

# MMDS Challenge

## Raport I

Phi Long Nguyen 116834  
Kajetan Gulbierz 134494

24 listopada 2017

## 1 Analiza danych ogłoszeń

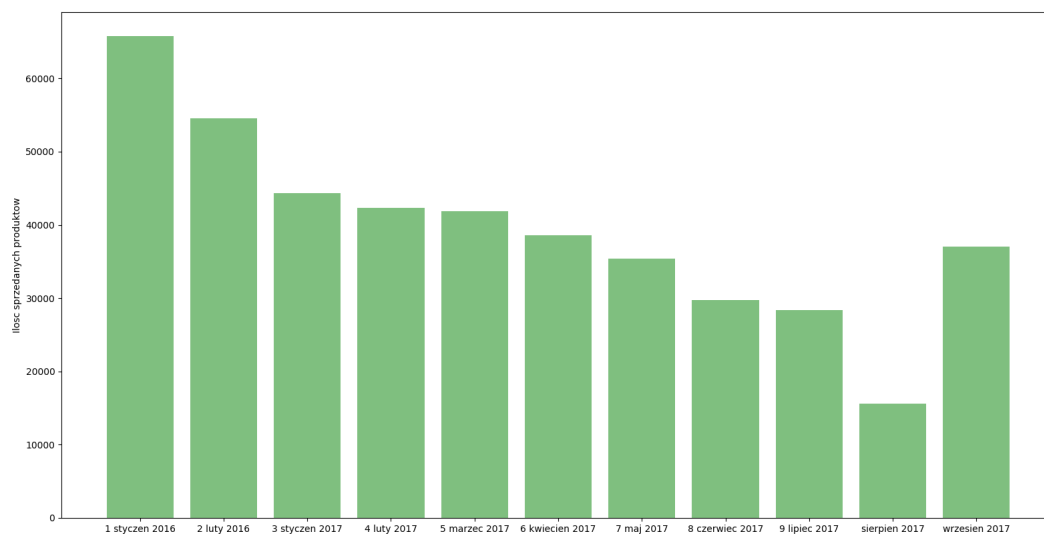
**Ogłoszenia** dostarczone są w formie plików, gdzie każdy z nich reprezentuje jeden miesiąc. W pliku zawierają się ogłoszenia z tego miesiąca. Każde ogłoszenie ma ściśle określoną strukturę, dzięki czemu jesteśmy w stanie je analizować. Celem zadania jest zrozumienie zjawiska jakiego dotyczą dane. Do przeprowadzenia analizy użyliśmy języka *python* wraz z narzędziem *pandas*, *numpy* oraz *matplotlib*, który służy do generowania wykresów.

### 1.1 Histogramy

**Wykresy** mają za zadanie pokazać trendy, jakie się utrzymują na przestrzeni roku (histogram sprzedaży). Oprócz tego pozwalają zobaczyć ile ogłoszeń miało dużą liczbę wyświetleń oraz dużą liczbę odpowiedzi.

#### 1.1.1 Histogram dla sprzedaży

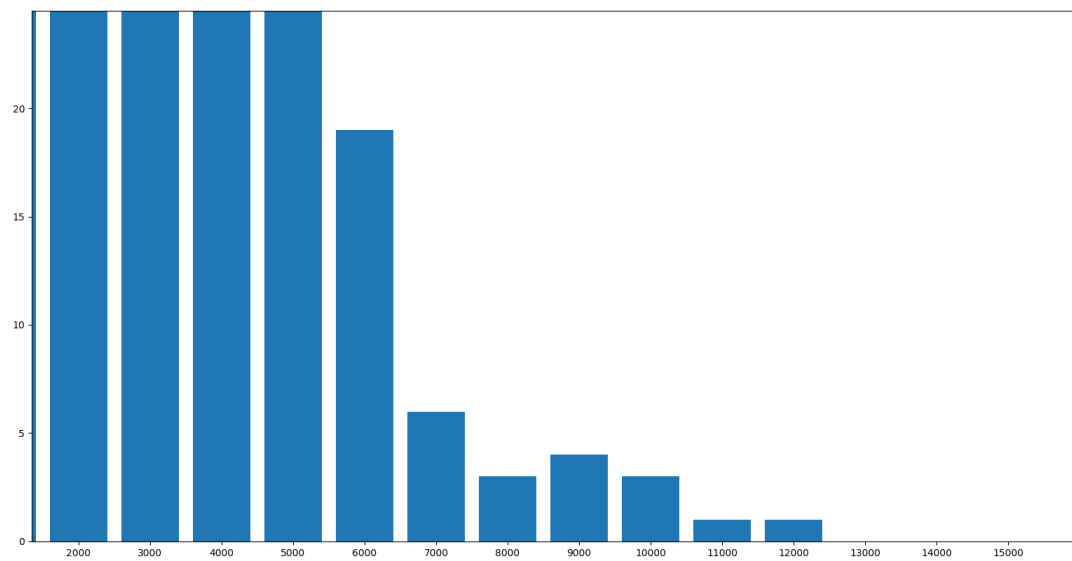
Rysunek 1 pokazuje histogram liczby sprzedaży ogółem dla każdego z miesięcy. Można zauważyć, że ogół sprzedaży jest największy dla pierwszego miesiąca w roku i stopniowo spada. Taki stan rzeczy utrzymuje się mniej więcej do sierpnia, gdzie liczba gwałtownie się podnosi. Na takie wyniki może wpływać fakt, że wrzesień jest miesiącem rozpoczynającym rok szkolny. Najmniejsze zaś wyniki można odnotować w okresie wakacyjnym.



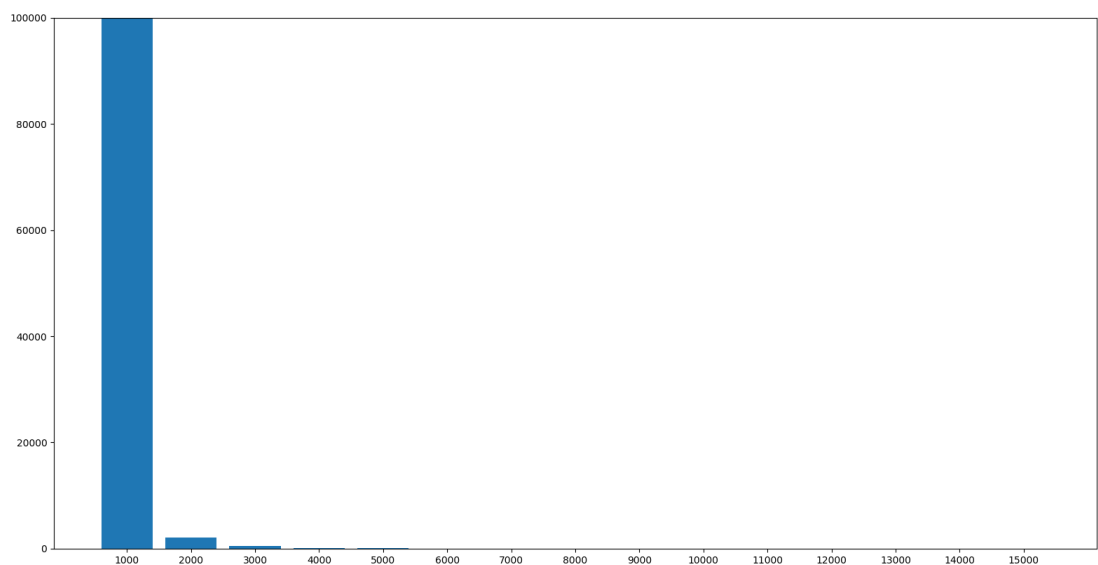
Rysunek 1: Histogram sprzedaży

### 1.1.2 Histogram liczby wyświetleń

Rysunek 2 oraz 3 pokazują wykres liczby ogłoszeń mających mniej niż  $X$  wyświetleń. Przedziały  $X$  są oznaczone na osi-x. W tym przypadku widać bardzo dużą różnicę między liczbą ogłoszeń, które zostały mało razy wyświetlone (mniej niż 1000/2000) do tych, które tych wyświetleń miało dużo.



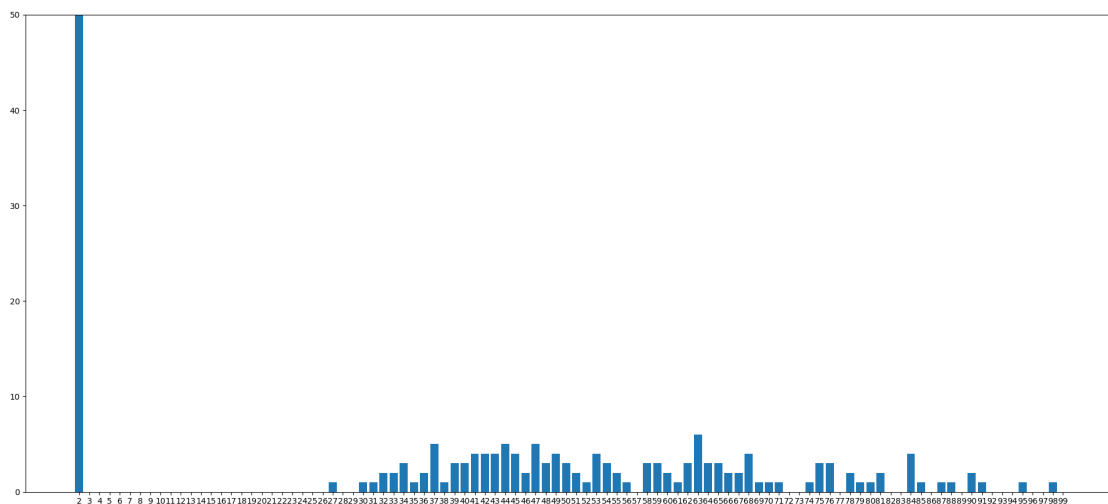
Rysunek 2: Wykres wyświetleń



Rysunek 3: Wykres wyświetleń

### 1.1.3 Histogram liczby odpowiedzi

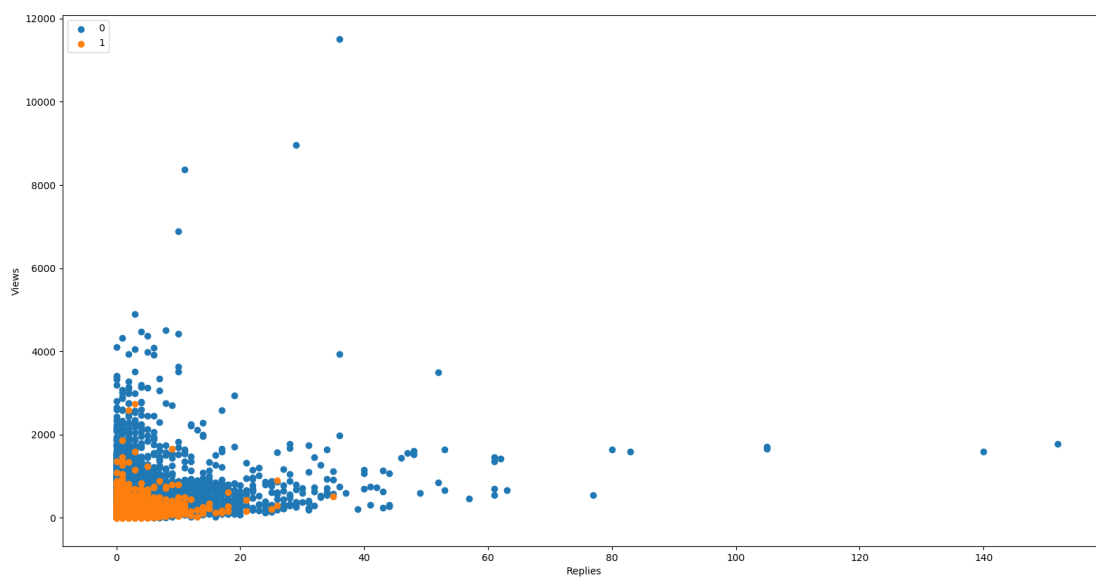
Rysunek 4 pokazuje wykres liczby ogłoszeń mających mniej niż  $X$  odpowiedzi. Przedziały  $X$  są oznaczone na osi-x.



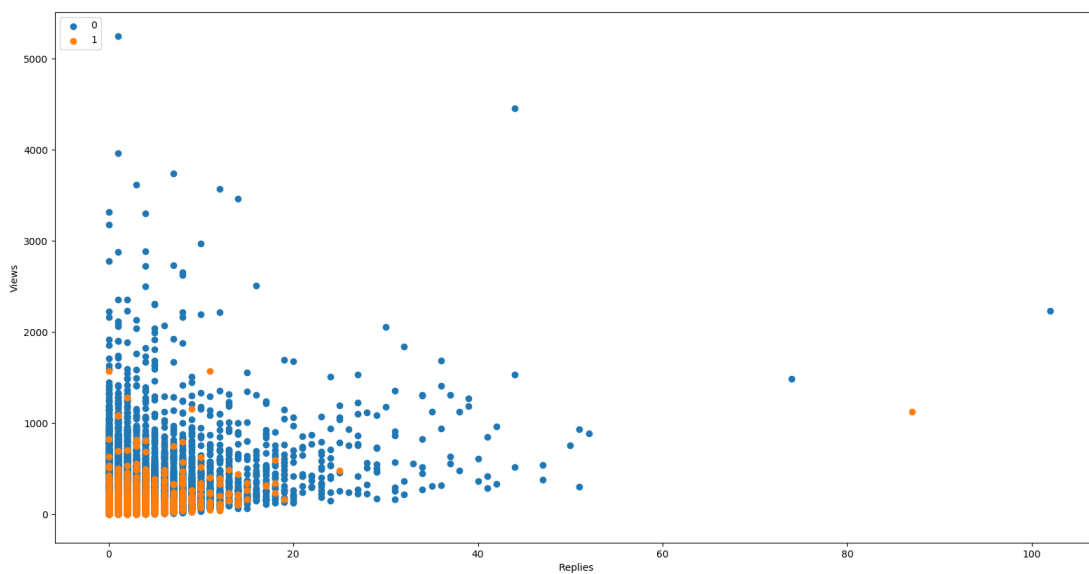
Rysunek 4: Wykres wyświetleń

#### 1.1.4 Wykres korelacji

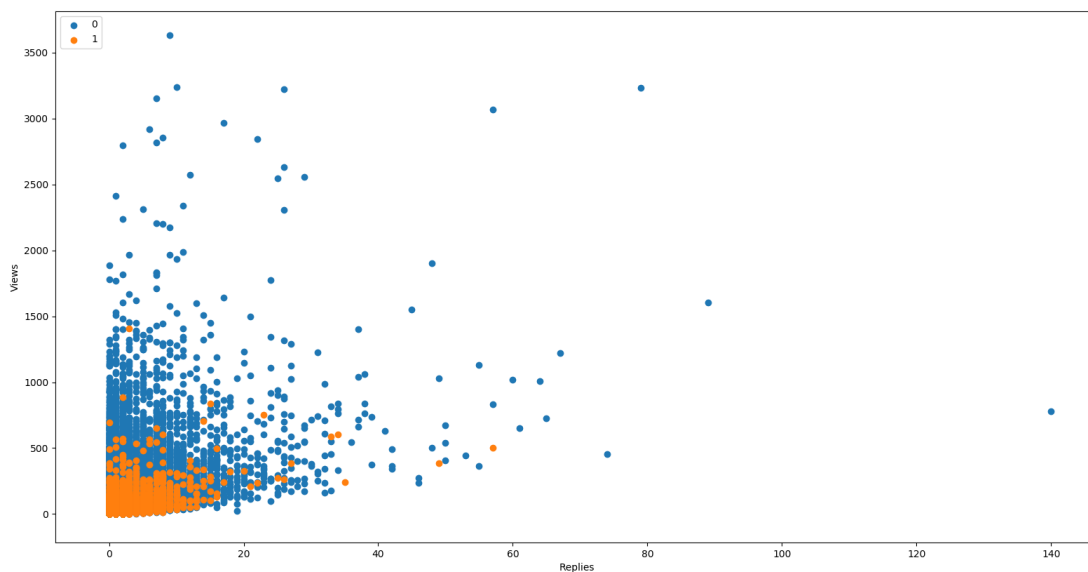
Rysunek 5, 6 oraz 7 pokazuje zależność między wyświetleniami, odpowiedzią a sprzedażą w danym miesiącu. Można zauważyć o wiele większą rozpiętość zarówno wyświetleń jak i odpowiedzi dla ogłoszeń, które nie zostały zakończone sprzedażą w stosunku do tych, którym sprzedaż się powiodła. Należy również zwrócić uwagę na fakt, że dla każdego zbadanego miesiąca, wzorzec jest podobny.



Rysunek 5: Wykres korelacji pierwszy



Rysunek 6: Wykres korelacji drugi



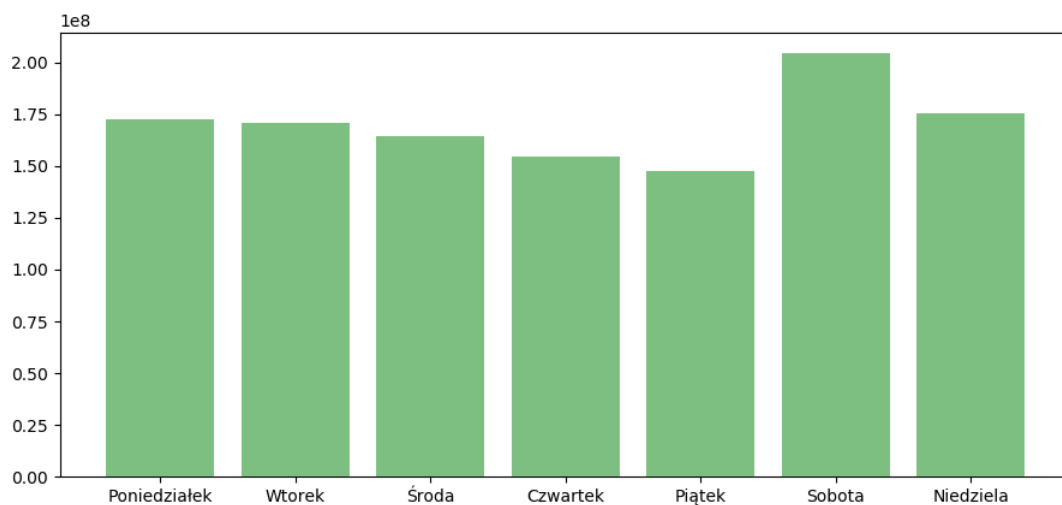
Rysunek 7: Wykres korelacji trzeci

## 2 Analiza danych zapytań

**Pliki zapytań** reprezentują liczbę sesji dla danych kategorii. W ramach sesji tworzone są zapytania. Liczba sesji *de facto* oznacza ile aktywnych użytkowników w danym momencie wystosowało dane zapytanie.

### 2.1 Cykliczność tygodniowa

Badanie cykliczności liczby zapytań ma za zadanie pomóc dostrzec wzorzec jaki spełniają dane w tygodniu. Na rysunku 8 można zauważyć wzrost liczby zapytań w ostatnie dni tygodnia. Najmniejszy zaś wynik przypada w piątek, co może oznaczać, że początek weekendu użytkownicy spędzają inaczej, niż na przeglądaniu ofert.



Rysunek 8: Wykres liczby zapytań w danym dniu tygodnia

## 2.2 Wykres popularności

Rysunek 9 pokazuje jak zmieniała się popularność - w tym wypadku dla kategorii *iPhone*. Można zauważyć, znaczny wzrost liczby wyszukiwań w okresie wakacyjnym. Sporą liczbę zapytań można również zauważyć w grudniu - można przypuszczać, że ma to związek z okresem świątecznym.



Rysunek 9: Wykres popularności zapytania dla iPhone