

R science !

Super_Kret working with R !

Nad czym pracujemy !

Jesteśmy grupą data science, zajmujemy się sondażami, jednak też testujemy możliwości R, stąd praca nad oceanem i cukierkami!

Co Wam

pokażemy!

1. Shiny - Mateusz
2. Sondaże wyborcze - Ewa
3. Tweety - Mateusz
4. Regresja logistyczna - Grzegorz
5. Regresja liniowa - Marta
6. Podsumowanie- Ewa

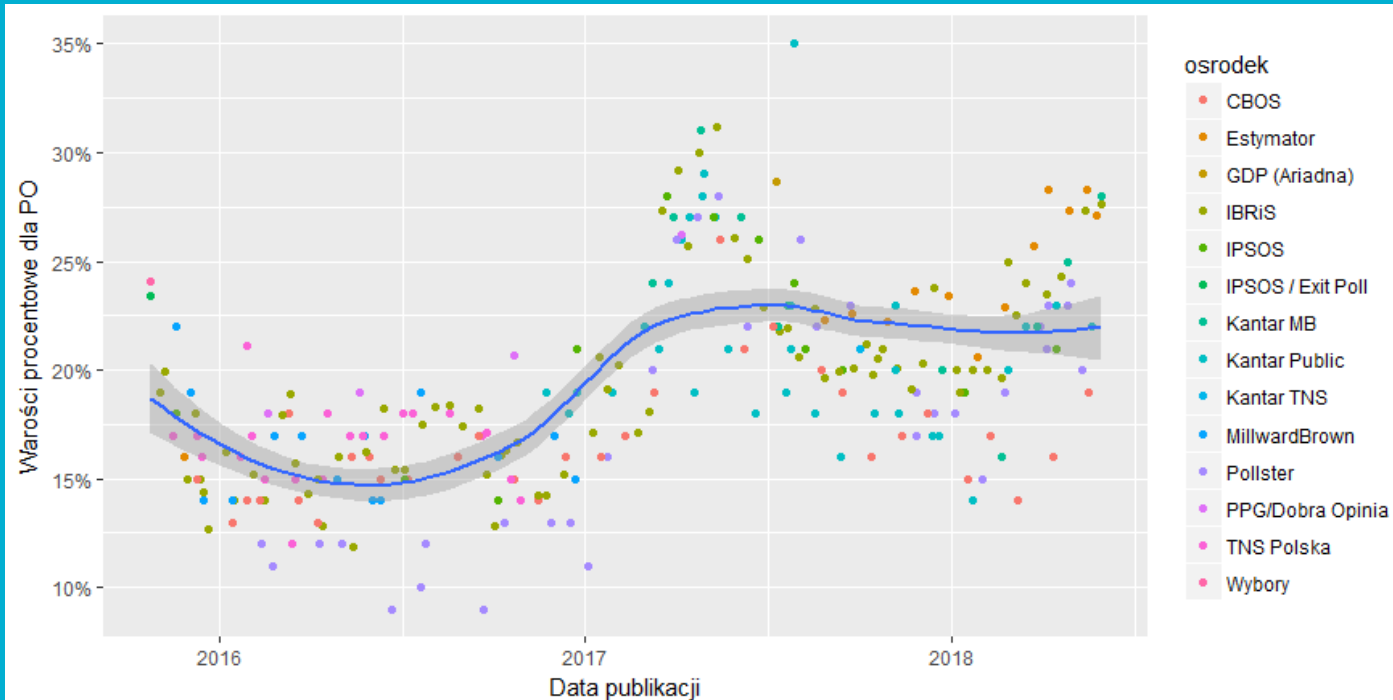
SHINY

Podsumowanie shiny

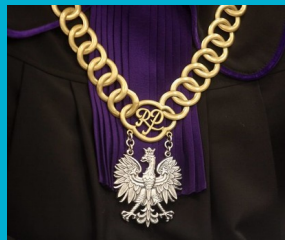
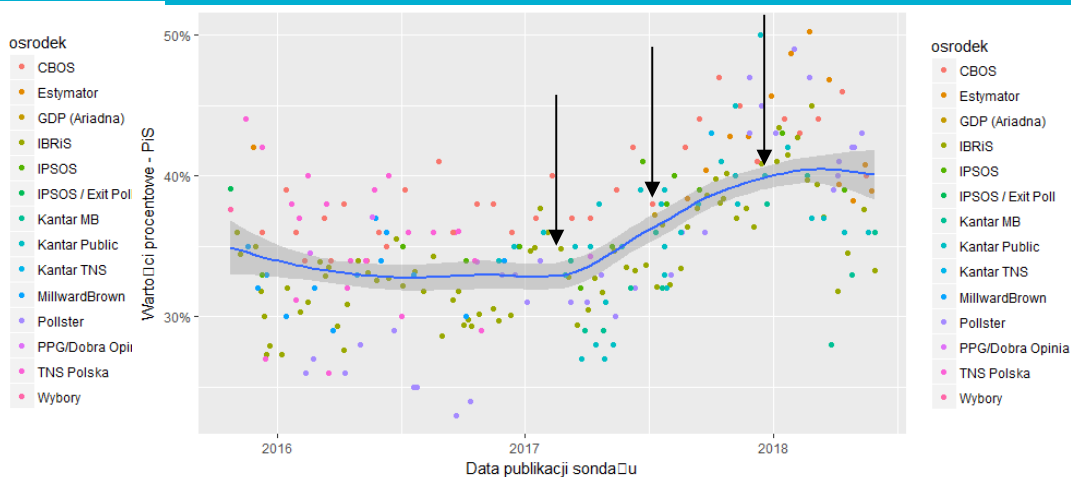
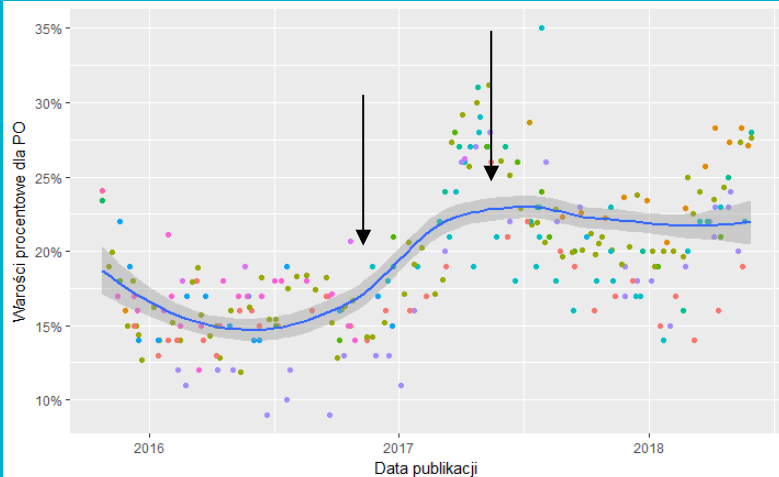
https://mateuszkieszkowski.shinyapps.io/my_shiny_portfolio/

- Analiza wyników sondaży dla partii politycznych
- Wczytanie tekstu z pliku, text mining oraz stworzenie wordcloud/analiza sentymentu
- Wpisanie twitta oraz przedstawienie danych w tabeli

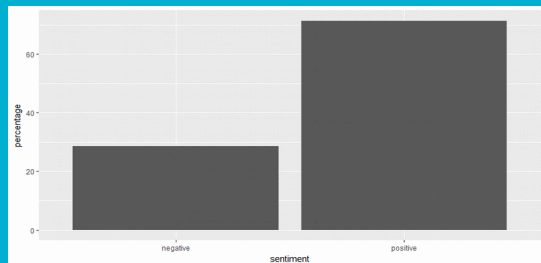
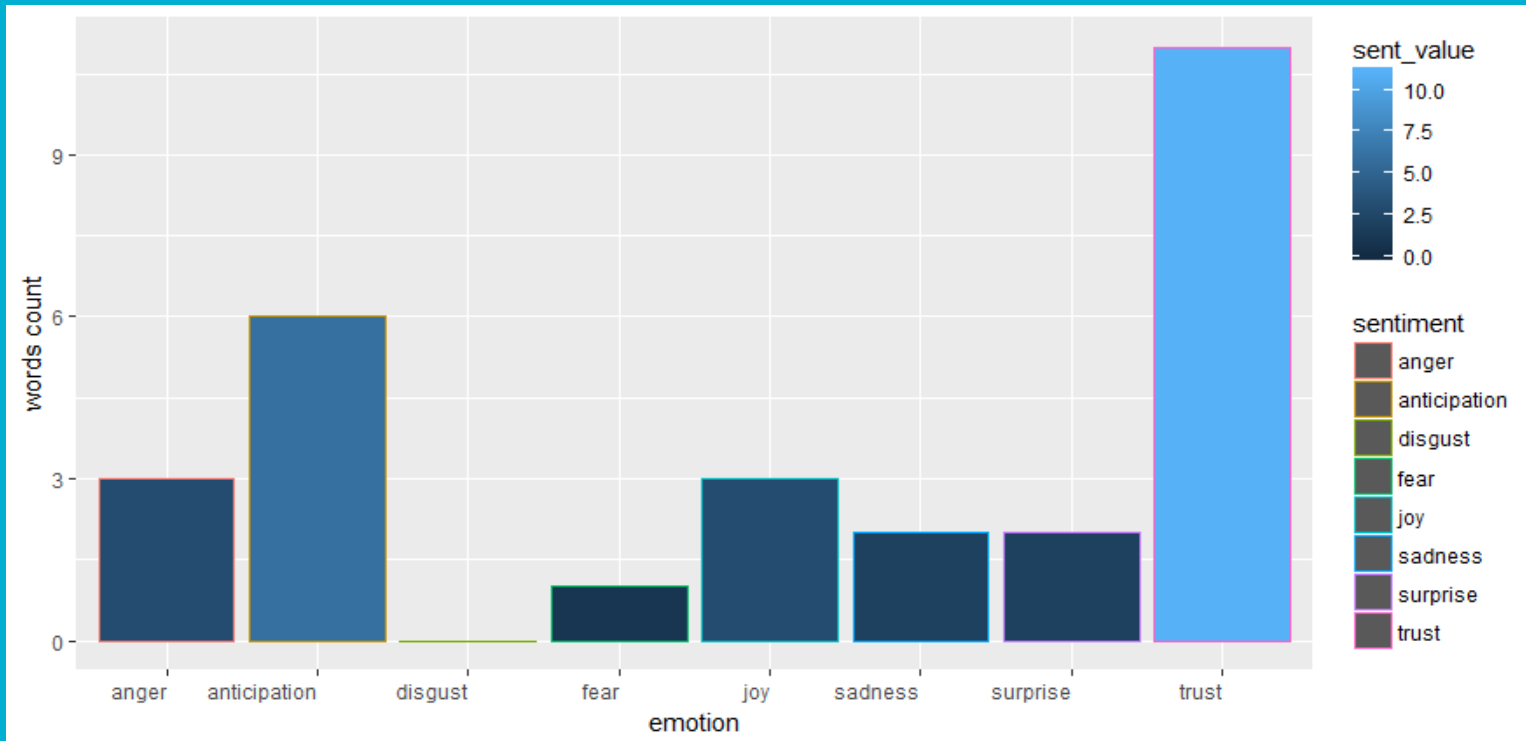
SONDAŽE



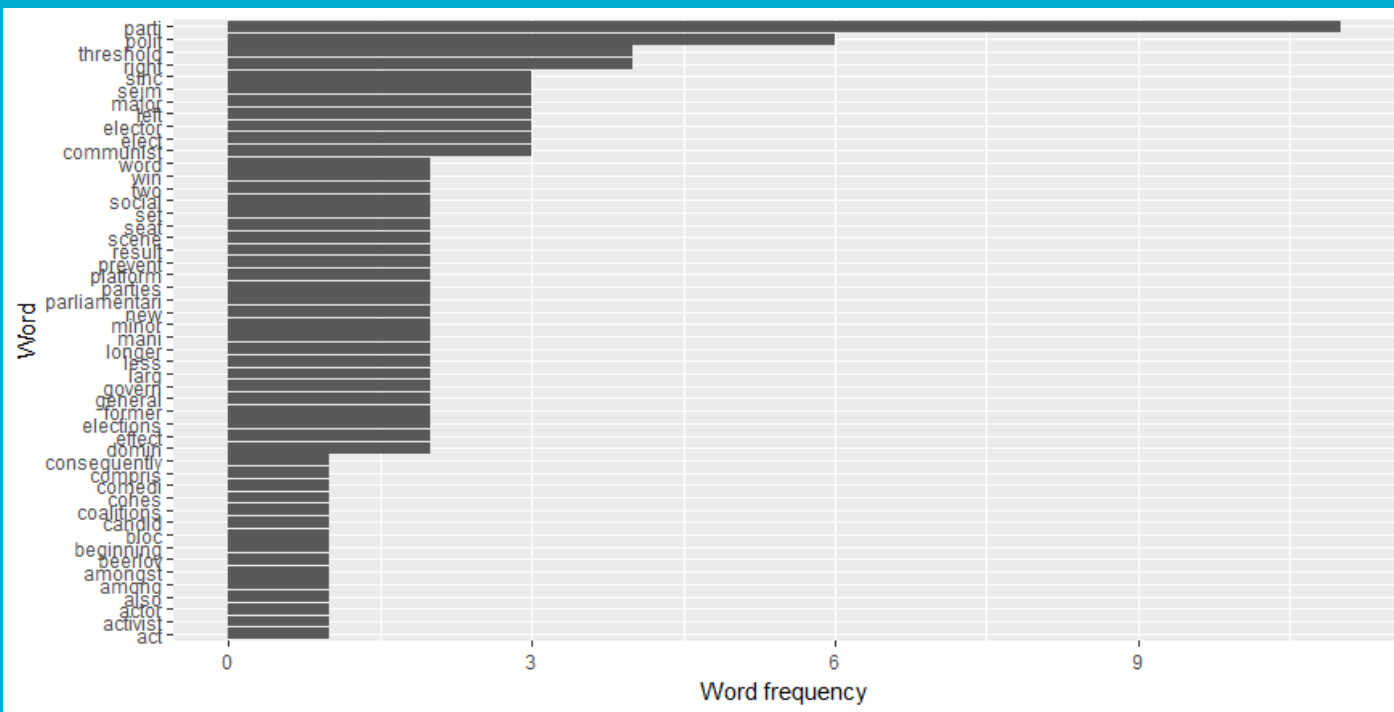
Kilka wykresów?



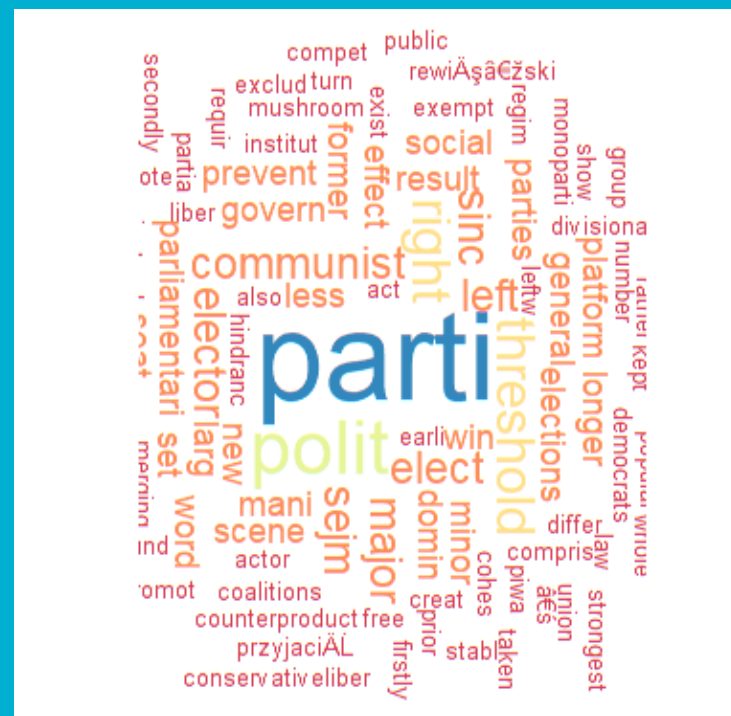
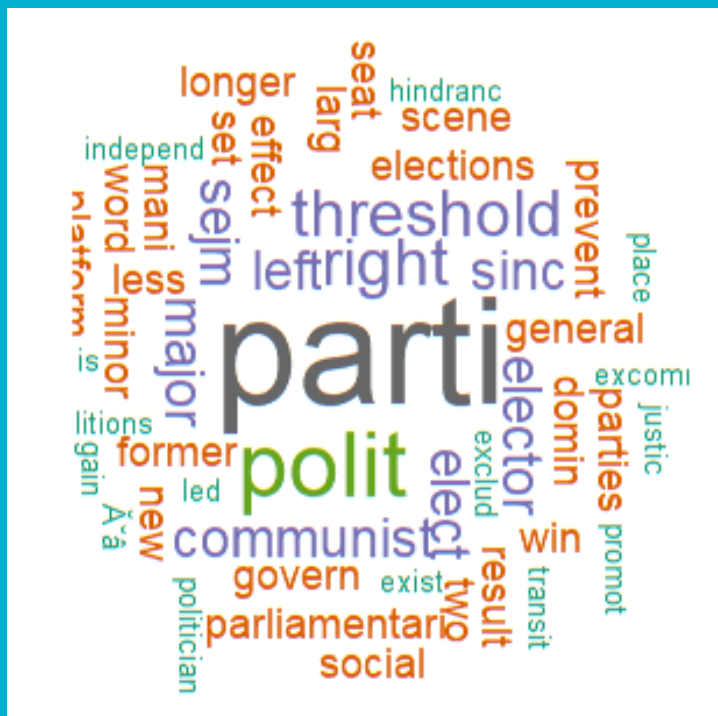
**Najwięksi polityczni adwersarze
PO -PiS**



Sentymment



**Word
Frequency**



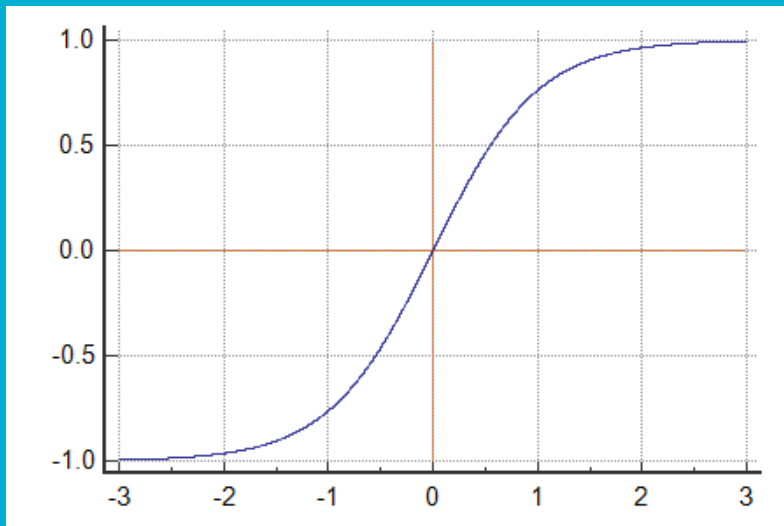
WordCloud

parti

//polit

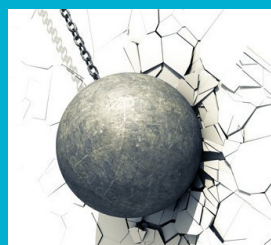
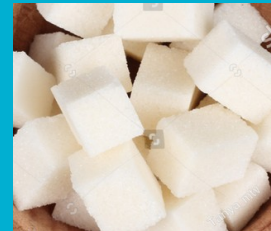
REGRESJA LOGISTYCZNA

Want some candy?



Regresja logistyczna

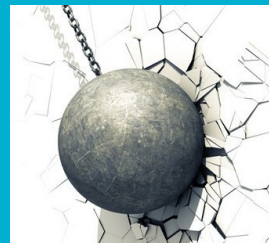
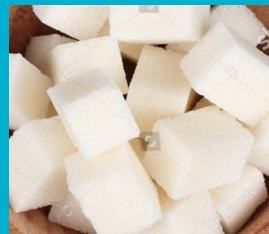
Candy wants you!



Regresja logistyczna



u!



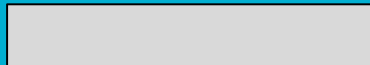
Regresja logistyczna



Regre



Candy wants you!

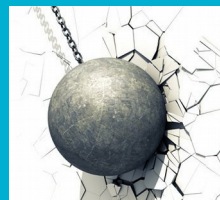
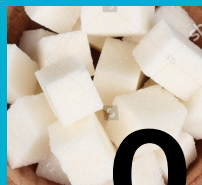


Regresja logistyczna

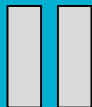
Candy wants you!




Regresja logistyczna



0.88

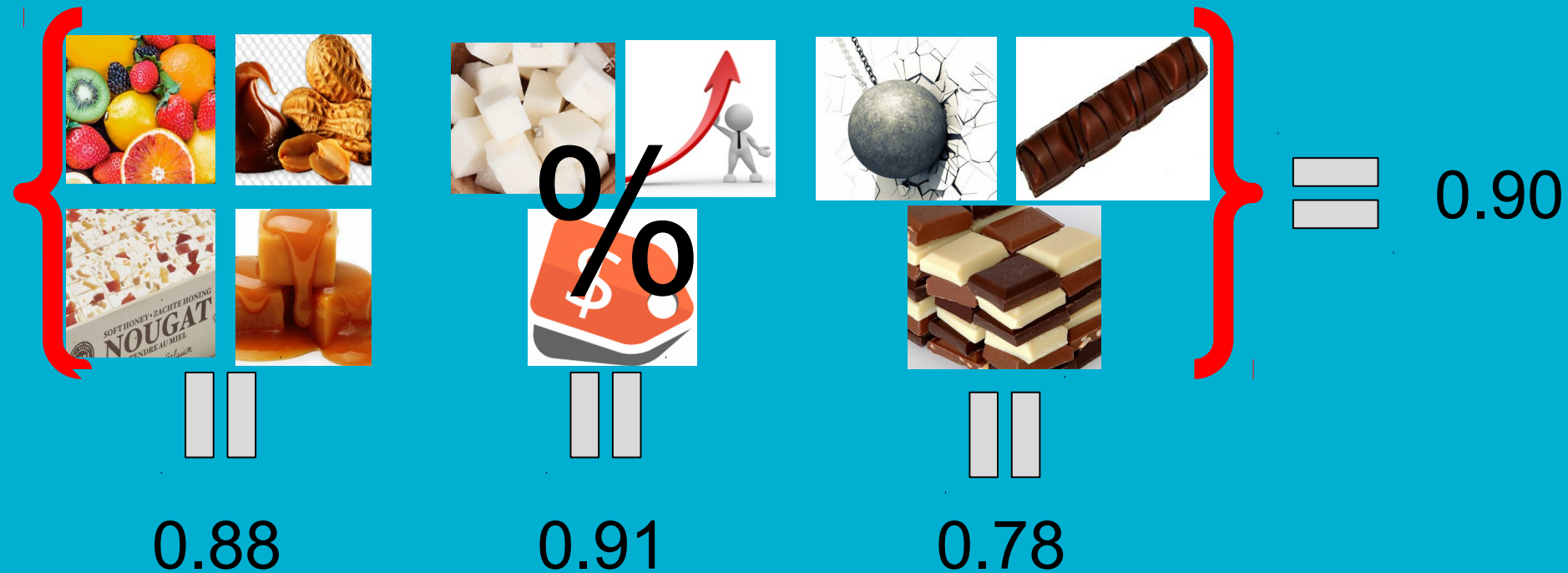


0.91



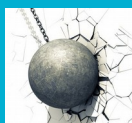
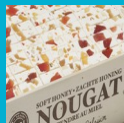
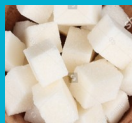
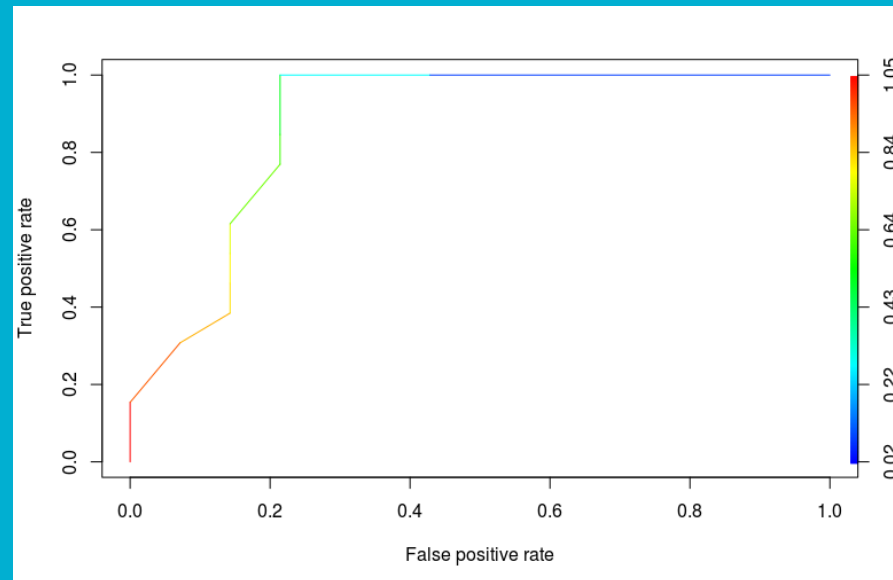
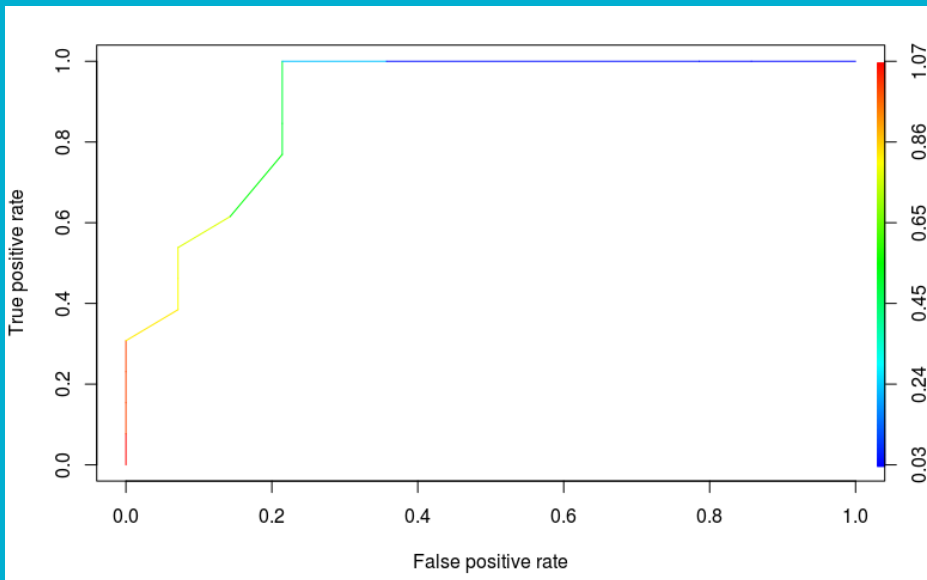
0.78

Regresja logistyczna

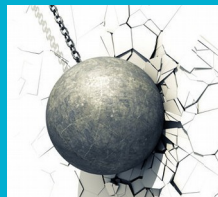
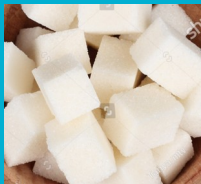
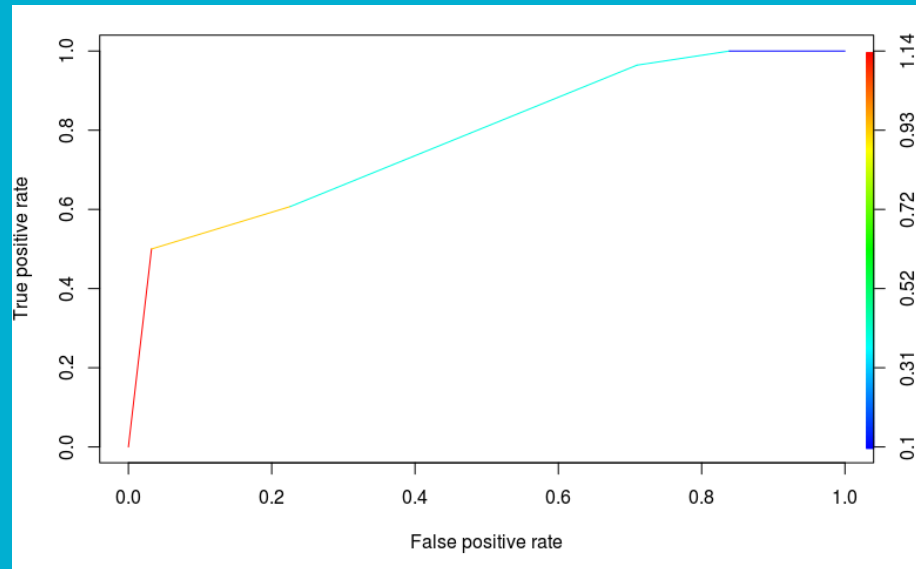
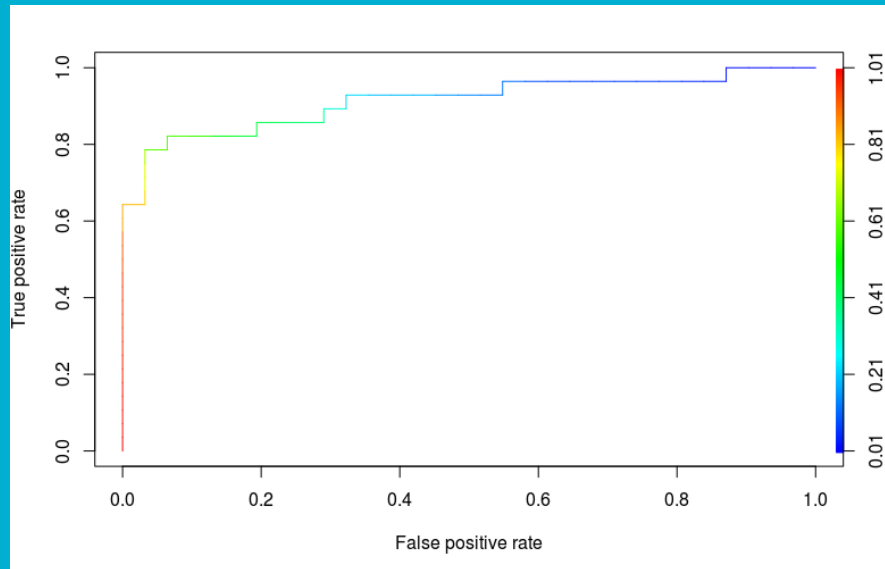


Regresja logistyczna

Podsumujmy:



Regresja logistyczna



Po słodyczach

Nurkujemy w oceanie?

REGRESJA LINIOWA

Regresja liniowa

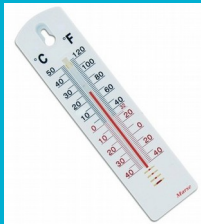
Dane:

- Dataset CalCOFI ze strony kaggle.com
- Oceanograficzne dane o rybach na świecie
- Czas obserwacji: 1949-obecnie
- Liczba obserwacji: 864 863
- Liczba zmiennych 74
- Dane zbierane w regularnych odstępach czasu na określonej przestrzeni



Zagadnienie:

Czy na podstawie zasolenia możemy określić temperaturę wody?



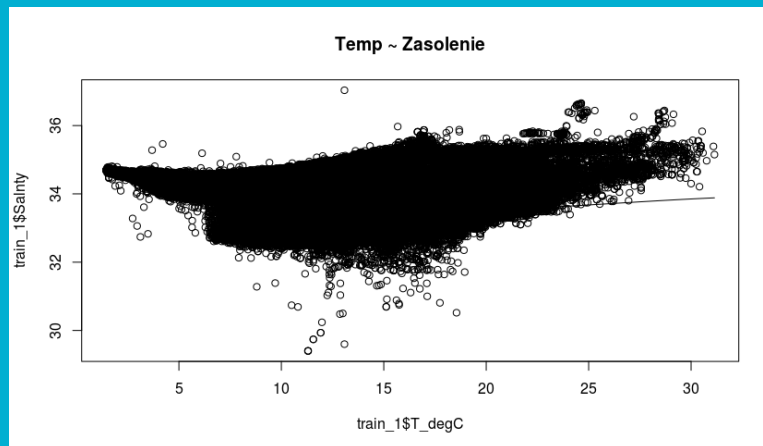
Eksploracja danych



Zmienna endogeniczna

Zmienna egzogeniczna

Min	1,44
Mediana	10,06
Średnia	10,8
Max	31,14
Puste wartości	10 963
Outlier?	Tak



Min	28,43
Mediana	33,86
Średnia	33,84
Max	37,03
Puste wartości	47 354
Outlier?	Tak

Budujemy 2 modele!

Budowa modeli

Model 1

Założenie: usunięcie obserwacji z brakującymi wartościami



Model 2

Założenie: zastąpienie brakujących obserwacji medianą

	1. Podział zbiorów na treningowy i testowy 70% / 30%	
- 0,6542004	2. Współczynnik korelacji	- 0,6429004
0,2532	3. Budowa modelu liniowego - ocena parametrów Współczynnik determinacji R ²	0,2403
13.19388	4. Sprawdzenie modelu na danych testowych Metoda najmniejszych kwadratów MSE	13.52719

Model 1!

RETROSPEKTYW A



Retrospektywa

- Co poszło dobrze?
 - Każdy wiedział co ma zrobić - podział zadań - ok
 - Współpraca w grupie - bez względu na czas odpowiedzi.
 - Na pierwszym etapie - lepsza komunikacja
 - Pomoc grupowa
 - Wykorzystanie wiedzy zdobytej na zajęciach
- Co poszło źle?
 - Ograniczona komunikacja na drugim etapie pracy
 - Dotrzymywanie terminów
 - Brak bieżących odpowiedzi ze strony trenerów spowalniało pracę
- Co możemy zrobić lepiej?
 - Dotrzymywać terminów które ustalimy!
 - Popracować nad komunikacją
 - Zorganizować spotkania / slack, real/



Ewa

Marta

Grzesiek

Mateusz

KONIEC!