

SQL - projekt - prezentacja

2018-03-11

Kurs Junior Data Scientist Zaoczne 1 (JDSZ1)

Raczki

Bartosz, Filip, Monika Kucal, Piotr

- 1. Czym się zajmujemy?**
- 2. Diagram bazy danych**
- 3. Raporty informacyjne**
- 4. Raporty analityczne**
- 5. Predykcje**

1.

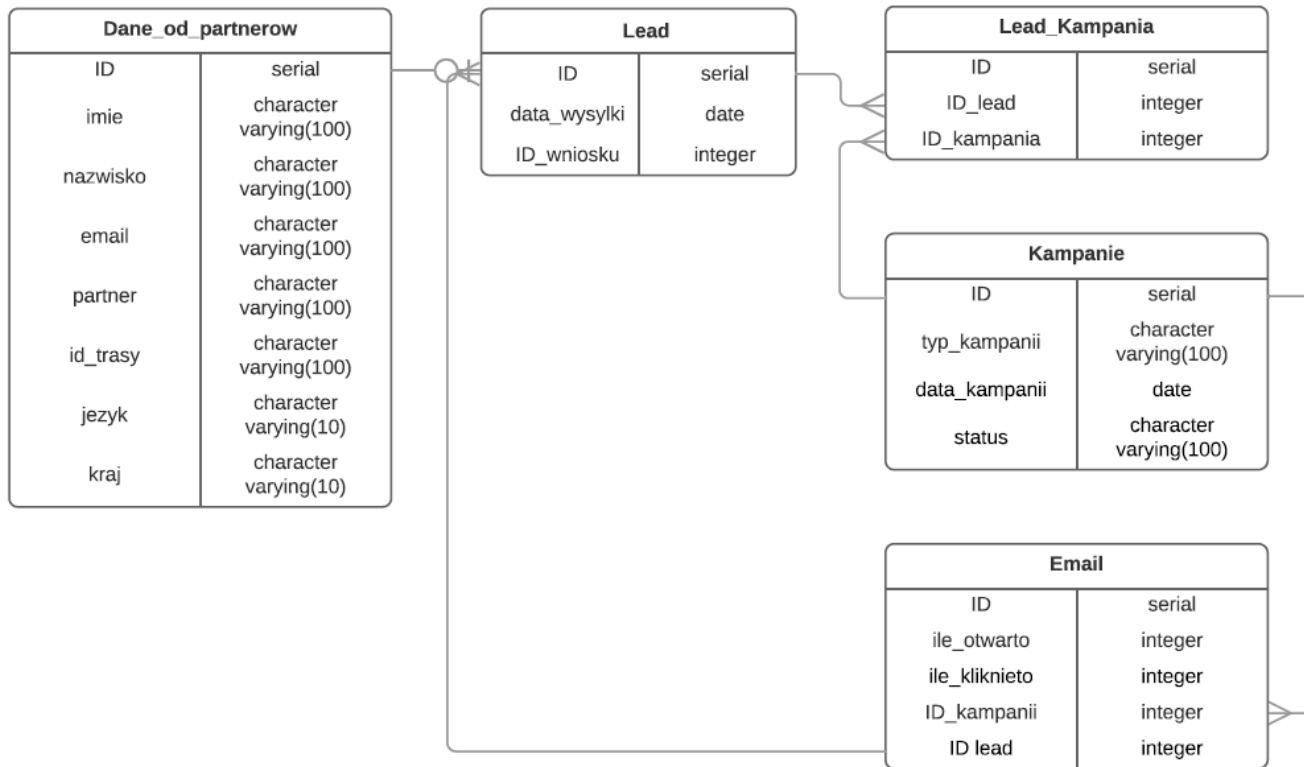
Czym się zajmujemy jako grupa?

Czym się zajmujemy jako grupa?

- Analitycy **działu marketingu** firmy opoznionypociag.pl;
- **Baza klientów** na podstawie danych o delegacjach pracowników firm zewnętrznych (w ramach modelu B2B);
- **Kampanie marketingowe** obejmujące klientów (lead'y), którym należy się rekompensata ze względu na zakłócenia w podróży; **kampanie przypominające** w przypadku braku skuteczności kampanii głównej;
- Raporty:
 - statystyki lead'ów i wysłanych maili,
 - efektywność kampanii i partnerów współpracujących,
 - prognoza lead'ów w kolejnych miesiącach.

2. Diagram bazy danych

Diagram bazy danych



3.

Raporty informacyjne (3)

Raport 1 - Lead'y przeterminowane

- **Cel**
 - Bieżące śledzenie listy lead'ów bliskich przeterminowania, tak aby zdążyć wysłać maila do klienta przed upływem terminu 3 lat od daty wyjazdu;
- **Dla kogo**
 - Zespół analizujący lead'y.

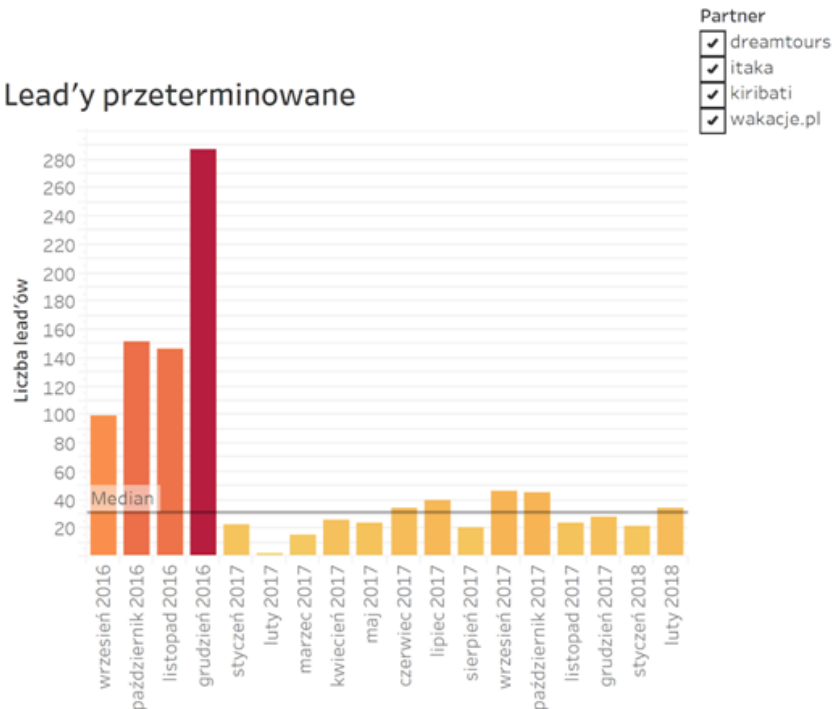
Raport 1 - Lead'y przeterminowane - Tableau

Analiza przeterminowania lead'ów

Lead'y bliskie przeterminowania

ID lead	Kraj	Partner	Month of deadline	Liczba dni
75696	PL	wakacje.pl	marzec 2018	0
75738	PL	dreamtours	marzec 2018	0
75740	PL	dreamtours	marzec 2018	0
75741	PL	dreamtours	marzec 2018	0
75753	PL	wakacje.pl	marzec 2018	0
75767	PL	dreamtours	marzec 2018	0
75791	PL	wakacje.pl	marzec 2018	0
75802	PL	dreamtours	marzec 2018	0
75807	PL	dreamtours	marzec 2018	0
381971	US	kiribati	marzec 2018	0
75832	PL	wakacje.pl	marzec 2018	1
75860	PL	dreamtours	marzec 2018	1
75877	PL	dreamtours	marzec 2018	1
75880	PL	dreamtours	marzec 2018	1
75883	PL	dreamtours	marzec 2018	1
75885	PL	dreamtours	marzec 2018	1
75896	PL	wakacje.pl	marzec 2018	2
75917	PL	wakacje.pl	marzec 2018	2
110947	PL	wakacje.pl	marzec 2018	2

Lead'y przeterminowane



Liczba dni do przeterminowania

0 118

Liczba lead'ów przeterminowanych

2 287

Raport 1 - Lead'y przeterminowane

- **Impakt**

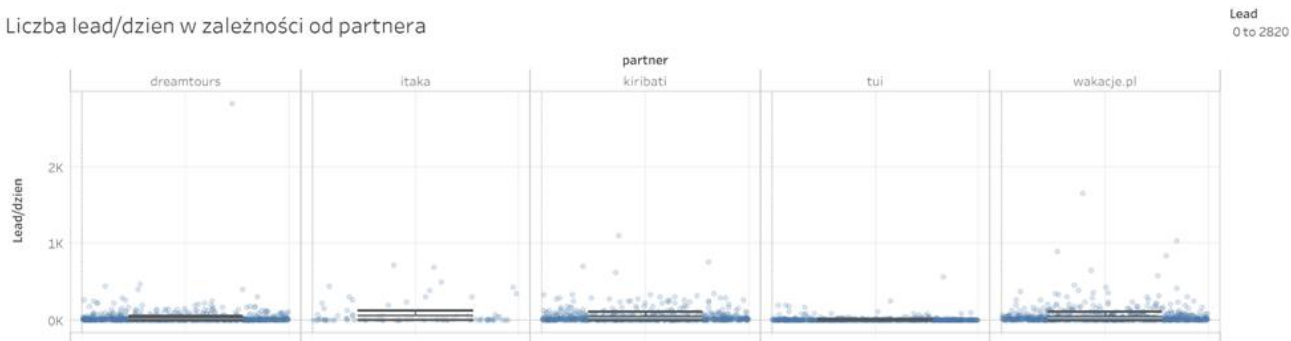
- Przeterminowany lead - utrata 25% prowizji od ewentualnie zaakceptowanego:
 - 30 lead'ów przeterminowanych w każdym miesiącu;
 - Średnia wartość lead'a: 300 zł;
 - $30 * 300 \text{ zł} * 12 \text{ miesięcy} * 25\% = \mathbf{27\ 000 \text{ zł prowizji}}$;
- Priorytetyzacja wysyłki lead'ów:
 - wzrost satysfakcji klientów;
 - minimalizacja utraty prowizji.

Raport 2 - Statystyki leadów

- **Cel**
 - Monitorowanie **ilości leadów od poszczególnych partnerów** w postaci średniej ilości leadów **na dzień** oraz zmian ilości leadów **na przestrzeni czasu**;
- **Dla kogo**
 - Zespół do spraw **współpracy z partnerami** i nawiązywania współpracy z nowymi partnerami.

Raport 2 - Statystyki leadów

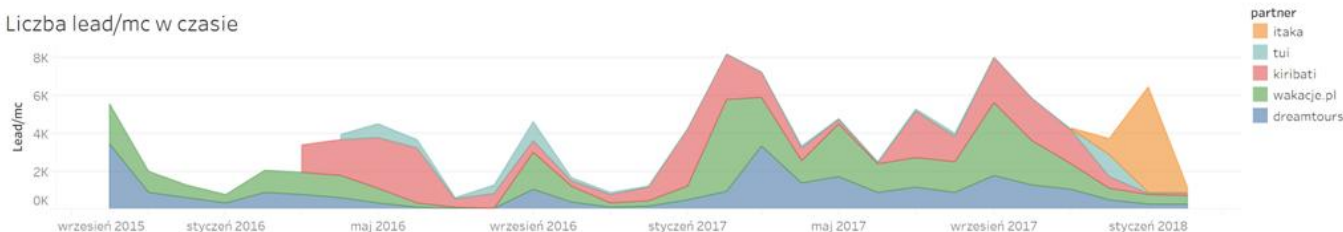
Liczba lead/dzien w zależności od partnera



Statystyki lead/dzien w zależności od partnera

	partner				
	dreamtours	itaka	kiribati	tui	wakacje.pl
Q1	3	2	4	0	5
Mediana	8	8	14	1	14
Q3	27	51	46	4	46
Średnia	30	69	43	8	45
Max lead/dzien	2 820	722	1 106	561	1 648

Liczba lead/mc w czasie



Raport 2 - Statystyki leadów

- **Impakt**

- Informacja o ilości leadów pozwala na **zidentyfikowanie kluczowych partnerów** firmy.
- Zysk z jednego leada:
 - $300 \text{ zł} * 25\% \text{ prowizji} * 15\% \text{ win rate} = \mathbf{11,5 \text{ zł/otrzymany lead}}$
- Porównując informację o ilości leadów do wielkości firmy partnerskiej, ilości realizowanych przez nią podróży oraz informacji o przewoźnikach z którymi współpracuje można określić potencjalny zysk w zależności od partnera.

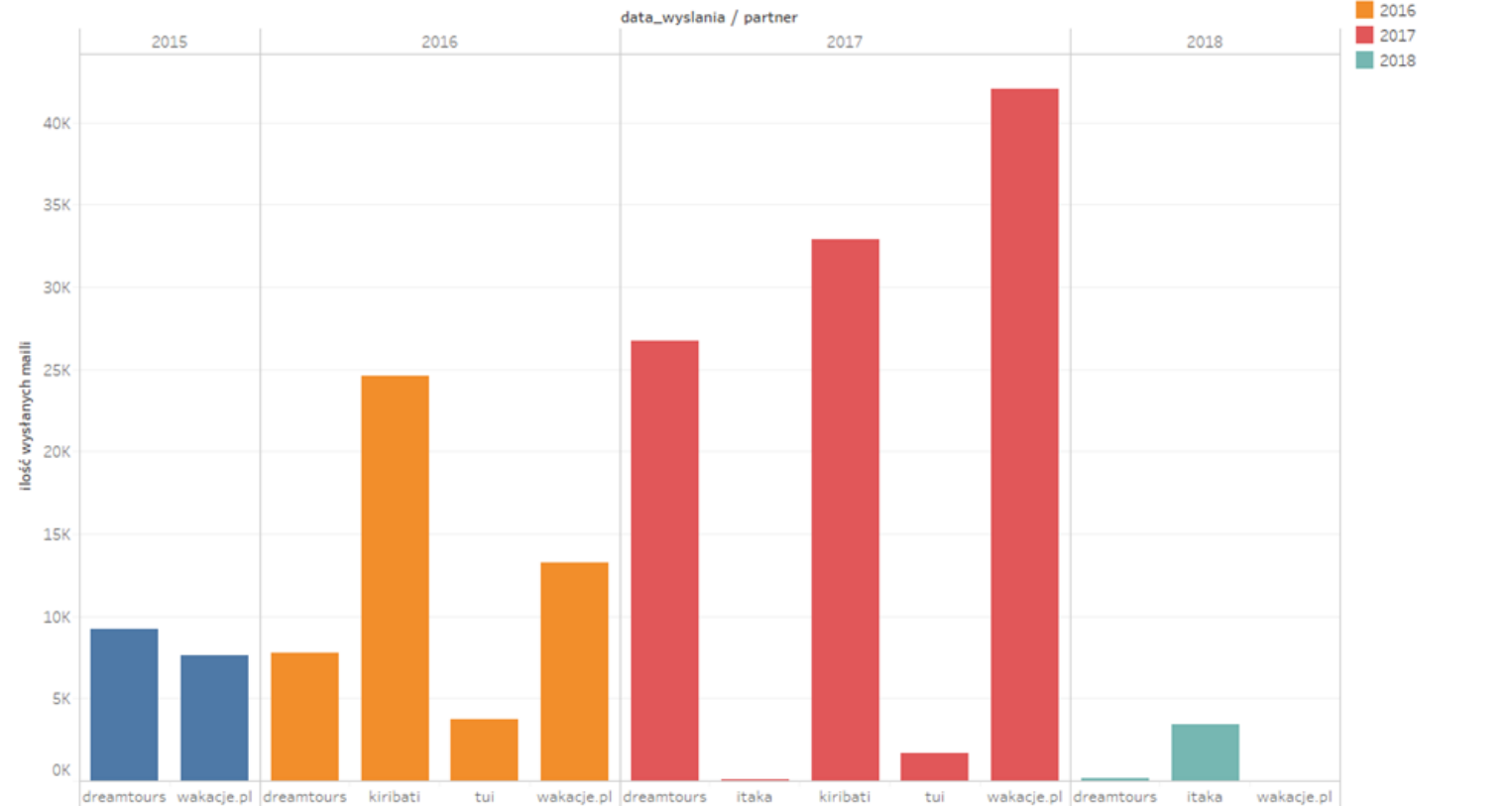
Partner	LEAD [n/dzień]	ZYSK [PLN/dzień]
<i>dreamtours</i>	30	345,00
<i>wakacje.pl</i>	69	793,50
<i>kiribati</i>	43	494,50
<i>tui</i>	8	92,00
<i>itaka</i>	45	517,50

Raport 3 - Email

- **Cel**
 - Monitorowanie **ilości wysyłanych maili dla poszczególnych partnerów na dzień / miesiąc / kwartał / rok** celem sprawdzania opłacalności współpracy z danym partnerem z uwagi na ilość wygranych spraw;
- **Dla kogo**
 - Zespół do spraw **współpracy z partnerami** i nawiązywania współpracy z nowymi partnerami.

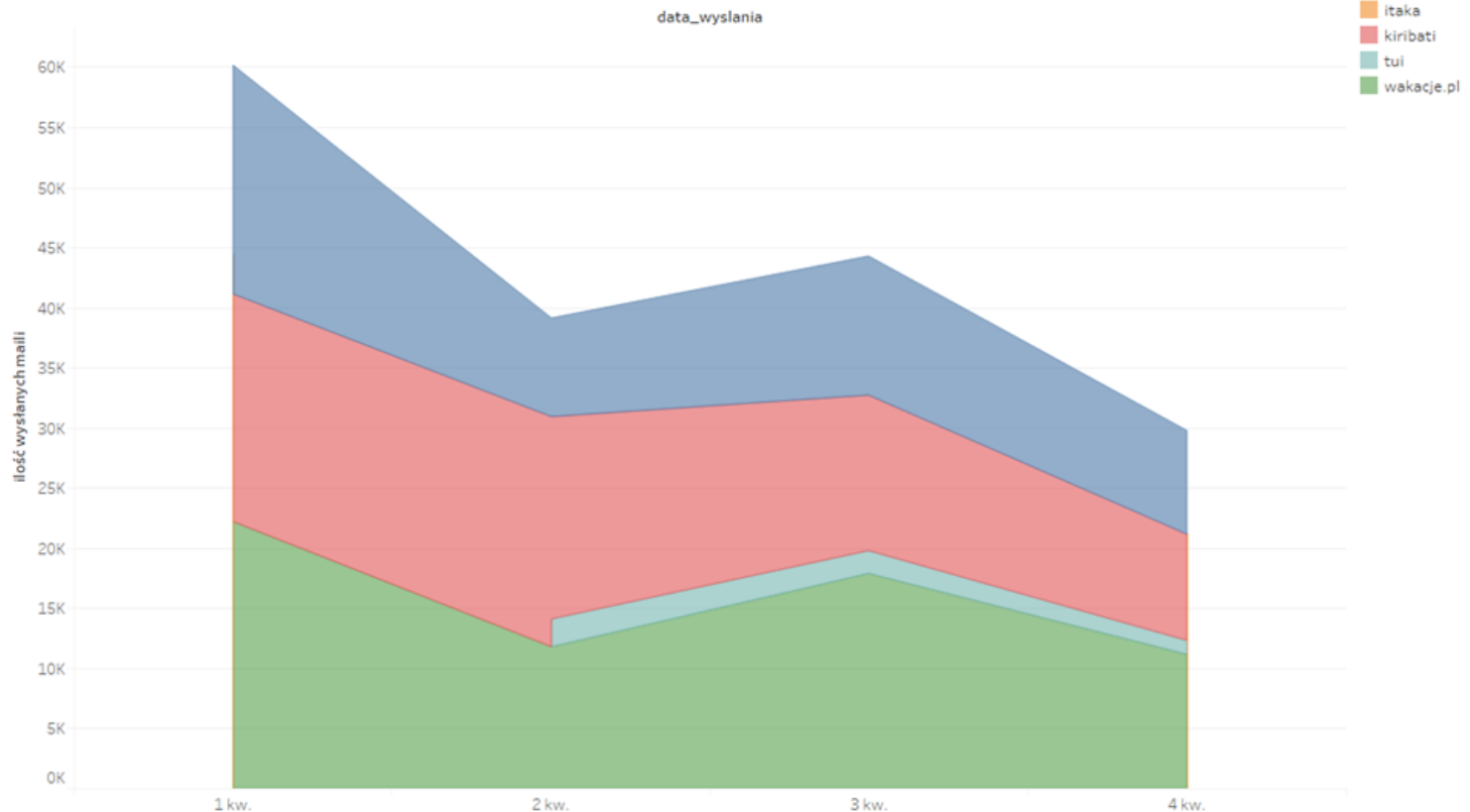
Raport 3 - Email

Ilość wysłanych maili w roku w zależności od partnera



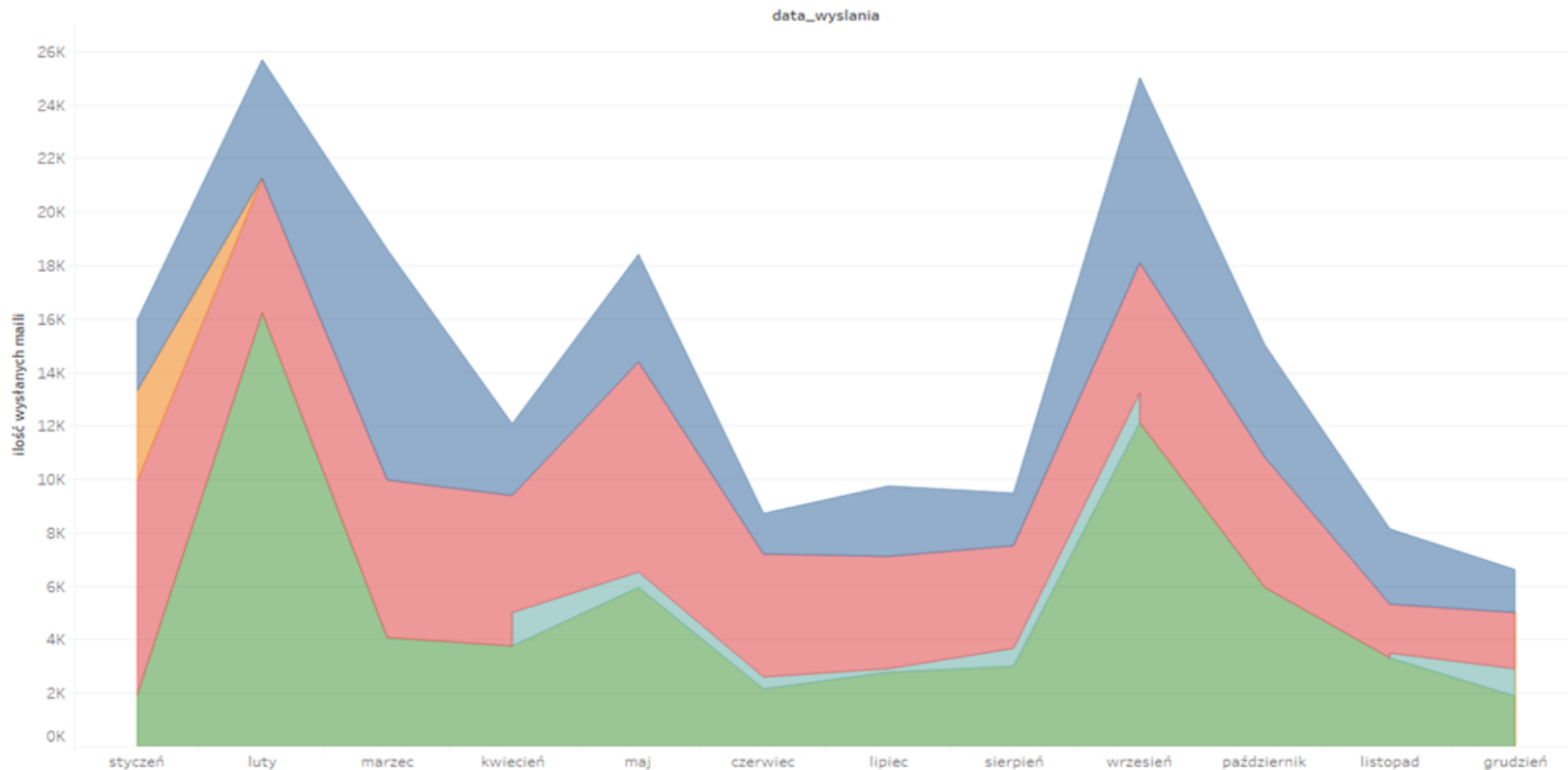
Raport 3 - Email

Ilość wysłanych maili w poszczególnych kwartałach



Raport 3 - Email

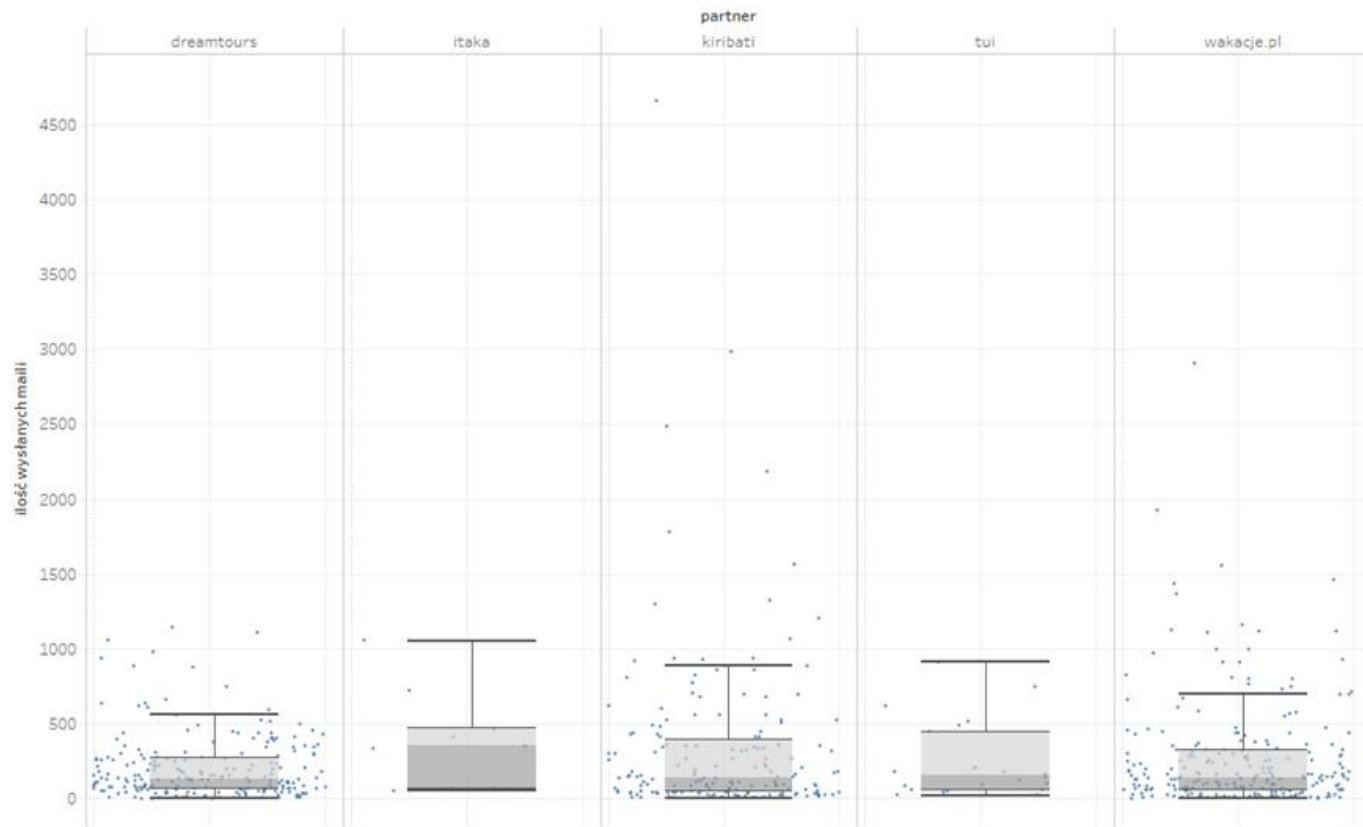
Ilość wysłanych maili w poszczególnych miesiącach



Raport 3 - Email

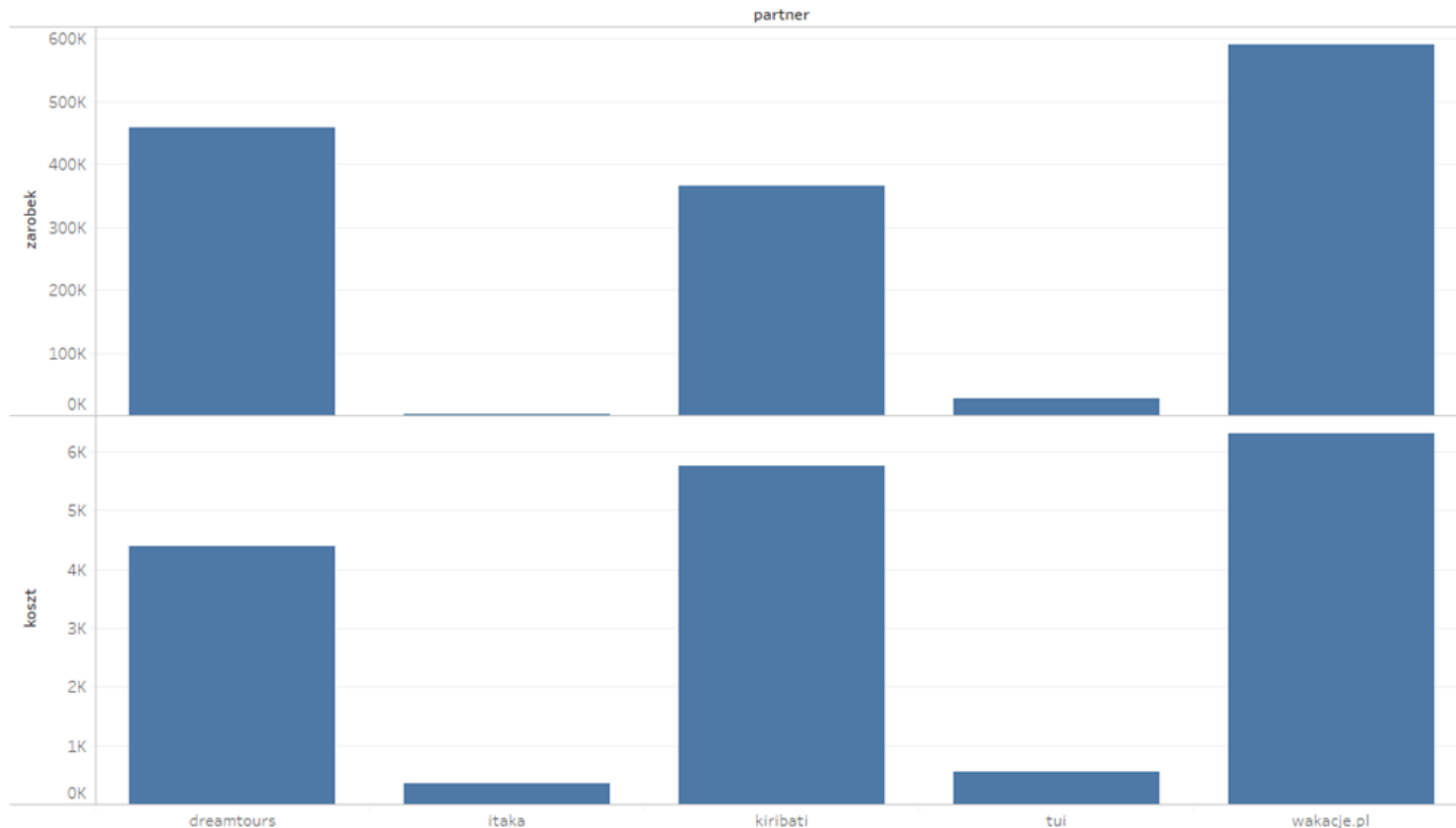
Ilość wysłanych mail według dni przez partnerów

ilość wysłanych maili
1 to 4653



Raport 3 - Email

Partnerzy - zarobek wynikający z wygranych spraw vs koszt wysłania maila



Raport 3 - Email

- **Impakt**

- Informacja o wartości przychodu od poszczególnych partnerów i wygenerowanych kosztach pozwala na **zidentyfikowanie kluczowych partnerów** firmy;
- ITAKA jest partnerem, który generuje 2 663 zł przychodu przy kosztach 352 zł stanowi to 13%. Jest to partner, z którym niedawno rozpoczęliśmy współpracę, jego analiza wymaga dłuższej współpracy;
- Średni roczny koszt wysłania maili w skali roku to **13 200 zł**.

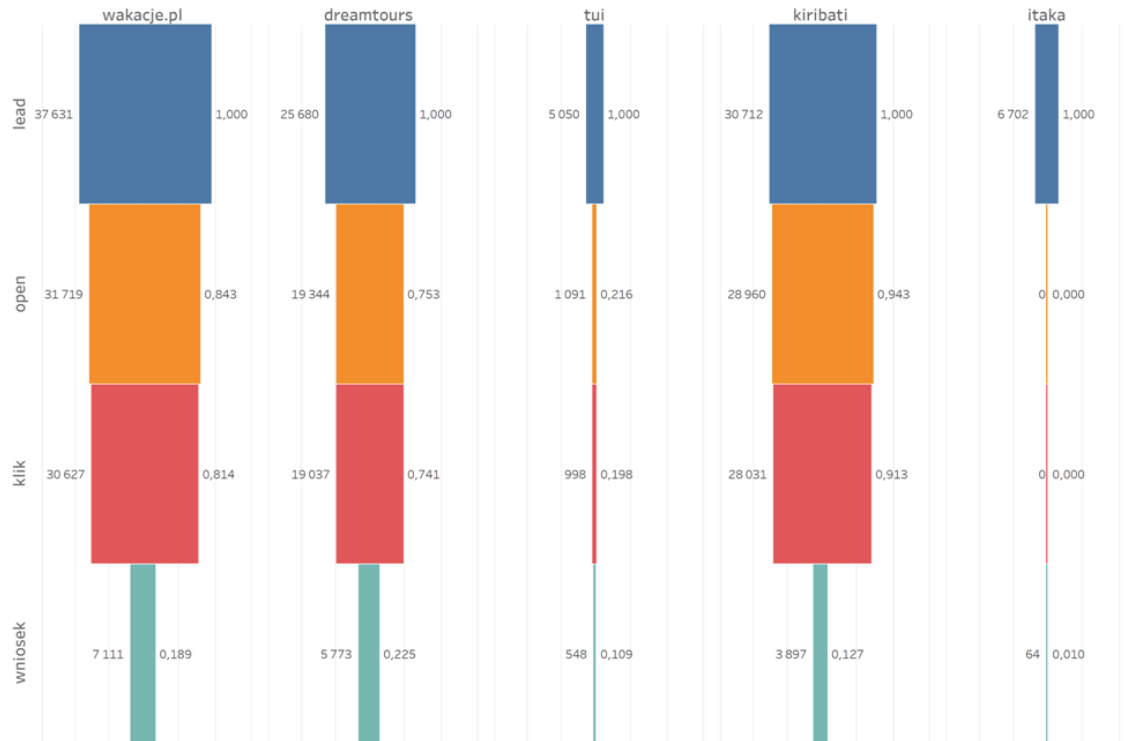
4. Raporty analityczne (4)

Raport 4 - Efektywność partnerów

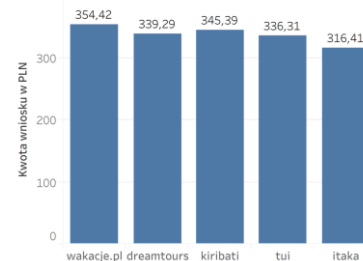
- **Cel**
 - Określenie **procedowania lead'u do wniosku** zależności od partnera od którego pochodzi lead poprzez sprawdzenie wskaźników: **otwarcia wiadomości, kliknięcie odnośnika i złożenia wniosku.**
- **Dla kogo**
 - Zespół odpowiedzialny za **tworzenie kampanii.**

Raport 4 - Efektywność partnerów

Konwersja lead'ów do wniosków w zależności od partnera



Średnia kwota wniosku w zależności od partnera



Raport 4 - Efektywność partnerów

- **Impakt**

- **Niskie wskaźniki konwersji** lead'ów do wniosków wpływają na **wzrost kosztów współpracy** z partnerami oraz **nieuzyskaniem przychodów** od procedowanych lead'ów.
- Najlepszym współpracującym partnerem jest **dreamtours** o współczynniku konwersji leadów **22,5%**;
- Zwiększenie współczynnika konwersji leadów, np. dedykowane kampanie może przynieść dodatkowy przychód:
 - Wakacje.pl: 25185 wniosków/rok * 3,6% wzrost konwersji * 300 zł * 25% prowizji = **68 000 zł**
 - Tui: 2920 wniosków/rok * 11,7% wzrost konwersji * 300 zł * 25% prowizji = **25 623 zł**
 - Kiribati: 15695 wniosków/rok * 9,8% wzrost konwersji * 300 zł * 25% prowizji = **115 360 zł**
 - Itaka: 16425 wniosków/rok * 21,6% wzrost konwersji * 300 zł * 25% prowizji = **266 085 zł**

Raport 5 - Efektywność kampanii

- **Cel**
 - Określenie **procedowania leadu** w zależności od kampanii od którego pochodzi lead poprzez sprawdzenie wskaźników: **otwarcia wiadomości, kliknięcie odnośnika i złożenia wniosku;**
- **Dla kogo**
 - Zespół odpowiedzialny za **tworzenie kampanii.**

Raport 5 - Efektywność kampanii

Efektywność kampanii



Raport 5 - Efektywność kampanii

- **Impakt**

- **Niskie wskaźniki konwersji** dla pierwszego i drugiego remindera wpływają na **wzrost kosztów wynikających** z tworzenia kolejnych kampanii przypominających;
- Najlepszym współczynnikiem konwersji ma reminder trzeci wynoszący **21%**;
- Poprawienie kampani przypominającej pierwszej albo drugiej może mieć znaczący wpływ na zwiększenie przychodu:
 - $\text{Zysk z 1 leda} = 300 \text{ zł} * 25\% \text{ prowizji} * 15\% \text{ win rate} = 11,5 \text{ zł/otrzymany lead};$
 - **Reminder 1** - zwiększenie konwersji do 15% (wzrost o ok. 5%, co daje ok 1 400 leadów więcej) zwiększy przychód o 16 000 zł/rok;
 - **Reminder 2** - zwiększenie konwersji do 20% (wzrost o ok. 10%, co daje ok 1 730 leadów więcej) zwiększy przychód o 20 000 zł/rok.

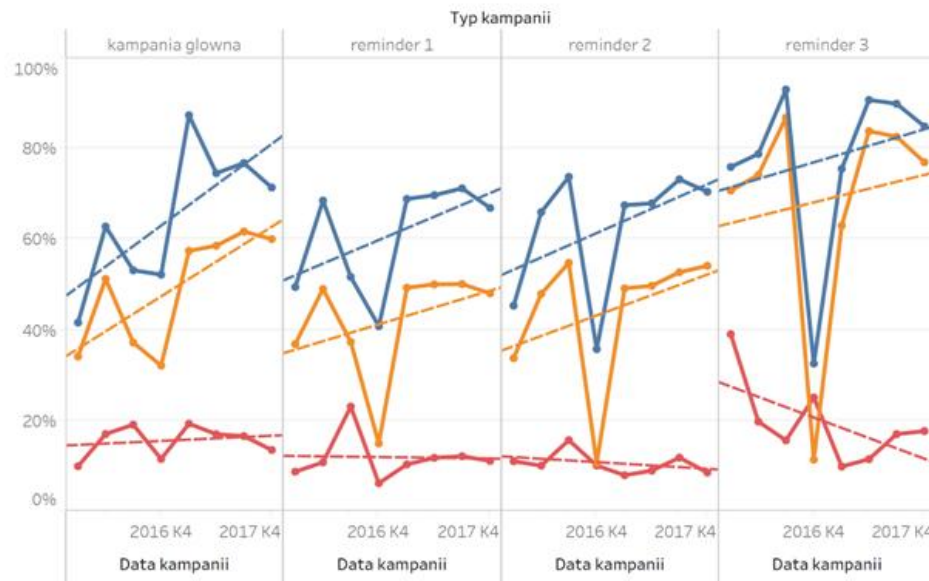
Raport 6 - Open rate & Click rate & Konwersja

- **Cel**
 - Kontrolowanie poziomu wskaźników Open Rate, Click Rate i Konwersji oraz obserwacja ich trendów w czasie;
- **Dla kogo**
 - Zespół monitorujący skuteczność wysłanych maili do klientów.

Raport 6 - Open rate & Click rate & Konwersja - Tableau

Open Rate & Click Rate & Konwersja

Wskaźniki w zależności od daty kampanii



Wskaźniki
■ Open Rate
■ Click Rate
■ Convert Rate

Test t - Zmienność wskaźników w czasie

	2017 K1	2017 K2	2017 K3	2017 K4
Open Rate	79,86%	73,99%	75,85%	70,94%
Click Rate	55,19%	57,66%	59,37%	56,81%
Convert Rate	15,11%	14,16%	15,28%	12,26%

W celu zbadania zależności wskaźników Open Rate, Click Rate i Konwersji od daty wysłania maila (bez względu na typ kampanii) przeprowadzony został Test t na parach kwartałów w 2017 r.: K1 vs. K2, K2 vs. K3, K3 vs. K4.

Wyniki testu

Open Rate - średnie między kolejnymi kwartałami są istotnie różne od siebie
 Click Rate - średnie między kolejnymi kwartałami są istotnie różne od siebie
 Konwersja - średnie między kolejnymi kwartałami są istotnie różne od siebie (z wyjątkiem pary K1 vs. K2)

Interpretacja

Wartości wszystkich wskaźników zależą od daty wysłania maila.

Typ kampanii
All

Język
All

Raport 6 - Open rate & Click rate & Konwersja

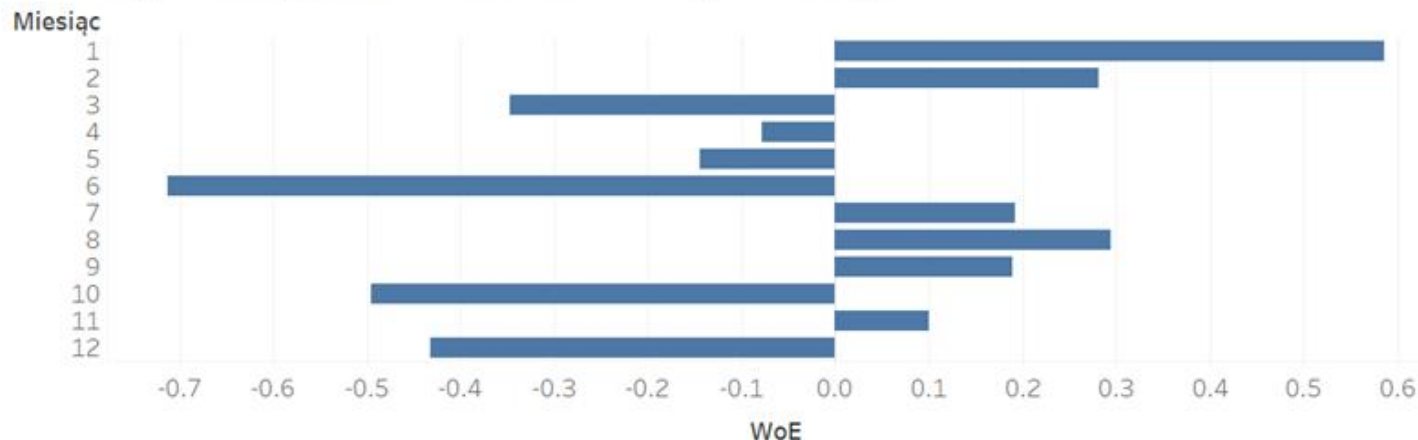
- **Impakt**

- W 2017 K4 zmniejszenie poziomu Open Rate o 6% i Click Rate o 4% powoduje spadek Konwersji o 20%;
- W celu poprawy wskaźników planowana jest nowa formuła kampanii (atrakcyjny tytuł i treść maila), która obejmie 25% maili. Spodziewany wzrost Konwersji spowoduje wzrost Win Rate do poziomu 15%:
 - Roczny koszt: $25\% * 11\ 000 * 0,10\ \text{zł} * 12 = 3\ 300\ \text{zł}$
 - Roczny przychód: $25\% * 11\ 000 * 300\ \text{zł} * 25\% * 15\% * 12 = 371\ 250\ \text{zł}$
 - Roczny zysk: $371\ 250\ \text{zł} - 3\ 300\ \text{zł} = \mathbf{367\ 950\ \text{zł zysku}}$

Raport 7 - Wpływ języka, kraju i daty kampanii na konwersję do wniosku

- **Cel:**
 - Analiza wpływu zmiennych na wynik kampanii, w celu poprawy konwersji;
- **Dla kogo**
 - Zespół odpowiedzialny za analizę skuteczności kampanii;
 - Zespół odpowiedzialny za tworzenie nowych kampanii.

Miesiąc kampanii - WoE & Interpretacja



IV = 0.130498

W celu sprawdzenia wpływu języka na konwersję wniosków wykorzystano technikę Information Value.

Miesiąc	Pozytywne	Negatywne	Wszystkie	DG	DB	DG - DB	WoE	IV_part
1	6,548	17,472	24,020	0.141127	0.078568	0.062559	0.585696	0.036640
2	6,133	22,197	28,330	0.132182	0.099815	0.032367	0.280863	0.009091
3	2,672	18,115	20,787	0.057589	0.081459	-0.023871	-0.346777	0.008278
4	2,205	11,436	13,641	0.047524	0.051425	-0.003902	-0.078903	0.000308
5	3,151	17,454	20,605	0.067912	0.078487	-0.010575	-0.144713	0.001530
6	909	8,881	9,790	0.019591	0.039936	-0.020345	-0.712189	0.014489
7	2,660	10,519	13,179	0.057330	0.047302	0.010028	0.192279	0.001928
8	2,673	9,553	12,226	0.057610	0.042958	0.014652	0.293482	0.004300
9	7,461	29,587	37,048	0.160804	0.133046	0.027758	0.189490	0.005260
10	6,050	47,569	53,619	0.130394	0.213908	-0.083514	-0.494987	0.041338
11	4,670	20,250	24,920	0.100651	0.091060	0.009591	0.100140	0.000960
12	1,266	9,348	10,614	0.027286	0.042036	-0.014750	-0.432164	0.006375

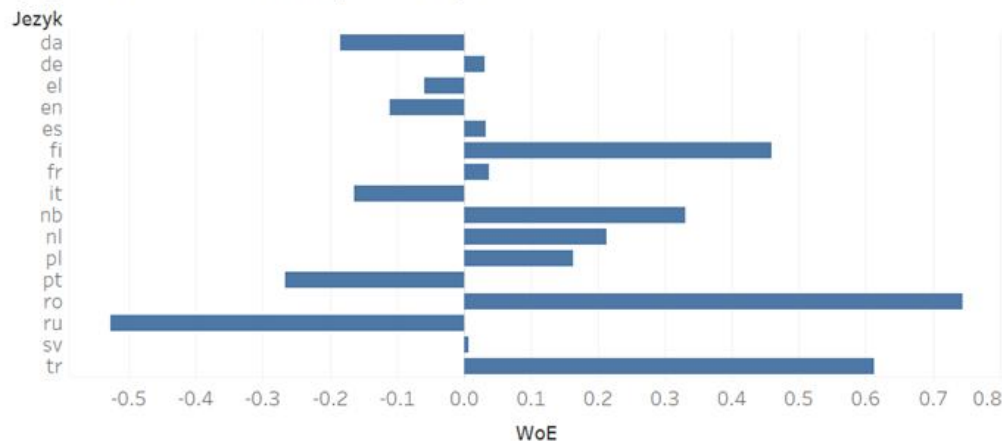
Interpretacja WoE:

> 0 - DG > DB

< 0 - DG < DB

Interpretacja: wartość IV wskazuje na słabą wartość predykcyjną.

Język - WoE & Interpretacja



IV = 0.0342439

W celu sprawdzenia wpływu języka na konwersję wniosków wykorzystano technikę Information Value.

Język	Pozytywne	Negatywne	Wszystkie	DG	DB	DG-DB	WoE	IV_part
en	24,101	129,029	153,130	0.519440	0.580216	-0.060775	-0.110648	0.006725
sv	4,161	19,813	23,974	0.089681	0.089095	0.000586	0.006553	0.000004
nb	5,207	17,948	23,155	0.112225	0.080708	0.031516	0.329661	0.010390
fr	3,295	15,204	18,499	0.071016	0.068369	0.002647	0.037983	0.000101
fi	3,394	10,283	13,677	0.073150	0.046240	0.026909	0.458653	0.012342
es	1,547	7,183	8,730	0.033342	0.032300	0.001042	0.031736	0.000033
de	1,206	5,607	6,813	0.025992	0.025213	0.000779	0.030429	0.000024
nl	1,128	4,372	5,500	0.024311	0.019660	0.004651	0.212361	0.000988
it	755	4,261	5,016	0.016272	0.019161	-0.002889	-0.163406	0.000472
pl	569	2,318	2,887	0.012263	0.010424	0.001840	0.162556	0.000299
da	394	2,270	2,664	0.008492	0.010208	-0.001716	-0.184048	0.000316
pt	278	1,740	2,018	0.005992	0.007824	-0.001833	-0.266884	0.000489
ru	211	1,713	1,924	0.004548	0.007703	-0.003155	-0.527008	0.001663
el	104	529	633	0.002241	0.002379	-0.000137	-0.059462	0.000008
ro	43	98	141	0.000927	0.000441	0.000486	0.743368	0.000361
tr	5	13	18	0.000108	0.000058	0.000049	0.611624	0.000030

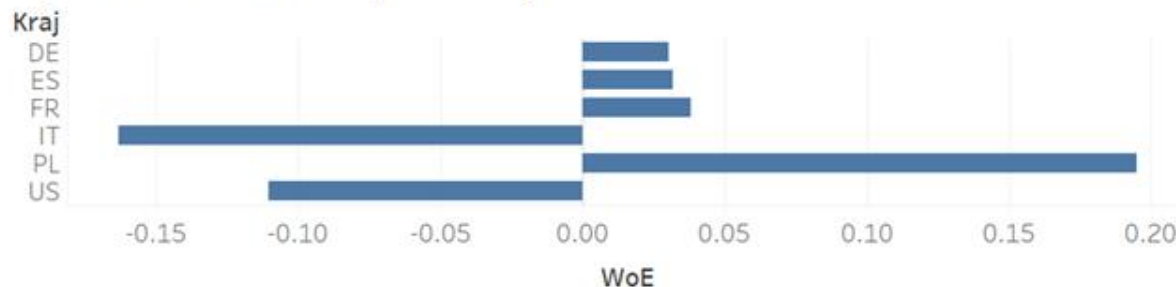
Interpretacja WoE:

>0 - DG > DB

<0 - DG < DB

Interpretacja: wartość IV wskazuje na słabą wartość predykcyjną.

Kraj - WoE & Interpretacja



IV = 0.018905

W celu sprawdzenia wpływu kraju na konwersję wniosków wykorzystano technikę Information Value.

Kraj	Pozytyw..	Negatyw..	Wszystkie	DG	DB	DG - DB	WoE	IV_part
DE	1,206	5,607	6,813	0.025992	0.025213	0.000779	0.030429	0.00002
ES	1,547	7,183	8,730	0.033342	0.032300	0.001042	0.031736	0.00003
FR	3,295	15,204	18,499	0.071016	0.068369	0.002647	0.037983	0.00010
IT	755	4,261	5,016	0.016272	0.019161	-0.002889	-0.163406	0.00047
PL	15,494	61,097	76,591	0.333937	0.274740	0.059197	0.195126	0.01155
US	24,101	129,029	153,130	0.519440	0.580216	-0.060775	-0.110648	0.00672

Interpretacja WoE:

>0 - DG > DB

<0 - DG < DB

Interpretacja: Wartość IV wskazuje na nieużyteczność w predykcji.

Raport 7 - Wpływ języka, kraju i daty kampanii na konwersję do wniosku

- **Impakt**

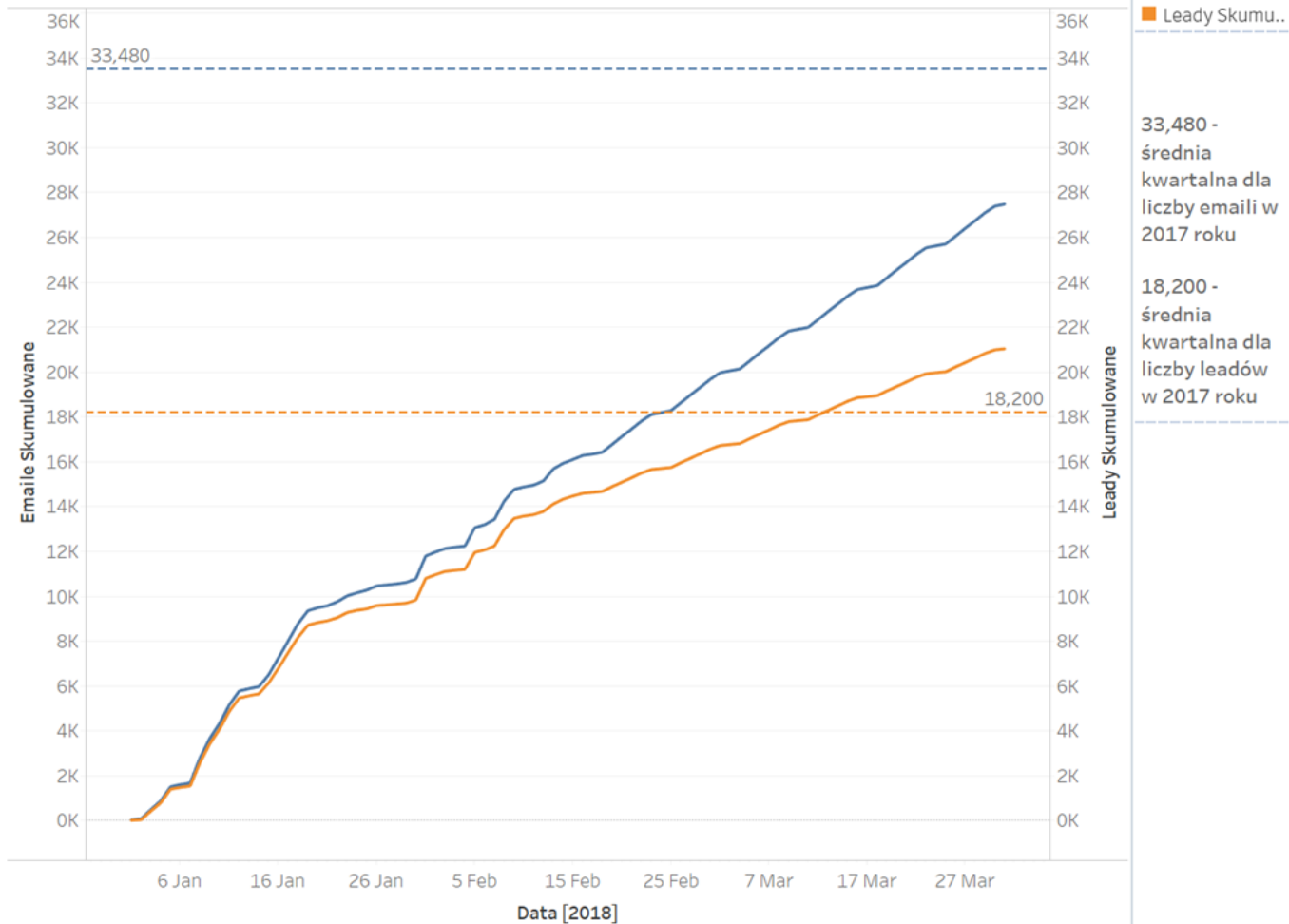
- W celu skonwertowania większej liczby wniosków, leady z grudnia (o ile nie są przeterminowane) zostaną przełożone na kampanię prowadzoną w styczniu;
- Różnica w konwersji między grudniem a styczniem: $27,26\% - 11,92\% = 15,34\%$;
- Potencjalny zysk z przeniesienia kampanii: $((10614 * 27,26\%) - 1266) * 300 * 25\% / 3\text{lata} = \underline{\text{40 675 zł}}$;
- Przesunięcie kampanii wpłynie na ilość pracy w grudniu.

5. Predykcje (1)

Predykcje liczby emaili i leadów

- **Cel:**
 - Kontrolowanie bieżących wartości leadów i emaili oraz obserwacja ich trendów w celu zapobiegania spadkowi ilości leadów/emaili;
- **Dla kogo:**
 - Zespół odpowiedzialny za tworzenie kampanii.

Predykcja liczby leadów i emaili dla Q1-2018



Predykcje liczby emaili i leadów

- **Impakt:**

- Przewidywany przychód z leadów w okresie (17.02-31.03):
 - $300 \text{ zł} * 25\% \text{ prowizji} * 15\% \text{ win rate} * 6359 = \underline{\underline{73\ 128\text{zł}\ 50\ \text{gr}}}$
- Przewidywany koszt z emaili w okresie (17.02-31.03):
 - $11023 * 0,1\text{zł} = \underline{\underline{1\ 105\text{zł}\ 70\ \text{gr}}}$
- Przewidywany zysk do końca Q1:
 - $73128\text{zł}\ 50\text{gr} - 1105\text{zł}\ 70\text{gr} = \underline{\underline{72\ 022\text{zł}\ 80\text{gr}}}$
- Predykcja pozwoli lepiej alokować zasoby, będzie wiadomo ilu ludzi jest potrzebnych do tworzenia kampanii/przetwarzania wniosków.

Podsumowanie raportów - całkowity impact

Raport	Impact na opinie klienta	Impact na czas procesowania	Impact na przychód
1 - Leady przeterminowane	✓ pozytywny	✓ pozytywny	✓ 27 000 PLN / rok
2 - Statystyki leadów	✗ brak	✗ brak	✓ 818 000 PLN / rok
3 - Emaille (+ koszty wysyłki emaili)	✗ brak	✗ brak	✗ - 13 200 PLN / rok
4 - Efektywność partnerów	✗ brak	✗ brak	✓ 475 000 PLN / rok
5 - Efektywność kampanii	✗ brak	✓ pozytywny	36 000 PLN / rok
6 - Open rate + Click rate + Konwersja	✓ pozytywny	✓ pozytywny	✓ 367 950 PLN / rok
7 - Czy język/kraj/data kampanii mają wpływ na konwersję do wniosku?	✗ brak	✗ negatywny	✓ 40 675 PLN / rok
8 - Predykcja leadów i maili	✗ brak	✓ pozytywny	✓ 72 023 PLN / Q1

SUMA: 1 213 963 PLN / rok + 72 023 PLN / Q1



Linki

- **Jira:**
 - Nasz board: [JDSZ1 Raczki Kanban Board](#)
- **Github**
 - Nasze repozytorium: [JDSZ1 Raczki Github](#)
- **Tableau** - linki do naszych analiz:
 - [Raport 1 - Lead'y przeterminowane](#)
 - [Raport 2 - Statystyki wniosków](#)
 - Raport 3 - Email - [wykres 1](#), [wykres 2](#), [wykres 3](#), [wykres 4](#), [wykres 5](#)
 - [Raport 4 - Efektywność partnerów](#)
 - [Raport 5 - Efektywność kampanii](#)
 - [Raport 6 - Open Rate & Click Rate & Konwersja](#)
 - Raport 7 - Wpływ języka, kraju i daty kampanii na konwersję do wniosku:
 - [Kraj](#)
 - [Język](#)
 - [Data kampanii](#)
 - [Raport 8 - Predykcje liczby leadów i emaili](#)



Dziękuję

Pytania?
Slack / email