

# PANDAS ZADANIA (BEZ WYKRESU)

1. WYŚWIETL tylko te rekordy gdzie liczba nadanych imion była większa niż 1000 w danym roku

```
df=pd.read_excel('imiona.xlsx')
print(df)
liczbawiekszaod1000=df[df.Liczba>1000]
print(liczbawiekszaod1000)
```

2. Wyświetl tylko te rekordy gdzie imię jest takie jak twoje

```
imiejakja=df[df.Imie=='PIOTR']
print(imiejakja)
```

3. Suma ludzi urodzonych w całym okresie

```
suma=df.Liczba.sum()
print(suma)
```

4. Suma dzieci urodzonych w latach 2000-2005

```
sumawlatach20002005=df[df.Rok<2006].groupby('Rok').agg({'Liczba':['sum']})
print(sumawlatach20002005)
```

5. suma urodzonych chłopów i bab

```
chlopaki=df[df.Plec=='M'].agg({'Liczba':['sum']})
print(chlopaki)
babeczki=df[df.Plec=='K'].agg({'Liczba':['sum']})
print(babeczki)
```

6. najbardziej popularne imię dla dziecka m/k po 2 rekordy na rok

```
print(df.sort_values('Liczba', ascending=False).groupby(['Rok', 'Plec']).nth(0))
```

7. najpopularniejsze imię chłop/baba

```
print(df[df['Plec'] == 'M'].groupby(['Imie']).agg({'Liczba': ['sum']}).sort_values(('Liczba', 'sum'), ascending=False).head(1))
print(df[df['Plec'] == 'K'].groupby(['Imie']).agg({'Liczba': ['sum']}).sort_values(('Liczba', 'sum'), ascending=False).head(1))
```

8. unikalni sprzedawcy

```
unikalne=df.Sprzedawca.unique()
print(unikalne)
```

9. 5 najwyższych utargów

```
print(df.sort_values('Utag', ascending=False).head(5))
```

10. ilość zamówień złożonych przez każdego sprzedawcę

```
print(df.groupby('Sprzedawca').size())
```

11. suma zamówień dla każdego kraju

```
print(df.groupby('Kraj').size())
```

12.suma zamowien dla roku 2005 dla sprzedawcow z polski

```
print(df[(df['Kraj']=='Polska')&(df['Data zamowienia']>='2005-01-01')&(df['Data zamowienia']<='2005-12-31')].agg({'Uzartg':['sum']}))
```

13.srednia kwota zamówienia w 2004 roku

```
print(df[(df['Data zamowienia']>='2004-01-01')&(df['Data zamowienia']<='2004-12-31')].agg({'Uzartg':['mean']}))
```

14.zapisz dane za 2004 rok do pliku zamówienia\_2004.csv a dane za 2005 do pliku zamówienia\_2005.csv

```
dane2004=df[(df['Data zamowienia']>='2004-01-01')&(df['Data zamowienia']<='2004-12-31')]  
print(dane2004)  
dane2004.to_csv('zamowienia_2004.csv',index=False)  
dane2005=df[(df['Data zamowienia']>='2005-01-01')&(df['Data zamowienia']<='2005-12-31')]  
dane2005.to_csv('zamowienia_2005.csv',index=False)
```

Przydatne komendy pandasa

```
# Usuwanie kolumny  
df.drop('Kolumna1', axis=1, inplace=True)  
  
# Zmiana nazwy kolumny  
df.rename(columns={'Kolumna2': 'NowaNazwa'}, inplace=True)  
  
# Grupowanie danych i obliczanie statystyk grupowych  
df.groupby('Kolumna2').mean()
```

```
# Zapisywanie DataFrame do pliku CSV  
df.to_csv('nazwa_pliku.csv', index=False)  
  
# Odczytywanie danych z pliku Excel  
df = pd.read_excel('nazwa_pliku.xlsx')  
  
# Zapisywanie DataFrame do pliku Excel  
df.to_excel('nazwa_pliku.xlsx', index=False)
```