

```
In [1]: # 1. Wczytywanie danych i wyświetlanie podstawowych informacji
import pandas as pd

file_path = 'opiekuny.csv'
df = pd.read_csv(file_path)
print(df.head())
print(df.info())
print(df.describe())
```

	Imię	Nazwisko	Województwo	Powiat	Gmina \
0	Iwona	Dziubak	MAZOWIECKIE	garwoliński	Żelechów
1	Karolina	Olejniczak	POMORSKIE	kartuski	Żukowo
2	MONIKA	ADAMOWSKA	MAZOWIECKIE	Warszawa	Warszawa - Mokotów
3	Emilia	Lis	MAZOWIECKIE	Warszawa	Warszawa - Wola
4	Natalia	Kadyszewska	MAZOWIECKIE	Warszawa	Warszawa - Białołęka

	Kod pocztowy	Liczba miejsc	Liczba dzieci zapisanych \
0	08-430	8	8.0
1	80-297	8	8.0
2	00-777	5	5.0
3	01-258	8	5.0
4	03-044	8	8.0

	Liczba miejsc ze środków FERS	Liczba dzieci zapisanych na miejsca z FERS \
0	8.0	8.0
1	0.0	0.0
2	NaN	NaN
3	NaN	NaN
4	NaN	NaN

	Opłata miesięczna za pobyt	Opłata miesięczna na miejscach z KPO/FERS \
0	2000.0 zł	2000.0 zł
1	1400.0 zł	0.0 zł
2	0.0 zł	0.0 zł
3	2100.0 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	NaN
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	17.0 zł
1	NaN	19.5 zł
2	NaN	22.0 zł
3	NaN	22.0 zł
4	NaN	18.0 zł

```
<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 KPO/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 wy żywien ie - miesięczna
339 non-null    object
15 Opłata za wy żywien ie - dzienna
2090 non-null    object
dtypes: float64(3), int64(1), object(12)
memory usage: 312.0+ KB
None

```

	Liczba miejsc	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

```
In [2]: # 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']))
print("2.3. Suma miejsc ze środków FERS: "+str(df['Liczba miejsc ze środków FERS']))
df['Opłata miesięczna za pobyt'] = pd.to_numeric(df['Opłata miesięczna za pobyt'])
print("2.4. Maksymalna opłata miesięczna za pobyt: "+str(df['Opłata miesięczna za pobyt'].max()))
df['Opłata za wy żywien ie - dzienna'] = pd.to_numeric(df['Opłata za wy żywien ie - dzienna'])
print("2.5. Minimalna dzienna opłata za wy żywien ie: "+str(df['Opłata za wy żywien ie - dzienna'].min()))
```

- 2.1. Średnia liczba miejsc: 6.238877755511022
- 2.2. Średnia liczba zapisanych dzieci: 4.54491833030853
- 2.3. Suma miejsc ze środków FERS: 515.0
- 2.4. Maksymalna opłata miesięczna za pobyt: 3700.0
- 2.5. Minimalna dzienna opłata za wyżywienie: 0.0

In [3]: # 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 średni
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 miejscowościach z KPO/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 (&gt; 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	Iwona	Dziubak	MAZOWIECKIE	garwoliński	Żelechów	
1	Karolina	Olejniczak	POMORSKIE	kartuski	Żukowo	
2	MONIKA	ADAMOWSKA	MAZOWIECKIE	Warszawa	Warszawa - Mokotów	
3	Emilia	Lis	MAZOWIECKIE	Warszawa	Warszawa - Wola	
4	Natalia	Kadyszewska	MAZOWIECKIE	Warszawa	Warszawa - Białołęka	

## Kod pocztowy Liczba miejsc Liczba dzieci zapisanych \

0	08-430	8	8.0
1	80-297	8	8.0
2	00-777	5	5.0
3	01-258	8	5.0
4	03-044	8	8.0

	Opłata miesięczna za pobyt	Opłata miesięczna na miejscach z KPO/FERS
0	2000.0	2000.0 zł
1	1400.0	0.0 zł
2	0.0	0.0 zł
3	2100.0	0.0 zł
4	1500.0	0.0 zł

	Opłata za wyżywienie - dzienna
0	17.0
1	19.5
2	22.0
3	22.0
4	18.0

```
In [4]: # 4. Wykrywanie wartości odstających
import numpy as np
import matplotlib.pyplot as plt

kolumny_analizowane = ['Opłata miesięczna za pobyt'] # Można dopisać więcej kolumn

# Metoda 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}, Górska 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

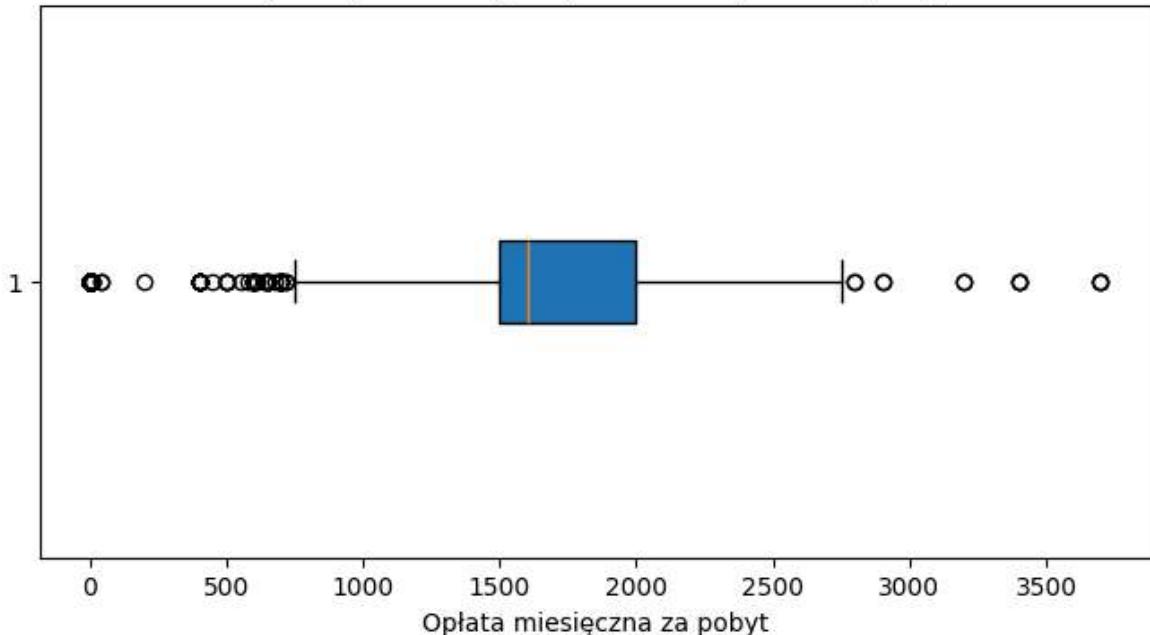
Wartości odstające w kolumnie 'Opłata miesięczna za pobyt':

```
2      0.0
16     0.0
18     0.0
35     0.0
59     0.0
...
2447    0.0
2450    0.0
2454   400.0
2479    0.0
2489   600.0
```

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

Dolna granica: 750.0, Górska granica: 2750.0

Wykres pudełkowy - Opłata miesięczna za pobyt



```
In [5]: # 5. Analiza zależności między kolumnami
import seaborn as sns
import matplotlib.pyplot as plt

kolumny = ['Liczba miejsc', 'Liczba dzieci zapisanych', 'Liczba miejsc ze środkami']

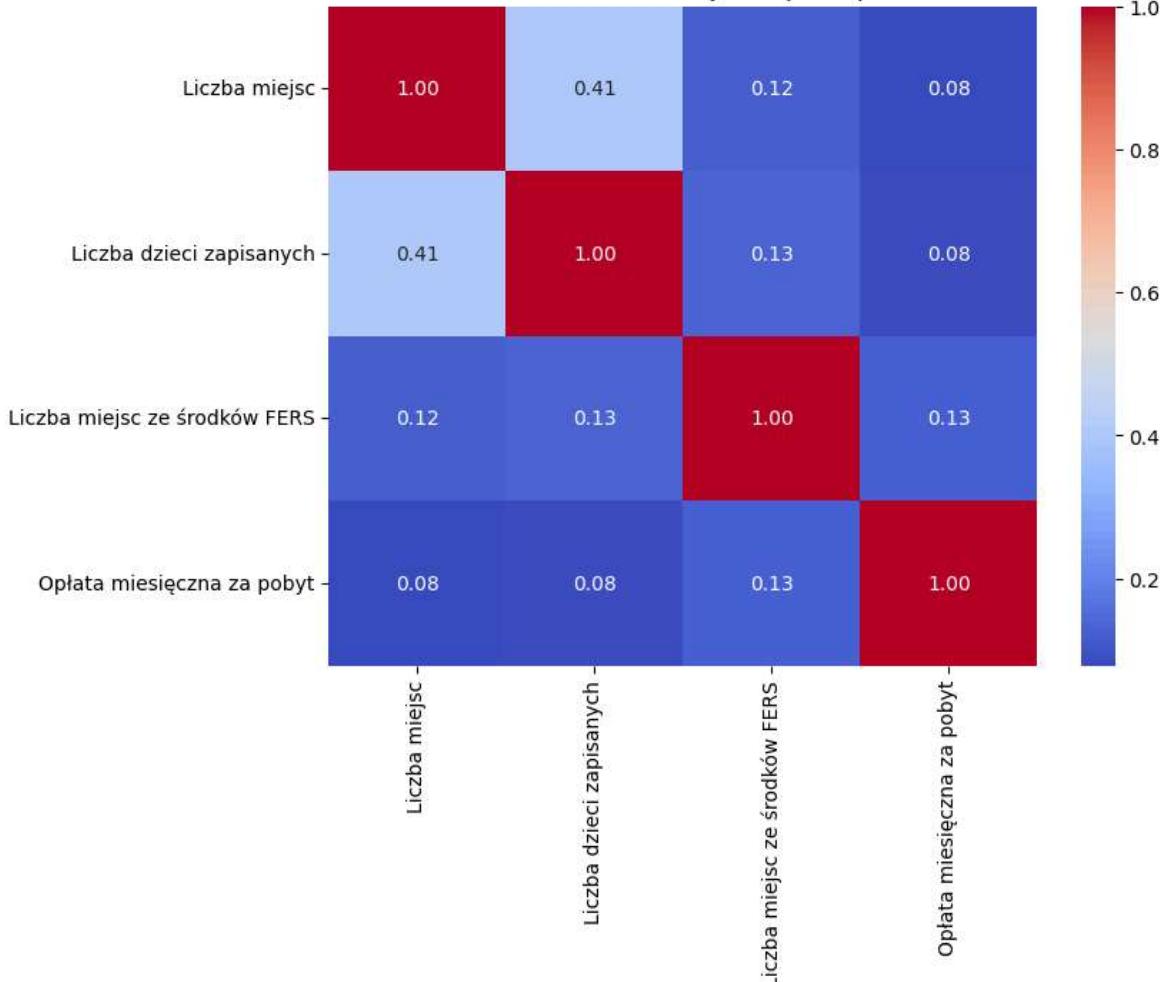
# Obliczanie macierzy korelacji
macierz_korelacji = df[kolumny].corr()
print("5.1. Macierz korelacji:")
print(macierz_korelacji)
plt.figure(figsize=(8, 6))
sns.heatmap(macierz_korelacji, annot=True, cmap='coolwarm', fmt=".2f")
plt.title("5.2. Macierz korelacji - mapa ciepła")
plt.show()

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:

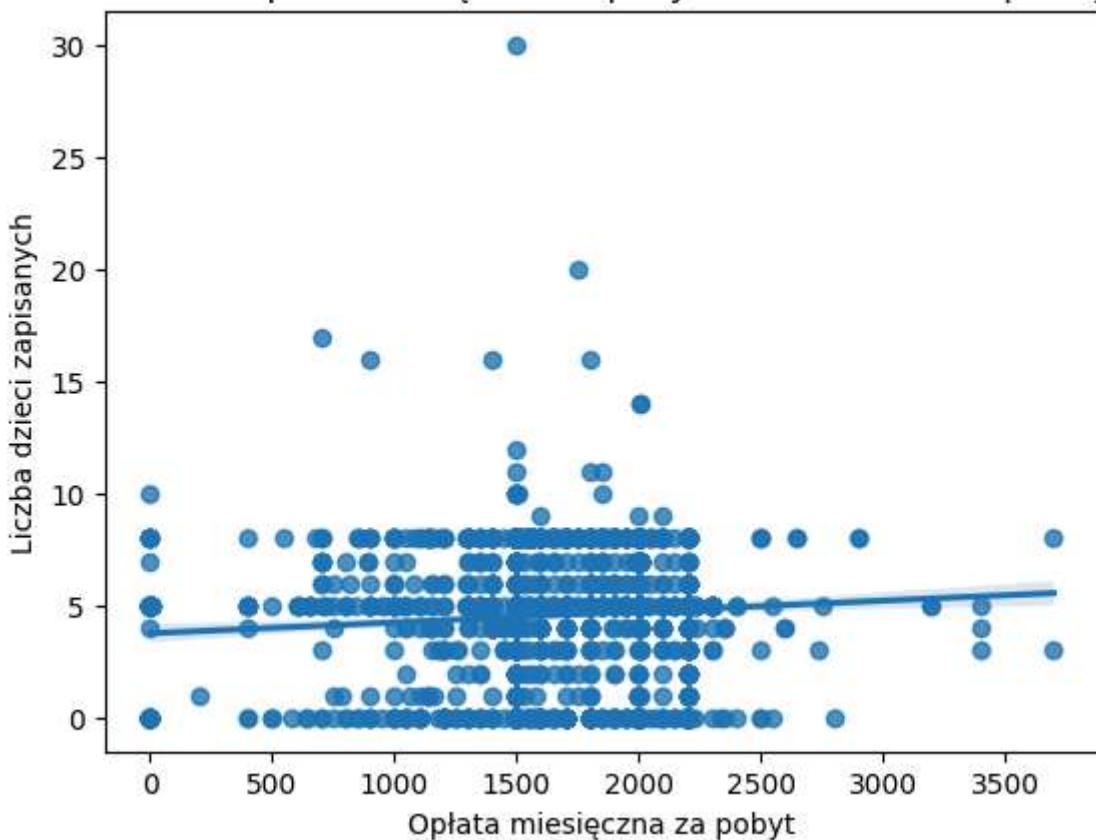
	Liczba miejsc	Liczba dzieci zapisanych	\
Liczba miejsc	1.000000	0.406514	
Liczba dzieci zapisanych	0.406514	1.000000	
Liczba miejsc ze środków FERS	0.120113	0.130622	
Opłata miesięczna za pobyt	0.078668	0.083243	
Liczba miejsc ze środków FERS \			
Liczba miejsc		0.120113	
Liczba dzieci zapisanych		0.130622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

### 5.2. Macierz korelacji - mapa ciepła



### 5.3. Zależność: Opłata miesięczna za pobyt a Liczba dzieci zapisanych

### Zależność: Opłata miesięczna za pobyt a Liczba dzieci zapisanych



```
In [6]: # 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'], as
print(sorted_multi.head(5))
```

## 6.1. Nowa kolumna 'Suma miejsc i dzieci':

	Imię	Nazwisko	Województwo	Powiat	Gmina \
0	Iwona	Dziubak	MAZOWIECKIE	garwoliński	Żelechów
1	Karolina	Olejniczak	POMORSKIE	kartuski	Żukowo
2	MONIKA	ADAMOWSKA	MAZOWIECKIE	Warszawa	Warszawa - Mokotów
3	Emilia	Lis	MAZOWIECKIE	Warszawa	Warszawa - Wola
4	Natalia	Kadyszewska	MAZOWIECKIE	Warszawa	Warszawa - Białołęka

	Kod pocztowy	Liczba miejsc	Liczba dzieci zapisanych \
0	08-430	8	8.0
1	80-297	8	8.0
2	00-777	5	5.0
3	01-258	8	5.0
4	03-044	8	8.0

	Liczba miejsc ze środków FERS	Liczba dzieci zapisanych na miejsca z FERS \
0	8.0	8.0
1	0.0	0.0
2	NaN	NaN
3	NaN	NaN
4	NaN	NaN

	Opłata miesięczna za pobyt	Opłata miesięczna na miejscach z KPO/FERS \
0	2000.0	2000.0 zł
1	1400.0	0.0 zł
2	0.0	0.0 zł
3	2100.0	0.0 zł
4	1500.0	0.0 zł

	Stawka godzinowa za każdą godzinę powyżej 10 godzin \
0	NaN
1	NaN
2	NaN
3	NaN
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	17.0
1	NaN	19.5
2	NaN	22.0
3	NaN	22.0
4	NaN	18.0

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

## 6.2. Grupowanie po województwie i obliczanie średnich:

	Imię	Nazwisko	Województwo	Powiat \
0	Iwona	Dziubak	MAZOWIECKIE	garwoliński

1	Karolina	Olejniczak	POMORSKIE	kartuski
2	MONIKA	ADAMOWSKA	MAZOWIECKIE	Warszawa
3	Emilia	Lis	MAZOWIECKIE	Warszawa
4	Natalia	Kadyszewska	MAZOWIECKIE	Warszawa
...	...	...	...	...
410	Angelina	Daria Kwiecień	ŚWIĘTOKRZYSKIE	Kielce
538	Agnieszka	Dębnicka	PODLASKIE	zambrowski
592	Magda	Mańkiewicz	PODLASKIE	zambrowski
657	Natalia	Gołębiowska	PODKARPACKIE	lubaczowski
1508	Marta	Popów	OPOLSKIE	brzeski

	Gmina	Kod pocztowy	Liczba miejsc	\
0	Żelechów	08-430	8	
1	Żukowo	80-297	8	
2	Warszawa - Mokotów	00-777	5	
3	Warszawa - Wola	01-258	8	
4	Warszawa - Białołęka	03-044	8	
...	...	...	...	...
410	Kielce	25-003	8	
538	Rutki	18-312	5	
592	Rutki	18-312	5	
657	Lubaczów	37-600	8	
1508	Brzeg	49-306	8	

	Liczba dzieci zapisanych	Liczba miejsc ze środków FERS	\
0	8.0	8.0	8.0
1	8.0	8.0	0.0
2	5.0	5.0	NaN
3	5.0	5.0	NaN
4	8.0	8.0	NaN
...	...	...	...
410	8.0	8.0	NaN
538	5.0	5.0	5.0
592	5.0	5.0	5.0
657	8.0	8.0	0.0
1508	8.0	8.0	NaN

	Liczba dzieci zapisanych na miejsca z FERS	Opłata miesięczna za pobyt	\
0	8.0	2000.0	2000.0
1	0.0	1400.0	1400.0
2	NaN	0.0	0.0
3	NaN	2100.0	2100.0
4	NaN	1500.0	1500.0
...	...	...	...
410	NaN	1500.0	1500.0
538	5.0	2200.0	2200.0
592	5.0	2200.0	2200.0
657	0.0	900.0	900.0
1508	NaN	1500.0	1500.0

	Opłata miesięczna na miejscach z KPO/FERS	\
0	2000.0 zł	
1	0.0 zł	
2	0.0 zł	
3	0.0 zł	
4	0.0 zł	
...	...	
410	0.0 zł	
538	2200.0 zł	
592	2200.0 zł	

657		0.0 zł
1508		0.0 zł

Stawka godzinowa za każdą godzinę powyżej 10 godzin \

0		NaN
1		NaN
2		NaN
3		NaN
4		NaN
...		...
410		NaN
538		NaN
592		NaN
657		NaN
1508		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
...		...
410		NaN
538		NaN
592		NaN
657		NaN
1508		NaN

0	Opłata za wyżywienie - miesięczna	Opłata za wyżywienie - dzienna	\
1		NaN	17.00
2		NaN	19.50
3		NaN	22.00
4		NaN	22.00
...		...	18.00
410		NaN	...
538		NaN	NaN
592		NaN	5.05
657		NaN	5.05
1508		17.0 zł	16.00
			NaN

Suma miejsc i dzieci

0		16.0
1		16.0
2		10.0
3		13.0
4		16.0
...		...
410		16.0
538		10.0
592		10.0
657		16.0
1508		16.0

[80 rows x 17 columns]

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

Imię	Nazwisko	Województwo	Powiat	Gmina	\
2335	Maria	Połec	MAZOWIECKIE	Warszawa	Warszawa - Mokotów

76	Angelika	Wróblewska	MAZOWIECKIE	Warszawa	Warszawa - Praga-Południe
422	JUSTYNA	DEJKO	MAZOWIECKIE	Warszawa	Warszawa - Białołęka
59	Marlena	Podlaska	MAZOWIECKIE	Warszawa	Warszawa - Praga-Południe
623	Yuliia	Cheban	MAZOWIECKIE	Warszawa	Warszawa - Mokotów

	Kod pocztowy	Liczba miejsc	Liczba dzieci zapisanych	\
2335	02-933	5	5.0	
76	03-813	5	0.0	
422	03-151	5	5.0	
59	03-813	5	0.0	
623	00-777	5	5.0	

	Liczba miejsc ze środków FERS	\
2335	0.0	
76	Nan	
422	Nan	
59	Nan	
623	Nan	

	Liczba dzieci zapisanych na miejsca z FERS	Opłata miesięczna za pobyt	\
2335	0.0	0.0	
76	Nan	0.0	
422	Nan	0.0	
59	Nan	0.0	
623	Nan	0.0	

	Opłata miesięczna na miejscach z KPO/FERS	\
2335	0.0 zł	
76	0.0 zł	
422	0.0 zł	
59	0.0 zł	
623	0.0 zł	

	Stawka godzinowa za każdą godzinę powyżej 10 godzin	\
2335	Nan	
76	Nan	
422	Nan	
59	Nan	
623	Nan	

	Opłata godzinowa - podstawowa opłata ponoszona przez rodziców za pobyt dziecka (bez zniżek i bez wyżywienia)	\
2335	Nan	
76	Nan	
422	Nan	
59	Nan	
623	Nan	

	Opłata za wyżywienie - miesięczna	Opłata za wyżywienie - dzienna	\
2335	Nan	21.0	
76	Nan	24.0	
422	Nan	20.0	
59	Nan	24.0	
623	Nan	22.0	

	Suma miejsc i dzieci	
2335	10.0	
76	5.0	
422	10.0	
59	5.0	

623 10.0

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

	Imię	Nazwisko	Województwo	Powiat	Gmina	\
1496	URSZULA	STACHOWICZ	DOLNOŚLĄSKIE	oławski	Jelcz-Laskowice	
1702	Paulina	Duda	DOLNOŚLĄSKIE	wrocławski	Długołęka	
169	Agnieszka	Jacewicz	DOLNOŚLĄSKIE	wrocławski	Długołęka	
1304	Gabriela	Gredes	DOLNOŚLĄSKIE	złotoryjski	Wojcieszów	
63	Małgorzata	Grębowicz	DOLNOŚLĄSKIE	lwówecki	Lwówek Śląski	

	Kod pocztowy	Liczba miejsc	Liczba dzieci zapisanych	\
1496	55-220	8	16.0	
1702	55-093	8	12.0	
169	55-093	8	10.0	
1304	59-550	5	10.0	
63	59-600	8	8.0	

	Liczba miejsc ze środków FERS	\
1496	0.0	
1702	NaN	
169	NaN	
1304	0.0	
63	NaN	

	Liczba dzieci zapisanych na miejsca z FERS	Opłata miesięczna za pobyt	\
1496	0.0	1400.0	
1702	NaN	1500.0	
169	NaN	1500.0	
1304	0.0	1500.0	
63	NaN	1300.0	

	Opłata miesięczna na miejscach z KPO/FERS	\
1496	0.0 zł	
1702	0.0 zł	
169	0.0 zł	
1304	0.0 zł	
63	0.0 zł	

	Stawka godzinowa za każdą godzinę powyżej 10 godzin	\
1496	NaN	
1702	NaN	
169	NaN	
1304	NaN	
63	NaN	

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

1496	NaN
1702	NaN
169	NaN
1304	NaN
63	NaN

	Opłata za wyżywienie - miesięczna	Opłata za wyżywienie - dzienna	\
1496	NaN	NaN	
1702	NaN	22.0	
169	NaN	22.0	
1304	NaN	NaN	
63	NaN	14.5	

Suma miejsc i dzieci

1496	24.0
1702	20.0
169	18.0
1304	15.0
63	16.0

In [ ]: