

# Credit Card Fraud



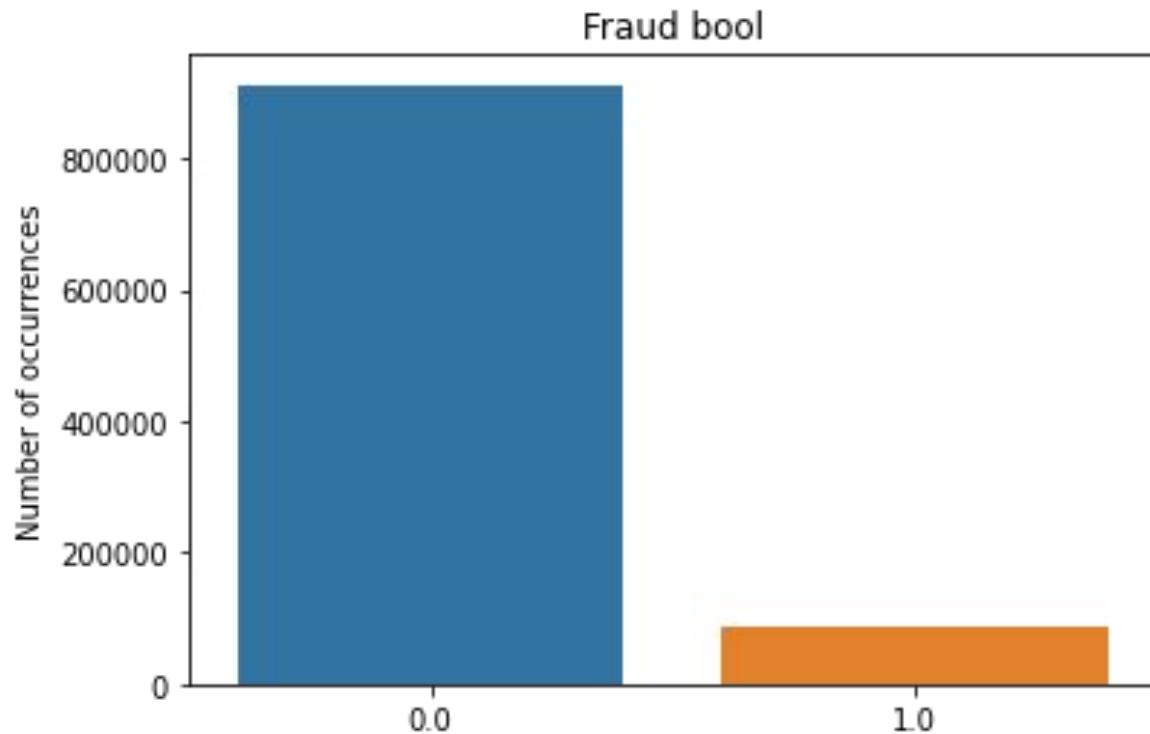


## About Dataset

- Zbiór zawiera milion przypadków
- 8 atrybutów
- Zbiór niezbalansowany
  - niecałe 10% przypadków pozytywnych

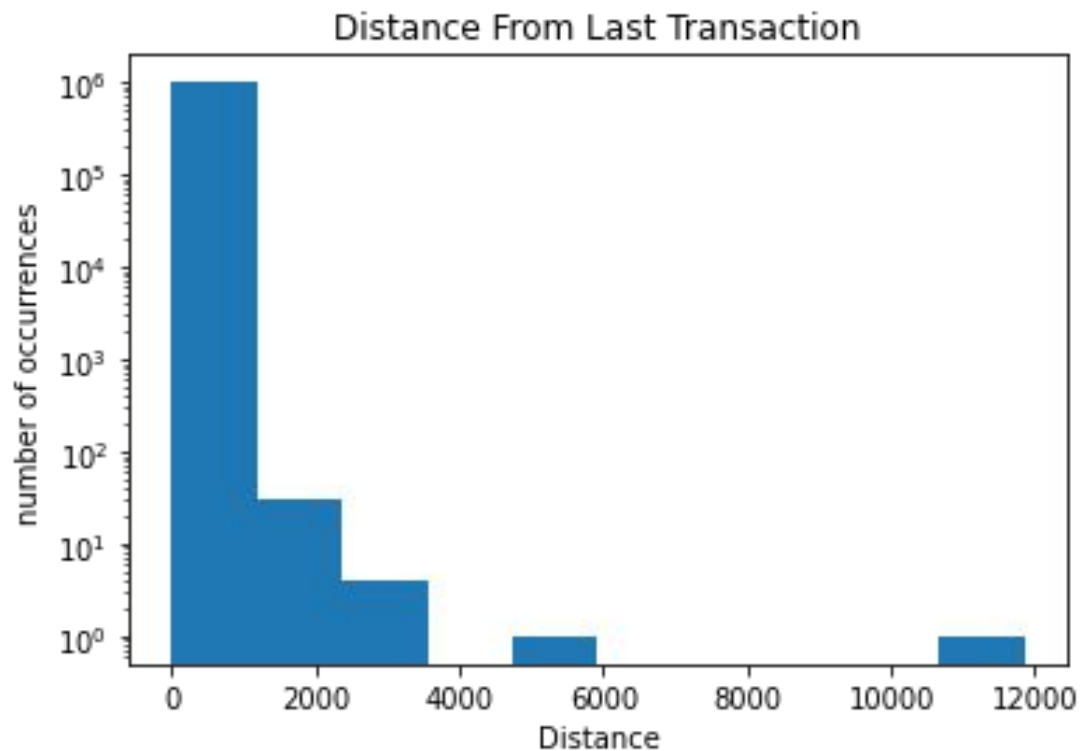


# Zbiór niezbalansowany





# Atrybuty ciągłe



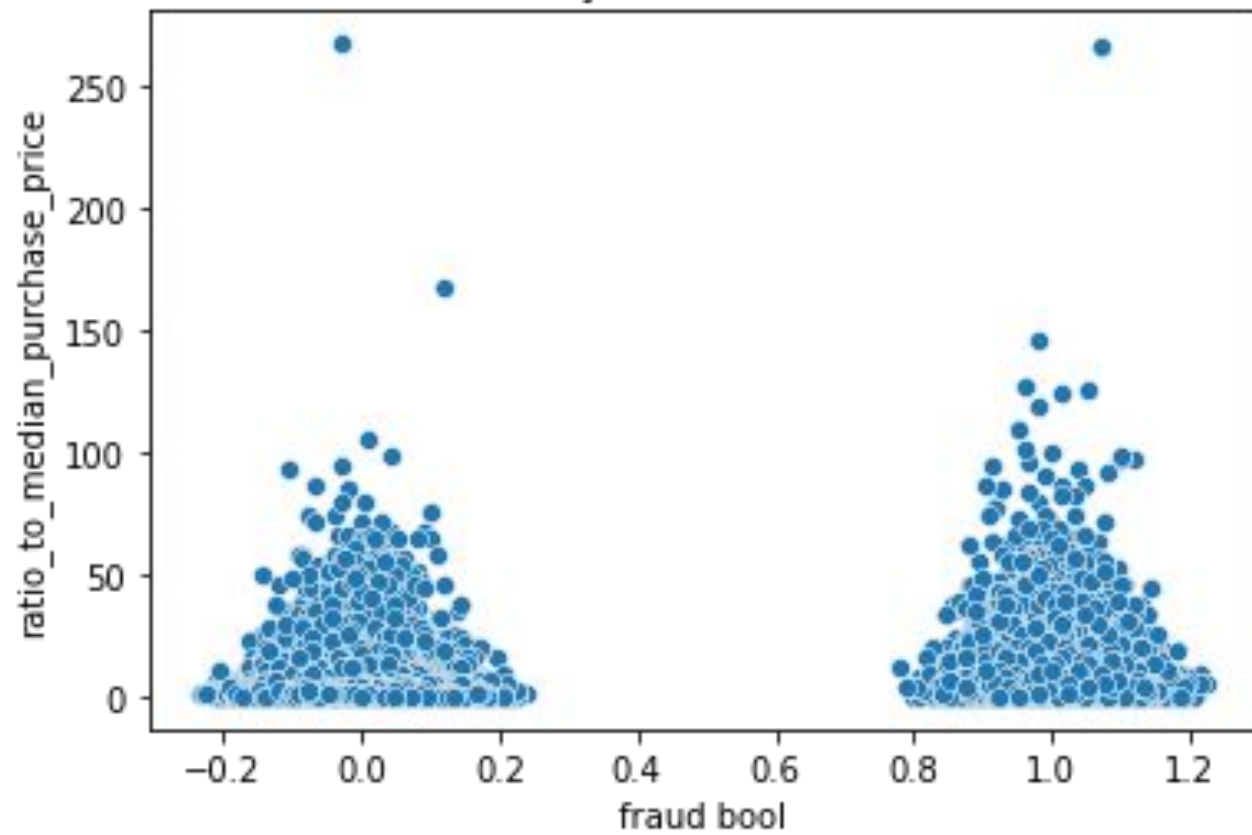


# Analiza zbioru

Macierz korelacji

1	0.00019	-0.0014	0.14	-0.0007	-0.0016	-0.0013	0.19
0.00019	1	0.001	-0.00093	0.0021	-0.0009	0.00014	0.092
-0.0014	0.001	1	0.0014	0.00059	0.00094	-0.00033	0.46
0.14	-0.00093	0.0014	1	-0.0013	-0.00042	-0.00053	-0.0014
-0.0007	0.0021	0.00059	-0.0013	1	-0.0014	-0.00022	-0.061
-0.0016	-0.0009	0.00094	-0.00042	-0.0014	1	-0.00029	-0.1
-0.0013	0.00014	-0.00033	-0.00053	-0.00022	-0.00029	1	0.19
0.19	0.092	0.46	-0.0014	-0.061	-0.1	0.19	1

Wykres rozrzutu





# **Próby predykcji**



# Gradient Boosting

F1	precision	recall	accuracy
0.9972	0.9999	0.9945	0.9995



Confusion Matrix Gradient Boosting

