



Projekt 1 - EDA

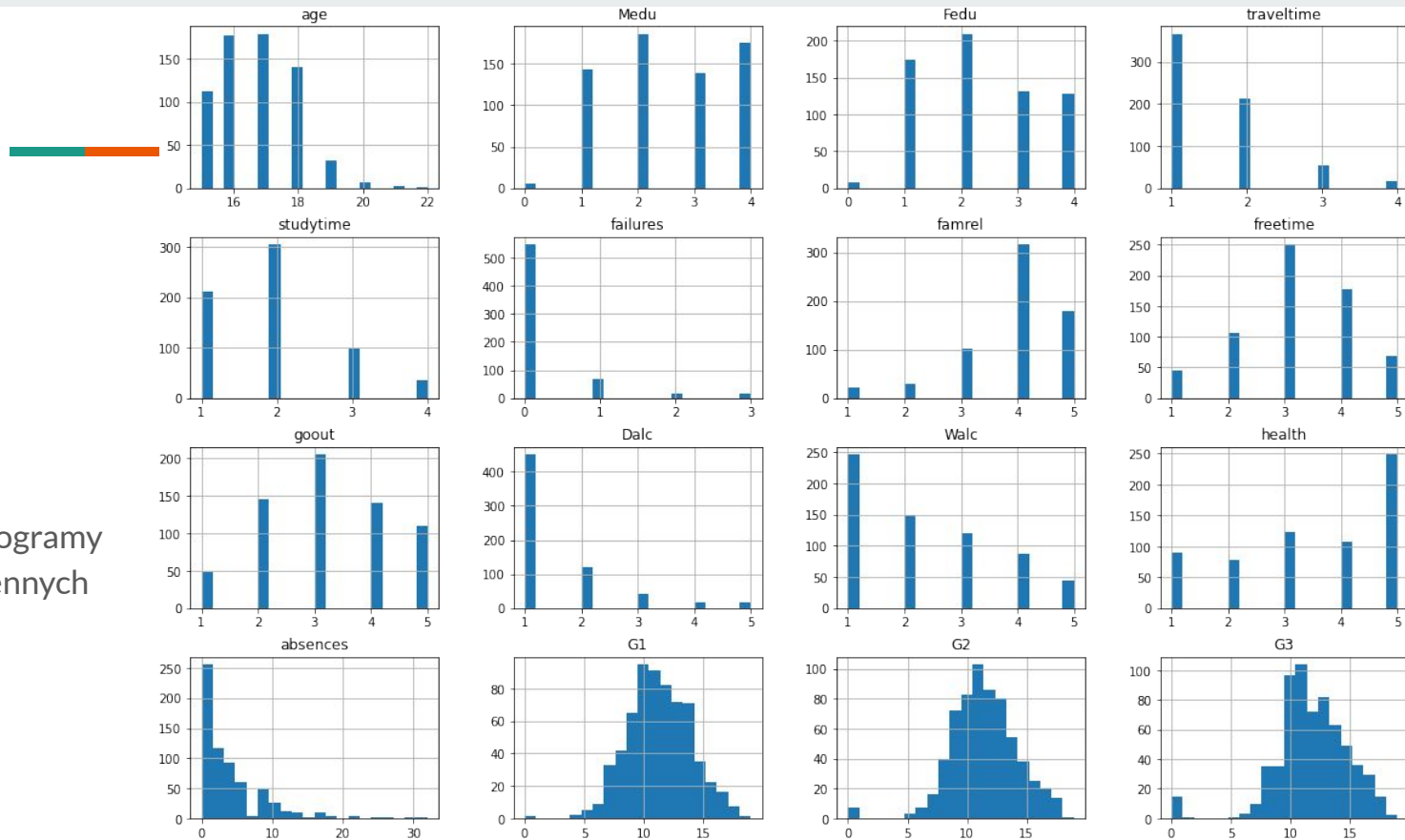
Mikołaj Spytek
Artur Żółkowski

#	Column	Non-Null Count	Dtype
0	school	649 non-null	object
1	sex	649 non-null	object
2	age	649 non-null	int64
3	address	649 non-null	object
4	famsize	649 non-null	object
5	Pstatus	649 non-null	object
6	Medu	649 non-null	int64
7	Fedu	649 non-null	int64
8	Mjob	649 non-null	object
9	Fjob	649 non-null	object
10	reason	649 non-null	object
11	guardian	649 non-null	object
12	traveltime	649 non-null	int64
13	studytime	649 non-null	int64
14	failures	649 non-null	int64
15	schoolsup	649 non-null	object
16	famsup	649 non-null	object
17	paid	649 non-null	object
18	activities	649 non-null	object
19	nursery	649 non-null	object
20	higher	649 non-null	object
21	internet	649 non-null	object
22	romantic	649 non-null	object
23	famrel	649 non-null	int64
24	freetime	649 non-null	int64
25	goout	649 non-null	int64
26	Dalc	649 non-null	int64
27	walc	649 non-null	int64
28	health	649 non-null	int64
29	absences	649 non-null	int64
30	G1	649 non-null	int64
31	G2	649 non-null	int64
32	G3	649 non-null	int64

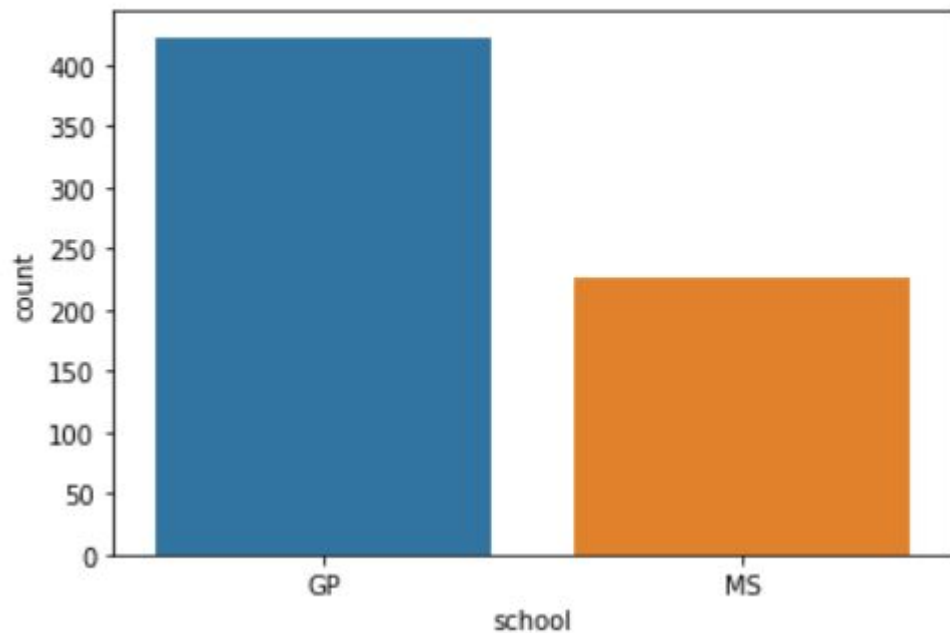
	age	failures	absences	G1	G2	G3
count	649.000000	649.000000	649.000000	649.000000	649.000000	649.000000
mean	16.744222	0.221880	3.659476	11.399076	11.570108	11.906009
std	1.218138	0.593235	4.640759	2.745265	2.913639	3.230656
min	15.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	16.000000	0.000000	0.000000	10.000000	10.000000	10.000000
50%	17.000000	0.000000	2.000000	11.000000	11.000000	12.000000
75%	18.000000	0.000000	6.000000	13.000000	13.000000	14.000000
max	22.000000	3.000000	32.000000	19.000000	19.000000	19.000000

Podstawowe statystyki zmiennych ciągłych

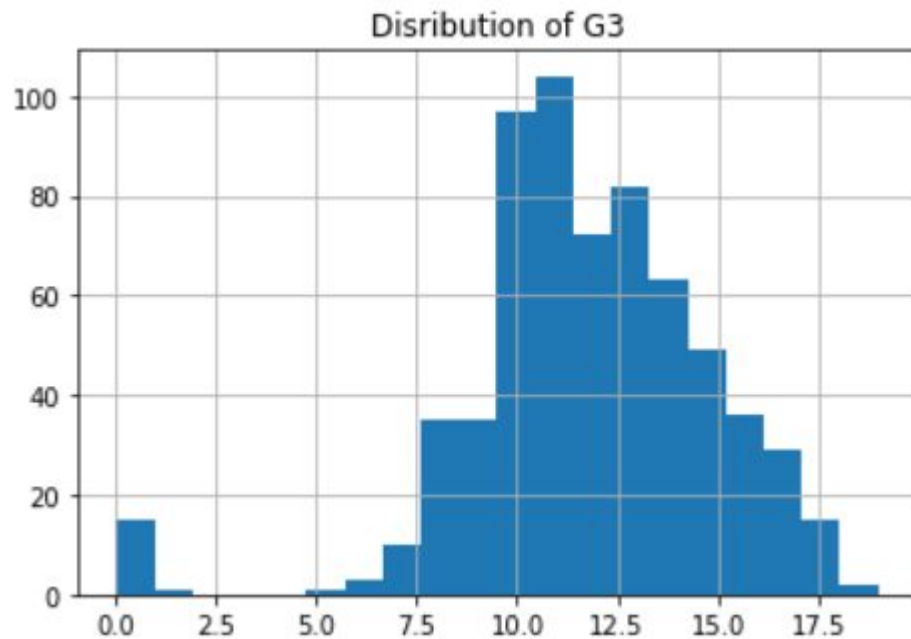
Histogramy zmiennych



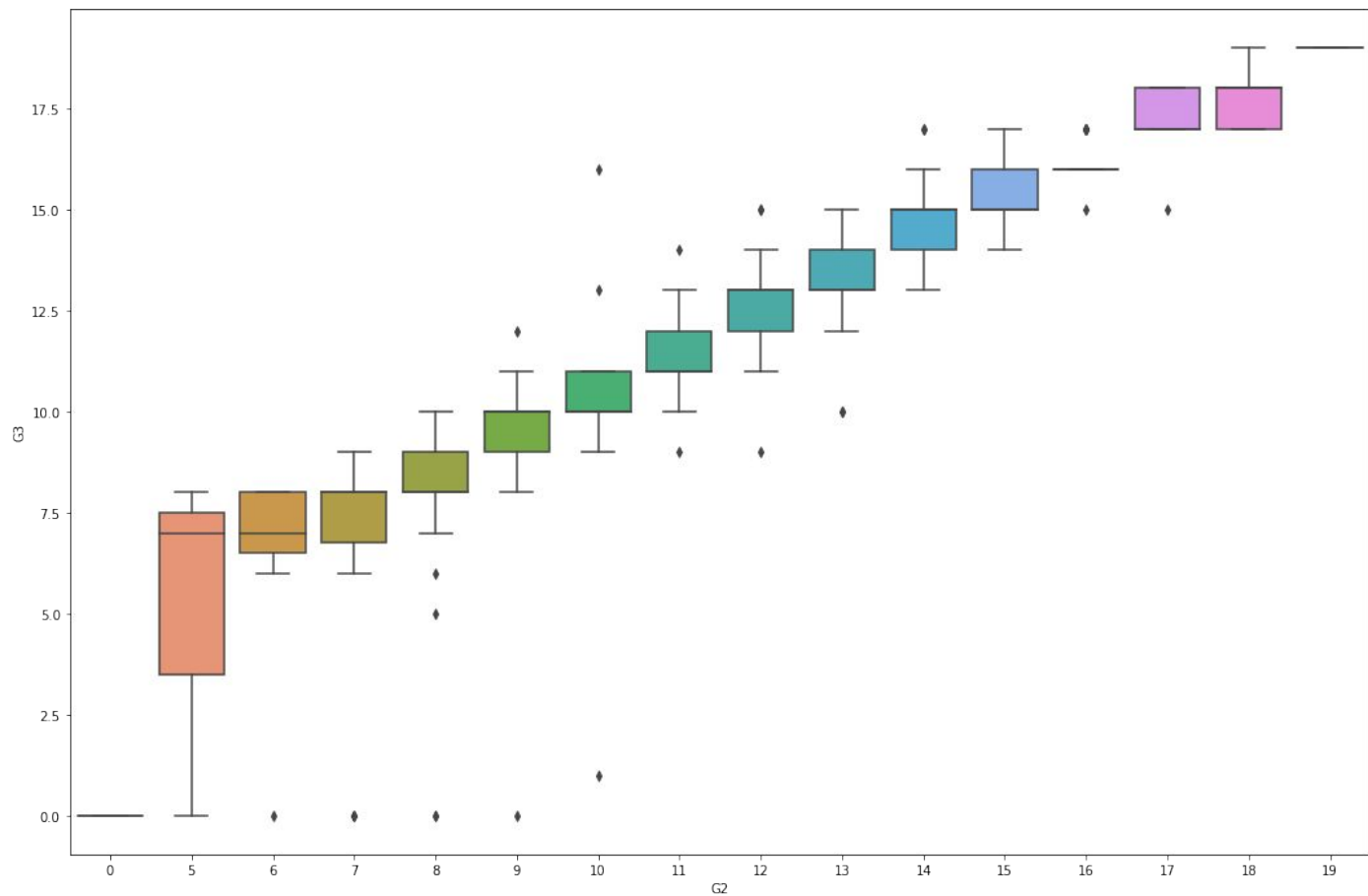
```
sns.countplot(data=df, x="school")  
plt.show()
```



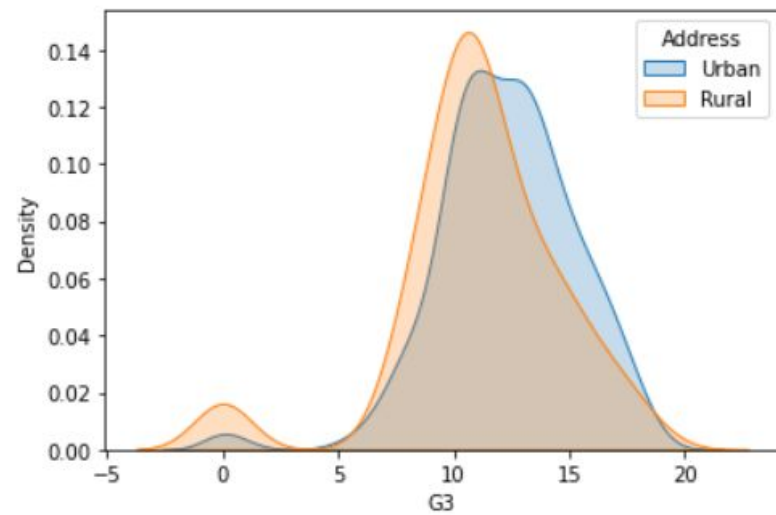
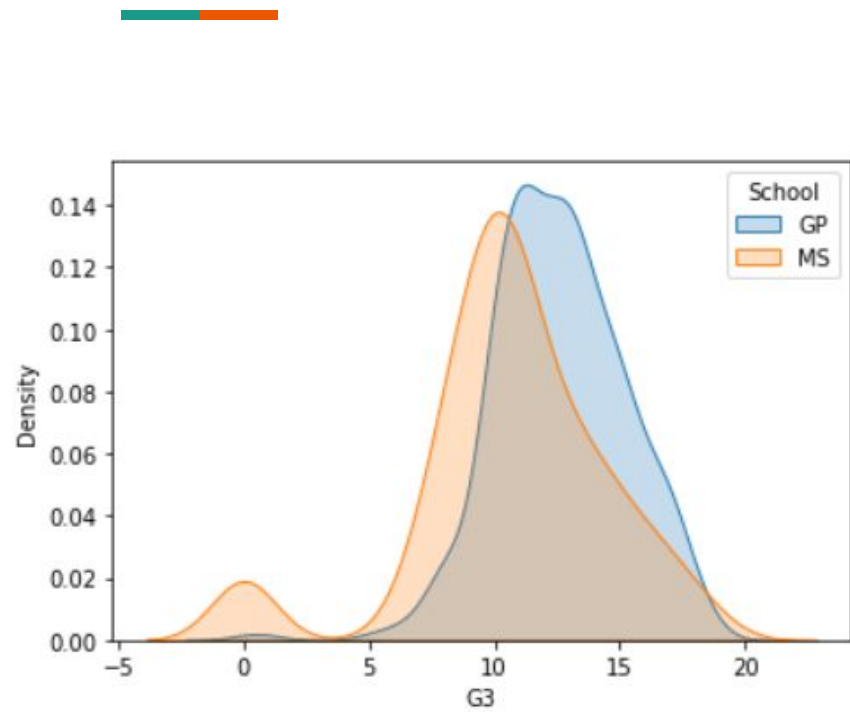
Dysproporcja w ilości danych z poszczególnych szkół

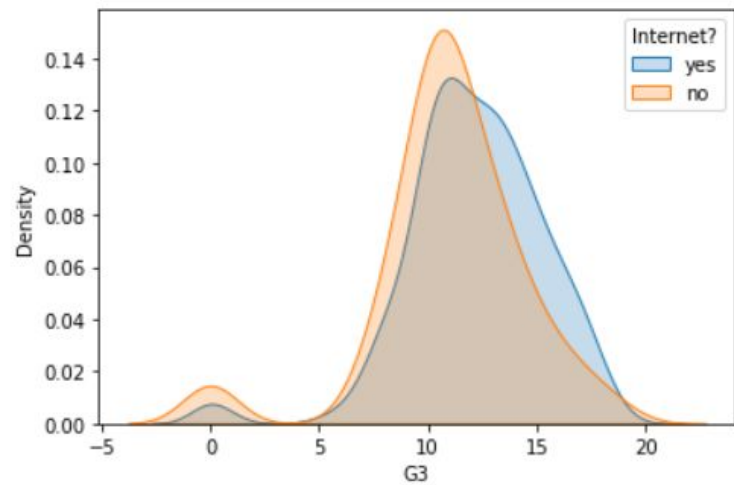
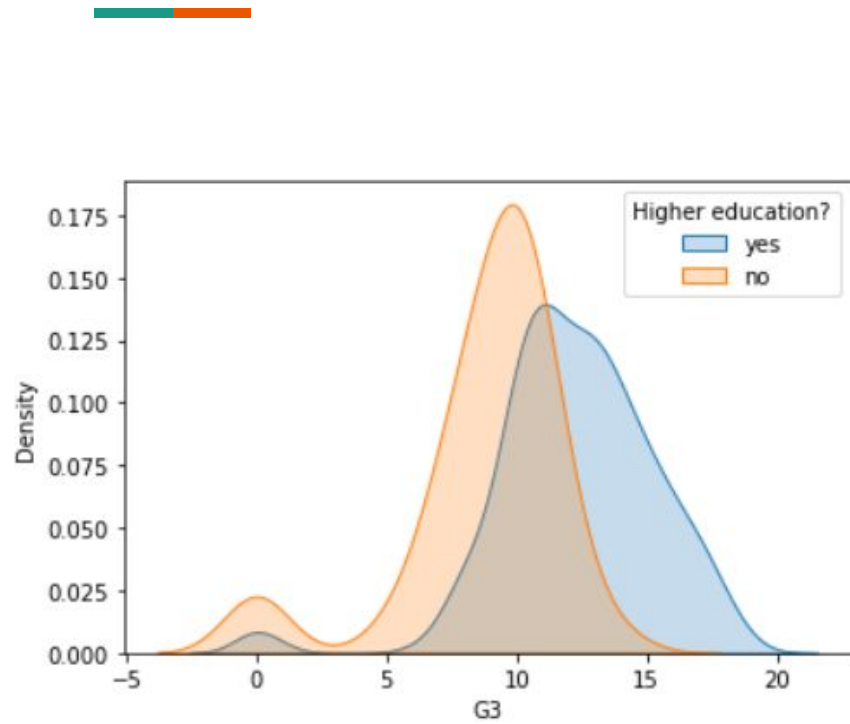


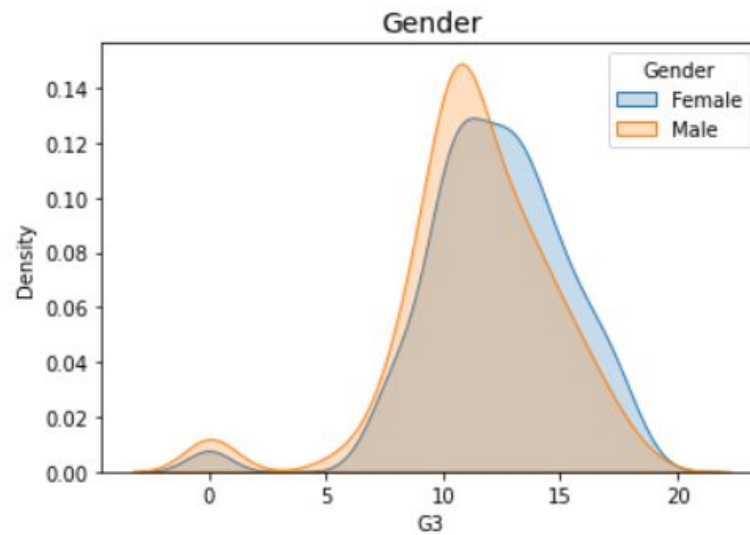
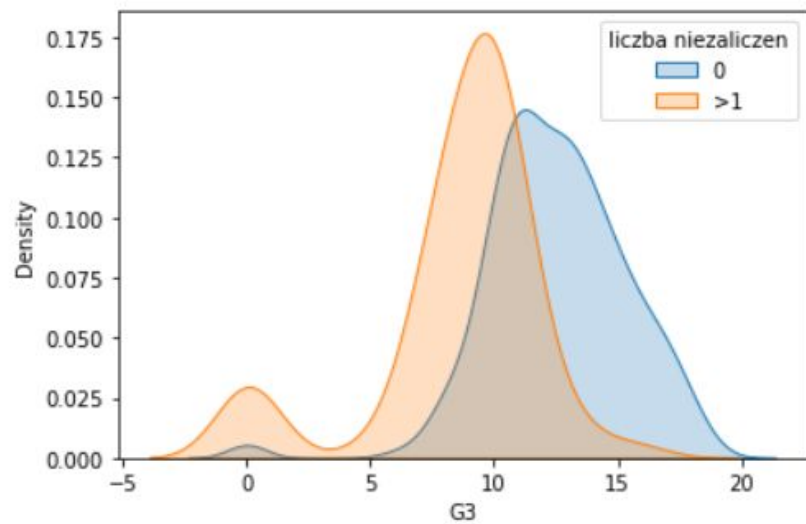
Rozkład zmiennej
wyjaśnianej

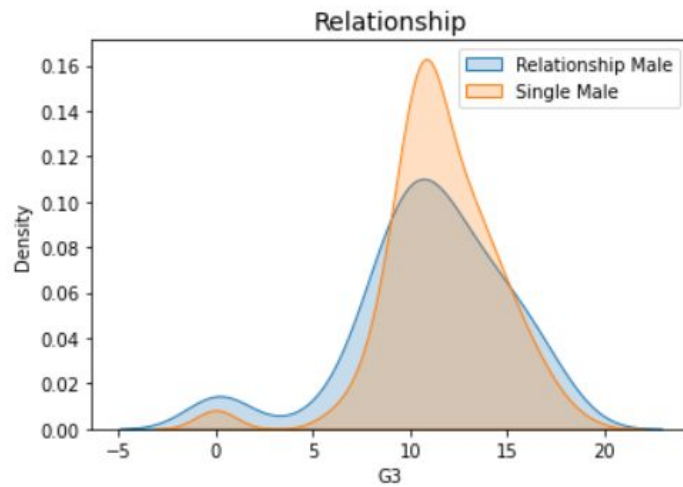
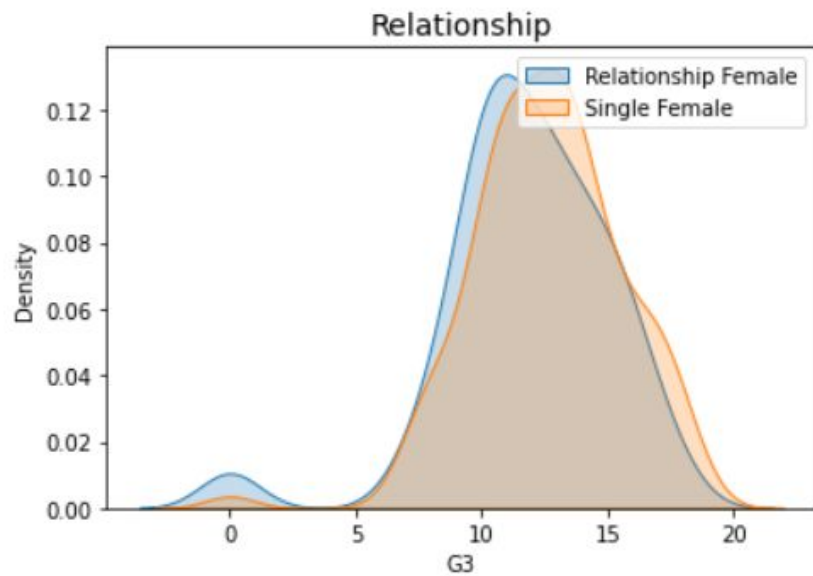


Korelacja z oceną
z poprzedniego
semestru

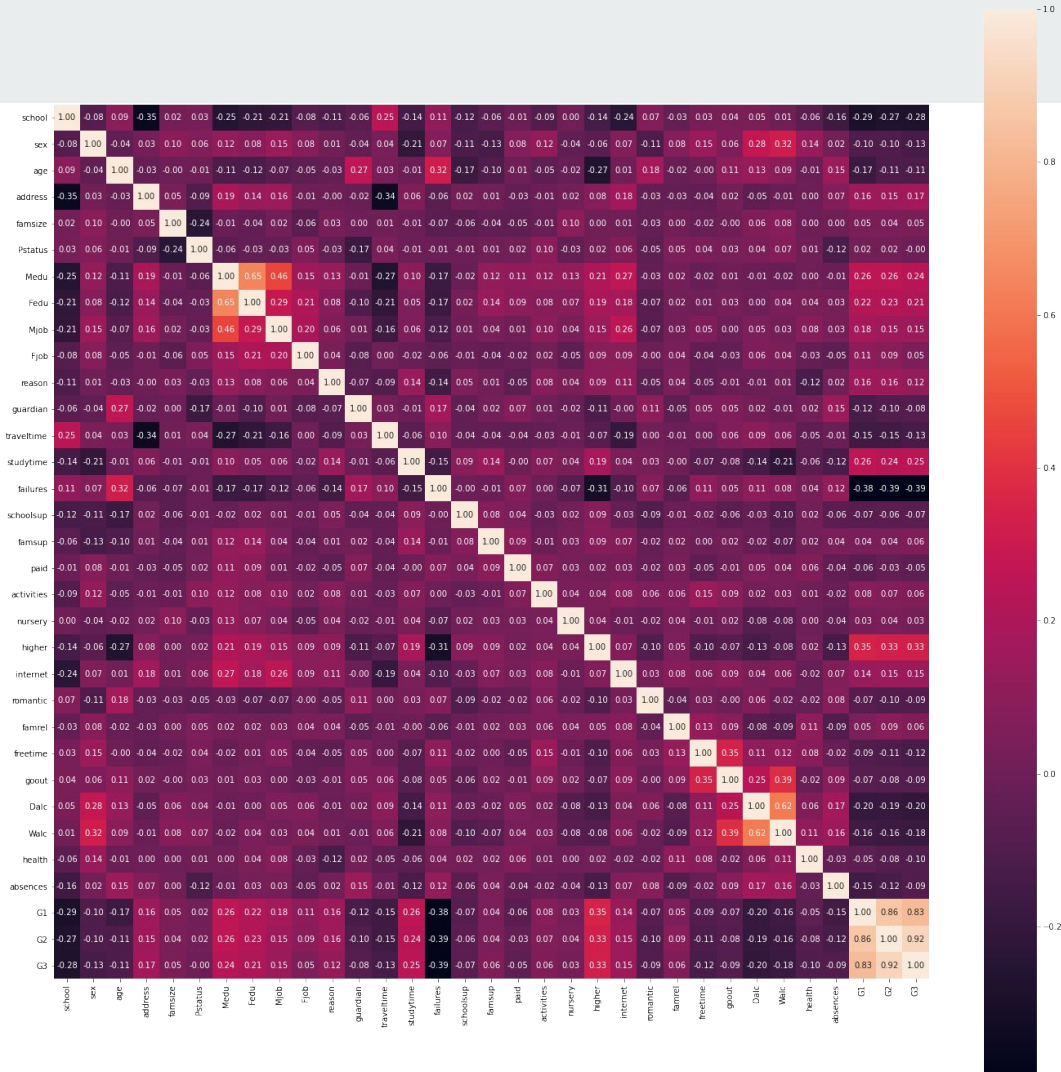


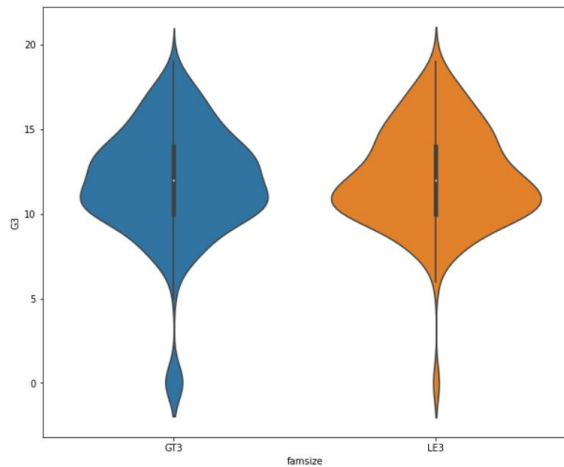




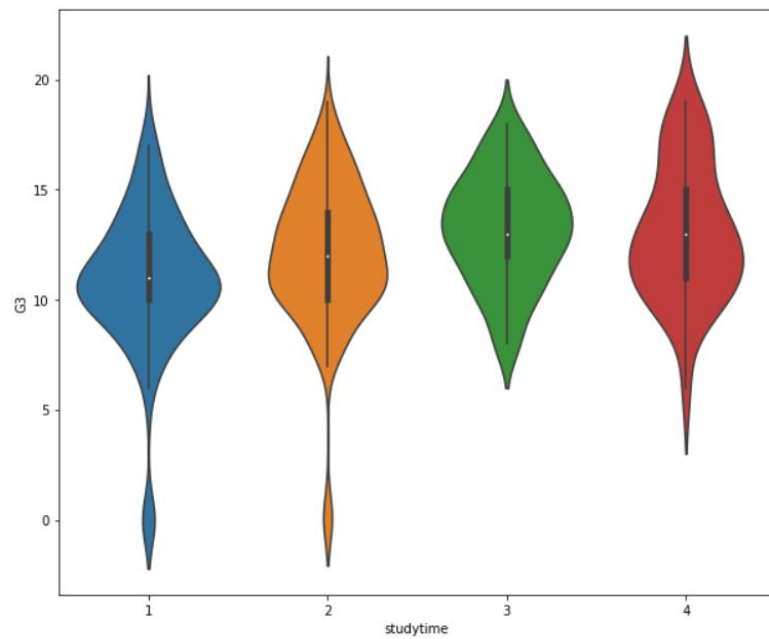
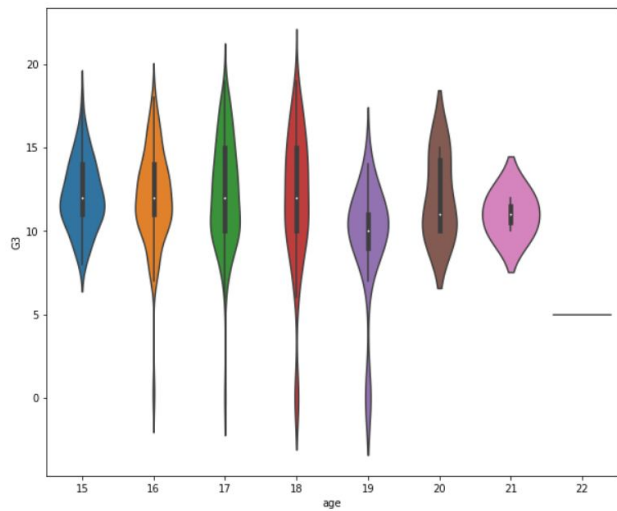


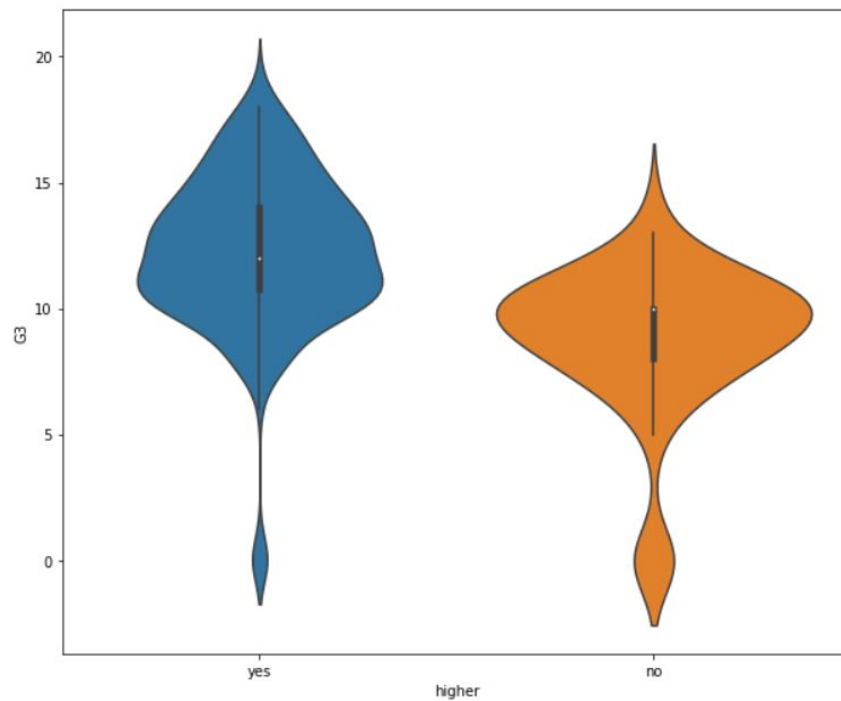
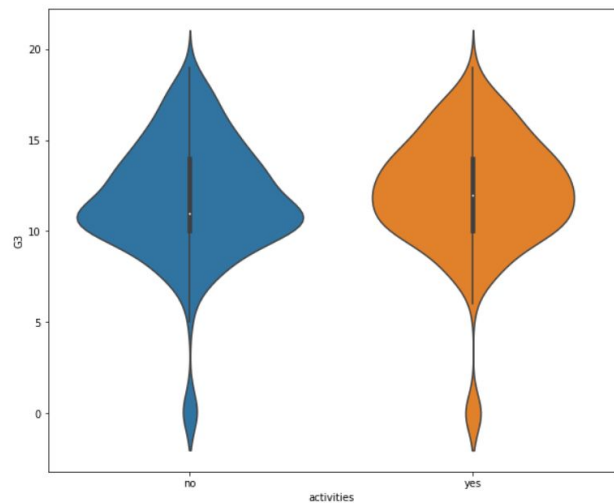
Macierz korelacji (Pearsona)





Rozkłady G3 w rozbiciu na niektóre zmienne





Rozkłady G3 w rozbiciu na niektóre zmienne