

# Zadanie 1

Damian Kędzierski 260493

## 1. Wprowadzenie

### ⇒ Cel

- „sformułowanie problemów biznesowych i analitycznych dla dwóch przykładowych sytuacji biznesowych, łącznie ze wstępnym określeniem wskaźników realizacji celów biznesowych (KPI)”
- „wstępna **eksploracyjna analiza i przygotowanie danych** (EDA) do budowy wizualizacji w Excelu i Tableau oraz budowy modeli predykcyjnych w Tableau”

„Wykonanie zadania jest wstępem do opracowania założeń i projektów pulpitów menedżerskich - zestawów wizualizacji, przeznaczonych dla decydentów (zadanie 2), a następnie budowy pulpitów w Excelu z Power Pivot lub w Power BI (zadanie 3) oraz w Tableau (zadania 5-6)”

### ⇒ Zbiory danych

- Zbiór danych o klientach („klienci”) – Arkusz „260493\_klienci.xlsx”
- Zbiór danych o lokatach („lokaty”) – Arkusz „260493\_lokaty.xlsx”

## 2. Sformułowanie problemów biznesowych i problemów analitycznych

### ⇒ Zbiór – klienci:

Ustalenie dla jakiego przedstawiciela, przeprowadzane rozmowy były najskuteczniejsze w określonym miesiącu.

WSKAŹNIKI:

$$\text{Skuteczność} = \frac{\text{kwota zakupu}}{\text{czas rozmowy}}$$

- skuteczność jest wprost proporcjonalna do kwoty zakupu i odwrotnie proporcjonalna do czasu trwania rozmowy

**Średnia skuteczność rozmów przedstawicieli w określonym miesiącu**

Etykiety wierszy	sierpień	wrzesień	październik	Średnia
P01	3,99	6,09	4,78	<b>4,95</b>
P02	2,68	3,39	2,80	<b>2,97</b>
P03	3,51	2,71	2,58	<b>2,91</b>
P04	2,74	4,79	2,20	<b>3,10</b>
P05	5,06	3,49	3,92	<b>4,15</b>
Średnia	<b>3,56</b>	<b>3,99</b>	<b>3,22</b>	<b>3,59</b>

### ⇒ Zbiór – lokaty:

Ustalenie do których grup klientów należy kierować akcje marketingowe

- Do jakiej grupy wiekowej jest kierowanych najwięcej akcji marketingowych?
- Która grupa wiekowa najchętniej zgadza się na założenie lokaty?

## WSKAŹNIKI:

- Liczba lokat założonych w określonej grupie wiekowej
- Procent zawartych umów w określonej grupie wiekowej

**Liczba klientów reprezentujących określoną grupę wiekową, którzy zdecydowali się zapisać na lokatę terminową**

Etykiety wierszy	no	yes	Suma końcowa	Procent yes
<b>2008</b>	<b>26351</b>	<b>1339</b>	<b>27690</b>	<b>4,84%</b>
do 20	2		2	0,00%
20-30	2835	163	2998	5,44%
<b>30-40</b>	<b>10852</b>	<b>591</b>	<b>11443</b>	<b>5,16%</b>
40-50	7513	334	7847	4,26%
50-60	4986	241	5227	4,61%
60-70	163	10	173	5,78%
<b>2009</b>	<b>9212</b>	<b>2228</b>	<b>11440</b>	<b>19,48%</b>
do 20	29	23	52	44,23%
20-30	1628	494	2122	23,28%
<b>30-40</b>	<b>4016</b>	<b>779</b>	<b>4795</b>	<b>16,25%</b>
40-50	2071	347	2418	14,35%
50-60	1057	321	1378	23,29%
60-70	240	142	382	37,17%
70-80	120	85	205	41,46%
80-90	47	35	82	42,68%
ponad 90	4	2	6	33,33%
<b>2010</b>	<b>985</b>	<b>1073</b>	<b>2058</b>	<b>52,14%</b>
do 20	10	11	21	52,38%
20-30	243	231	474	48,73%
<b>30-40</b>	<b>355</b>	<b>345</b>	<b>700</b>	<b>49,29%</b>
40-50	108	153	261	58,62%
50-60	122	135	257	52,53%
60-70	70	99	169	58,58%
70-80	55	59	114	51,75%
80-90	21	37	58	63,79%
ponad 90	1	3	4	75,00%
<b>Suma końcowa</b>	<b>36548</b>	<b>4640</b>	<b>41188</b>	<b>11,27%</b>

## 3. Wstępna analiza i przygotowanie danych – dane klienci

Zbiór posiada 1000 rekordów.

Fragment tabeli danych klientów

numer klienta	czas rozmowy	kwota zakupu	przedstawiciel	data rozmowy		dzień	miesiąc	rok	Skutecz.	Region
0006	129	64	P03	czwartek, 19 wrzesień 2013		19	9	2013	50%	Południe
0012	150	128	P03	poniedziałek, 7 październik 2013		7	10	2013	85%	Południe
0012	140	172	P02	piątek, 4 październik 2013		4	10	2013	123%	Zachód
0009	175	122	P05	niedziela, 20 październik 2013		20	10	2013	70%	Centrum
0003	11	195	P01	środa, 28 sierpień 2013		28	8	2013	1773%	Centrum
0007	69	154	P04	sobota, 7 wrzesień 2013		7	9	2013	223%	Centrum
0007	101	183	P02	sobota, 5 październik 2013		5	10	2013	181%	Zachód

### Tabela danych przedstawicieli

przedstawiciel	Imię i nazwisko	Oddział	Region
P01	Jan Kowalski	Warszawa	Centrum
P02	Krzysztof Bąk	Wrocław	Zachód
P03	Paweł Panas	Rzeszów	Południe
P04	Aleksandra Matulak	Warszawa	Centrum
P05	Krystyna Noczulska	Łódź	Centrum

### Struktura, rozmiar, typy zmiennych zbioru klientów

```
tibble [1,000 × 10] (S3: tbl_df/tbl/data.frame)
 $ numer klienta : num [1:1000] 6 12 12 9 3 7 7 11 9 2 ...
 $ czas rozmowy  : num [1:1000] 129 150 140 175 11 69 101 135 171 97 ...
 $ kwota zakupu  : num [1:1000] 64 128 172 122 195 154 183 159 69 205 ...
 $ przedstawiciel: chr [1:1000] "P03" "P03" "P02" "P05" ...
 $ data rozmowy  : POSIXct[1:1000], format: "2013-09-19" "2013-10-07" "2013-10-04" "2013-10-20" ...
 $ dzien         : num [1:1000] 19 7 4 20 28 7 5 25 26 14 ...
 $ miesiac       : num [1:1000] 9 10 10 10 8 9 10 10 8 8 ...
 $ rok           : num [1:1000] 2013 2013 2013 2013 2013 ...
 $ Skuteczność   : num [1:1000] 0.496 0.853 1.229 0.697 17.727 ...
 $ Region        : chr [1:1000] "Południe" "Południe" "Zachód" "Centrum" ...
```

### Dane statystyczne (istotnych) zmiennych ilościowych

czas rozmowy	kwota zakupu	Skuteczność
Min. : 1.00	Min. : 20.0	Min. : 0.1282
1st Qu.: 46.00	1st Qu.: 70.0	1st Qu.: 0.7494
Median : 95.00	Median :124.0	Median : 1.3398
Mean : 92.57	Mean :122.9	Mean : 3.5921
3rd Qu.:137.00	3rd Qu.:173.2	3rd Qu.: 2.5765
Max. :179.00	Max. :225.0	Max. :158.0000

### Ilość wystąpień danego przedstawiciela w zbiorze (zmienna kategoryczna)

P01	P02	P03	P04	P05
189	218	207	186	200

Po wstępnej analizie można zaobserwować, że głównie, że najwięcej rozmów odbył przedstawiciel Krzysztof Bąk (P02), natomiast najlepszą skuteczność posiada przedstawiciel Jan Kowalski (P01). Ponadto najdłuższy czas rozmowy to nie całe 3 minuty, a najkrótszy to 1 sekunda, przy czym średnia długość to ok. 1.5 minuty. Najwyższa kwota zakupu wynosiła 225, najmniejsza 20, a średnia 124. Najwyższa uzyskana skuteczność to 158, najmniejsza 0.13, a średnia 3.59. Zbiór nie posiada braków danych.

## 4. Wstępna analiza i przygotowanie danych – dane lokaty

Zbiór posiada 41 188 rekordów.

Fragment tabeli danych lokaty

age	job	marital	education	default	housing	loan	contact	month	day_of_week	duration	campaign	pdays	previous	poutcome	emp.var.rate	cons.price.idx	cons.conf.idx	euribor3m	nr.employed	y
56	housemaid	married	basic.4y	no	no	no	telephone	may	mon	261	1	999	0	nonexistent	1,1	93,99	-36,40	4,86	5191	no
57	services	married	high.school	unknown	no	no	telephone	may	mon	149	1	999	0	nonexistent	1,1	93,99	-36,40	4,86	5191	no
37	services	married	high.school	no	yes	no	telephone	may	mon	226	1	999	0	nonexistent	1,1	93,99	-36,40	4,86	5191	no
40	admin.	married	basic.6y	no	no	no	telephone	may	mon	151	1	999	0	nonexistent	1,1	93,99	-36,40	4,86	5191	no
56	services	married	high.school	no	no	yes	telephone	may	mon	307	1	999	0	nonexistent	1,1	93,99	-36,40	4,86	5191	no
45	services	married	basic.9y	unknown	no	no	telephone	may	mon	198	1	999	0	nonexistent	1,1	93,99	-36,40	4,86	5191	no

  

duration	day_of_w	month (num)	year	month	year	age range	CPI range	CCI range	euribor
4,35	1	5	2008	05.2008	50-60	4-5	93,5-94	-40 do -35	4-5
2,48	1	5	2008	05.2008	50-60	2-3	93,5-94	-40 do -35	4-5
3,77	1	5	2008	05.2008	30-40	3-4	93,5-94	-40 do -35	4-5
2,52	1	5	2008	05.2008	40-50	2-3	93,5-94	-40 do -35	4-5
5,12	1	5	2008	05.2008	50-60	5-6	93,5-94	-40 do -35	4-5
3,30	1	5	2008	05.2008	40-50	3-4	93,5-94	-40 do -35	4-5

Struktura, rozmiar, typy zmiennych zbioru lokaty

```
tibble [41,188 × 31] (S3: tbl_df/tbl/data.frame)
 $ age          : num [1:41188] 56 57 37 40 56 45 59 41 24 25 ...
 $ job          : chr [1:41188] "housemaid" "services" "services" "admin." ...
 $ marital      : chr [1:41188] "married" "married" "married" "married" ...
 $ education    : chr [1:41188] "basic.4y" "high.school" "high.school" "basic.6y" ...
 $ default      : chr [1:41188] "no" "unknown" "no" "no" ...
 $ housing      : chr [1:41188] "no" "no" "yes" "no" ...
 $ loan         : chr [1:41188] "no" "no" "no" "no" ...
 $ contact      : chr [1:41188] "telephone" "telephone" "telephone" "telephone" ...
 $ month        : chr [1:41188] "may" "may" "may" "may" ...
 $ day_of_week  : chr [1:41188] "mon" "mon" "mon" "mon" ...
 $ duration     : num [1:41188] 261 149 226 151 307 198 139 217 380 50 ...
 $ campaign     : num [1:41188] 1 1 1 1 1 1 1 1 1 1 ...
 $ pdays        : num [1:41188] 999 999 999 999 999 999 999 999 999 999 ...
 $ previous     : num [1:41188] 0 0 0 0 0 0 0 0 0 0 ...
 $ poutcome     : chr [1:41188] "nonexistent" "nonexistent" "nonexistent" "nonexistent" ...
 $ emp.var.rate : num [1:41188] 1.1 1.1 1.1 1.1 1.1 1.1 1.1 1.1 1.1 1.1 ...
 $ cons.price.idx : num [1:41188] 94 94 94 94 94 94 ...
 $ cons.conf.idx : num [1:41188] -36.4 -36.4 -36.4 -36.4 -36.4 -36.4 -36.4 -36.4 -36.4 ...
 $ euribor3m    : num [1:41188] 4.86 4.86 4.86 4.86 4.86 ...
 $ nr.employed  : num [1:41188] 5191 5191 5191 5191 5191 ...
 $ y           : chr [1:41188] "no" "no" "no" "no" ...
 $ duration [min] : num [1:41188] 4.35 2.48 3.77 2.52 5.12 ...
 $ day_of_week [num] : num [1:41188] 1 1 1 1 1 1 1 1 1 1 ...
 $ month [num]    : num [1:41188] 5 5 5 5 5 5 5 5 5 5 ...
 $ year          : num [1:41188] 2008 2008 2008 2008 2008 ...
 $ date          : POSIXct[1:41188], format: "2008-05-01" "2008-05-01" "2008-05-01" "2008-05-01" ...
 $ age range     : chr [1:41188] "50-60" "50-60" "30-40" "40-50" ...
 $ duration range [min]: chr [1:41188] "4-5" "2-3" "3-4" "2-3" ...
 $ CPI range     : chr [1:41188] "93,5-94" "93,5-94" "93,5-94" "93,5-94" ...
 $ CCI range     : chr [1:41188] "-40 do -35" "-40 do -35" "-40 do -35" "-40 do -35" ...
 $ euribor3m range : chr [1:41188] "4-5" "4-5" "4-5" "4-5" ...
```

Dane statystyczne (istotnych) zmiennych ilościowych

age	duration	emp.var.rate	cons.price.idx	cons.conf.idx	euribor3m	nr.employed
Min.: 17.00	Min.: 0.0	Min.: -3.40000	Min.: 92.20	Min.: -50.8	Min.: 0.634	Min.: 4964
1st Qu.: 32.00	1st Qu.: 102.0	1st Qu.: -1.80000	1st Qu.: 93.08	1st Qu.: -42.7	1st Qu.: 1.344	1st Qu.: 5099
Median: 38.00	Median: 180.0	Median: 1.10000	Median: 93.75	Median: -41.8	Median: 4.857	Median: 5191
Mean: 40.02	Mean: 258.3	Mean: 0.08189	Mean: 93.58	Mean: -40.5	Mean: 3.621	Mean: 5167
3rd Qu.: 47.00	3rd Qu.: 319.0	3rd Qu.: 1.40000	3rd Qu.: 93.99	3rd Qu.: -36.4	3rd Qu.: 4.961	3rd Qu.: 5228
Max.: 98.00	Max.: 4918.0	Max.: 1.40000	Max.: 94.77	Max.: -26.9	Max.: 5.045	Max.: 5228

Liczba klientów, którzy zdecydowali się zapisać na lokatę terminową (zmienna kategoryjna)

no	yes
36548	4640

Podział grup wiekowych na określone przedziały (zmienna kategoryjna)

20-30	30-40	40-50	50-60	60-70	70-80	80-90	do 20	ponad 90
5594	16938	10526	6862	724	319	140	75	10

Po wstępnej analizie można zauważyć, że z 41 188 klientów, tylko 4640 z nich zdecydowało się zapisać na lokatę terminową. Najmłodszy potencjalny klient miał 17 lat, a najstarszy 98. Natomiast największą grupę klientów docelowych stanowiła grupa od 30 do 40 lat.