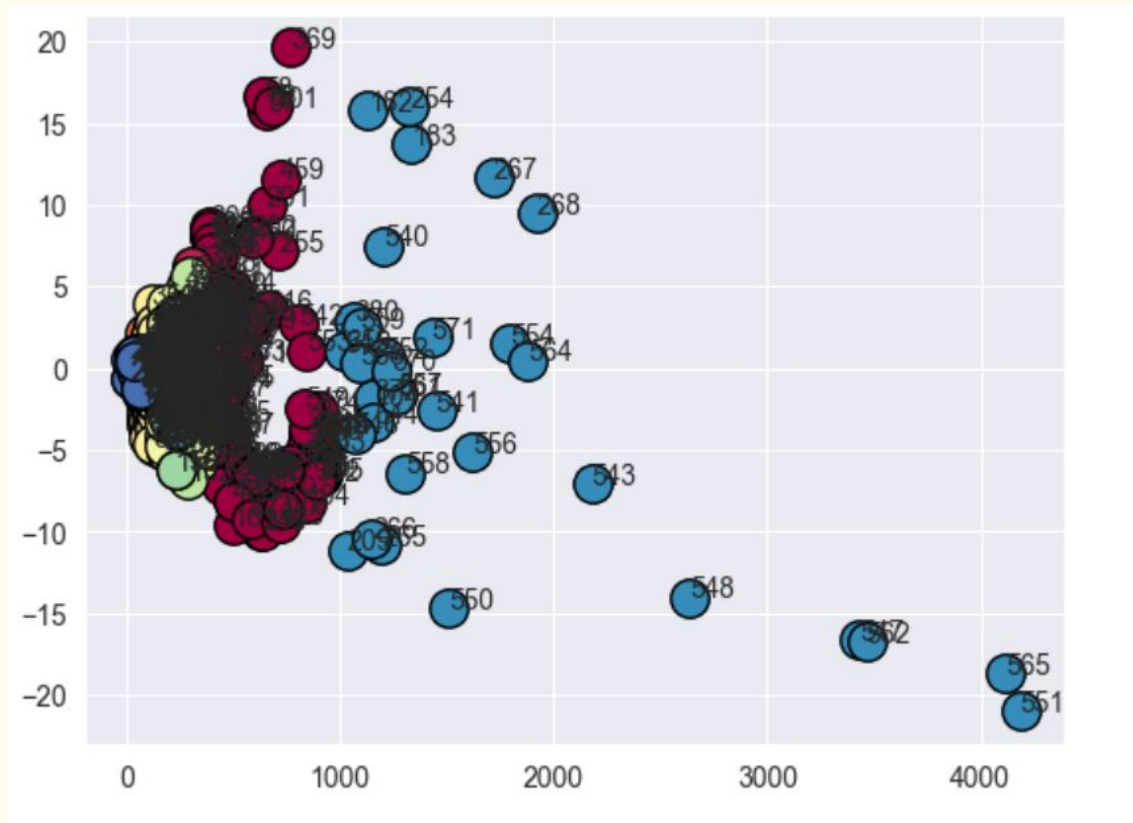
Wankomycyna nie wchłania się dobrze z przewodu pokarmowego, dlatego jest podawana dożylnie w celu osiągnięcia odpowiednich stężeń terapeutycznych we krwi.





Wyniki:

DBSCAN



# Wyniki:

Alprazolam jest stosowany w leczeniu zaburzeń lękowych lub krótkotrwałych objawów lęku.

## Alprazolam

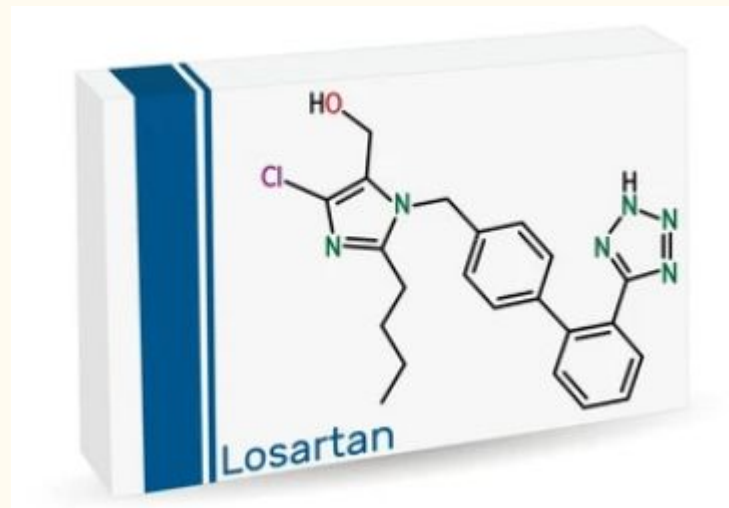


	MolecularWeight	XLogP	HBondDonorCount	HBondAcceptorCount	cluster	CID
1	418.60	4.7	1	5	0	2118.0

# Wyniki:

## Losartan

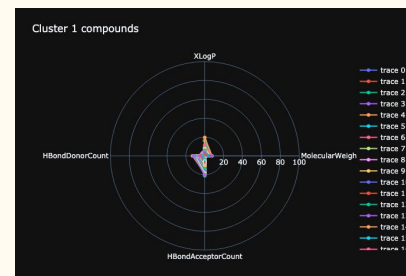
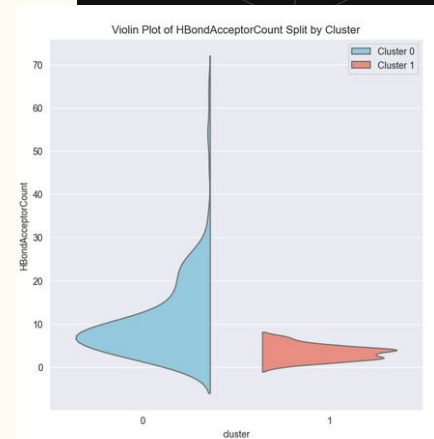
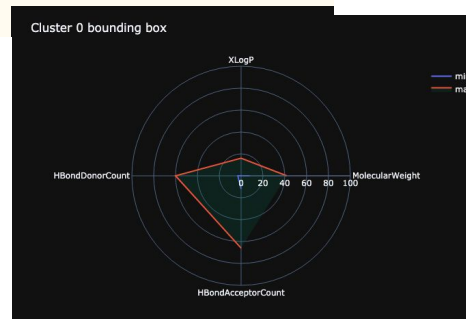
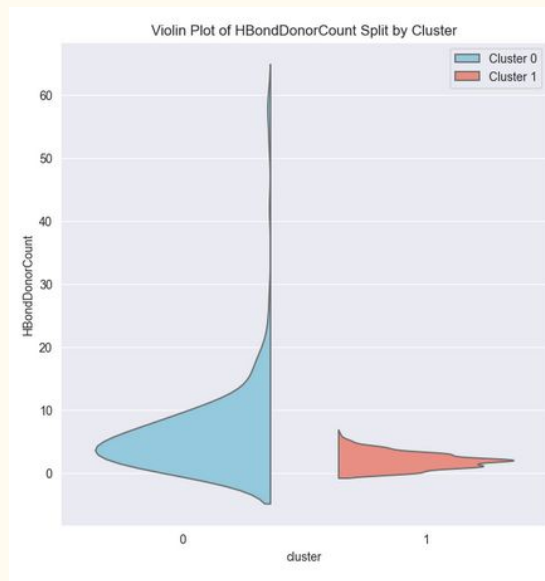
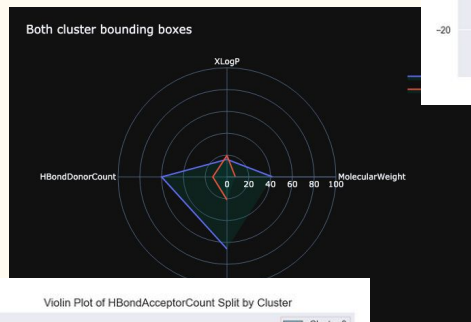
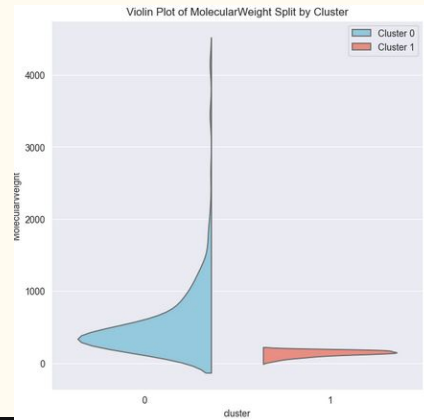
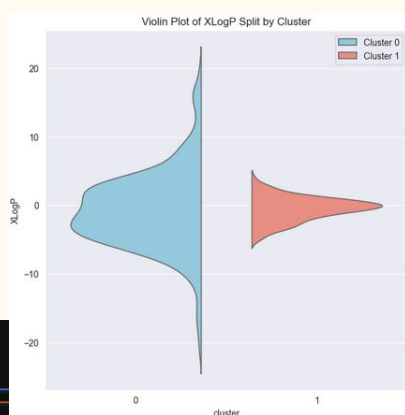
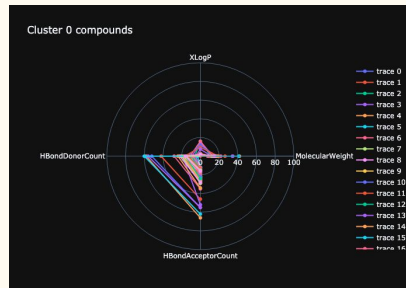
Stosowany głównie w leczeniu wysokiego ciśnienia krwi oraz w celu ochrony nerek przed uszkodzeniem u osób z cukrzycą typu 2.



	MolecularWeight	XLogP	HBondDonorCount	HBondAcceptorCount	cluster	CID
3	203.24	0.4	0	4	0	3961.0

Wnioski:

Leki spełniające cztery kryteria Lipińskiego mają szansę stać się skutecznym lekiem doustnym.



# Źródła:

- Rishal Hurbans, *Algorytmy sztucznej inteligencji*
- Bogocz Jacek. (2016). *Metody eksploracji baz danych w poszukiwaniu nowych reguł projektowania leków*. Praca doktorska. Katowice: Uniwersytet Śląski
- [https://en.m.wikipedia.org/wiki/File:Spotify\\_logo\\_with\\_text.svg](https://en.m.wikipedia.org/wiki/File:Spotify_logo_with_text.svg)
- [https://pl.m.wikipedia.org/wiki/Plik:Netflix\\_2015\\_logo.svg](https://pl.m.wikipedia.org/wiki/Plik:Netflix_2015_logo.svg)
- <https://pl.wikipedia.org/wiki/Alkaloidy>
- [https://pl.wikipedia.org/wiki/Mak\\_lekarski](https://pl.wikipedia.org/wiki/Mak_lekarski)
- <https://extrapolations.com/what-is-chemical-space/>
- <https://www.herbapol-polana.com/vademecum-botaniczne/mak-lekarski>
- <https://www.theguardian.com/film/2019/jan/21/from-red-pills-to-red-white-and-blue-brex-it-how-the-matrix-shaped-our-reality>
- <https://pl.pinterest.com/pin/563794447074793442/>
- <https://pl.pinterest.com/pin/442197257153939610/>
- <https://pubchem.ncbi.nlm.nih.gov/compound/Morphine>
- <https://www.artstation.com/artwork/6bNG65>
- <https://www.ebi.ac.uk/chembl/>
- <https://pubchem.ncbi.nlm.nih.gov/>
- <https://www.slideshare.net/100006619533516/zinc-database>
- <https://go.drugbank.com/>
- <https://www.genome.jp/kegg/pathway.html>
- <https://www.rcsb.org/>
- <https://ieeexplore.ieee.org/abstract/document/7036702>
- [https://fulmanski.pl/zajecia/seminarium\\_katedry/ten/TymoszczukPrace/old/1.pdf](https://fulmanski.pl/zajecia/seminarium_katedry/ten/TymoszczukPrace/old/1.pdf)
- <https://www.nature.com/articles/s41598-023-32790-3>
- <https://pubs.acs.org/doi/pdf/10.1021/ci00010a010>
- <https://www.etsy.com/in-en/listing/1095228459/pink-xanax-painting>
- <https://en.wikipedia.org/wiki/Mitragynine>
- <https://pl.wikipedia.org/wiki/Heroina>
- [https://www.researchgate.net/figure/The-structure-of-vancomycin\\_fig1\\_268349400](https://www.researchgate.net/figure/The-structure-of-vancomycin_fig1_268349400)
- <https://www.amazon.com/Activis-Sprite-Explicit-Kkaya/dp/B088FYFX16>
- <https://www.oakhurstvet.com/blog/how-do-you-care-for-a-guinea-pig/>
- <https://www.shutterstock.com/search/losartan>
-