

```
# 1. Wczytywanie danych i wyświetlanie podstawowych informacji
import pandas as pd
```

```
file_path = 'RZ-opiekunowie.csv'
df = pd.read_csv(file_path)
print(df.head())
print(df.info())
print(df.describe())
```

	Imię	Nazwisko	Województwo	Powiat	
Gmina \					
0	Izabela	Wójcik	MAZOWIECKIE	Warszawa	Warszawa - Praga-Północ
1	Paulina	Pomorska	MAZOWIECKIE	Warszawa	Warszawa - Mokotów
2	Anna	Spiechow	MAZOWIECKIE	otwocki	Otwock
3	Angelika	Rogaczewska	POMORSKIE	Gdańsk	Gdańsk
4	Paulina	Adamus	ŁÓDZKIE	Łódź	Łódź - Śródmieście

	Kod pocztowy	Liczba miejsc	Liczba dzieci zapisanych \
0	03-484	8	5.0
1	02-621	8	8.0
2	05-400	8	8.0
3	80-464	5	5.0
4	90-266	5	8.0

	Liczba miejsc ze środków FERS	Liczba dzieci zapisanych na miejsca z FERS \
0	0.0	0.0
1	0.0	0.0
2	NaN	NaN
3	NaN	NaN
4	0.0	0.0

	Opłata miesięczna za pobyt	Opłata miesięczna na miejscach z KP0/FERS \
0	NaN	0.0 zł
1	1950.0 zł	0.0 zł
2	1500.0 zł	0.0 zł

3	1950.0 zł	0.0 zł
4	1500.0 zł	0.0 zł

Stawka godzinowa za każdą godzinę powyżej 10 godzin \		
0		NaN
1		NaN
2		NaN
3	50.0 zł	
4		NaN

Opłata godzinowa - podstawowa opłata ponoszona przez rodziców za pobyt dziecka (bez zniżek i bez wyżywienia) \

0	NaN
1	NaN
2	NaN
3	NaN
4	NaN

Opłata za wyżywienie - miesięczna		Opłata za wyżywienie - dzienna
0	NaN	18.0 zł
1	300.0 zł	NaN
2	NaN	20.0 zł
3	NaN	20.0 zł
4	200.0 zł	NaN

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 2495 entries, 0 to 2494

Data columns (total 16 columns):

```
#      Column
Non-Null Count  Dtype
---  -
0      Imię
2495 non-null   object
1      Nazwisko
2495 non-null   object
2      Województwo
2495 non-null   object
3      Powiat
2495 non-null   object
4      Gmina
2495 non-null   object
5      Kod pocztowy
2495 non-null   object
```

```

6   Liczba miejsc
2495 non-null   int64
7   Liczba dzieci zapisanych
2204 non-null   float64
8   Liczba miejsc ze środków FERS
1094 non-null   float64
9   Liczba dzieci zapisanych na miejsca z FERS
1094 non-null   float64
10  Opłata miesięczna za pobyt
2356 non-null   object
11  Opłata miesięczna na miejscach z KP0/FERS
2495 non-null   object
12  Stawka godzinowa za każdą godzinę powyżej 10 godzin
344 non-null    object
13  Opłata godzinowa - podstawowa opłata ponoszona przez rodziców za
pobyt dziecka (bez zniżek i bez wyżywienia) 97 non-null    object
14  Opłata za wyżywienie - miesięczna
339 non-null    object
15  Opłata za wyżywienie - dzienna
2090 non-null   object
dtypes: float64(3), int64(1), object(12)
memory usage: 312.0+ KB
None

```

	Liczba miejsc środków FERS	Liczba dzieci zapisanych	Liczba miejsc ze środków FERS \
count	2495.000000	2204.000000	1094.000000
mean	6.238878	4.544918	0.470750
std	2.061565	3.066560	1.767492
min	0.000000	0.000000	0.000000
25%	5.000000	2.000000	0.000000
50%	5.000000	5.000000	0.000000
75%	8.000000	8.000000	0.000000
max	30.000000	30.000000	8.000000

	Liczba dzieci zapisanych na miejsca z FERS
count	1094.000000
mean	0.436015
std	1.677858
min	0.000000
25%	0.000000
50%	0.000000

75%	0.000000
max	8.000000

2. Obliczanie podstawowych statystyk

```
print("2.1. Średnia liczba miejsc: "+str(df['Liczba miejsc'].mean()))
print("2.2. Średnia liczba zapisanych dzieci: "+str(df['Liczba dzieci
zapisanych'].mean()))
print("2.3. Suma miejsc ze środków FERS: "+str(df['Liczba miejsc ze
środków FERS'].sum()))
df['Opłata miesięczna za pobyt'] = pd.to_numeric(df['Opłata miesięczna
za pobyt'].str.replace(' zł', ''), errors='coerce')
print("2.4. Maksymalna opłata miesięczna za pobyt: "+str(df['Opłata
miesięczna za pobyt'].max()))
df['Opłata za wyżywienie - dzienna'] = pd.to_numeric(df['Opłata za
wyżywienie - dzienna'].str.replace(' zł', ''), errors='coerce')
print("2.5. Minimalna dzienna opłata za wyżywienie: "+str(df['Opłata
za wyżywienie - dzienna'].min()))
```

1. Średnia liczba miejsc: 6.238877755511022
2. Średnia liczba zapisanych dzieci: 4.54491833030853
3. Suma miejsc ze środków FERS: 515.0
4. Maksymalna opłata miesięczna za pobyt: 3700.0
5. Minimalna dzienna opłata za wyżywienie: 0.0

3. Identyfikacja i obsługa brakujących danych

```
print("3.1. Liczba brakujących wartości w każdej kolumnie:
"+str(df.isnull().sum()))
missing_percentage = (df.isnull().sum() / len(df)) * 100
print("3.2. Procent brakujących danych w każdej kolumnie:")
print(missing_percentage.head(5))
```

Usuwanie danych gdzie brakuje ponad 50% kolumn

```
threshold = 50
columns_to_drop = missing_percentage[missing_percentage >
threshold].index
df_cleaned = df.drop(columns=columns_to_drop)
print(f"3.3. Kolumny usunięte z powodu dużej liczby braków (>
{threshold}%): ")
print(", ".join(columns_to_drop))
```

Uzupełnianie brakujących w kolumnie "Opłata miesięczna za pobyt" danych średnią wartością

```
df_cleaned['Opłata miesięczna za pobyt'] = df_cleaned['Opłata
miesięczna za pobyt'].fillna(df_cleaned['Opłata miesięczna za
pobyt'].mean())
```

```
print("3.4. Dane po uzupełnieniu braków w wybranych kolumnach
(pierwszych 5):")
print(df_cleaned.head(5))
```

```

3.1. Liczba brakujących wartości w każdej kolumnie: Imię
0
Nazwisko
0
Województwo
0
Powiat
0
Gmina
0
Kod pocztowy
0
Liczba miejsc
0
Liczba dzieci zapisanych
291
Liczba miejsc ze środków FERS
1401
Liczba dzieci zapisanych na miejsca z FERS
1401
Opłata miesięczna za pobyt
139
Opłata miesięczna na miejscach z KP0/FERS
0
Stawka godzinowa za każdą godzinę powyżej 10 godzin
2151
Opłata godzinowa - podstawowa opłata ponoszona przez rodziców za pobyt
dziecka (bez zniżek i bez wyżywienia) 2398
Opłata za wyżywienie - miesięczna
2156
Opłata za wyżywienie - dzienna
405
dtype: int64
3.2. Procent brakujących danych w każdej kolumnie:
Imię 0.0
Nazwisko 0.0
Województwo 0.0
Powiat 0.0
Gmina 0.0
dtype: float64
3.3. Kolumny usunięte z powodu dużej liczby braków (> 50%):
Liczba miejsc ze środków FERS, Liczba dzieci zapisanych na miejsca z
FERS, Stawka godzinowa za każdą godzinę powyżej 10 godzin, Opłata
godzinowa - podstawowa opłata ponoszona przez rodziców za pobyt
dziecka (bez zniżek i bez wyżywienia), Opłata za wyżywienie -
miesięczna
3.4. Dane po uzupełnieniu braków w wybranych kolumnach (pierwszych 5):
Imię Nazwisko Województwo Powiat
Gmina \
0 Izabela Wójcik MAZOWIECKIE Warszawa Warszawa - Praga-

```

Północ					
1	Paulina	Pomorska	MAZOWIECKIE	Warszawa	Warszawa -
	Mokotów				
2	Anna	Spiechow	MAZOWIECKIE	otwocki	
	Otwock				
3	Angelika	Rogaczewska	POMORSKIE	Gdańsk	
	Gdańsk				
4	Paulina	Adamus	ŁÓDZKIE	Łódź	Łódź -
	Śródmieście				

	Kod pocztowy	Liczba miejsc	Liczba dzieci zapisanych	\
0	03-484	8	5.0	
1	02-621	8	8.0	
2	05-400	8	8.0	
3	80-464	5	5.0	
4	90-266	5	8.0	

	Opłata miesięczna za pobyt	Opłata miesięczna na miejscach z
KP0/FERS	\	
0	1580.53472	0.0
zł		
1	1950.00000	0.0
zł		
2	1500.00000	0.0
zł		
3	1950.00000	0.0
zł		
4	1500.00000	0.0
zł		

	Opłata za wyżywienie - dzienna
0	18.0
1	NaN
2	20.0
3	20.0
4	NaN

4. Wykrywanie wartości odstających

```
import numpy as np
import matplotlib.pyplot as plt
```

```
kolumny_analizowane = ['Opłata miesięczna za pobyt'] # Można dopisać
więcej kolumn do przetwarzania i wizualizacji
```

Metodą IRQ

```
def detect_outliers_iqr(data, column):
    Q1 = data[column].quantile(0.25)
    Q3 = data[column].quantile(0.75)
    IQR = Q3 - Q1
    lower_bound = Q1 - 1.5 * IQR
```

```

    upper_bound = Q3 + 1.5 * IQR
    outliers = data[(data[column] < lower_bound) | (data[column] >
upper_bound)]
    return outliers, lower_bound, upper_bound

# Wykrywanie odstających dla każdej kolumny
for kolumna in kolumny_analizowane:
    print(f"\nAnaliza kolumny: {kolumna}")
    outliers, lower, upper = detect_outliers_iqr(df, kolumna)
    print(f"Wartości odstające w kolumnie '{kolumna}':")
    print(outliers[kolumna])
    print(f"Dolna granica: {lower}, Górna granica: {upper}")

# Wizualizacja za pomocą wykresów pudełkowych
for kolumna in kolumny_analizowane:
    plt.figure(figsize=(8, 4))
    plt.boxplot(df[kolumna].dropna(), vert=False, patch_artist=True)
    plt.title(f"Wykres pudełkowy - {kolumna}")
    plt.xlabel(kolumna)
    plt.show()

```

Analiza kolumny: Opłata miesięczna za pobyt

Wartości odstające w kolumnie 'Opłata miesięczna za pobyt':

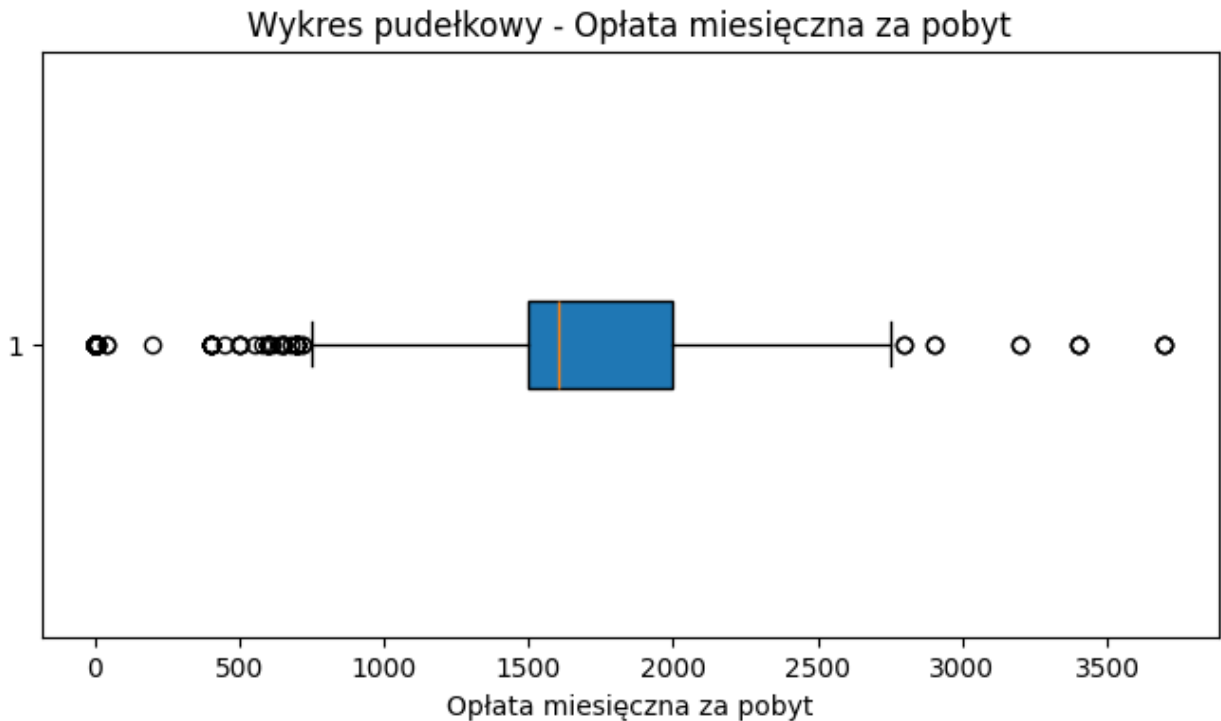
```

5          0.0
7          0.0
20         0.0
21         0.0
24         0.0
...
2424      700.0
2461       0.0
2465      700.0
2474       0.0
2485       0.0

```

Name: Opłata miesięczna za pobyt, Length: 205, dtype: float64

Dolna granica: 750.0, Górna granica: 2750.0



5. Analiza zależności między kolumnami

```
import seaborn as sns
import matplotlib.pyplot as plt
```

```
kolumny = ['Liczba miejsc', 'Liczba dzieci zapisanych', 'Liczba miejsc ze środków FERS', 'Opłata miesięczna za pobyt']
```

Obliczanie macierzy korelacji

```
macierz_korelacji = df[kolumny].corr()
```

```
print("5.1. Macierz korelacji:")
```

```
print(macierz_korelacji)
```

```
plt.figure(figsize=(8, 6))
```

```
sns.heatmap(macierz_korelacji, annot=True, cmap='coolwarm', fmt=".2f")
```

```
plt.title("5.2. Macierz korelacji - mapa ciepła")
```

```
plt.show()
```

```
print("5.3. Zależność: Opłata miesięczna za pobyt a Liczba dzieci zapisanych")
```

```
sns.regplot(data=df, x='Opłata miesięczna za pobyt', y='Liczba dzieci zapisanych')
```

```
plt.title("Zależność: Opłata miesięczna za pobyt a Liczba dzieci zapisanych")
```

```
plt.xlabel("Opłata miesięczna za pobyt")
```

```
plt.ylabel("Liczba dzieci zapisanych")
```

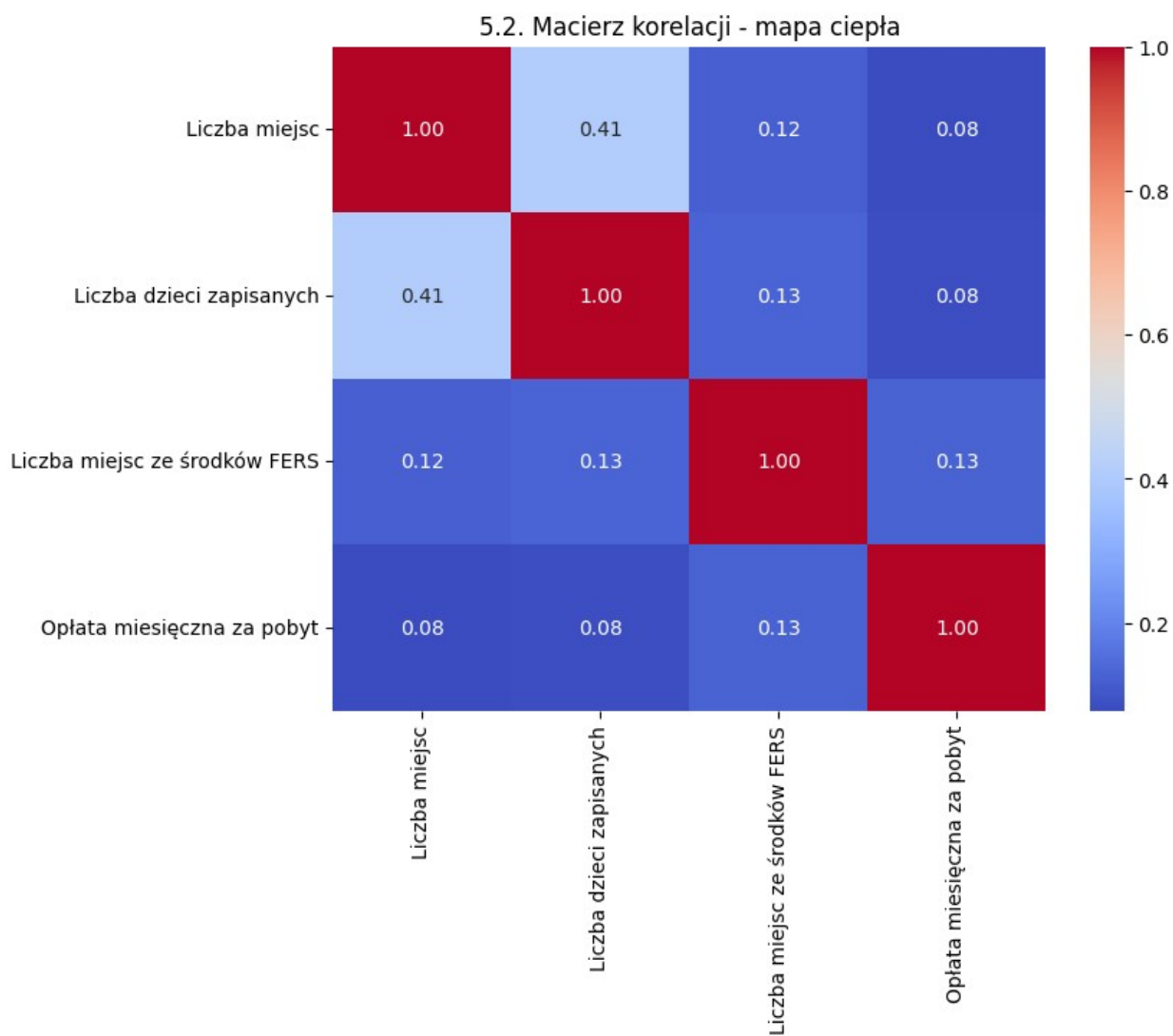
```
plt.show()
```


5.1. Macierz korelacji:

	Liczba miejsc	Liczba dzieci zapisanych
\		
Liczba miejsc	1.000000	0.406514
Liczba dzieci zapisanych	0.406514	1.000000
Liczba miejsc ze środków FERS	0.120113	0.130622
Opłata miesięczna za pobyt	0.078668	0.083243

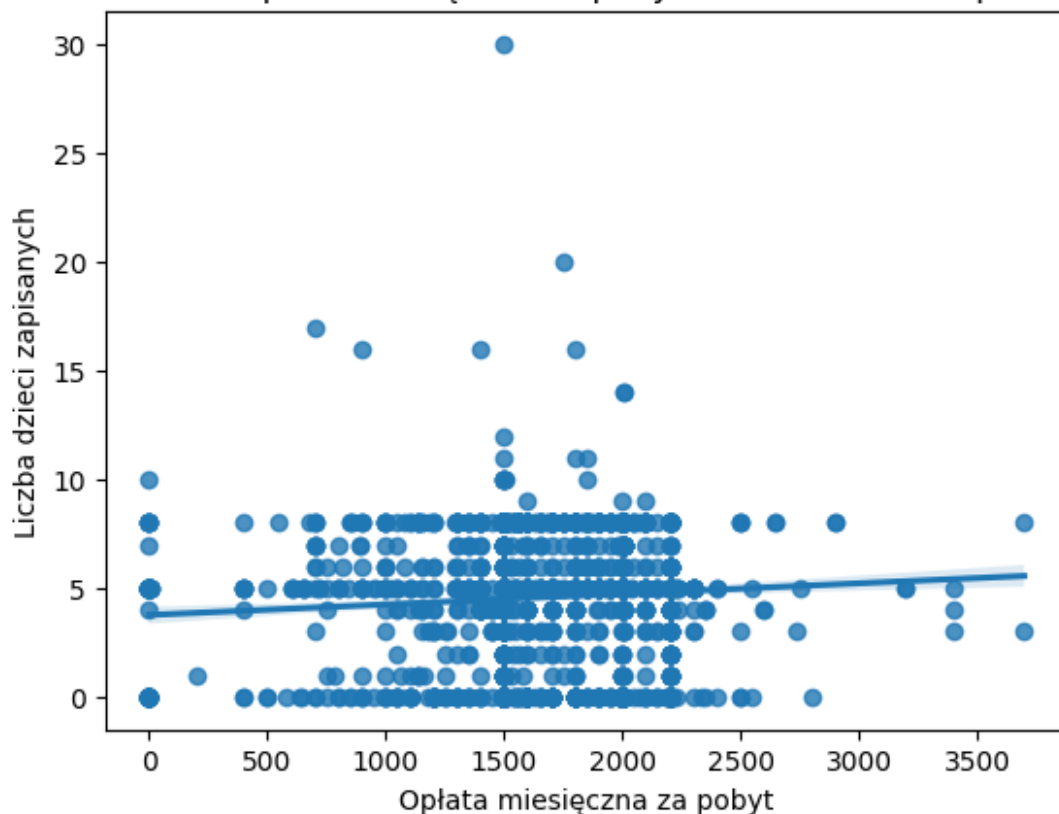
	Liczba miejsc ze środków FERS	\
Liczba miejsc	0.120113	
Liczba dzieci zapisanych	0.130622	
Liczba miejsc ze środków FERS	1.000000	
Opłata miesięczna za pobyt	0.128428	

	Opłata miesięczna za pobyt	
Liczba miejsc	0.078668	
Liczba dzieci zapisanych	0.083243	
Liczba miejsc ze środków FERS	0.128428	
Opłata miesięczna za pobyt	1.000000	



5.3. Zależność: Оплата miesięczna za pobyt a Liczba dzieci zapisanych

Zależność: Opłata miesięczna za pobyt a Liczba dzieci zapisanych



```
# 6. Przekształcanie danych
```

```
df['Suma miejsc i dzieci'] = df['Liczba miejsc'] + df['Liczba dzieci zapisanych']
```

```
print("6.1. Nowa kolumna 'Suma miejsc i dzieci':")
```

```
print(df.head())
```

```
print("6.2. Grupowanie po województwie i obliczanie średnich:")
```

```
grouped_by_wojewodztwo = df.groupby('Województwo')
```

```
print(grouped_by_wojewodztwo.head(5))
```

```
print("6.3. Sortowanie po kolumnie 'Opłata miesięczna za pobyt':")
```

```
sorted_df = df.sort_values(by='Opłata miesięczna za pobyt',  
                           ascending=True)
```

```
print(sorted_df.head(5))
```

```
print("6.4. Sortowanie po 'Województwo' i 'Liczba dzieci zapisanych':")
```

```
sorted_multi = df.sort_values(by=['Województwo', 'Liczba dzieci zapisanych'],  
                              ascending=[True, False])
```

```
print(sorted_multi.head(5))
```

6.1. Nowa kolumna 'Suma miejsc i dzieci':

	Imię	Nazwisko	Województwo	Powiat	
Gmina \					
0	Izabela	Wójcik	MAZOWIECKIE	Warszawa	Warszawa - Praga-Północ
1	Paulina	Pomorska	MAZOWIECKIE	Warszawa	Warszawa - Mokotów
2	Anna	Spiechow	MAZOWIECKIE	otwocki	Otwock
3	Angelika	Rogaczewska	POMORSKIE	Gdańsk	Gdańsk
4	Paulina	Adamus	ŁÓDZKIE	Łódź	Łódź-Śródmieście

	Kod pocztowy	Liczba miejsc	Liczba dzieci zapisanych \
0	03-484	8	5.0
1	02-621	8	8.0
2	05-400	8	8.0
3	80-464	5	5.0
4	90-266	5	8.0

	Liczba miejsc ze środków FERS	Liczba dzieci zapisanych na miejsca z FERS \
0	0.0	0.0
1	0.0	0.0
2	NaN	NaN
3	NaN	NaN
4	0.0	0.0

	Opłata miesięczna za pobyt	Opłata miesięczna na miejscach z KP0/FERS \
0	NaN	0.0
zł		
1	1950.0	0.0
zł		
2	1500.0	0.0
zł		
3	1950.0	0.0
zł		
4	1500.0	0.0
zł		

	Stawka godzinowa za każdą godzinę powyżej 10 godzin \
0	NaN
1	NaN

2	NaN
3	50.0 zł
4	NaN

Opłata godzinowa - podstawowa opłata ponoszona przez rodziców za pobyt dziecka (bez zniżek i bez wyżywienia) \

0	NaN
1	NaN
2	NaN
3	NaN
4	NaN

	Opłata za wyżywienie - miesięczna	Opłata za wyżywienie - dzienna \
0	NaN	18.0
1	300.0 zł	NaN
2	NaN	20.0
3	NaN	20.0
4	200.0 zł	NaN

	Suma miejsc i dzieci
0	13.0
1	16.0
2	16.0
3	10.0
4	13.0

6.2. Grupowanie po województwie i obliczanie średnich:

	Imię	Nazwisko	Województwo	Powiat
Gmina \				
0	Izabela	Wójcik	MAZOWIECKIE	Warszawa Warszawa -
Praga-Północ				
1	Paulina	Pomorska	MAZOWIECKIE	Warszawa Warszawa -
Mokotów				
2	Anna	Spiechow	MAZOWIECKIE	otwocki
Otwock				
3	Angelika	Rogaczewska	POMORSKIE	Gdańsk
Gdańsk				
4	Paulina	Adamus	ŁÓDZKIE	Łódź Łódź -
Śródmieście				
...
...				
791	Klaudia	Krajewska	PODLASKIE	sejneński
Krasnopol				
865	Edyta	Suchowierska	PODLASKIE	Białystok
Białystok				
1120	Jolanta	Czaja	OPOLSKIE	prudnicki

Głogówek				
1207	Julia	Milanowska	PODLASKIE	Białystok
Białystok				
1681	Marta	Popów	OPOLSKIE	brzeski
Brzeg				

	Kod pocztowy	Liczba miejsc	Liczba dzieci zapisanych	\
0	03-484	8	5.0	
1	02-621	8	8.0	
2	05-400	8	8.0	
3	80-464	5	5.0	
4	90-266	5	8.0	
...	
791	16-503	0	5.0	
865	15-062	8	8.0	
1120	48-250	5	5.0	
1207	15-698	5	5.0	
1681	49-306	8	8.0	

	Liczba miejsc ze środków FERS	\
0	0.0	
1	0.0	
2	NaN	
3	NaN	
4	0.0	
...	...	
791	NaN	
865	NaN	
1120	0.0	
1207	5.0	
1681	NaN	

	Liczba dzieci zapisanych na miejsca z FERS	Opłata miesięczna za pobyt	\
0	0.0		
NaN			
1	0.0		
1950.0			
2	NaN		
1500.0			
3	NaN		
1950.0			
4	0.0		
1500.0			
...	...		
...			
791	NaN		
NaN			
865	NaN		
1800.0			

1120	0.0
1200.0	
1207	5.0
2010.0	
1681	NaN
1500.0	

Opłata miesięczna na miejscach z KP0/FERS \	
0	0.0 zł
1	0.0 zł
2	0.0 zł
3	0.0 zł
4	0.0 zł
...	...
791	0.0 zł
865	0.0 zł
1120	0.0 zł
1207	2010.0 zł
1681	0.0 zł

Stawka godzinowa za każdą godzinę powyżej 10 godzin \	
0	NaN
1	NaN
2	NaN
3	50.0 zł
4	NaN
...	...
791	NaN
865	NaN
1120	NaN
1207	NaN
1681	40.0 zł

Opłata godzinowa - podstawowa opłata ponoszona przez rodziców za pobyt dziecka (bez zniżek i bez wyżywienia) \

0	NaN
1	NaN
2	NaN
3	NaN
4	NaN
...	...
791	7.0 zł
865	NaN

1120	NaN
1207	NaN
1681	NaN

	Opłata za wyżywienie - miesięczna	Opłata za wyżywienie - dzienna
\		
0	NaN	18.0
1	300.0 zł	NaN
2	NaN	20.0
3	NaN	20.0
4	200.0 zł	NaN
...
791	NaN	7.0
865	NaN	16.0
1120	0.0 zł	NaN
1207	NaN	18.0
1681	17.0 zł	NaN

	Suma miejsc i dzieci
0	13.0
1	16.0
2	16.0
3	10.0
4	13.0
...	...
791	5.0
865	16.0
1120	10.0
1207	10.0
1681	16.0

[80 rows x 17 columns]

6.3. Sortowanie po kolumnie 'Opłata miesięczna za pobyt':

	Imię	Nazwisko	Województwo	Powiat	Gmina
\					
20	Katarzyna	Adamska	MAZOWIECKIE	Warszawa	Warszawa - Mokotów

25	Karolina	Goczał	MAZOWIECKIE	Warszawa	Warszawa - Mokotów
24	Karolina	Józwik	MAZOWIECKIE	Warszawa	Warszawa - Mokotów
2461	Karolina	Bartoszuk	MAZOWIECKIE	Warszawa	Warszawa - Wawer
1215	Sandra	Wszołek	MAZOWIECKIE	Warszawa	Warszawa - Wawer

	Kod pocztowy	Liczba miejsc	Liczba dzieci zapisanych	\
20	02-677	5	5.0	
25	02-677	5	5.0	
24	02-677	5	5.0	
2461	04-964	5	5.0	
1215	04-964	5	5.0	

	Liczba miejsc ze środków FERS	\
20	0.0	
25	0.0	
24	0.0	
2461	NaN	
1215	NaN	

	Liczba dzieci zapisanych na miejsca z FERS	Opłata miesięczna za pobyt	\
20	0.0	0.0	
25	0.0	0.0	
24	0.0	0.0	
2461	NaN	NaN	
1215	NaN	NaN	

	Opłata miesięczna na miejscach z KP0/FERS	\
20	0.0 zł	
25	0.0 zł	
24	0.0 zł	
2461	0.0 zł	
1215	0.0 zł	

	Stawka godzinowa za każdą godzinę powyżej 10 godzin	\
20	NaN	
25	NaN	
24	NaN	
2461	NaN	
1215	NaN	

Opłata godzinowa - podstawowa opłata ponoszona przez rodziców za pobyt dziecka (bez zniżek i bez wyżywienia) \

20 NaN

25 NaN

24 NaN

2461 NaN

1215 NaN

Opłata za wyżywienie - miesięczna Opłata za wyżywienie - dzienna

\ 20 NaN 22.0

25 NaN 22.0

24 NaN 22.0

2461 NaN 25.0

1215 NaN 25.0

Suma miejsc i dzieci

20 10.0

25 10.0

24 10.0

2461 10.0

1215 10.0

6.4. Sortowanie po 'Województwo' i 'Liczba dzieci zapisanych':

Imię Nazwisko Województwo Powiat

Gmina \

566 URSZULA STACHOWICZ DOLNOŚLĄSKIE oławski Jelcz-

Laskowice

1307 Paulina Duda DOLNOŚLĄSKIE wrocławski

Długołęka

621 Gabriela Gredes DOLNOŚLĄSKIE złotoryjski

Wojcieszów

1310 Agnieszka Jacewicz DOLNOŚLĄSKIE wrocławski

Długołęka

52 Anna Prażmowska DOLNOŚLĄSKIE lwówecki Lwówek

Śląski

Kod pocztowy Liczba miejsc Liczba dzieci zapisanych \

566 55-220 8 16.0

1307 55-093 8 12.0

621 59-550 5 10.0

1310	55-093	8	10.0
52	59-600	8	8.0
Liczba miejsc ze środków FERS \			
566		0.0	
1307		NaN	
621		0.0	
1310		NaN	
52		0.0	
Liczba dzieci zapisanych na miejsca z FERS Opłata miesięczna za			
pobyt \			
566		0.0	
1400.0			
1307		NaN	
1500.0			
621		0.0	
1500.0			
1310		NaN	
1500.0			
52		0.0	
1100.0			
Opłata miesięczna na miejscach z KP0/FERS \			
566		0.0 zł	
1307		0.0 zł	
621		0.0 zł	
1310		0.0 zł	
52		0.0 zł	
Stawka godzinowa za każdą godzinę powyżej 10 godzin \			
566		NaN	
1307		NaN	
621		NaN	
1310		NaN	
52		NaN	
Opłata godzinowa - podstawowa opłata ponoszona przez rodziców za			
pobyt dziecka (bez zniżek i bez wyżywienia) \			
566		NaN	
1307		NaN	
621		NaN	
1310		NaN	
52		NaN	
Opłata za wyżywienie - miesięczna Opłata za wyżywienie - dzienna			

\		
566	NaN	NaN
1307	NaN	22.0
621	NaN	NaN
1310	NaN	22.0
52	NaN	15.0
Suma miejsc i dzieci		
566	24.0	
1307	20.0	
621	15.0	
1310	18.0	
52	16.0	