

Dokumentacja modelu predykcyjnego

Wybrany model: XGBoost

Wyniki najlepszego modelu:

Model: XGBoost

MSE: 47.901353493023365

R²: 0.3683257711992005

MAE: 5.700117308379226

Pierwsze 5 wierszy danych:

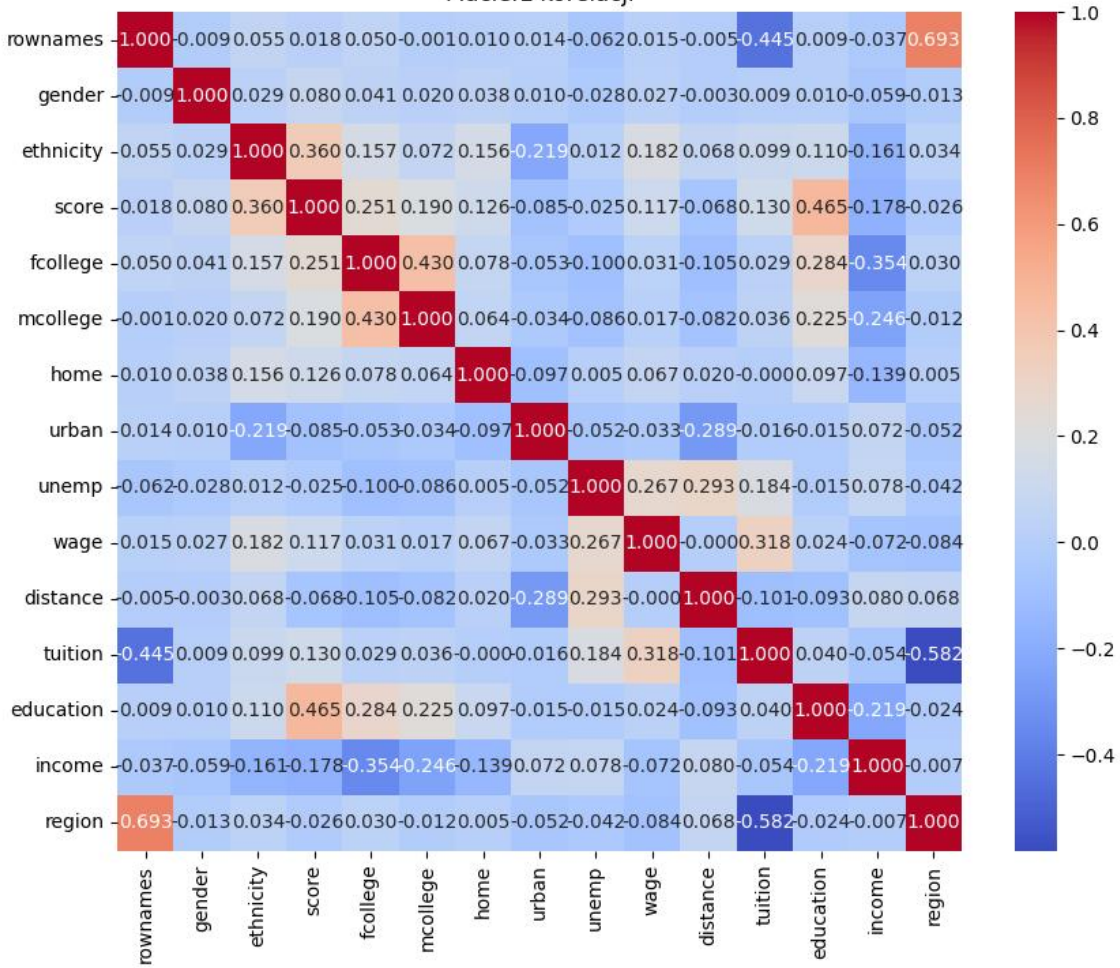
[1.	1.	2.	39.15000153	1.	0.	1.	1.	6.19999981	8.09000015	0.2	0.88915002	0.	0.
[2.	0.	2.	48.86999893	0.	0.	1.	1.	6.19999981	8.09000015	0.2	0.88915002	0.	1.
[3.	1.	2.	48.74000168	0.	0.	1.	1.	6.19999981	8.09000015	0.2	0.88915002	0.	1.
[4.	1.	0.	40.40000153	0.	0.	1.	1.	6.19999981	8.09000015	0.2	0.88915002	0.	1.
[5.	0.	2.	40.47999954	0.	0.	0.	1.	5.59999999	8.09000015	0.40000001	0.88915002	1.	

Statystyki:

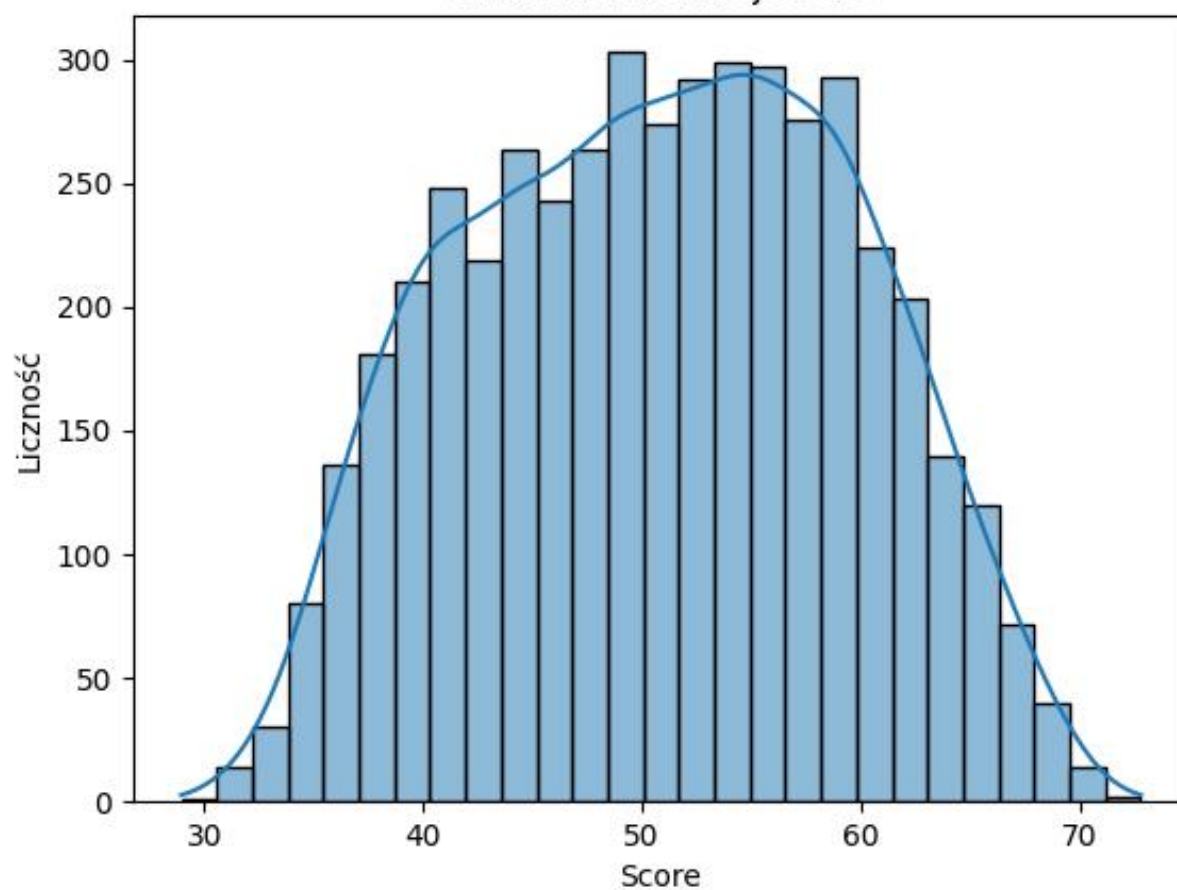
Statystyka	rownames	gender	ethnicity	score	fcollege	mcollege
count	4739.0	4739.0	4739.0	4739.0	4739.0	4739.0
mean	3954.639	0.451	1.478	50.889	0.208	0.137
std	5953.828	0.498	0.762	8.702	0.406	0.344
min	1.0	0.0	0.0	28.95	0.0	0.0
25%	1185.5	0.0	1.0	43.925	0.0	0.0
50%	2370.0	0.0	2.0	51.19	0.0	0.0
75%	3554.5	1.0	2.0	57.77	0.0	0.0
max	37810.0	1.0	2.0	72.81	1.0	1.0

Wykresy:

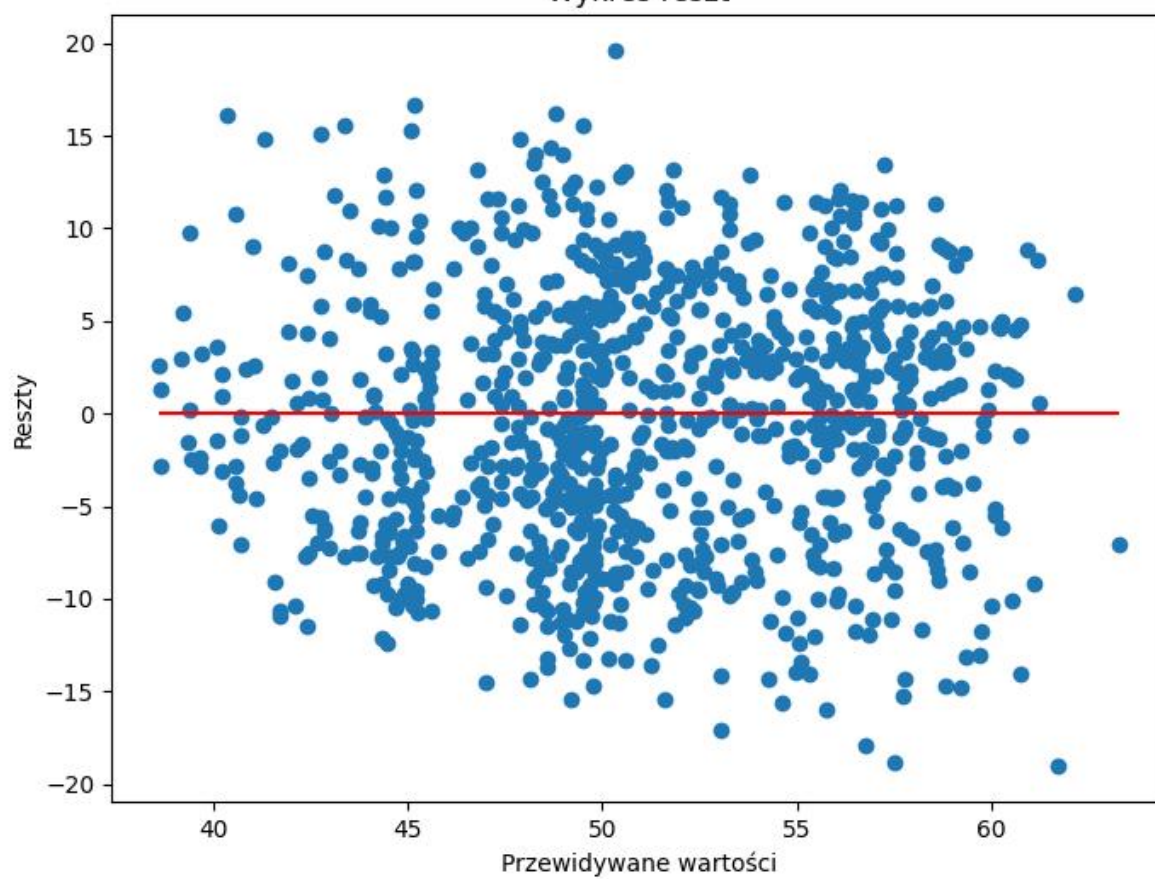
Macierz korelacji



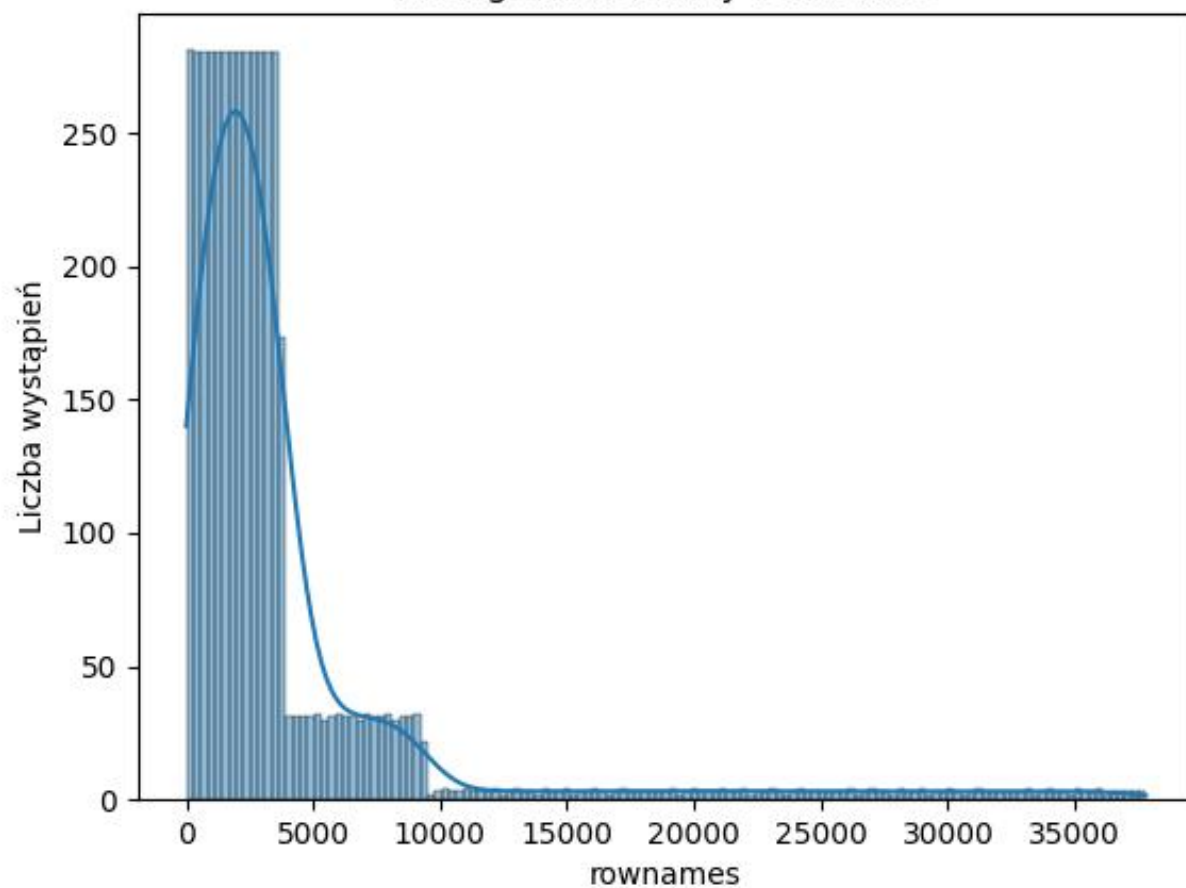
Rozkład zmiennej score



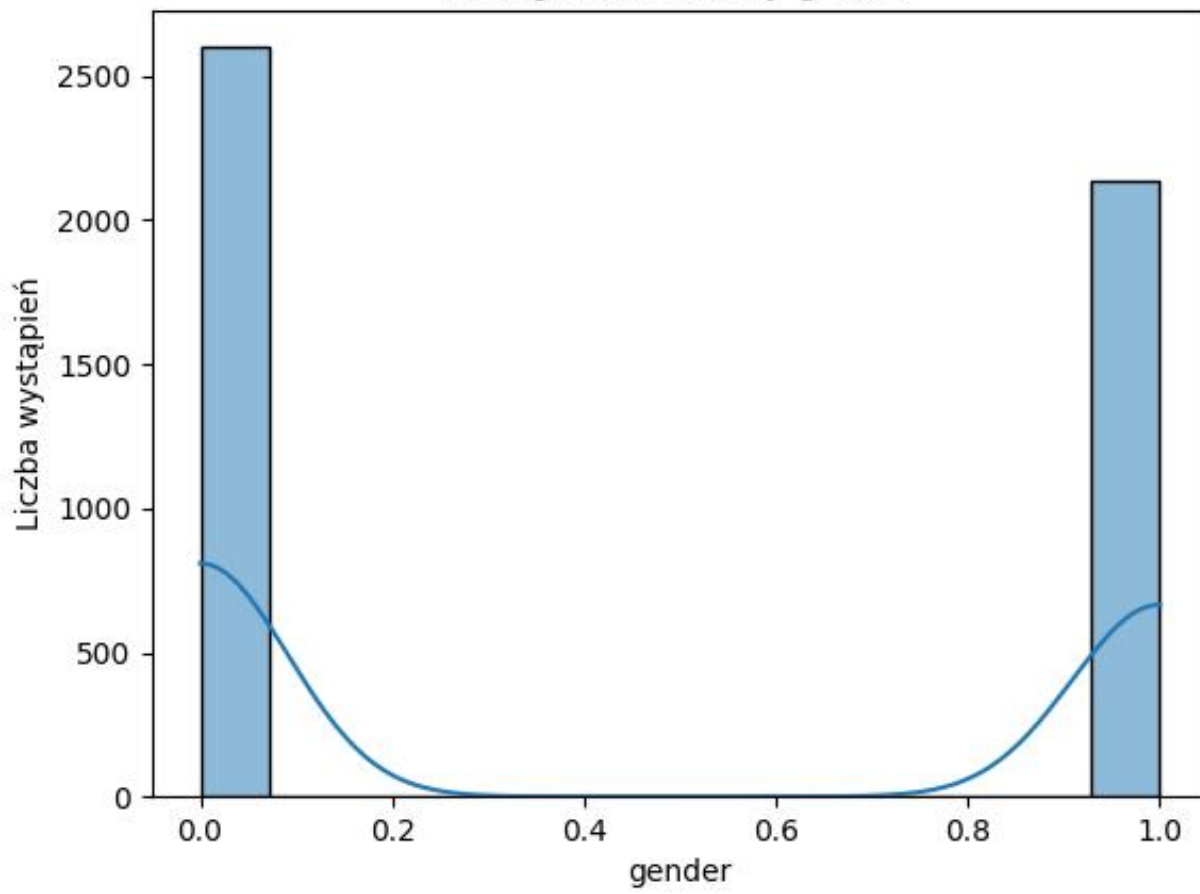
Wykres reszt



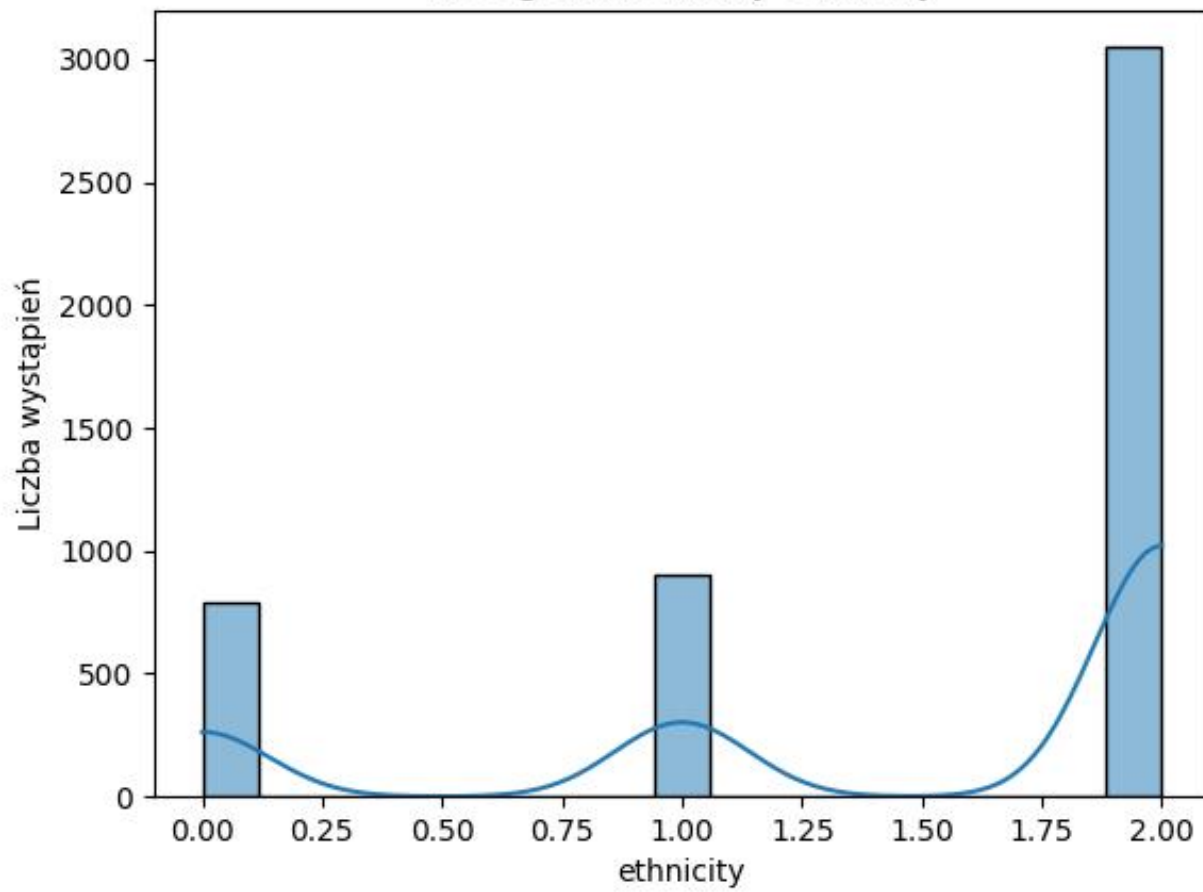
Histogram kolumny rownames



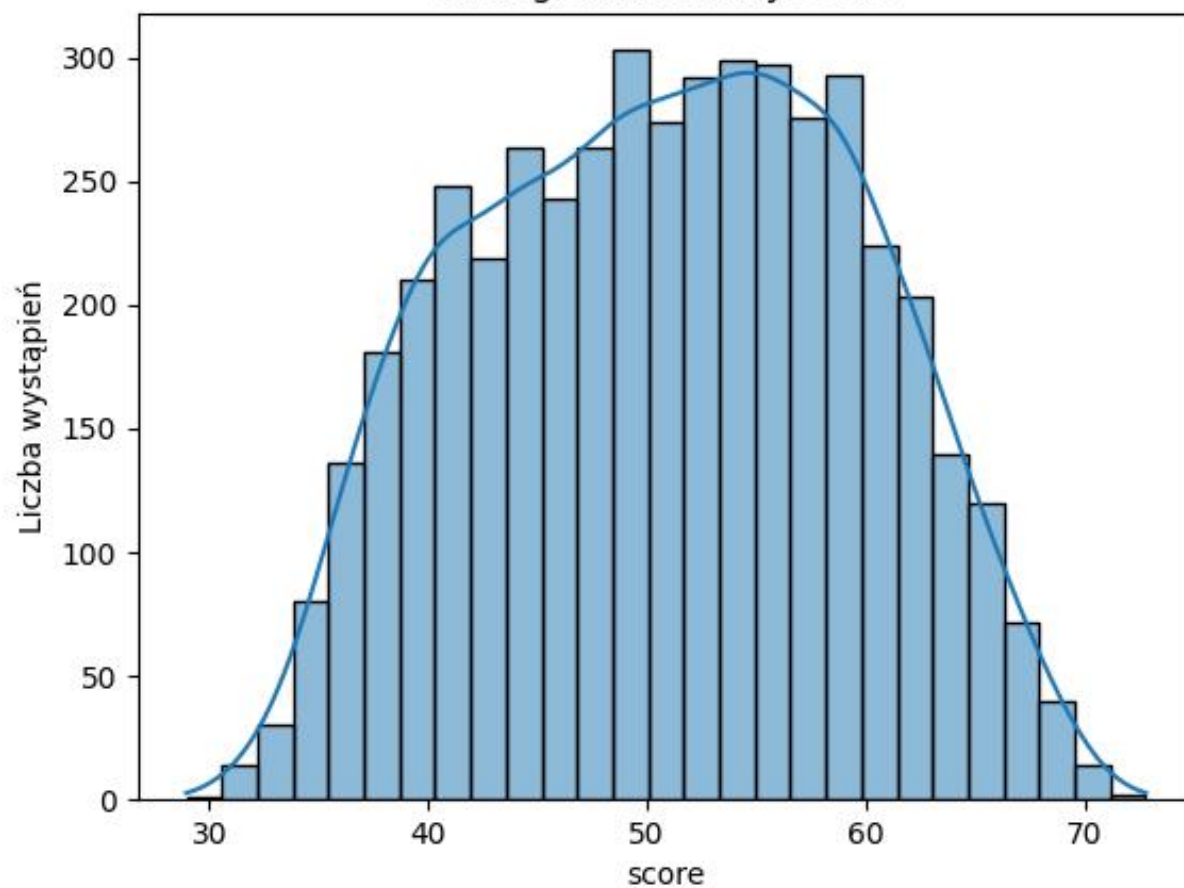
Histogram kolumny gender



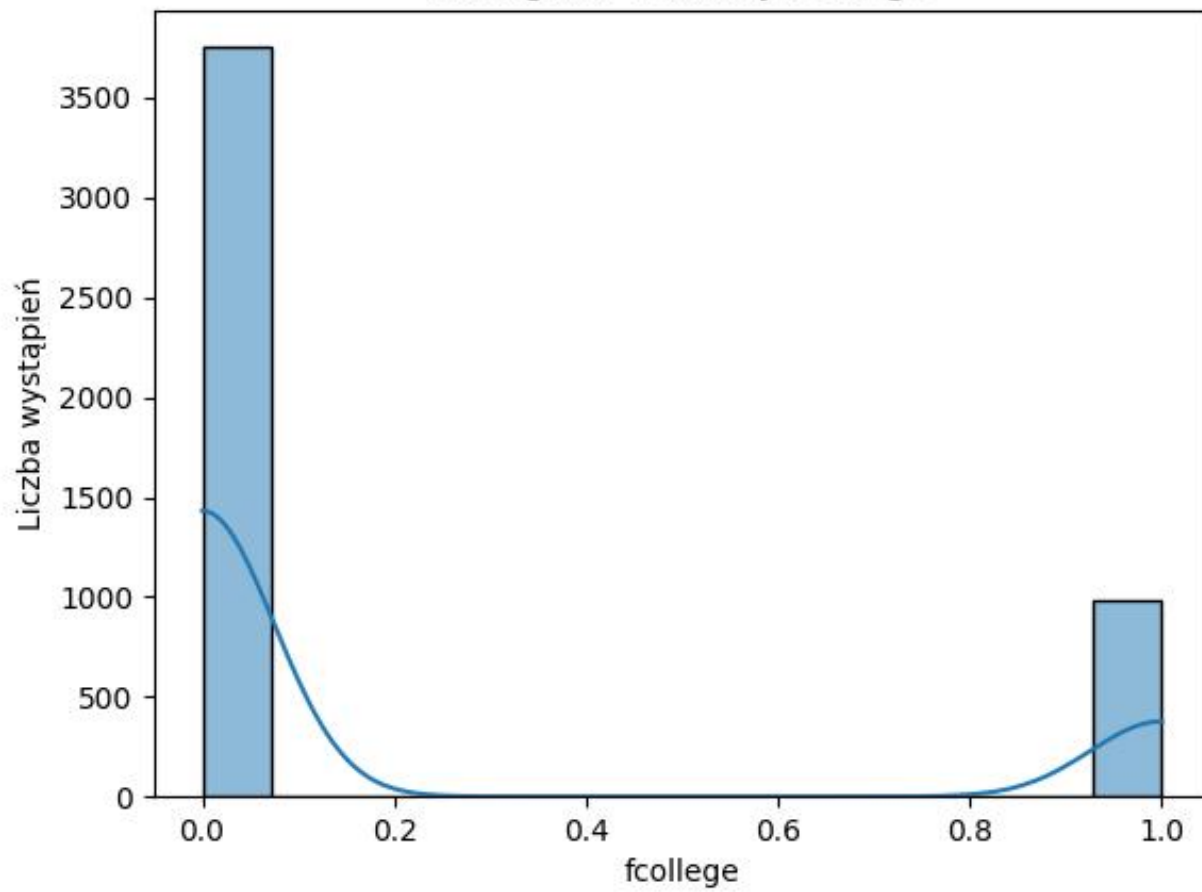
Histogram kolumny ethnicity



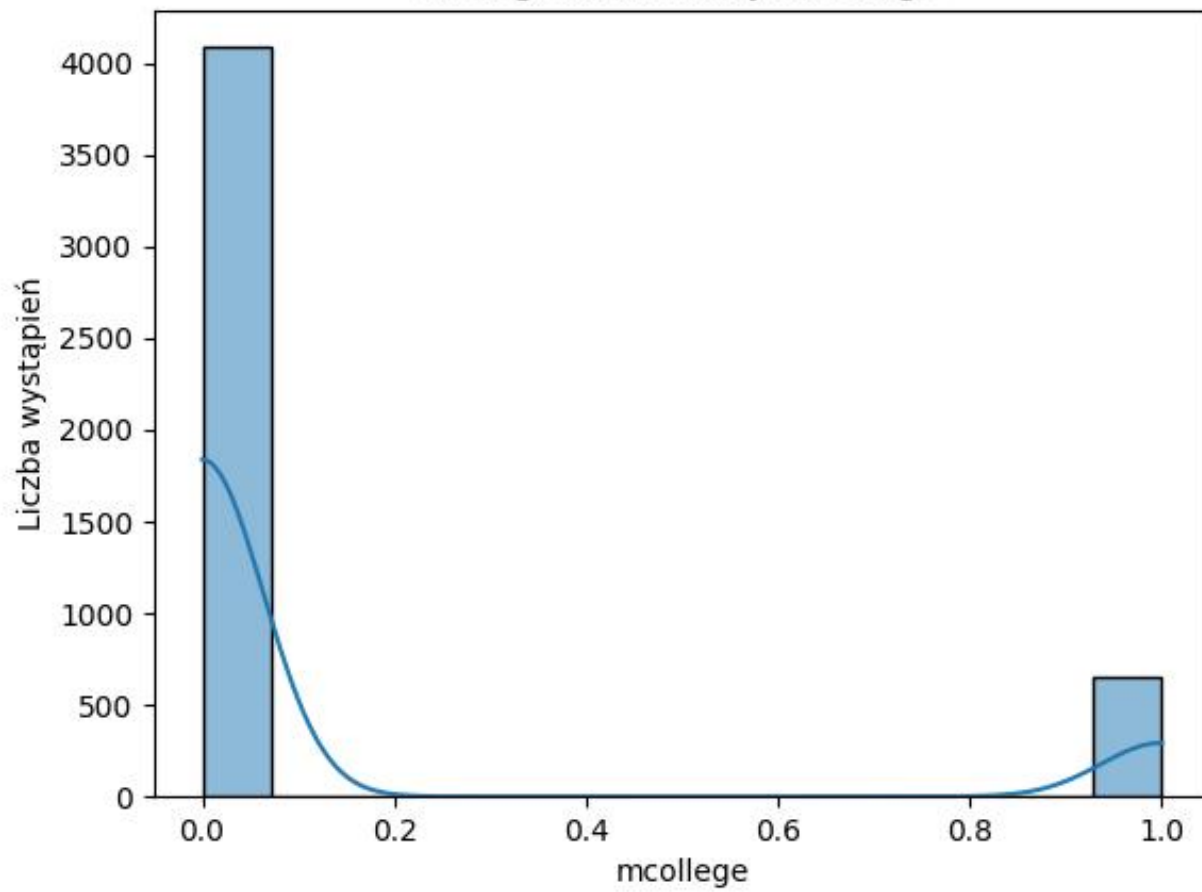
Histogram kolumny score



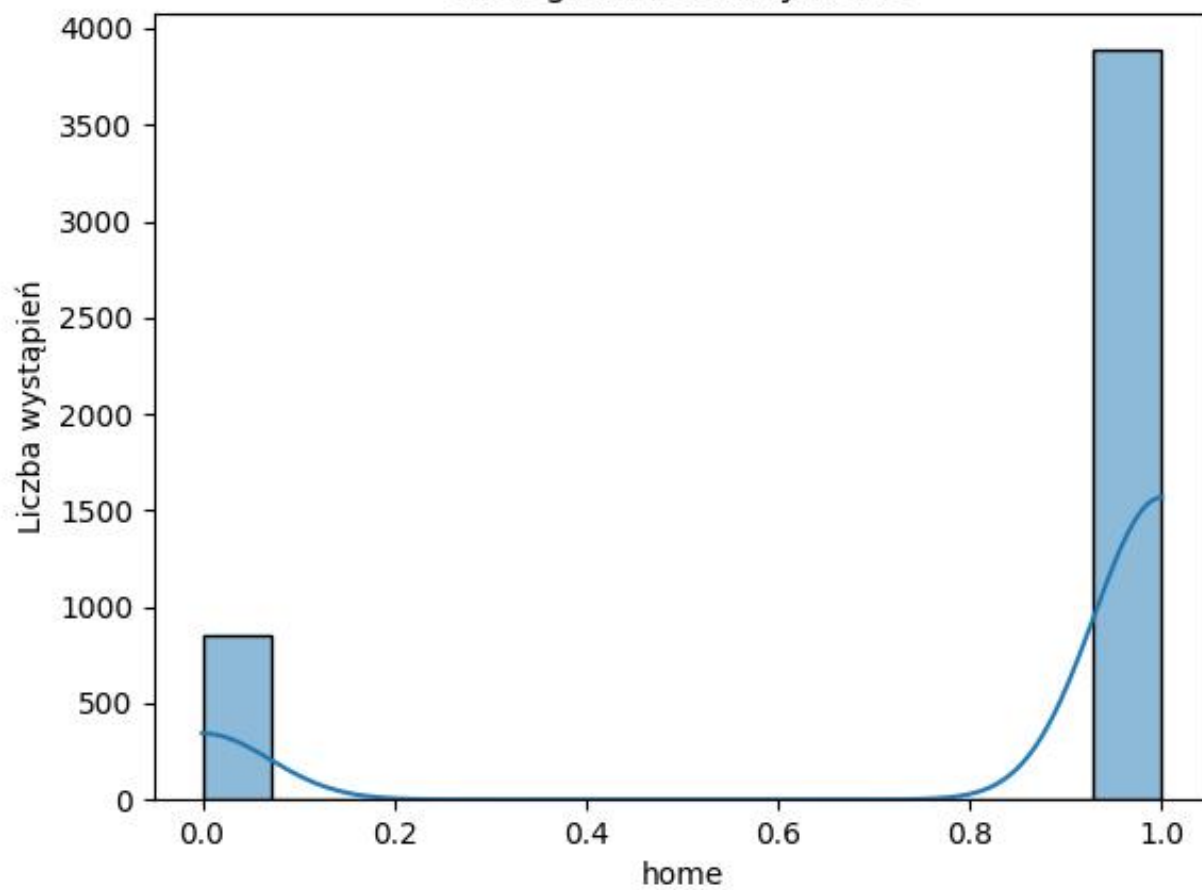
Histogram kolumny fcollege



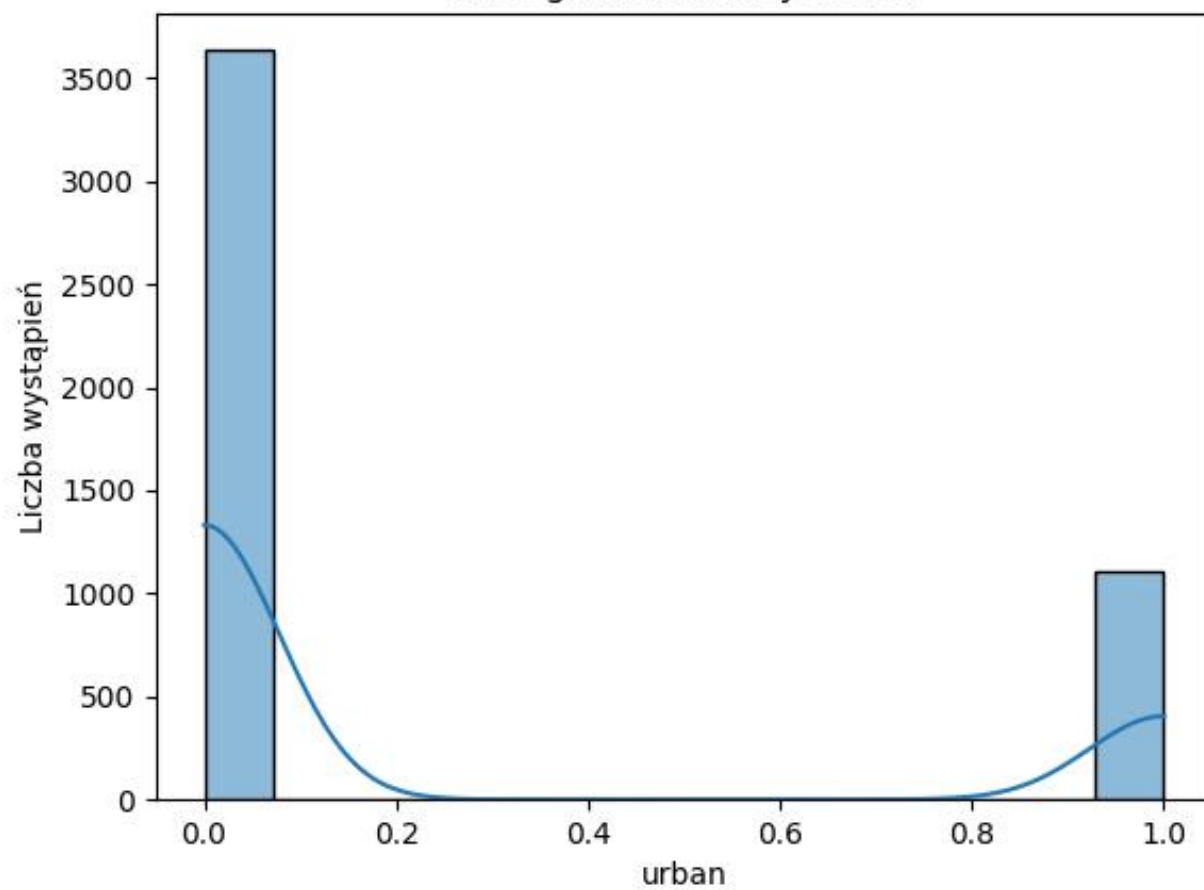
Histogram kolumny mcollege



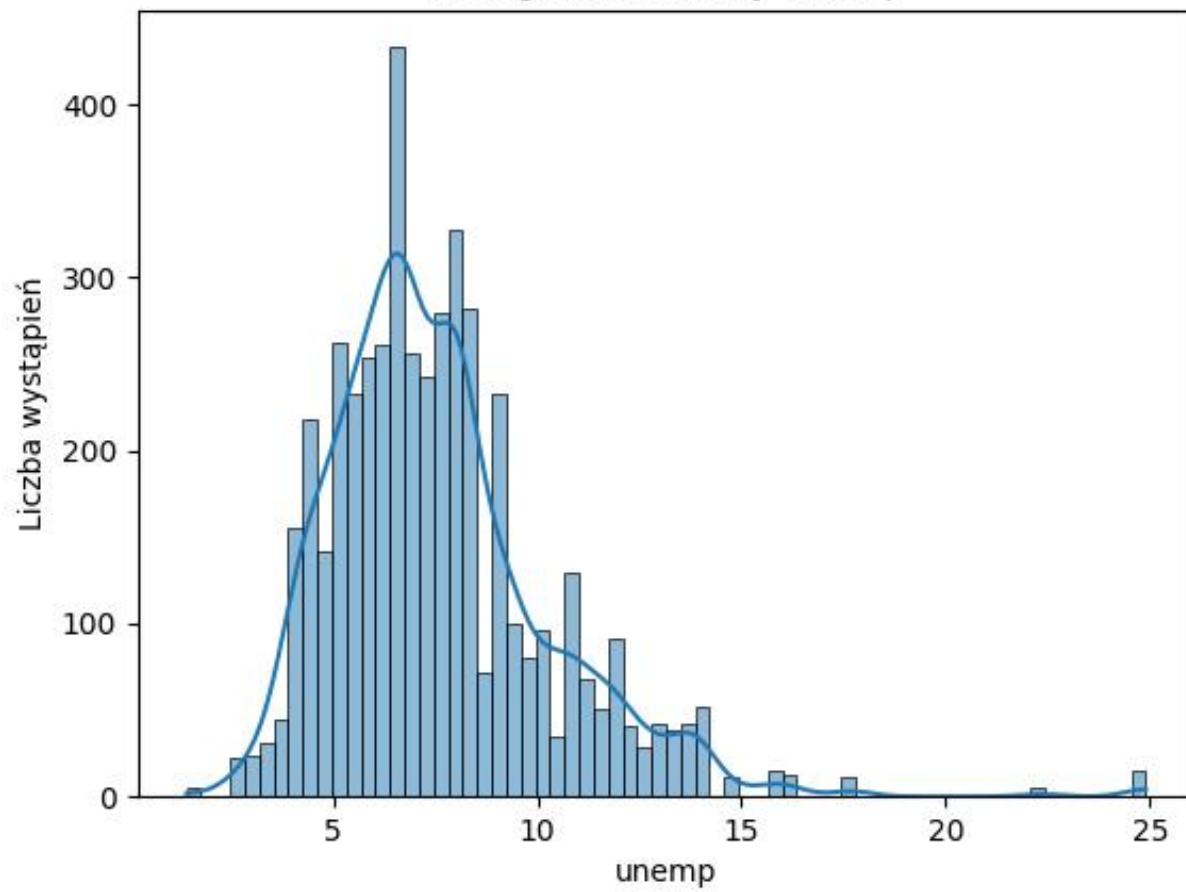
Histogram kolumny home



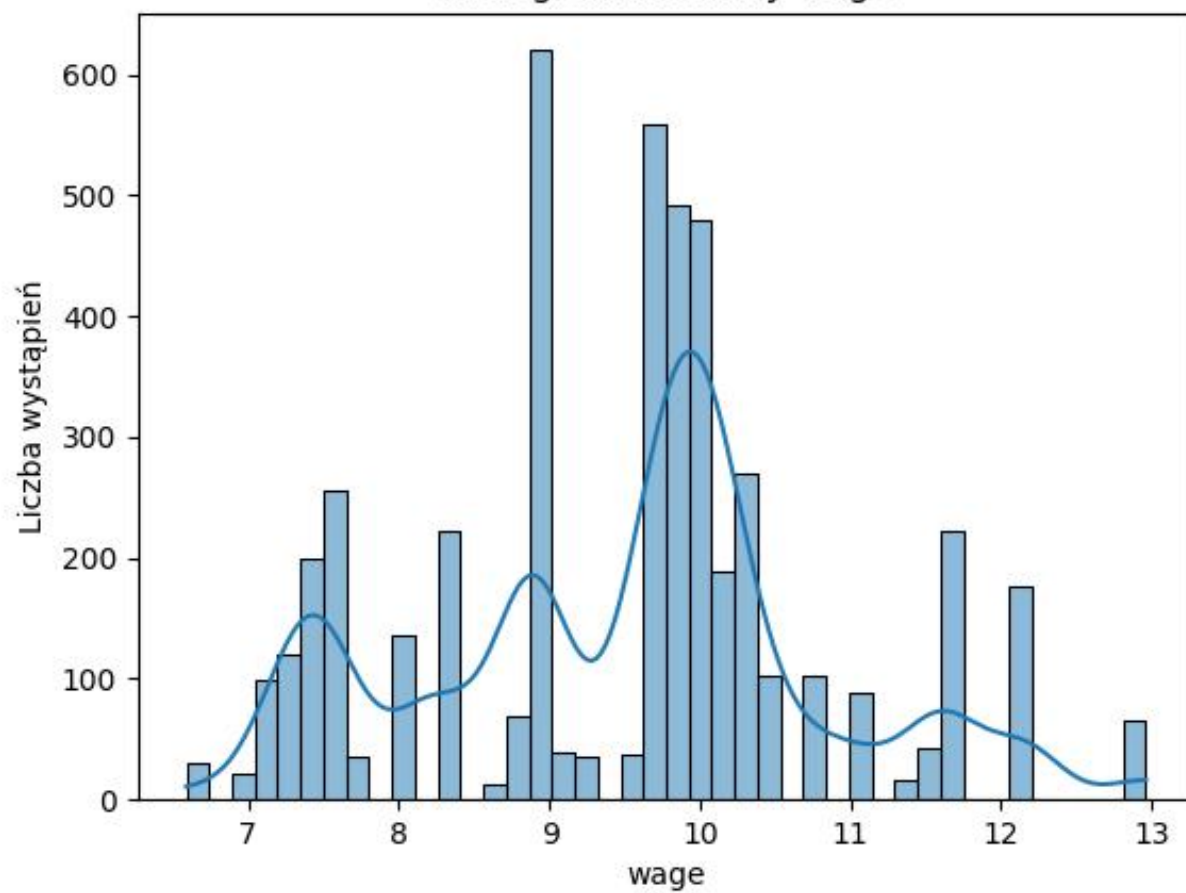
Histogram kolumnny urban



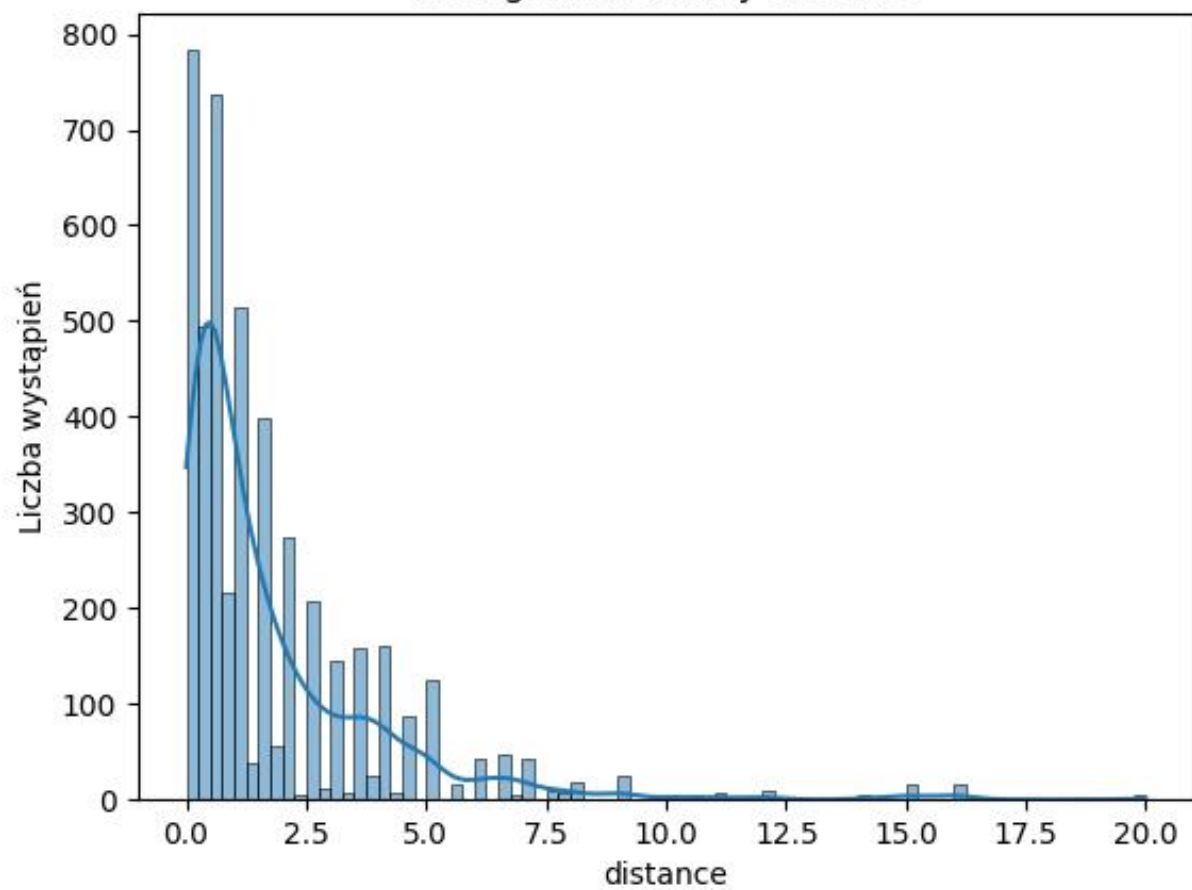
Histogram kolumny unemp



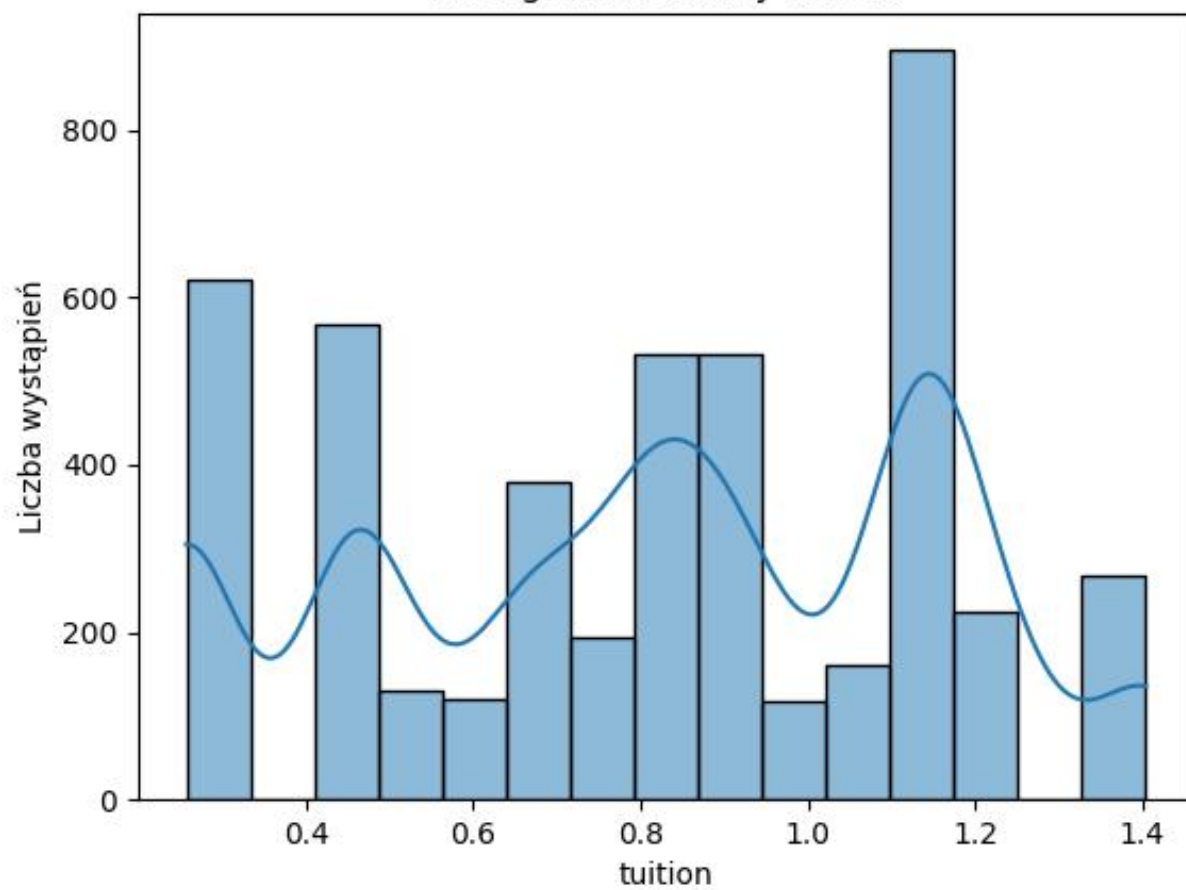
Histogram kolumny wage



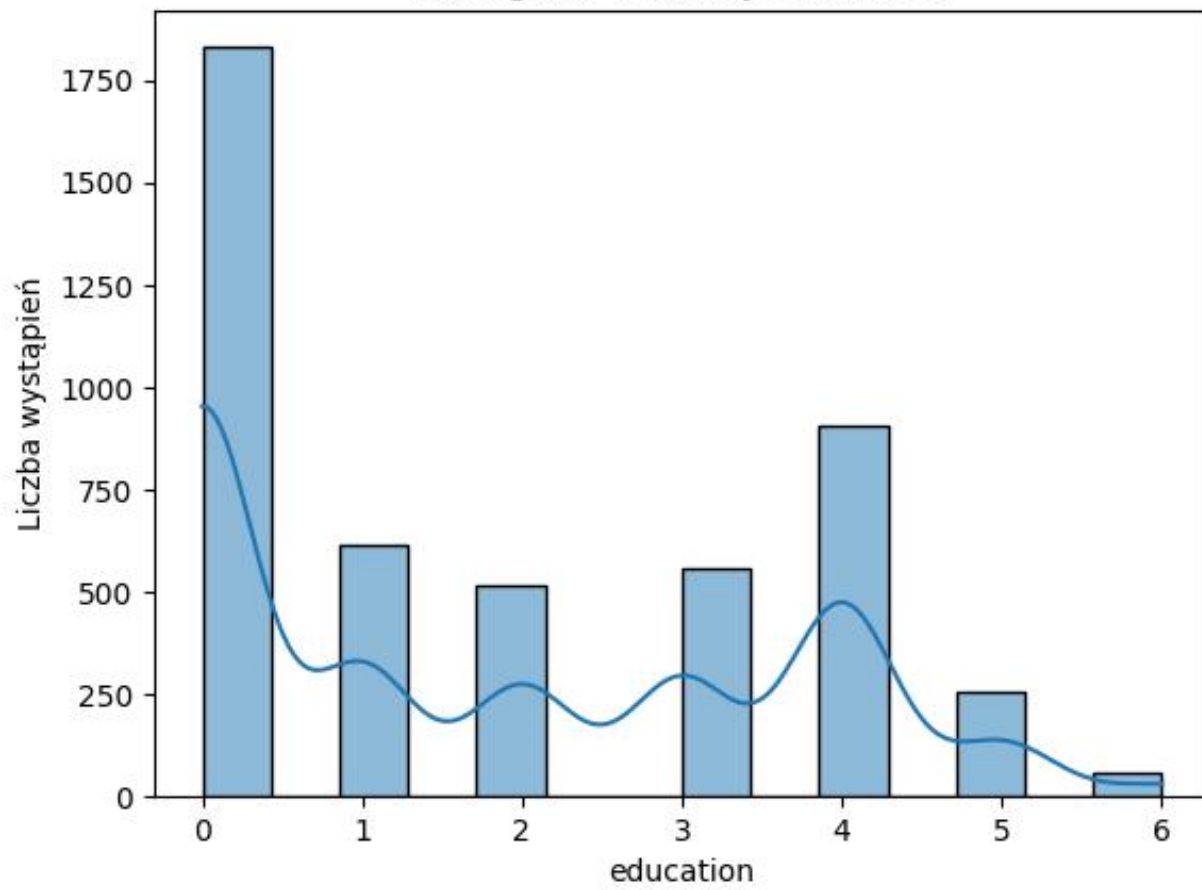
Histogram kolumny distance



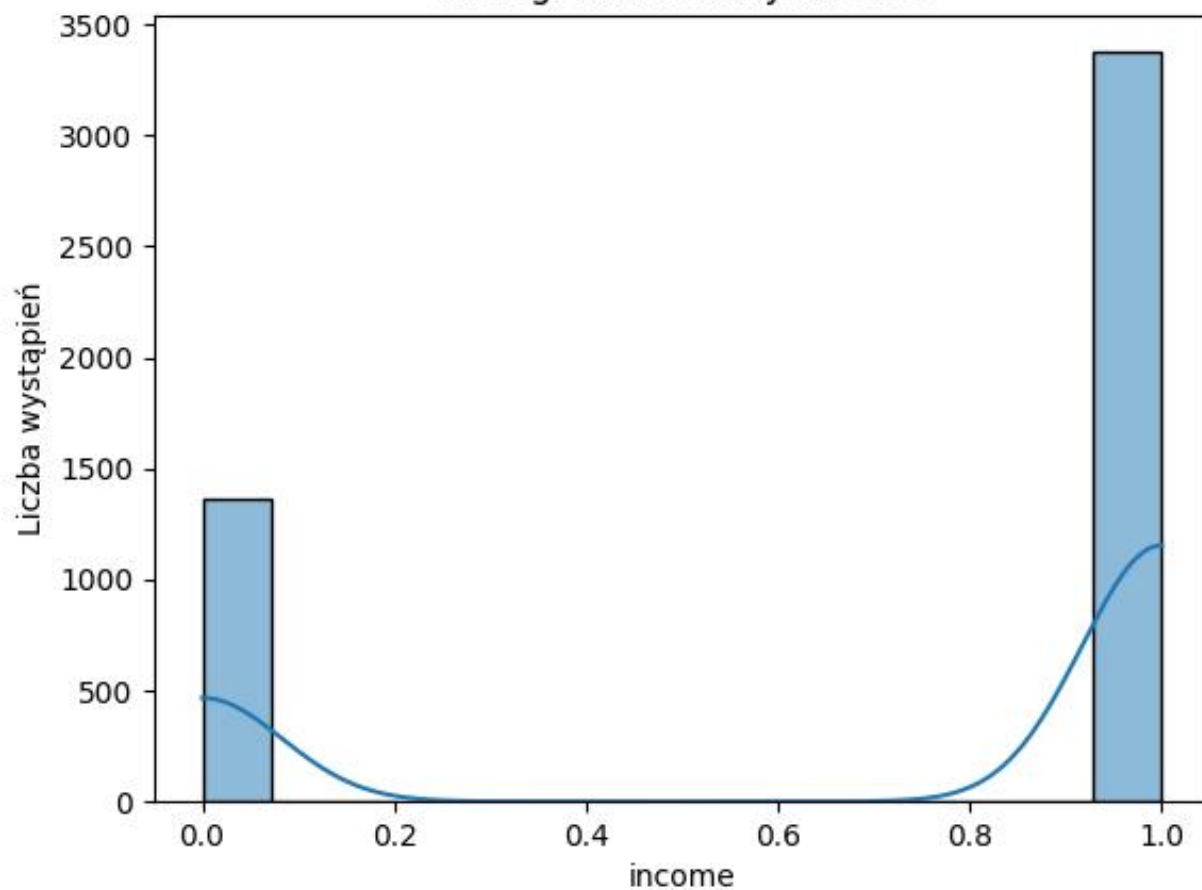
Histogram kolumnowy tuition



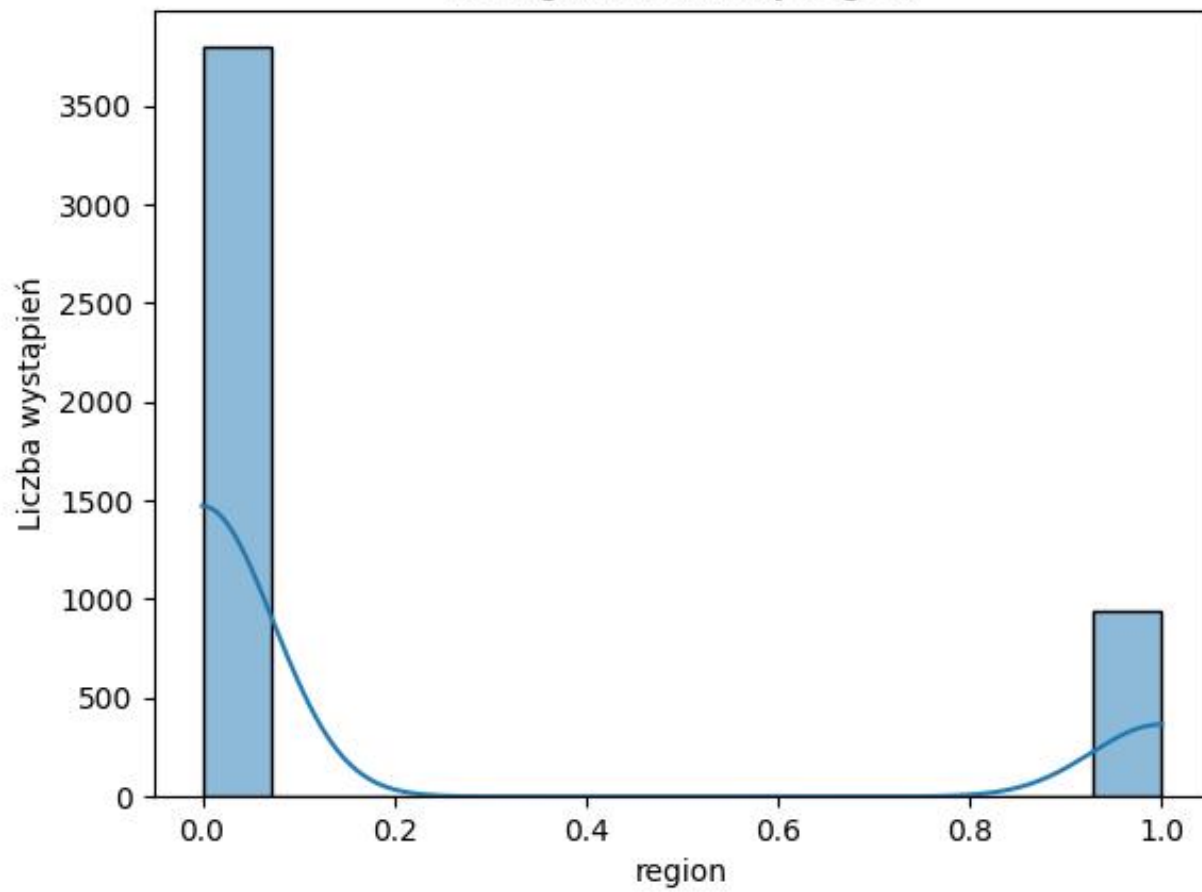
Histogram kolumnowy education



Histogram kolumny income



Histogram kolumny region



Koniec dokumentacji