

Algorytm działania programu, na przykładzie dwóch tabel zapisanych w plikach CSV:

	Panstwo	Miasto
0	Hiszpania	Barcelona
1	Hiszpania	Matryt
2	Grecja	Ateny
3	Irlandia	Dublin
4	Litwa	Wilno
5	Polska	Kielce
6	Polska	Warszawa

Miasto.csv

	Panstwo	Zaludnienie(mln)
0	Litwa	2.8
1	Francja	67.5
2	Hiszpania	47.1
3	Polska	38.2
4	Rumunia	21.2

Zaludnienie.csv

Najpierw program wybiera tabelę główną oraz tabelę poboczną. Tabela główna charakteryzuje się tym, że w kolumnie łączącej mogą znajdować się powtórzenia. W podanym tutaj przykładzie, kolumną łączącą jest kolumna Państwo. W tabeli Miasto.csv znajdują się powtórzenia w kolumnie łączącej (Hiszpania i Polska powtarzają się 2 razy), a w tabeli Zaludnienie.csv nie ma powtórzeń w kolumnie łączącej. W związku z tym program wybierze Miasto.csv jako tabelę główną. W kolejnych krokach tabela główna będzie modyfikowana, żeby otrzymać wynik łączenia.

Następne kroki program wykonuje w zależności od tego, jaki mamy typ łączenia. Możemy w związku z tym opisać 3 przypadki:

1. Tabela_główna Tabela_poboczna left/Tabela_poboczna Tabela_główna right
2. Tabela_główna Tabela_poboczna inner/Tabela_poboczna Tabela_główna inner
3. Tabela_główna Tabela_poboczna right/Tabela_poboczna Tabela_główna left

Pierwszy z tych przypadków jest najprostsz, a trzeci najbardziej skomplikowany.

Pierwszy przypadek:

1. Program dodaje do tabeli głównej puste kolumny — ilość dodanych kolumn jest równa ilości kolumn w tabeli pobocznej minus 1.

	Panstwo	Miasto	Zaludnienie(mln)
0	Hiszpania	Barcelona	0.0
1	Hiszpania	Matryt	0.0
2	Grecja	Ateny	0.0
3	Irlandia	Dublin	0.0
4	Litwa	Wilno	0.0
5	Polska	Kielce	0.0
6	Polska	Warszawa	0.0

2. Program tworzy 4 tablice Numpy:

a) pierwsza zawiera wartości w kolumnie łączącej w tabeli głównej (jest to tablica jednowymiarowa)

b) druga zawiera wartości w kolumnie łączącej w tabeli pobocznej (jest to tablica jednowymiarowa)

c) trzecia zawiera wartości z tabeli pobocznej, nie licząc wartości z kolumny łączącej

d) czwarta jest pustą tablicą, której rozmiar jest taki sam jak rozmiar, który tworzą puste kolumny dodane wcześniej do tabeli głównej (ponieważ wartości, które zostaną dodane do tej tablicy, następnie zostaną przeniesione do tych pustych kolumn)

```
['Hiszpania', 'Hiszpania', 'Grecja', 'Irlandia', 'Litwa', 'Polska', 'Polska']  
['Litwa', 'Francja', 'Hiszpania', 'Polska', 'Rumunia']  
[[2.8, 67.5, 47.1, 38.2, 21.2]]  
[['', '', '', '', '', '', '']]
```

3. Następnie w pętli program sprawdza, czy wartość z pierwszej tablicy jest taka sama jak wartość w drugiej. Jeśli tak, to wartości z trzeciej tablicy (które znajdują się w tym samym wierszu, co wartości z drugiej) dodawane są do czwartej tablicy.

```
[[47.1, 47.1, '', '', 2.8, 38.2, 38.2]]
```

4. Na końcu wartości z czwartej tablicy przenoszone są do pustych kolumn tabeli głównej.

```
Podaj komende: join zaludnienie.csv miasto.csv Panstwo right

    Panstwo    Miasto Zaludnienie(mln)
0  Hiszpania  Barcelona         47.1
1  Hiszpania   Matryt         47.1
2    Grecja    Ateny
3  Irlandia   Dublin
4    Litwa    Wilno         2.8
5    Polska   Kielce        38.2
6    Polska  Warszawa        38.2

naciśnij enter żeby zakończyć działanie programu
```

Drugi przypadek:

1. Program usuwa wiersze w tabeli głównej, jeśli wartości w kolumnie łączącej nie ma w tabeli pobocznej.

	Panstwo	Miasto
0	Hiszpania	Barcelona
1	Hiszpania	Matryt
2	Litwa	Wilno
3	Polska	Kielce
4	Polska	Warszawa

2. Program dodaje do tabeli głównej puste kolumny — ilość dodanych kolumn jest równa ilości kolumn w tabeli pobocznej minus 1.

	Panstwo	Miasto	Zaludnienie(mln)
0	Hiszpania	Barcelona	0.0
1	Hiszpania	Matryt	0.0
2	Litwa	Wilno	0.0
3	Polska	Kielce	0.0
4	Polska	Warszawa	0.0

3. Program tworzy 4 tablice Numpy:

a) pierwsza zawiera wartości w kolumnie łączącej w tabeli głównej (jest to tablica jednowymiarowa)

b) druga zawiera wartości w kolumnie łączącej w tabeli pobocznej (jest to tablica jednowymiarowa)

c) trzecia zawiera wartości z tabeli pobocznej, nie licząc wartości z kolumny łączącej

d) czwarta jest pustą tablicą, której rozmiar jest taki sam jak rozmiar, który tworzą puste kolumny dodane wcześniej do tabeli głównej (ponieważ wartości, które zostaną dodane do tej tablicy, następnie zostaną przeniesione do tych pustych kolumn)

```
['Hiszpania', 'Hiszpania', 'Litwa', 'Polska', 'Polska']  
['Litwa', 'Francja', 'Hiszpania', 'Polska', 'Rumunia']  
[[2.8, 67.5, 47.1, 38.2, 21.2]]  
[['', '', '', '', '']]
```

4. Następnie w pętli program sprawdza, czy wartość z pierwszej tablicy jest taka sama jak wartość w drugiej. Jeśli tak, to wartości z trzeciej tablicy (które znajdują się w tym samym wierszu, co wartości z drugiej) dodawane są do czwartej tablicy.

```
[[47.1, 47.1, 2.8, 38.2, 38.2]]
```

5. Na końcu wartości z czwartej tablicy przenoszone są do pustych kolumn tabeli głównej.

```
Podaj komende: join zaludnienie.csv miasto.csv Panstwo inner

    Panstwo    Miasto    Zaludnienie(mln)
0  Hiszpania  Barcelona    47.1
1  Hiszpania    Matryt    47.1
2      Litwa    Wilno     2.8
3   Polska    Kielce    38.2
4   Polska  Warszawa    38.2

naciśnij enter żeby zakończyć działanie programu
```

Trzeci przypadek

1. Program usuwa wiersze w tabeli głównej, jeśli wartości w kolumnie łączącej nie ma w tabeli pobocznej.

	Panstwo	Miasto
0	Hiszpania	Barcelona
1	Hiszpania	Matryt
4	Litwa	Wilno
5	Polska	Kielce
6	Polska	Warszawa

2. Program dodaje do tabeli głównej wiersze z wartościami, które znajdują się w tabeli pobocznej w kolumnie łączącej, ale nie ma ich tabeli głównej.

	Panstwo	Miasto
0	Hiszpania	Barcelona
1	Hiszpania	Matryt
2	Litwa	Wilno
3	Polska	Kielce
4	Polska	Warszawa
5	Francja	
6	Rumunia	

3. Program dodaje do tabeli głównej puste kolumny — ilość dodanych kolumn jest równa ilości kolumn w tabeli pobocznej minus 1.

	Panstwo	Miasto	Zaludnienie(mln)
0	Hiszpania	Barcelona	0.0
1	Hiszpania	Matryt	0.0
2	Litwa	Wilno	0.0
3	Polska	Kielce	0.0
4	Polska	Warszawa	0.0
5	Francja		0.0
6	Rumunia		0.0

4. Program tworzy 4 tablice Numpy:

a) pierwsza zawiera wartości w kolumnie łączącej w tabeli głównej (jest to tablica jednowymiarowa)

b) druga zawiera wartości w kolumnie łączącej w tabeli pobocznej (jest to tablica jednowymiarowa)

c) trzecia zawiera wartości z tabeli pobocznej, nie licząc wartości z kolumny łączącej

d) czwarta jest pustą tablicą, której rozmiar jest taki sam jak rozmiar, który tworzą puste kolumny dodane wcześniej do tabeli głównej (ponieważ wartości, które zostaną dodane do tej tablicy, następnie zostaną przeniesione do tych pustych kolumn)

```
['Hiszpania', 'Hiszpania', 'Litwa', 'Polska', 'Polska', 'Francja', 'Rumunia']  
['Litwa', 'Francja', 'Hiszpania', 'Polska', 'Rumunia']  
[[2.8, 67.5, 47.1, 38.2, 21.2]]  
[['', '', '', '', '', '', '']]
```

5. Następnie w pętli program sprawdza, czy wartość z pierwszej tablicy jest taka sama jak wartość w drugiej. Jeśli tak, to wartości z trzeciej tablicy (które znajdują się w tym samym wierszu, co wartości z drugiej) dodawane są do czwartej tablicy.

```
[[47.1, 47.1, 2.8, 38.2, 38.2, 67.5, 21.2]]
```

6. Na końcu wartości z czwartej tablicy przenoszone są do pustych kolumn tabeli głównej.

```
Podaj komende: join zaludnienie.csv miasto.csv Panstwo left  
  
   Panstwo   Miasto  Zaludnienie(mln)  
0  Hiszpania Barcelona    47.1  
1  Hiszpania  Matryt     47.1  
2    Litwa    Wilno      2.8  
3   Polska   Kielce     38.2  
4   Polska   Warszawa    38.2  
5   Francja              67.5  
6   Rumunia              21.2  
  
naciśnij enter żeby zakończyć działanie programu
```