

PADPY PD3

Aleksandra Miesiąc & Tymoteusz Makowski

21 stycznia 2019

Ekspert
●○○○○○

Najbardziej żyte ze sobą pary użytkowników
○○○○○○○○○

Aktywność a wydarzenia
○○○○○○○

Ekspert

Gdzież on się podział?

- 1 Wybieramy posty które mają zaakceptowaną odpowiedź:
`posts.dropna(subset = "AcceptedAnswerId")`
- 2 Robimy tabelkę użytkowników, którzy “zamknęli” post swoją odpowiedzią
- 3 Grupujemy po identyfikatorze konta użytkownika i zliczamy ilość wystąpień.
- 4 Sortujemy względem liczby zamkniętych postów.
- 5 Wybieramy pierwszy rekord
- 6 Mamy zwycięzcę!

Gdzież on się podział?

- 1 Wybieramy posty które mają zaakceptowaną odpowiedź:
`posts.dropna(subset = "AcceptedAnswerId")`
- 2 Robimy tabelkę użytkowników, którzy “zamknęli” post swoją odpowiedzią
- 3 Grupujemy po identyfikatorze konta użytkownika i zliczamy ilość wystąpień.
- 4 Sortujemy względem liczby zamkniętych postów.
- 5 Wybieramy pierwszy rekord
- 6 Mamy zwycięzcę!

Gdzież on się podział?

- 1 Wybieramy posty które mają zaakceptowaną odpowiedź:
`posts.dropna(subset = "AcceptedAnswerId")`
- 2 Robimy tabelkę użytkowników, którzy “zamknęli” post swoją odpowiedzią
- 3 Grupujemy po identyfikatorze konta użytkownika i zliczamy ilość wystąpień.
- 4 Sortujemy względem liczby zamkniętych postów.
- 5 Wybieramy pierwszy rekord
- 6 Mamy zwycięzcę!

Gdzież on się podział?

- 1 Wybieramy posty które mają zaakceptowaną odpowiedź:
`posts.dropna(subset = "AcceptedAnswerId")`
- 2 Robimy tabelkę użytkowników, którzy “zamknęli” post swoją odpowiedzią
- 3 Grupujemy po identyfikatorze konta użytkownika i zliczamy ilość wystąpień.
- 4 Sortujemy względem liczby zamkniętych postów.
- 5 Wybieramy pierwszy rekord
- 6 Mamy zwycięzcę!

Gdzie on się podział?

- 1 Wybieramy posty które mają zaakceptowaną odpowiedź:
`posts.dropna(subset = "AcceptedAnswerId")`
- 2 Robimy tabelkę użytkowników, którzy “zamknęli” post swoją odpowiedzią
- 3 Grupujemy po identyfikatorze konta użytkownika i zliczamy ilość wystąpień.
- 4 Sortujemy względem liczby zamkniętych postów.
- 5 Wybieramy pierwszy rekord
- 6 Mamy zwycięzcę!

Gdzież on się podział?

- 1 Wybieramy posty które mają zaakceptowaną odpowiedź:
`posts.dropna(subset = "AcceptedAnswerId")`
- 2 Robimy tabelkę użytkowników, którzy “zamknęli” post swoją odpowiedzią
- 3 Grupujemy po identyfikatorze konta użytkownika i zliczamy ilość wystąpień.
- 4 Sortujemy względem liczby zamkniętych postów.
- 5 Wybieramy pierwszy rekord
- 6 Mamy zwycięzcę!

“egreg”! Parę informacji:

Spójrzmy na aktywność na poszczególnych forach naszego Eksperta:

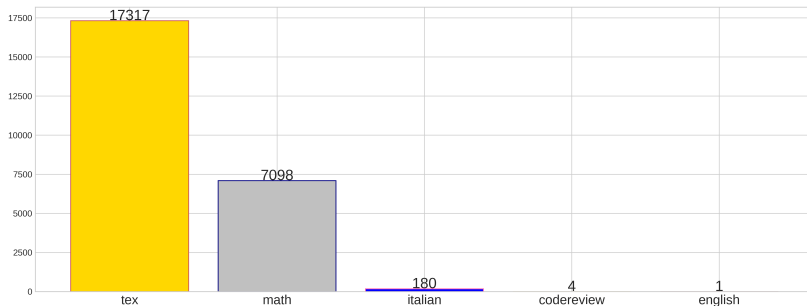


Figure 1: Aktywność Eksperta

Ciekawe fakty:

Jak wygląda zależność między pytaniami a odpowiedziami Eksperta?

Forum	PostTypeld	odp i pyt
tex	2	17317
math	2	7098
italian	2	180
codereview	2	4
english	2	1

Co można o nim wywnioskować?

- że jest mistrzem texa
- że jest matematykiem
- że lubi włoski/jest włochem
- klepie kod

Ciekawe fakty cd.

Zobaczmy więc, jego AboutMe:

I'm a mathematician working in algebra.
I've been using TeX since 1987, but couldn't write
my Ph.D. thesis with it. Sob!

I teach LaTeX at my university, besides the mathematics
courses. I'm the author of some LaTeX packages and
even of one Plain TeX package. At my site it's possible
to download my book on LaTeX programming (in Italian).

Wniosek?

-> Nie trzeba znać osoby, żeby ją znać :)

Coś bardziej ekscytującego?

Zobaczmy, gdzie się plasuje nasz zawodnik w turnieju na najwięcej napisanych postów: >Interaktywny wykres<

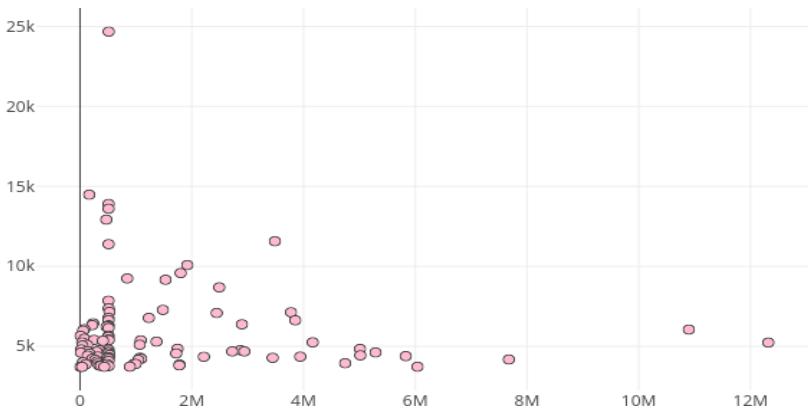
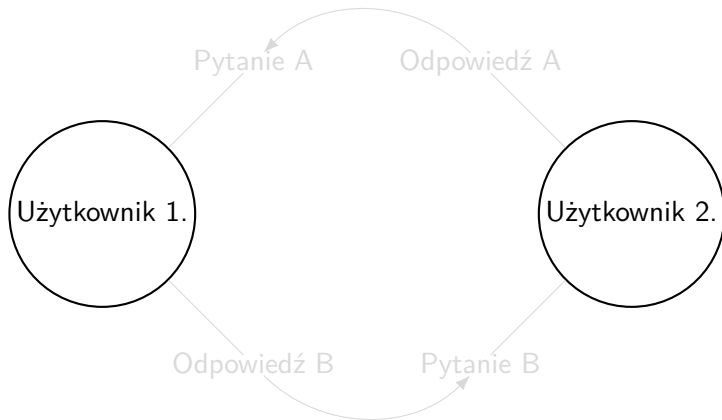


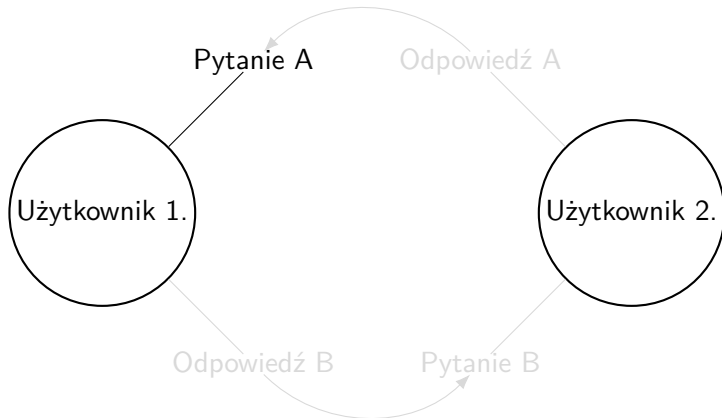
Figure 2: Aktywność 100 Najlepszych

Najbardziej zżyte ze sobą pary użytkowników

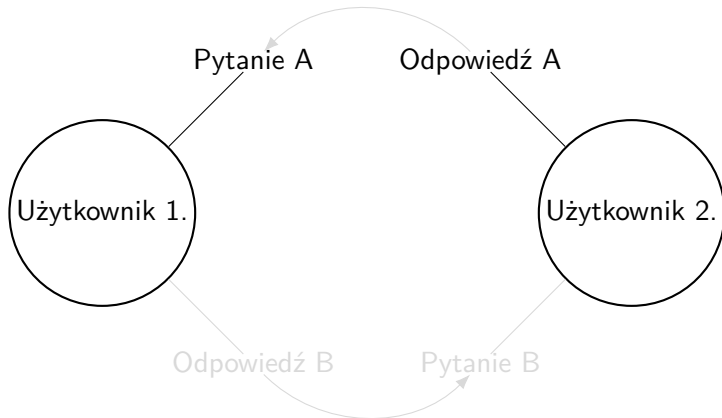
Co to znaczy być „żytym”?



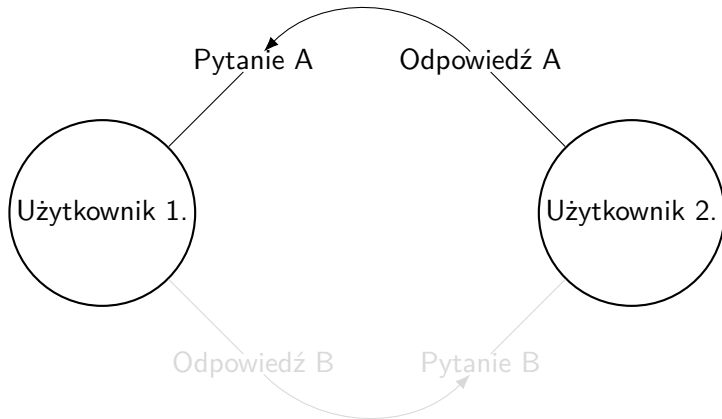
Co to znaczy być „żytym”?



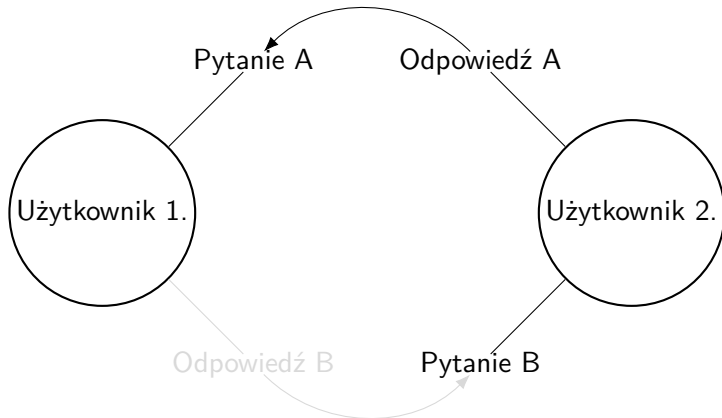
Co to znaczy być „żytym”?



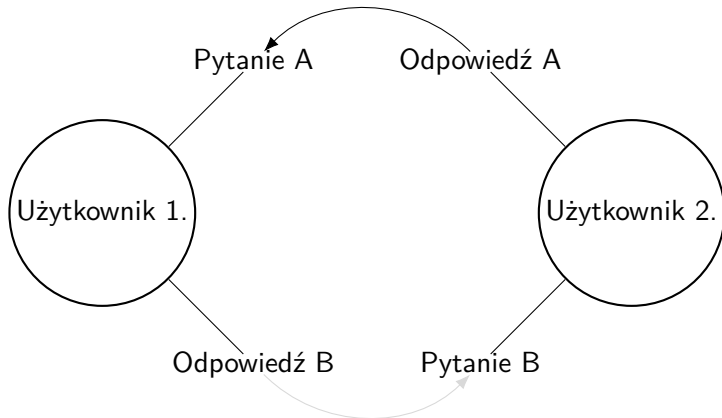
Co to znaczy być „żytym”?



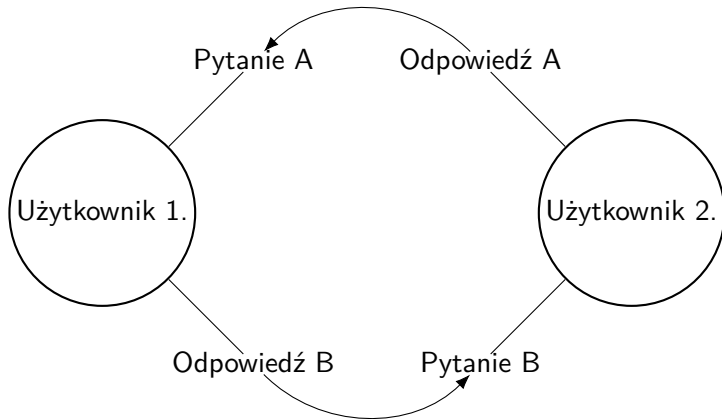
Co to znaczy być „żytym”?



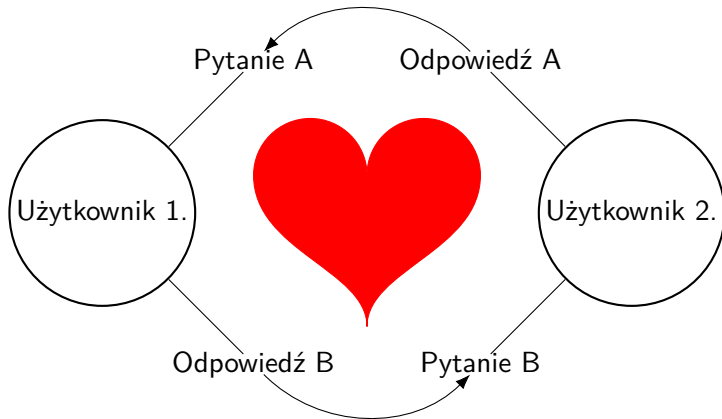
Co to znaczy być „żytym”?



Co to znaczy być „żytym”?



Co to znaczy być „żytym”?



Jak znaleźć takich kompanów?

- 1 Dzielimy posty na:
 - pytania, które mają zaakceptowane odpowiedzi,
 - odpowiedzi.
- 2 Scalamy te tabele ze sobą względem wartości:
 $(\text{Forum}, \text{Id}) == (\text{Forum}, \text{AcceptedAnswerId})$.
- 3 Tę tabelkę łączymy z użytkownikami dwukrotnie:
 - raz, żeby odczytać identyfikator konta osoby pytającej,
 - drugi, aby wydobyć identyfikator konta osoby odpowiadającej.

AccPytający	Forum	AccOdpowiadający
5581116	devops	1178508
1773383	security	61611

Jak znaleźć takich kompanów?

- 1 Dzielimy posty na:
 - pytania, które mają zaakceptowane odpowiedzi,
 - odpowiedzi.
- 2 Scalamy te tabele ze sobą względem wartości:
 $(\text{Forum}, \text{Id}) == (\text{Forum}, \text{AcceptedAnswerId})$.
- 3 Tę tabelkę łączymy z użytkownikami dwukrotnie:
 - raz, żeby odczytać identyfikator konta osoby pytającej,
 - drugi, aby wydobyć identyfikator konta osoby odpowiadającej.

AccPytający	Forum	AccOdpowiadający
5581116	devops	1178508
1773383	security	61611

Jak znaleźć takich kompanów?

- 1 Dzielimy posty na:
 - pytania, które mają zaakceptowane odpowiedzi,
 - odpowiedzi.
- 2 Scalamy te tabele ze sobą względem wartości:
 $(\text{Forum}, \text{Id}) == (\text{Forum}, \text{AcceptedAnswerId})$.
- 3 Tę tabelkę łączymy z użytkownikami dwukrotnie:
 - raz, żeby odczytać identyfikator konta osoby pytającej,
 - drugi, aby wydobyć identyfikator konta osoby odpowiadającej.

AccPytający	Forum	AccOdpowiadający
5581116	devops	1178508
1773383	security	61611

Jak znaleźć takich kompanów?

- 1 Dzielimy posty na:
 - pytania, które mają zaakceptowane odpowiedzi,
 - odpowiedzi.
- 2 Scalamy te tabele ze sobą względem wartości:
 $(\text{Forum}, \text{Id}) == (\text{Forum}, \text{AcceptedAnswerId})$.
- 3 Tę tabelkę łączymy z użytkownikami dwukrotnie:
 - raz, żeby odczytać identyfikator konta osoby pytającej,
 - drugi, aby wydobyć identyfikator konta osoby odpowiadającej.

AccPytający	Forum	AccOdpowiadający
5581116	devops	1178508
1773383	security	61611

Jak znaleźć takich kompanów?

4 Tabelkę z poprzedniego kroku:

- łączymy ze sobą po kluczach
`AccOdpowiadający == AccPytający`
- zostawiamy tylko te rekordy, gdzie
`AccPytający_Left == AccOdpowiadający_prawy`
- grupujemy po
(`Użytkownik1 = AccPytający_Left`, `Użytkownik2 = AccOdpowiadający_prawy`)

Użytkownik1	Użytkownik2	LiczbaWymian
5314232	6854495	2
6906961	6906961	1
8770670	1408765	4

Jak znaleźć takich kompanów?

- 4 Tabelkę z poprzedniego kroku:
- łączymy ze sobą po kluczach
`AccOdpowiadający == AccPytający`
 - zostawiamy tylko te rekordy, gdzie
`AccPytający_Left == AccOdpowiadający_prawy`
 - grupujemy po
(`Użytkownik1 = AccPytający_Left, Użytkownik2 =`
`AccOdpowiadający_prawy`)

Użytkownik1	Użytkownik2	LiczbaWymian
5314232	6854495	2
6906961	6906961	1
8770670	1408765	4

Jak znaleźć takich kompanów?

4 Tabelkę z poprzedniego kroku:

- łączymy ze sobą po kluczach
`AccOdpowiadający == AccPytający`
- zostawiamy tylko te rekordy, gdzie
`AccPytający_Left == AccOdpowiadający_prawy`
- grupujemy po
`(Użytkownik1 = AccPytający_Left, Użytkownik2 =
AccOdpowiadający_prawy)`

Użytkownik1	Użytkownik2	LiczbaWymian
5314232	6854495	2
6906961	6906961	1
8770670	1408765	4

Jak znaleźć takich kompanów?

- 4 Tabelkę z poprzedniego kroku:
- łączymy ze sobą po kluczach
`AccOdpowiadający == AccPytający`
 - zostawiamy tylko te rekordy, gdzie
`AccPytający_Left == AccOdpowiadający_prawy`
 - grupujemy po
(`Użytkownik1 = AccPytający_Left`, `Użytkownik2 = AccOdpowiadający_prawy`)

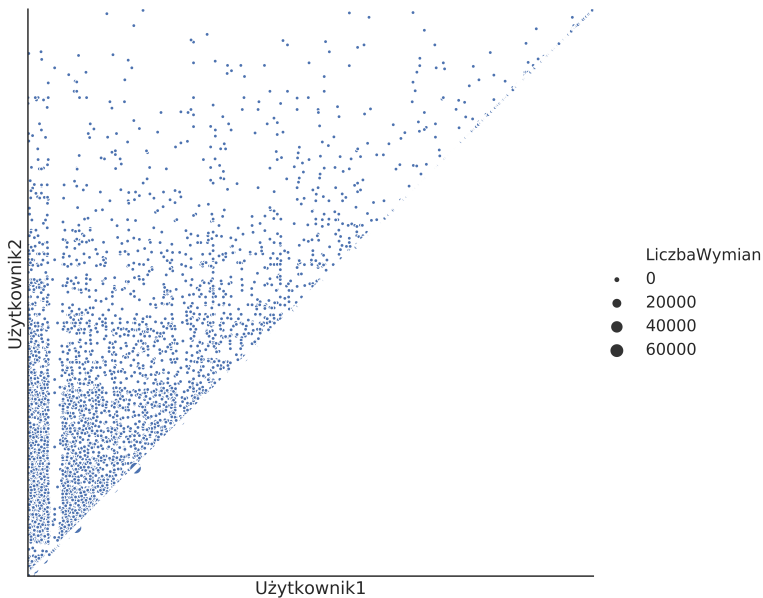
Użytkownik1	Użytkownik2	LiczbaWymian
5314232	6854495	2
6906961	6906961	1
8770670	1408765	4

Jak znaleźć takich kompanów?

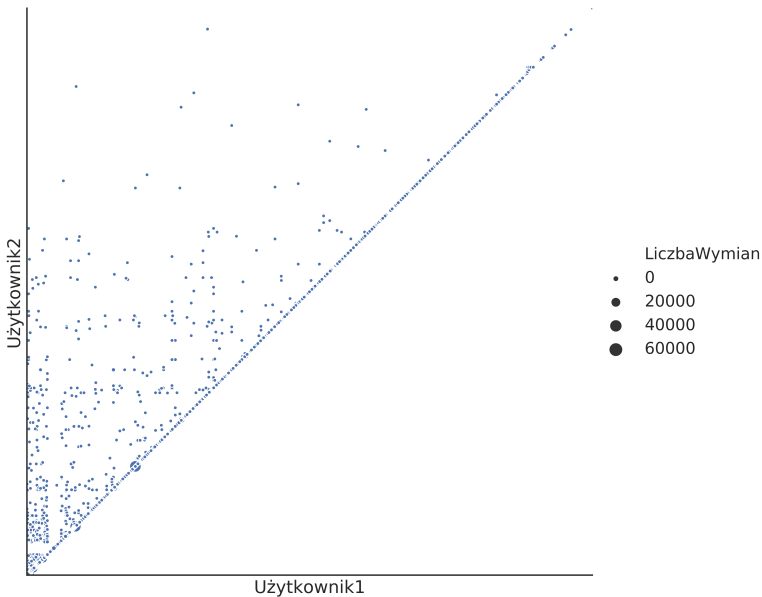
- 4 Tabelkę z poprzedniego kroku:
- łączymy ze sobą po kluczach
`AccOdpowiadający == AccPytający`
 - zostawiamy tylko te rekordy, gdzie
`AccPytający_Left == AccOdpowiadający_prawy`
 - grupujemy po
(`Użytkownik1 = AccPytający_Left`, `Użytkownik2 = AccOdpowiadający_prawy`)

Użytkownik1	Użytkownik2	LiczbaWymian
5314232	6854495	2
6906961	6906961	1
8770670	1408765	4

Kto z kim się lubi?



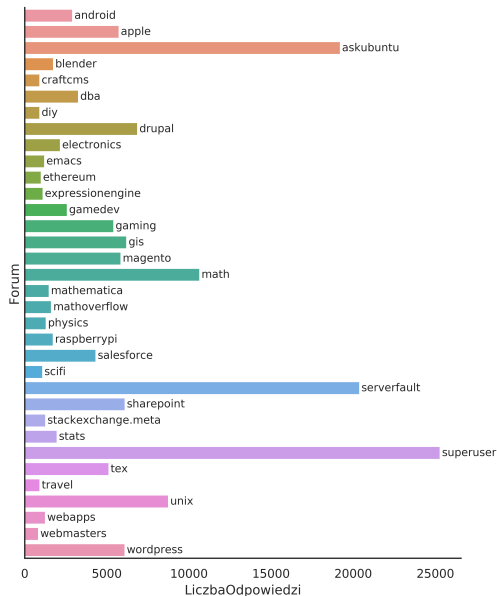
Kto z kim się lubi bardziej niż 10 razy?



Podsumowanie internetowych zbliżeń

Informacja	Soliści	Tandemy
Liczba użytkowników	93465	9094
Średnia liczba interakcji	19.63	5.91
Maksymalna liczba interakcji	53361	1716

Na których forach spotkamy najwięcej solistów?



Aktywność a wydarzenia

Jak to zrobić?

Każdy wynik będzie opierał się o podobny schemat działania:

1. Grupujemy ilość postów po dacie.
2. Zliczamy ilość wystąpień postów danego dnia.

Jak to zrobić?

Każdy wynik będzie opierał się o podobny schemat działania:

1. Grupujemy ilość postów po dacie.
2. Zliczamy ilość wystąpień postów danego dnia.

Jak to zrobić?

Każdy wynik będzie opierał się o podobny schemat działania:

1. Grupujemy ilość postów po dacie.
2. Zliczamy ilość wystąpień postów danego dnia.

Rozgrzewka

Zobaczmy jaka jest zależność, między posiadaną ilością kont na forach, a aktywnością użytkowników.

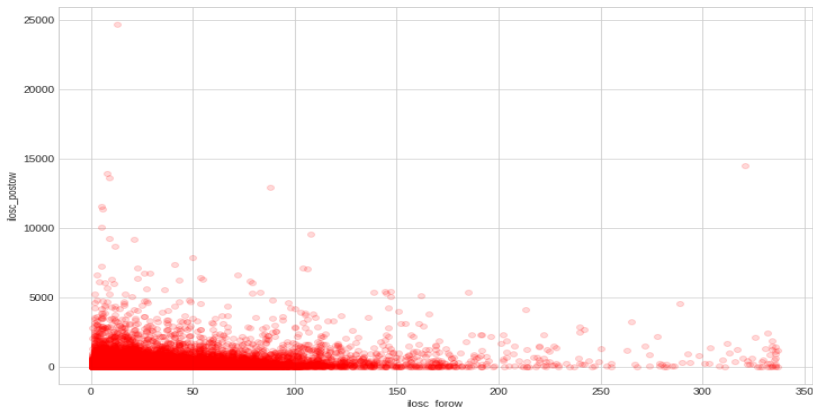


Figure 3: Aktywność vs Ilość forów

Ukraina

Mimo, że ukrainian forum ma 1806 użytkowników, to na aktywność na nim dba tylko 4 użytkowników. >Wykres Interaktywny<

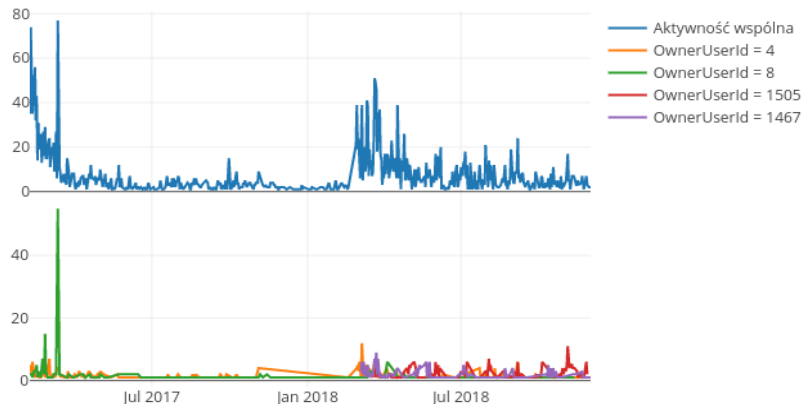


Figure 4: Ukrainian forum

Bitcoin

Możemy śmiało powiedzieć, że kurs bitcoina wpłynął na jego zainteresowanie. >Wykres interaktywny<

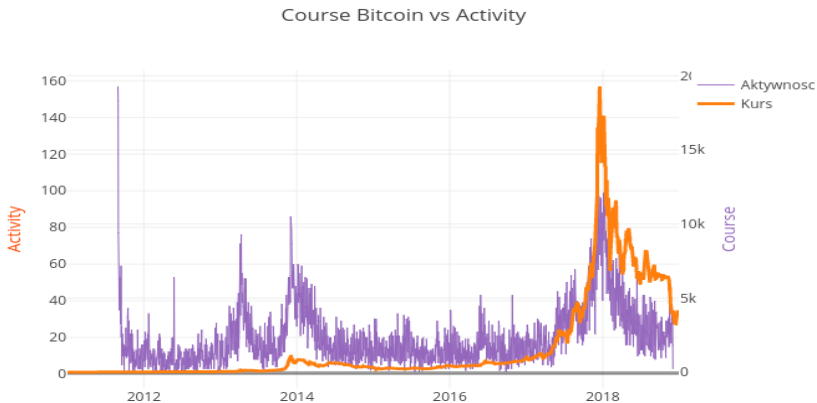


Figure 5: Bitcoin

Rolnictwo

Zauważmy jak bardzo zmienia się aktywność na forum, gdy zbliża się sezon. >Wykres Interaktywny<

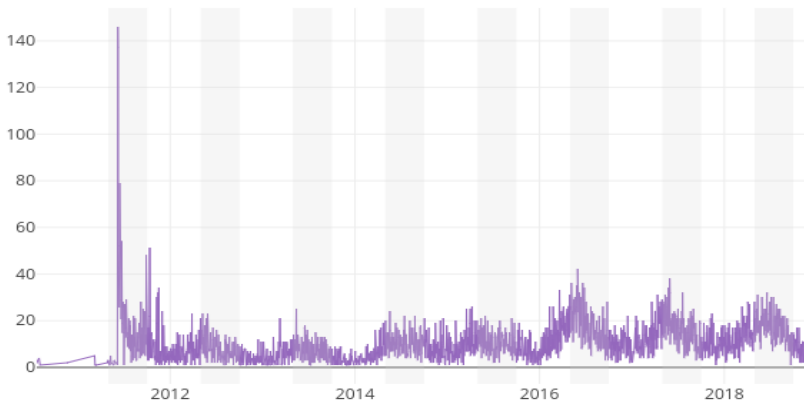


Figure 6: Rolnictwo