

# Spis treści

[Spis treści](#)

[Osoby w grupie:](#)

[Dane](#)

[Wymiary](#)

[Pytania biznesowe](#)

[Schemat Gwiazdy](#)

[ETL](#)

[Opis kostki danych](#)

[Miary wyliczalne:](#)

[Podzbiory](#)

[Key Performance Indicators](#)

[Raporty](#)

[Pytanie 1.](#)

[Pozytywne komentarze:](#)

[Neutralne komentarze](#)

[Negatywne komentarze](#)

[Wniosek](#)

[Pytanie 2](#)

[Pozytywny komentarz](#)

[Negatywny komentarz](#)

[Wniosek](#)

[Pytanie 3](#)

[Wniosek](#)

[Pytanie 4](#)

[Wniosek](#)

[Pytanie 5](#)

[Wniosek](#)

[Pytanie 6](#)

[Wniosek](#)

[Bonus - raport z KPI](#)

[Wniosek](#)

## Osoby w grupie:

1. Dastin Sandurs s10523
2. Diana Sandura s10776

## Dane

Dane zorientowane są na temat wystawionych ocen hoteli. Ocena składa się z numerycznej wartości oraz dwóch opcjonalnych tekstowych uwagach: negatywnej i pozytywnej. Osoby które wystawiają opinie, nie mają unikalnych tożsamości i są rozróżniane przez narodowości i ilość opinii które wystawiły. Każda wystawiona opinia posiada dane na temat typu pobytu w hotelu.

Źródło danych to Kaggle:

[https://www.kaggle.com/jiashenliu/515k-hotel-reviews-data-in-europe?fbclid=IwAR3G-0qkrR\\_ntkEaEM3dRMCwvytJpS7Spig-89hV69oPI18hujKbdYc3jFs](https://www.kaggle.com/jiashenliu/515k-hotel-reviews-data-in-europe?fbclid=IwAR3G-0qkrR_ntkEaEM3dRMCwvytJpS7Spig-89hV69oPI18hujKbdYc3jFs)

## Wymiary

1. Stay Description
2. Reviewer
3. Hotel
4. Comment Characteristic

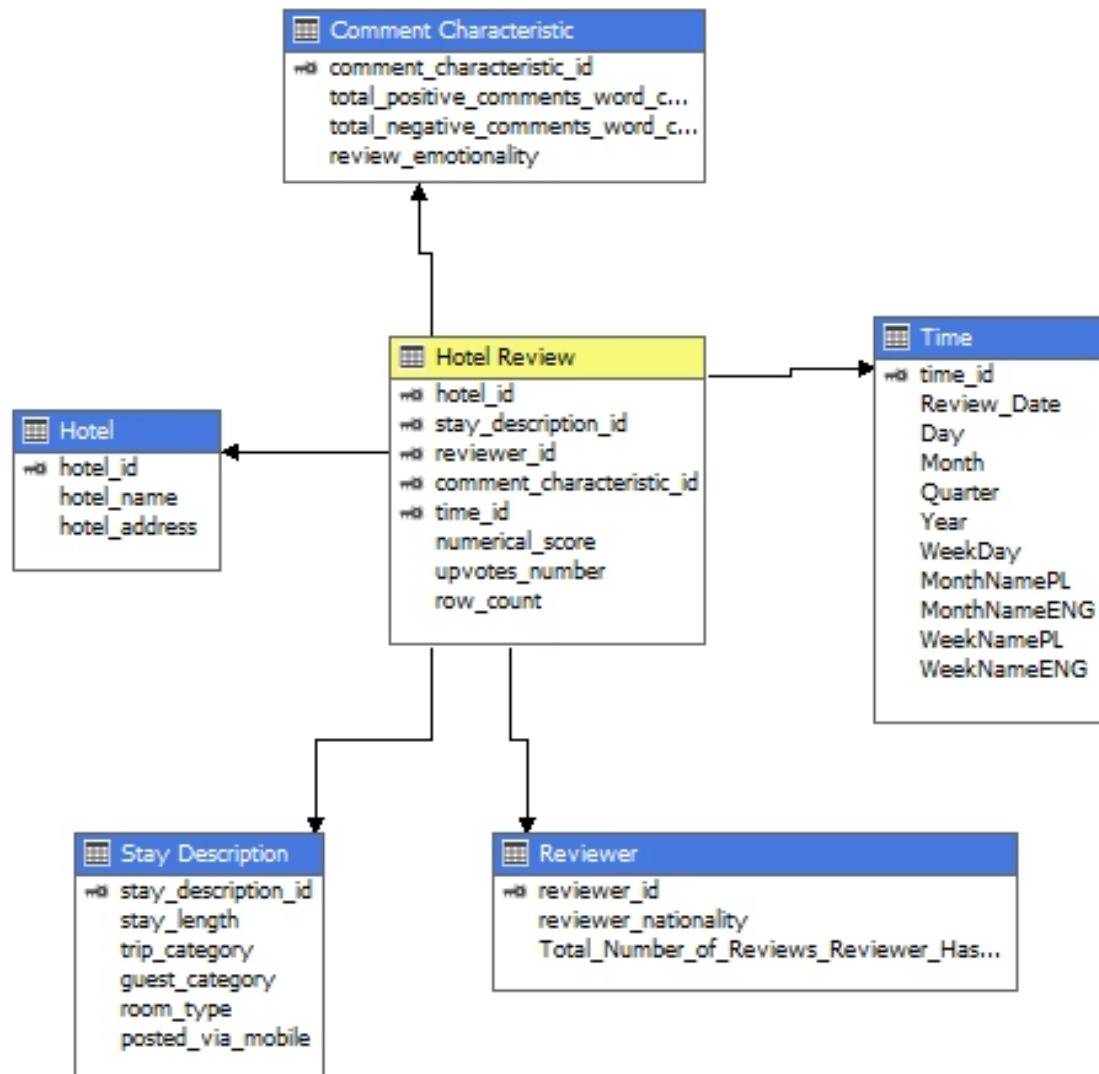
## Pytania biznesowe

Pytania biznesowe zorientowane są na odkrycie w danych zależności które pomogą managerom danego hotelu, polepszyć jakość świadczonych usług poprzez zwiększenie świadomości tego jak hotel jest postrzegany przez odwiedzających.

1. Jaki kraj daje nawiecej pozytywnych/negatywnych komentarzy.
2. Czy wieksza ilosc slow w komentarzu to wieksza ilosc polubień opinii?
3. Który hotel ma najlepsze oceny w kategori pobytów jednodniowych?
4. W którym miesiącu jest najwiecej wystawianych opinii?
5. Czy opinie wprowadzone przez telefon są przychylniejsze?
6. Jaki rodzaj gości wystawia najwiecej opini podczas pobytów?

## Schemat Gwiazdy

Gwiazda zorientowana jest wokół opinii (Hotel Review). Wszystkie wymiary opisują opinię i pozwalają wydobyć z danych odpowiedzi na interesujące nas pytania.



# ETL

Komponent ze skryptem rozdzielania kolumny *Tag* na atrybuty wymiaru Stay Description. Komponent ma nazwę "Split Tags into separate columns" i posiada następującą logikę:

```
public override void Input0_ProcessInputRow(Input0Buffer Row)
{
    HashSet<String> tripType = new HashSet<string> { "Leisure trip", "Business trip" };
    HashSet<String> guestCategories = new HashSet<string> {
        "Couple", "Family with older children", "Family with young children",
        "Group", "Solo traveler", "With a pet"
    };
    //There is no set of 'room type' categories because it has many values.
    //We will match 'room type' as the data which did not match other categories
    if (Row.Tags != null) {
        string[] words = Row.Tags.ToString().Split('\\'); // data is between single quotes
        for (int i = 2; i <= words.Length; i += 2)
        {
            String token = words[(i - 1)].Trim();
            if (token.Equals("Submitted from a mobile device"))
            {
                Row.postedviamobile = "Submitted from a mobile device";
            } else if (token.StartsWith("Stayed "))
            {
                Row.staylength = token;
            } else if (guestCategories.Contains(token))
            {
                Row.guestcategory = token;
            } else if (tripType.Contains(token))
            {
                Row.tripcategory = token;
            } else // if nothing matched, it is data with 'room type' information
            {
                Row.roomtype = token;
            }
        }
    }
}
```

Kod źródłowy:

```
public override void Input0_ProcessInputRow(Input0Buffer Row)
{
    HashSet<String> tripType = new HashSet<string> { "Leisure trip", "Business trip" };
    HashSet<String> guestCategories = new HashSet<string> { "Couple", "Family with
older children", "Family with young children", "Group", "Solo traveler", "With a pet" };
    //There is no set of 'room type' categories because it has many values. We will match
'room type' as the data which did not match other categories
    if (Row.Tags != null) {
        string[] words = Row.Tags.ToString().Split('\\'); // data is between single quotes

        for (int i = 2; i <= words.Length; i += 2)
        {
            String token = words[(i - 1)].Trim();
            if (token.Equals("Submitted from a mobile device"))
            {
                Row.postedviamobile = "Submitted from a mobile device";
```

```

    } else if (token.StartsWith("Stayed "))
    {
        Row.staylength = token;
    } else if (guestCategories.Contains(token))
    {
        Row.guestcategory = token;
    } else if (tripType.Contains(token))
    {
        Row.tripcategory = token;
    } else // if nothing matched, it is data with 'room type' information
    {
        Row.roomtype = token;
    }
}
}
}

```

## Opis kostki danych

### Miary wyliczalne:

Name:

[Avg Hotel Score]

Parent Properties

Parent hierarchy:

Measures

Parent member:

Expression

[Measures].[Numerical Score]/[Measures].[Row Count]

No issues found

Additional Properties

Format string:

Visible:

True

Non-empty behavior:

Associated measure group:

(Undefined)

Display folder:

Calculations

Name:

[Calculated Hotel Score]

Parent Properties

Parent hierarchy:

Measures

Parent member:

Change

Expression

```
([Measures].[Avg Hotel Score]^[Measures].[Avg Hotel Score])+([Measures].[Row Count])
```

✓ No issues found

Additional Properties

Format string:

▼

Visible:

True

Non-empty behavior:

▼

Associated measure group:

(Undefined)

Display folder:

Calculations

Name:

[Review Count Category]

Parent Properties

Parent hierarchy:

Measures

Parent member:

Expression

```
CASE
  When ([Measures].[Hotel Review Count] <105
  THEN "small"
  WHEN ([Measures].[Hotel Review Count] <305
  THEN "medium"
  ELSE ("large")
END
```

✓ No issues found

Additional Properties

Format string:

"Standard"

Visible:

True

Non-empty behavior:

▼

Associated measure group:

(Undefined)

Display folder:

## Podzbiory

Name:

[Couples]

⌄ Expression

```
[Stay Description].[Guest Category].&[Couple]
```

✓ No issues found

⌄ Additional Properties

Type:

Dynamic

Display folder:

Sets

Name:

[Other Guest Categories]

⌄ Expression

```
UNION (  
  [Stay Description].[Guest Category].&[Family with older children],  
  [Stay Description].[Guest Category].&[Family with young children],  
  [Stay Description].[Guest Category].&[Group],  
  [Stay Description].[Guest Category].&[Solo traveler]  
)
```

✓ No issues found

⌄ Additional Properties

Type:

Dynamic

Display folder:

Sets

# Key Performance Indicators

## ⌄ KPI

Name:

Ceiling

Associated measure group:

Hotel Review

## ⌄ Value Expression

[Measures].[Avg Hotel Score]

✓ No issues found


## ⌄ Goal Expression

9.5

✓ No issues found

## ⌄ Status

Status indicator:

 Gauge


Status expression:

```
CASE
  WHEN KpiValue( "Ceiling" )/KpiGoal( "Ceiling" ) >= 1 THEN 1
  WHEN KpiValue( "Ceiling" )/KpiGoal( "Ceiling" ) < 1
    AND KpiValue( "Ceiling" )/KpiGoal( "Ceiling" ) >= .85 THEN 0
  ELSE -1
END
```

✓ No issues found

## ⌄ Trend

Trend indicator:

 Status arrow

Trend expression:

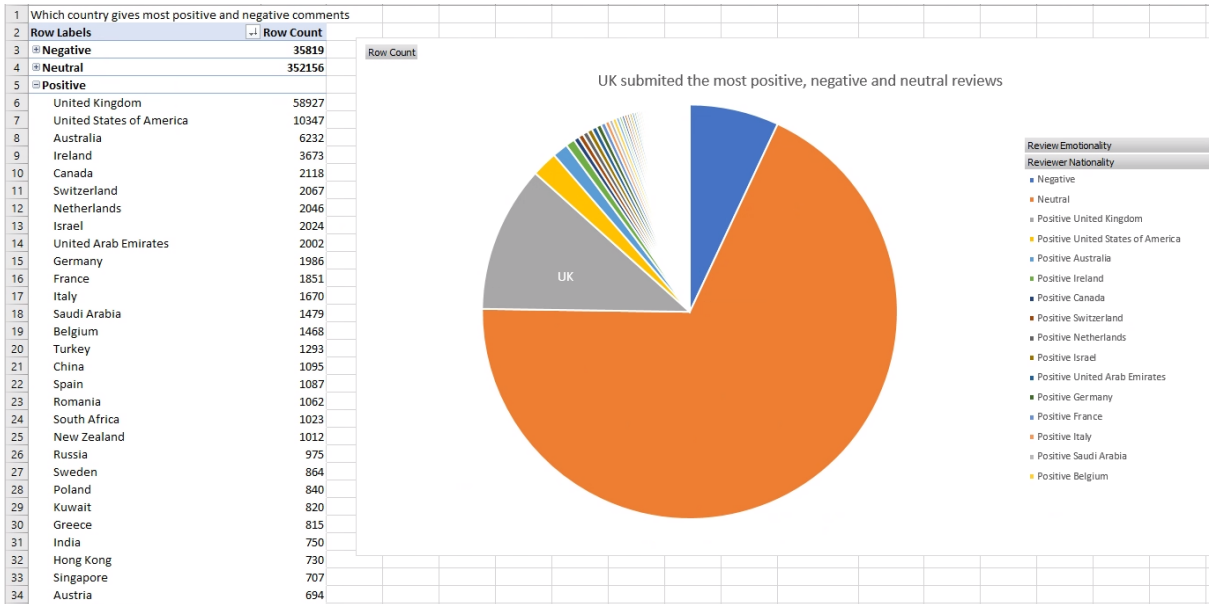
([Measures].[Avg Hotel Score]-9.5)/[Measures].[Avg Hotel Score]



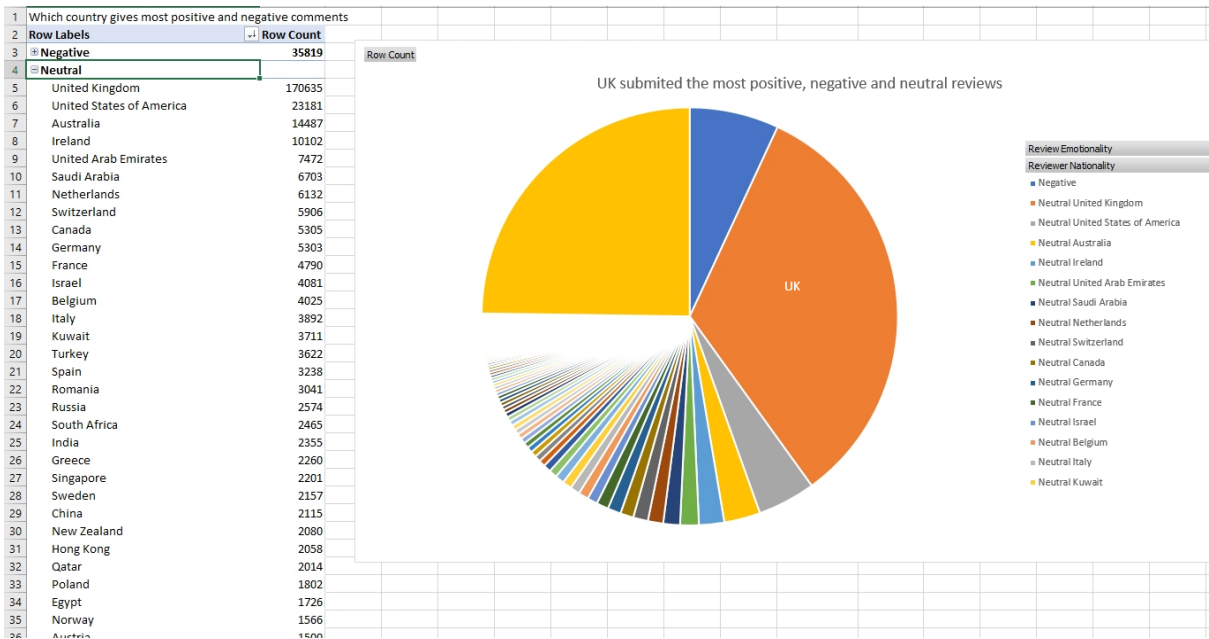
# Raporty

## Pytanie 1.

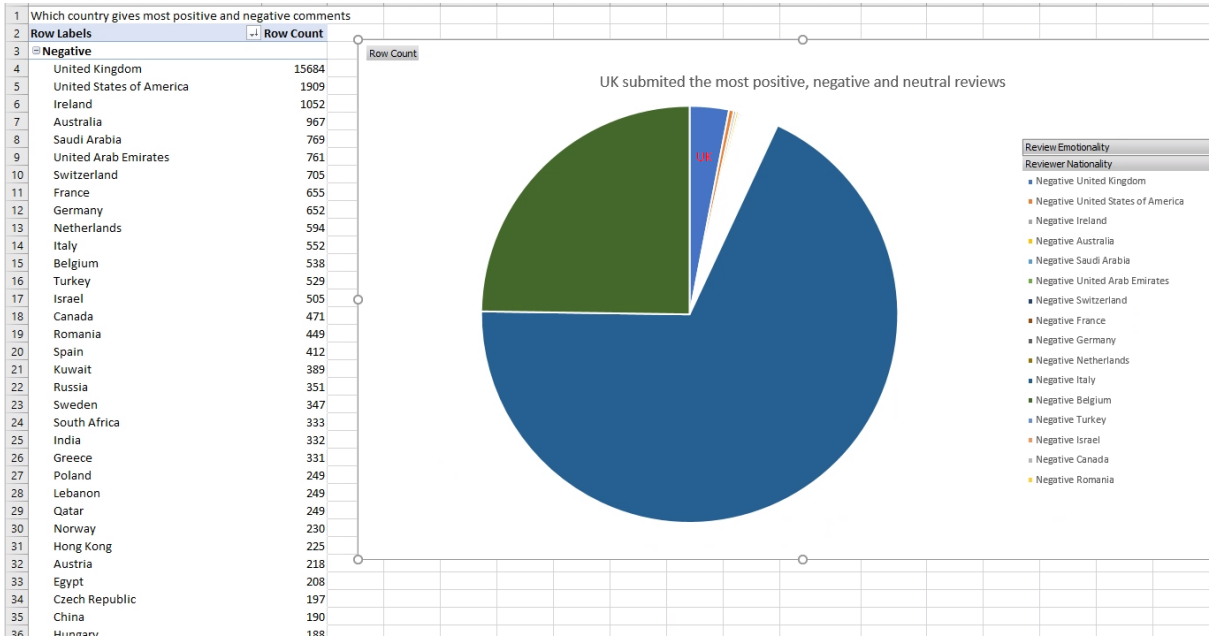
### Pozytywne komentarze:



### Neutralne komentarze



# Negatywne komentarze

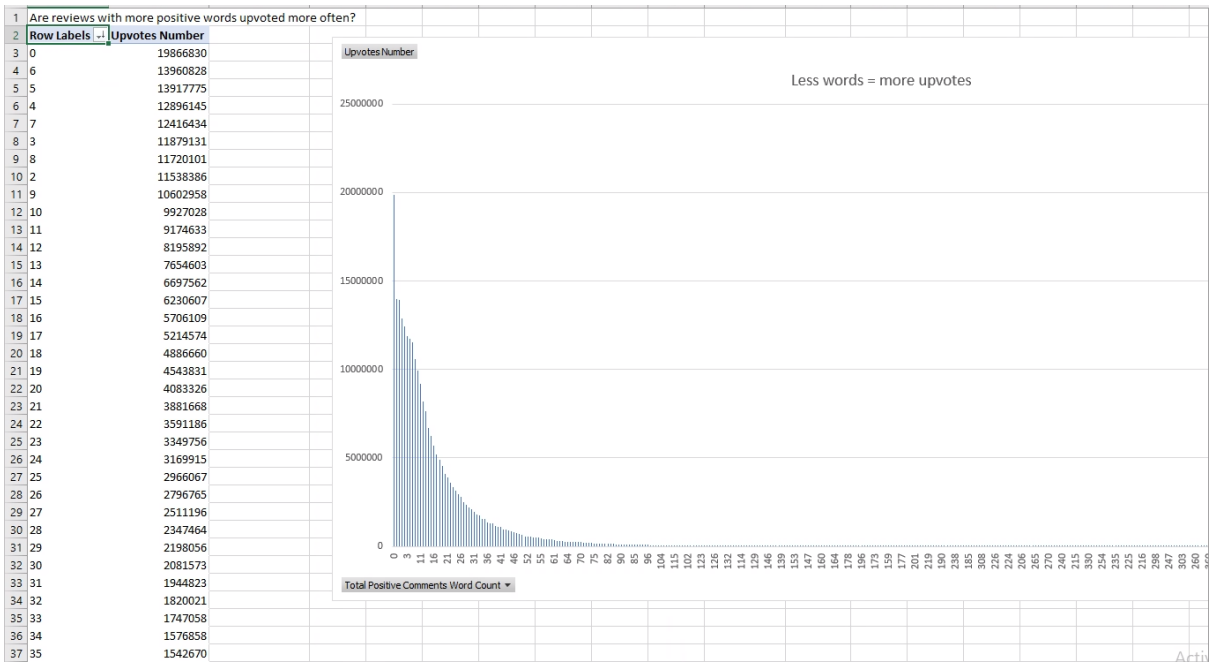


# Wniosek

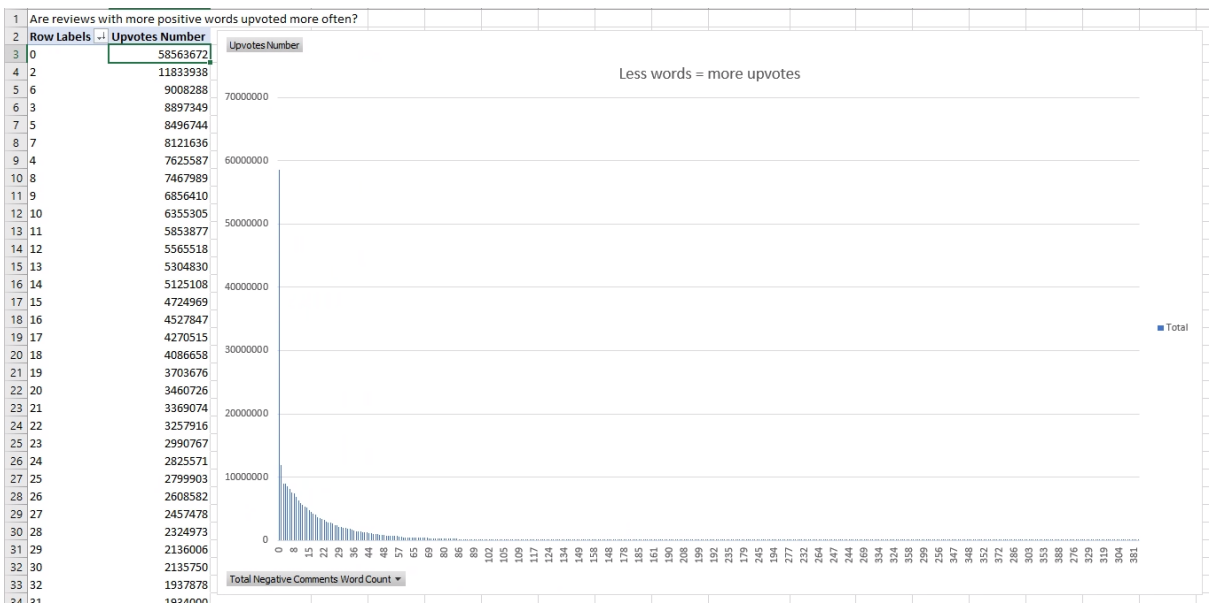
Wielka Brytania wystawiła najwięcej pozytywnych, negatywnych i neutralnych opinii.

# Pytanie 2

# Pozytywny komentarz



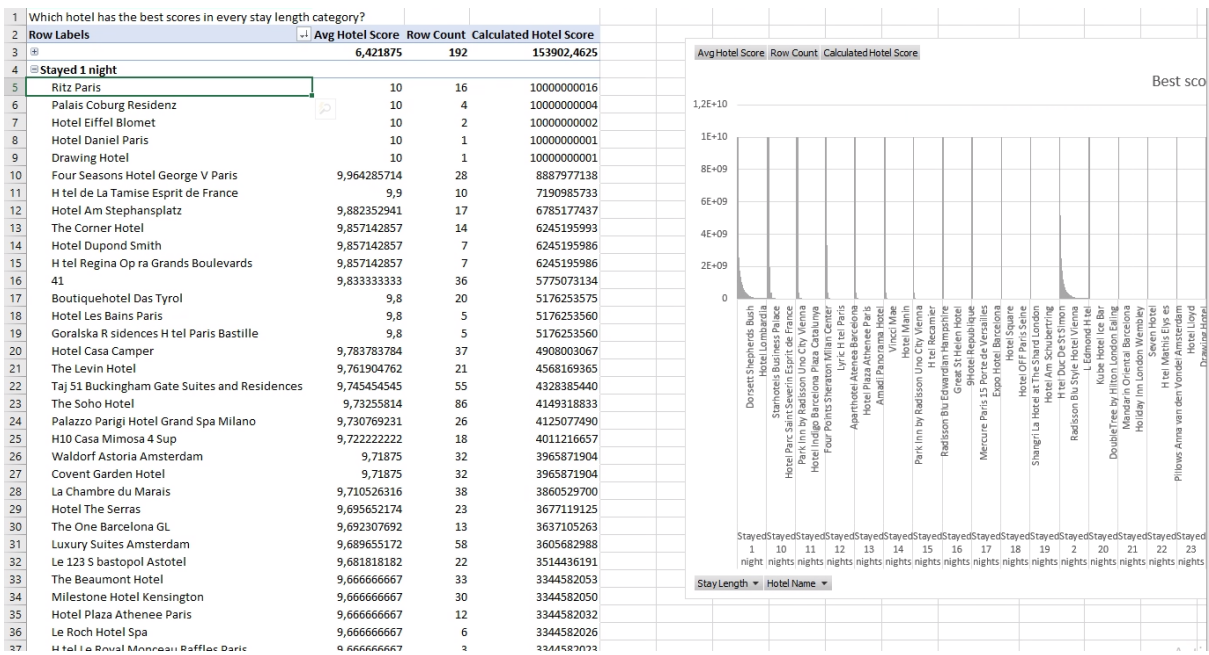
# Negatywny komentarz



# Wniosek

Im komentarz ma mniej słów, tym więcej czytelników ‘lubi’ opinię o hotelu.

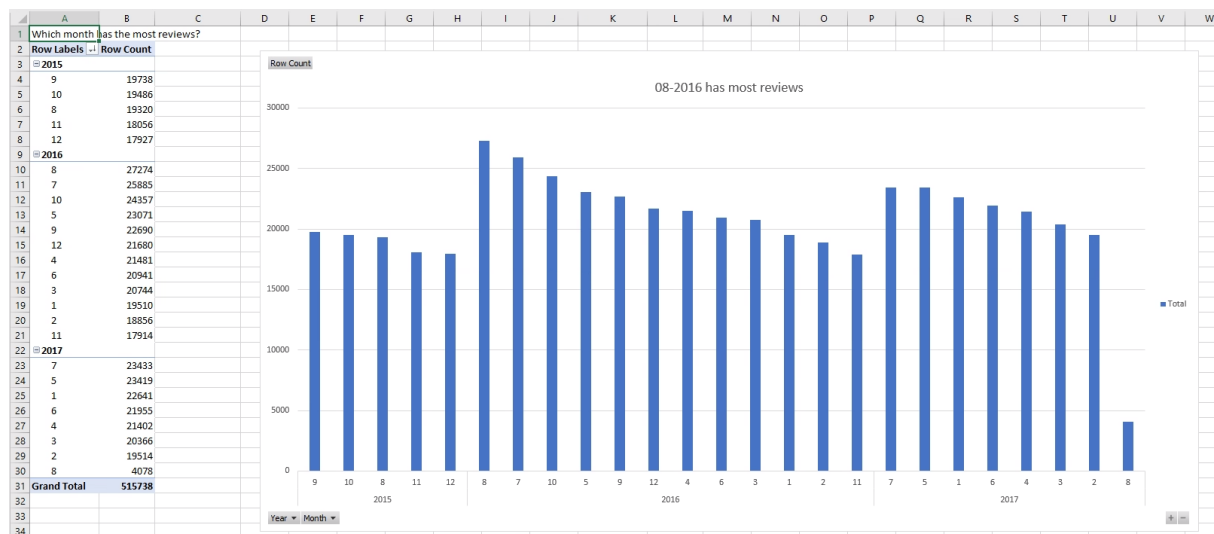
# Pytanie 3



# Wniosek

Ritz Paris ma najlepsze oceny w kategori pobyków jednodniowych.

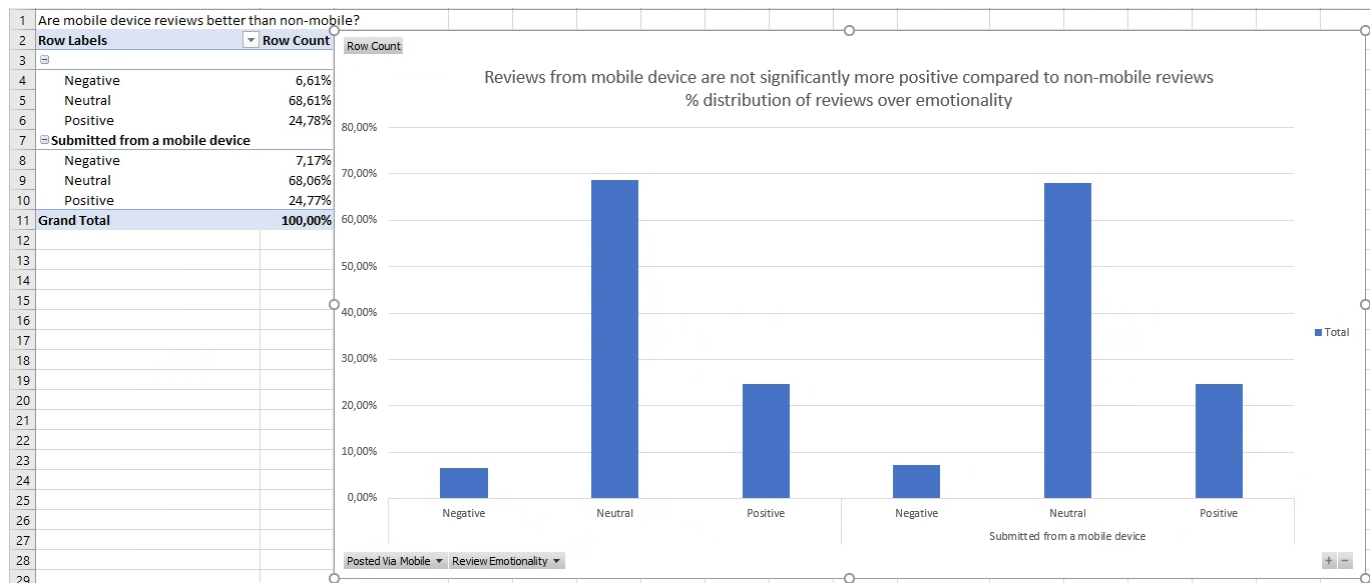
## Pytanie 4



## Wniosek

Sierpień 2016 ma najwięcej opini. Na drugim miejscu jest wrzesień 2015 a za nim lipiec 2017. Uogólniając, okres wakacyjny ma najwięcej opinii.

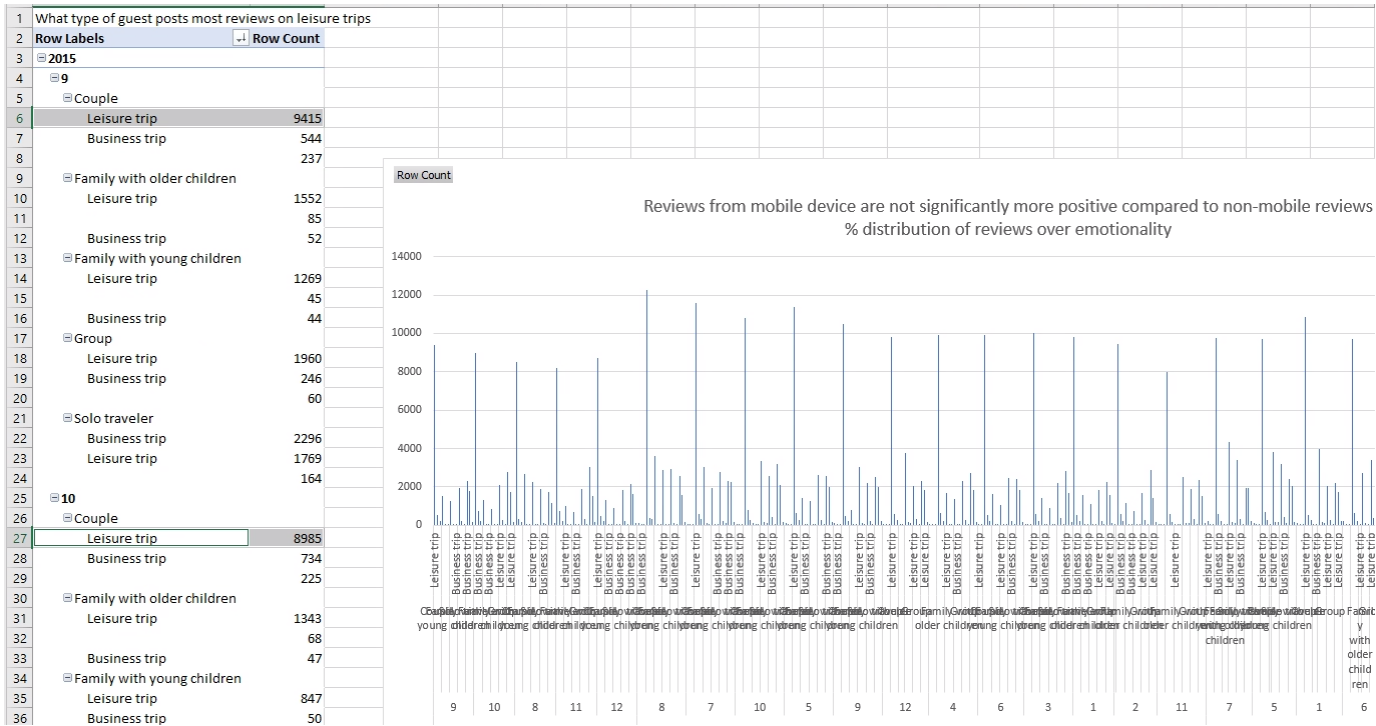
## Pytanie 5



## Wniosek

Opinie wystawione przez urządzenia mobilne nie są bardziej pozytywne od tych wystawionych przez urządzenia nie-mobilne.

Pytanie 6



Wniosek

Najwięcej opinii jest wystawianych przez pary, które odbywają pobyt wypoczynkowy.

## Bonus - raport z KPI

Row Labels	Review Count Category	Hotel Review Count	Ceiling Goal	Ceiling Status	Ceiling Trend	Avg Hotel Score
11 Cadogan Gardens	medium	159	9,5			8,893081761
1K Hotel	medium	148	9,5			7,885135135
25hours Hotel beim MuseumsQuartier	large	689	9,5			9,04934688
41	small	103	9,5			9,766990291
45 Park Lane Dorchester Collection	small	28	9,5			9,642857143
88 Studios	large	459	9,5			8,546840959
9Hotel Republique	medium	183	9,5			8,808743169
A La Villa Madame	small	41	9,5			8,975609756
ABaC Restaurant Hotel Barcelona GL Monumento	small	31	9,5			8,516129032
Abba Garden	medium	125	9,5			8,072
Abba Sants	medium	120	9,5			8,408333333
AC Hotel Barcelona Forum a Marriott Lifestyle Hotel	medium	289	9,5			8,076124567
AC Hotel Diagonal L Illa a Marriott Lifestyle Hotel	small	56	9,5			8,428571429
AC Hotel Irla a Marriott Lifestyle Hotel	small	16	9,5			8,1875
AC Hotel Milano a Marriott Lifestyle Hotel	small	72	9,5			8,25
AC Hotel Paris Porte Maillot by Marriott	small	75	9,5			8,52
AC Hotel Sants a Marriott Lifestyle Hotel	small	91	9,5			7,945054945
AC Hotel Victoria Suites a Marriott Lifestyle Hotel	small	45	9,5			8,888888889
Acad mie H tel Saint Germain	medium	255	9,5			9,125490196
Acca Palace	medium	209	9,5			8,641148325
Ace Hotel London Shoreditch	large	514	9,5			8,760700389
Acevi Villarroel	large	352	9,5			7,6875
Acta Atrium Palace	large	458	9,5			8,716157205
Acta CITY47	large	381	9,5			8,25984252
ADI Doria Grand Hotel	medium	147	9,5			8,43537415
ADI Hotel Poliziano Fiera	medium	107	9,5			8,065420561
Admiral Hotel	small	26	9,5			7,153846154
Adria Boutique Hotel	medium	231	9,5			8,818181818
Ako Suites Hotel	medium	131	9,5			8,786259542
Albus Hotel Amsterdam City Centre	medium	194	9,5			8,25257732
Alexandra Barcelona A DoubleTree by Hilton	small	77	9,5			8,051948052
Alma Barcelona GL	medium	179	9,5			9,396648045
Alma Boutique Hotel	medium	152	9,5			9,006578947
Aloft London Excel	large	1126	9,5			8,690941385
Am Spiegeln	small	21	9,5			8,285714286
Amadi Panorama Hotel	large	481	9,5			8,48024948
Amadi Park Hotel	large	396	9,5			8,247474747

## Wniosek

Raport pokazujący KPI pozwala łatwo wykryć hotele, które mają spadkowy trend w średniej ocenie liczbowej wystawianej hotelowi.