

```
In [109... import datetime as dt
import pandas as pd

import os
```

```
In [110... file_path1 = "../data/1._przecietne_miesieczne_wynagrodzenia_brutto_w_gos
file_path2 = "../data/3._przecietne_zatrudnienie_ogolem_i_w_sektorze_prze
os.path.isfile(file_path1), os.path.isfile(file_path2)
```

```
Out[110... (True, True)
```

```
In [111... df_wynag = pd.read_csv(file_path1, header=0, encoding='latin2', delimiter
df_wynag.head(5)
```

```
Out[111...
```

	Zmienna	Jednostka terytorialna	Jednostka	Rok	Kwartał	Wartość
0	Przeciętne miesięczne wynagrodzenia brutto w g...	Polska	zł	2010	1	3241,88
1	Przeciętne miesięczne wynagrodzenia brutto w g...	Polska	zł	2010	2	3270,87
2	Przeciętne miesięczne wynagrodzenia brutto w g...	Polska	zł	2010	3	3258,95
3	Przeciętne miesięczne wynagrodzenia brutto w g...	Polska	zł	2010	4	3371,07
4	Przeciętne miesięczne wynagrodzenia brutto w g...	Polska	zł	2011	1	3392

Możemy użyć metody `stepwise_selection` z pakietu *dmba*.

```
In [112... df_wynag["Zmienna"].unique()
```

```
Out[112... array(['Przeciętne miesięczne wynagrodzenia brutto w gospodarce narodowe
j',
      'Przeciętne miesięczne wynagrodzenia brutto w sektorze przedsiębi
orstw'],
      dtype=object)
```

```
In [113... for ind, gr in df_wynag.groupby(["Rok", "Kwartał"]):
    print(ind, gr)
    break
```

```
(2010, 1)
terytorialna \
0   Przeciętne miesięczne wynagrodzenia brutto w g...
ka
59  Przeciętne miesięczne wynagrodzenia brutto w s...
ka
```

	Jednostka	Rok	Kwartał	Wartość
0	zł	2010	1	3241,88
59	zł	2010	1	3382,45

```
In [114... df_zatrud = pd.read_csv(file_path2, header=0, encoding='latin2', delimit
df_zatrud.head(5)
```

Out[114...

	Zmienna	Jednostka terytorialna	Jednostka	Rok	Kwartał	Wartość
0	Przeciętne zatrudnienie ogółem	Polska	tysiŃce	2010	1	8034
1	Przeciętne zatrudnienie ogółem	Polska	tysiŃce	2010	2	8110
2	Przeciętne zatrudnienie ogółem	Polska	tysiŃce	2010	3	8176
3	Przeciętne zatrudnienie ogółem	Polska	tysiŃce	2010	4	8214
4	Przeciętne zatrudnienie ogółem	Polska	tysiŃce	2011	1	8264

```
In [115... for ind, gr in df_zatrud.groupby(["Rok", "Kwartał"]):
              print(ind, gr)
              break
```

```
(2010, 1)                                Zmienna Jednostka
terytorialna \
0          Przeciętne zatrudnienie ogółem          Pols
ka
58 Przeciętne zatrudnienie w sektorze przedsiębiorst...  Pols
ka

      Jednostka  Rok  Kwartał  Wartość
0      tysiŃce  2010        1      8034
58      tysiŃce  2010        1      5304
```

```
In [116... df = pd.merge(left=df_wynag, right=df_zatrud, on=["Rok", "Kwartał", "Jed
df.drop(['Jednostka terytorialna', 'Jednostka_x', 'Jednostka_y'], axis=1,
df.columns, df.head())
```

```
Out[116... (Index(['Zmienna_x', 'Rok', 'Kwartał', 'Wartość_x', 'Zmienna_y', 'Wartoć
y'], dtype='object'),

              Zmienna_x  Rok  Kwartał  War
toć_x \
0  Przeciętne miesięczne wynagrodzenia brutto w g...  2010        1  3
241,88
1  Przeciętne miesięczne wynagrodzenia brutto w g...  2010        1  3
241,88
2  Przeciętne miesięczne wynagrodzenia brutto w g...  2010        2  3
270,87
3  Przeciętne miesięczne wynagrodzenia brutto w g...  2010        2  3
270,87
4  Przeciętne miesięczne wynagrodzenia brutto w g...  2010        3  3
258,95

              Zmienna_y  Wartość_y
0          Przeciętne zatrudnienie ogółem      8034
1  Przeciętne zatrudnienie w sektorze przedsiębiorst...      5304
2          Przeciętne zatrudnienie ogółem      8110
3  Przeciętne zatrudnienie w sektorze przedsiębiorst...      5351
4          Przeciętne zatrudnienie ogółem      8176 )
```

```
In [117... df.columns
```

```
Out[117...] Index(['Zmienna_x', 'Rok', 'Kwartał', 'Wartość_x', 'Zmienna_y', 'Wartość_y'], dtype='object')
```

```
In [118...] df["Zmienna_x"].unique(), df["Zmienna_y"].unique())
```

```
Out[118...] (array(['Przeciętne miesięczne wynagrodzenia brutto w gospodarce narodowej',  
        'Przeciętne miesięczne wynagrodzenia brutto w sektorze przedsiębiorstw'],  
        dtype=object),  
        array(['Przeciętne zatrudnienie ogółem',  
        'Przeciętne zatrudnienie w sektorze przedsiębiorstw'], dtype=object))
```

```
In [119...] _dct = {  
    'Przeciętne miesięczne wynagrodzenia brutto w gospodarce narodowej':  
    'Przeciętne miesięczne wynagrodzenia brutto w sektorze przedsiębiorstw',  
    'Przeciętne zatrudnienie ogółem': "zatrud_ogolem",  
    'Przeciętne zatrudnienie w sektorze przedsiębiorstw': "zatrud_przed"  
}  
_dct
```

```
Out[119...] {'Przeciętne miesięczne wynagrodzenia brutto w gospodarce narodowej': 'wynag_gosp',  
            'Przeciętne miesięczne wynagrodzenia brutto w sektorze przedsiębiorstw': 'wynag_przed',  
            'Przeciętne zatrudnienie ogółem': 'zatrud_ogolem',  
            'Przeciętne zatrudnienie w sektorze przedsiębiorstw': 'zatrud_przed'}
```

```
In [120...] fds = {}  
for ind, gr in df.groupby(["Rok", "Kwartał"]):  
    date = dt.datetime(year=ind[0], month=1, day=1) + pd.offsets.QuarterEnd(1)  
    d = {v: None for v in _dct.values()}  
    for row in gr.itertuples():  
        if d[_dct[row[1]]] is None:  
            d[_dct[row[1]]] = row[4]  
        if d[_dct[row[5]]] is None:  
            d[_dct[row[5]]] = row[6]  
    fds[date] = d
```

```
In [121...] ndf = pd.DataFrame(data=fds).T  
ndf.head()
```

```
Out[121...]
```

	wynag_gosp	wynag_przed	zatrud_ogolem	zatrud_przed
2010-03-31	3241,88	3382,45	8034	5304
2010-06-30	3270,87	3420,32	8110	5351
2010-09-30	3258,95	3424,61	8176	5392
2010-12-31	3371,07	3499,96	8214	5449
2011-03-31	3392	3521,75	8264	5512

```
In [122...] ndf.shape, ndf.index, ndf.columns
```

```
Out[122...] ((59, 4),
             DatetimeIndex(['2010-03-31', '2010-06-30', '2010-09-30', '2010-12-31',
                             '2011-03-31', '2011-06-30', '2011-09-30', '2011-12-31',
                             '2012-03-31', '2012-06-30', '2012-09-30', '2012-12-31',
                             '2013-03-31', '2013-06-30', '2013-09-30', '2013-12-31',
                             '2014-03-31', '2014-06-30', '2014-09-30', '2014-12-31',
                             '2015-03-31', '2015-06-30', '2015-09-30', '2015-12-31',
                             '2016-03-31', '2016-06-30', '2016-09-30', '2016-12-31',
                             '2017-03-31', '2017-06-30', '2017-09-30', '2017-12-31',
                             '2018-03-31', '2018-06-30', '2018-09-30', '2018-12-31',
                             '2019-03-31', '2019-06-30', '2019-09-30', '2019-12-31',
                             '2020-03-31', '2020-06-30', '2020-09-30', '2020-12-31',
                             '2021-03-31', '2021-06-30', '2021-09-30', '2021-12-31',
                             '2022-03-31', '2022-06-30', '2022-09-30', '2022-12-31',
                             '2023-03-31', '2023-06-30', '2023-09-30', '2023-12-31',
                             '2024-03-31', '2024-06-30', '2024-09-30'],
                             dtype='datetime64[ns]', freq=None),
             Index(['wynag_gosp', 'wynag_przed', 'zatrud_ogolem', 'zatrud_przed'], dtype='object'))
```

```
In [123...] ndf["zatrud_gosp"] = ndf["zatrud_ogolem"] - ndf["zatrud_przed"]
ndf.head()
```

```
Out[123...]      wynag_gosp  wynag_przed  zatrud_ogolem  zatrud_przed  zatrud_gosp
2010-03-31      3241,88      3382,45           8034           5304           2730
2010-06-30      3270,87      3420,32           8110           5351           2759
2010-09-30      3258,95      3424,61           8176           5392           2784
2010-12-31      3371,07      3499,96           8214           5449           2765
2011-03-31       3392       3521,75           8264           5512           2752
```

```
In [124...] ndf.wynag_gosp = ndf.wynag_gosp.apply(lambda r: r.replace(",", ".")).astype(float)
ndf.wynag_przed = ndf.wynag_przed.apply(lambda r: r.replace(",", ".")).astype(float)
ndf.dtypes, ndf.head()
```

```
Out[124...]      wynag_gosp      float64
wynag_przed      float64
zatrud_ogolem     float64
zatrud_przed      float64
zatrud_gosp       float64
dtype: object,
      wynag_gosp  wynag_przed  zatrud_ogolem  zatrud_przed  zatrud_gosp
2010-03-31      3241.88      3382.45           8034.0           5304.0
2730.0
2010-06-30      3270.87      3420.32           8110.0           5351.0
2759.0
2010-09-30      3258.95      3424.61           8176.0           5392.0
2784.0
2010-12-31      3371.07      3499.96           8214.0           5449.0
2765.0
2011-03-31      3392.00      3521.75           8264.0           5512.0
2752.0)
```

```
In [125...] ndf["wynag"] = (ndf.wynag_gosp * ndf.zatrud_gosp + ndf.wynag_przed * ndf.
```

```
ndf.head()
```

	wynag_gosp	wynag_przed	zatrud_ogolem	zatrud_przed	zatrud_gosp	
2010-03-31	3241.88	3382.45	8034.0	5304.0	2730.0	333
2010-06-30	3270.87	3420.32	8110.0	5351.0	2759.0	336
2010-09-30	3258.95	3424.61	8176.0	5392.0	2784.0	336
2010-12-31	3371.07	3499.96	8214.0	5449.0	2765.0	345
2011-03-31	3392.00	3521.75	8264.0	5512.0	2752.0	347

```
ndf.tail()
```

	wynag_gosp	wynag_przed	zatrud_ogolem	zatrud_przed	zatrud_gosp	
2023-09-30	7351.07	7563.31	9400.0	6526.0	2874.0	749
2023-12-31	7566.93	7737.15	9407.0	6518.0	2889.0	768
2024-03-31	7848.93	8001.09	9404.0	6516.0	2888.0	795
2024-06-30	8107.04	8161.60	9405.0	6510.0	2895.0	814
2024-09-30	8349.35	8381.15	NaN	6516.0	NaN	

```
ndf = ndf.loc[~ndf["wynag"].isna()]
ndf.tail()
```

	wynag_gosp	wynag_przed	zatrud_ogolem	zatrud_przed	zatrud_gosp	
2023-06-30	7108.18	7353.18	9400.0	6535.0	2865.0	727
2023-09-30	7351.07	7563.31	9400.0	6526.0	2874.0	749
2023-12-31	7566.93	7737.15	9407.0	6518.0	2889.0	768
2024-03-31	7848.93	8001.09	9404.0	6516.0	2888.0	795
2024-06-30	8107.04	8161.60	9405.0	6510.0	2895.0	814

```
ndf.to_csv("../data/wynagrodzenie.csv")
```