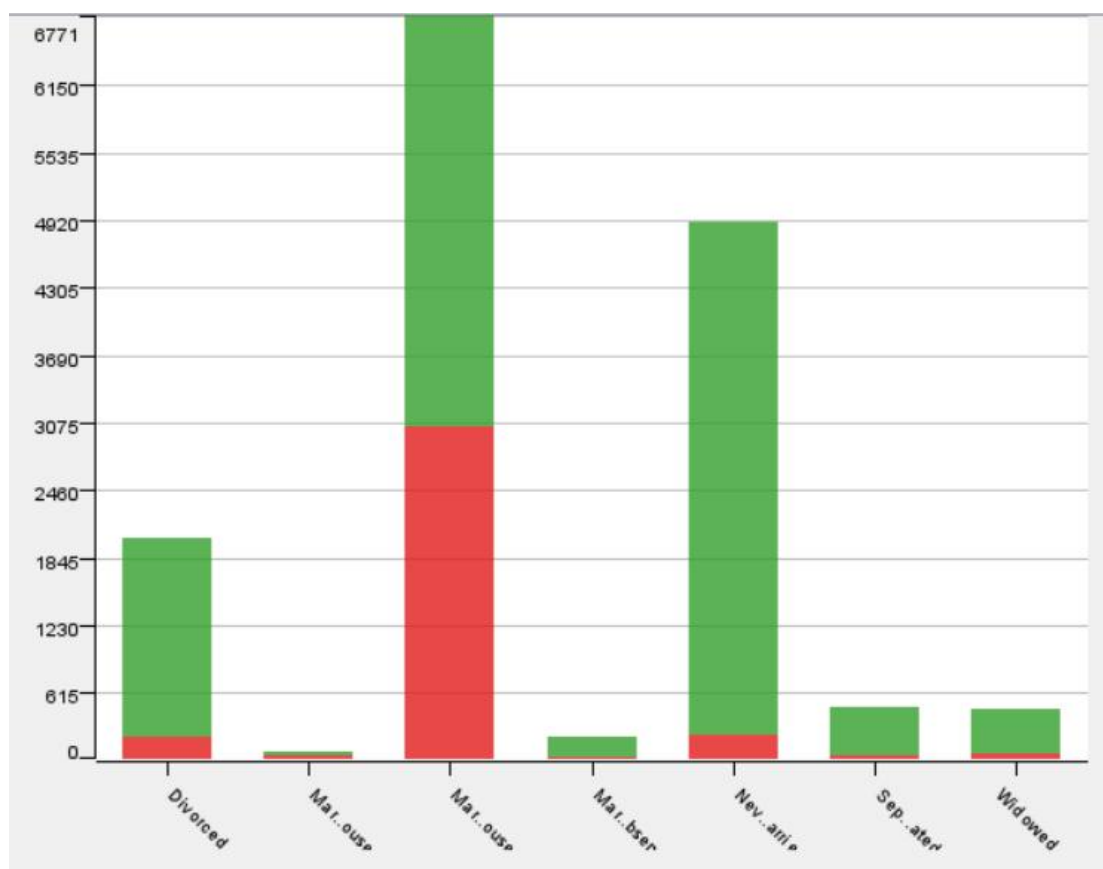


Raport nr 1

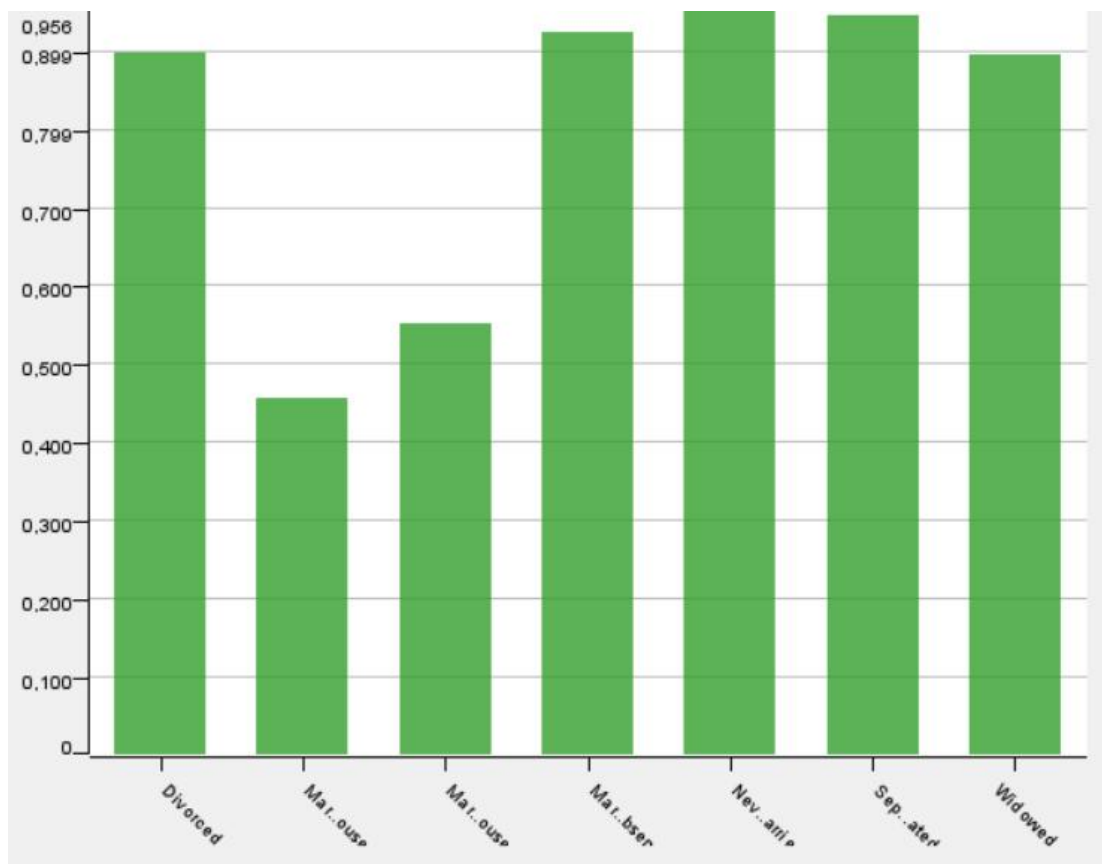
Plik `Adult_ch3_training` zawiera dane o 14797 respondentach takie jak wiek, grupa pracownicza, lata edukacji, stan cywilny, zawód, płeć oraz czy ich zarobki są powyżej lub nie więcej niż 50 tys. dolarów rocznie. Predyktorami były *maritalstatus* - stan cywilny oraz *education* - lata edukacji. Zmienna celu *income* - dochód. Analiza została przeprowadzona w programie KNIME. Kolorem zielonym zostały oznaczone dochody nie więcej niż 50 tys., a czerwonym powyżej 50 tys. dolarów rocznie.

Zestawiony wykres słupkowy dla zmiennej *maritalstatus* z wartościami zmiennej *income*



Najwięcej respondentów jest z grupy *Married-civ-spouse* - małżeństwo z cywilem. Na pierwszy rzut oka można przypuszczać że jest więcej respondentów zarabiających nie więcej niż 50 tys. dolarów rocznie w każdym stanie cywilnym. Największa różnica między osobami zarabiającymi poniżej 50 tys. dolarów rocznie a więcej niż 50 tys. jest w grupie *Never-married* - niezamężnych, natomiast najmniejsza różnica w grupie *Married-civ-spouse* - małżeństwo z cywilem.

Znormalizowany zestawiony wykres słupkowy dla zmiennej *maritalstatus* z wartościami zmiennej *income*



Po dokonaniu normalizacji można zauważyć że w grupie *Married-AF-spouse* - małżeństwo z osobą z sił zbrojnych więcej respondentów zarabia więcej niż 50 tys. dolarów rocznie (prawie 55%). W większości reszty grup (stan cywilny) około 90% zarabia nie więcej niż 50 tys. dolarów rocznie.

Wnioskować można że respondenci zamężni mają większe prawdopodobieństwo zarabiania powyżej 50 tys. dolarów rocznie.

Tabela krzyżowa dla zmiennych *maritalstatus* i *income*

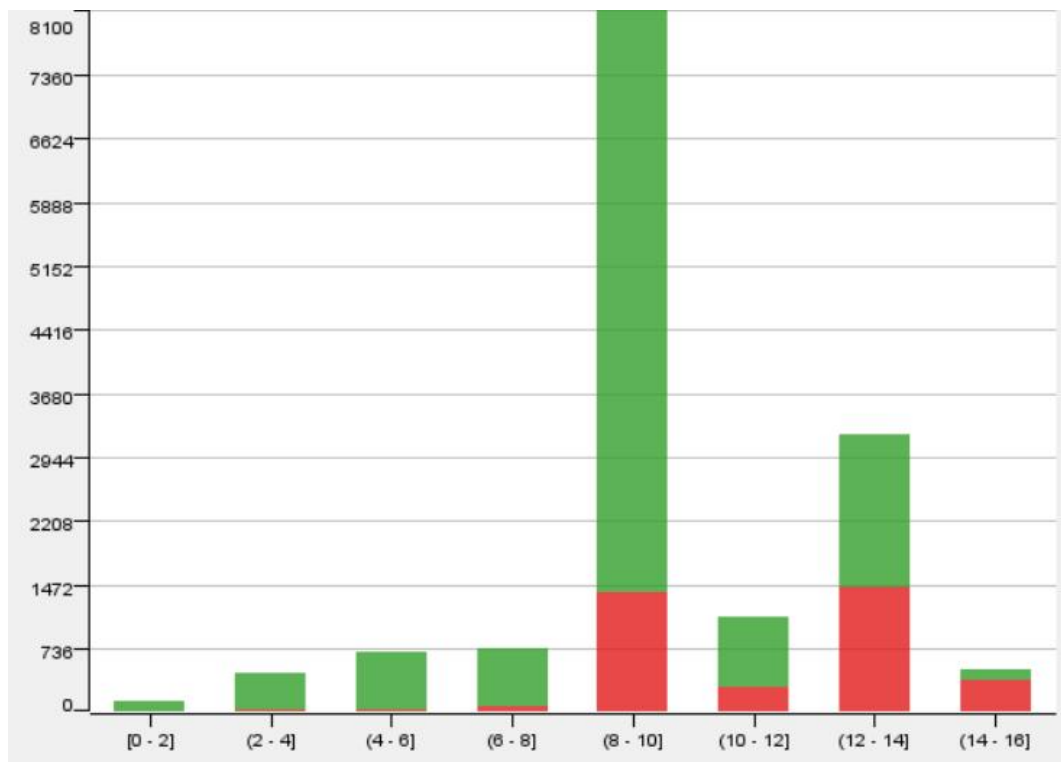
Column Percent	Divorced	Married-AF-spouse	Married-civ-spouse	Married-spouse-absent	Never-married	Separated	Widowed	Total
<=50K	89,6723%	45,4545%	55,147%	92,4623%	95,6121%	94,4915%	89,404%	
>50K	10,3277%	54,5455%	44,853%	7,5377%	4,3879%	5,5085%	10,596%	

W grupie *Married-AF-spouse* - małżeństwo z osobą z sił zbrojnych więcej respondentów zarabia więcej niż 50 tys. dolarów rocznie. We wszystkich innych więcej respondentów zarabia nie więcej niż 50 tys. dolarów rocznie.

Wnioskować można że respondenci zamężni mają większe prawdopodobieństwo zarabiania powyżej 50 tys. dolarów rocznie.

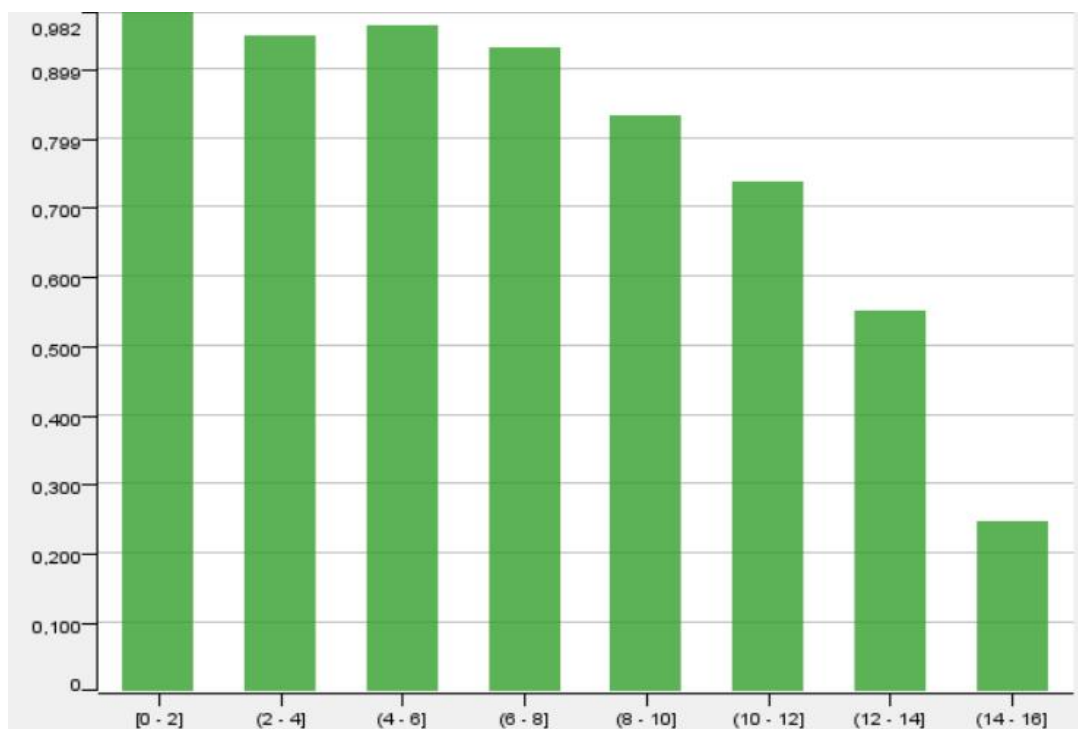
Obserwacji odstających ze względu na wartość zmiennej *education* - lata edukacji jest 552.

Histogram dla zmiennej *education* z wartościami zmiennej *income*



Największa ilość respondentów jest w grupie lata edukacji (8-10). Przy latach edukacji powyżej 8 lat można zauważyć większą ilość respondentów mających zarobki powyżej 50 tys niż przy latach edukacji poniżej 8 lat.

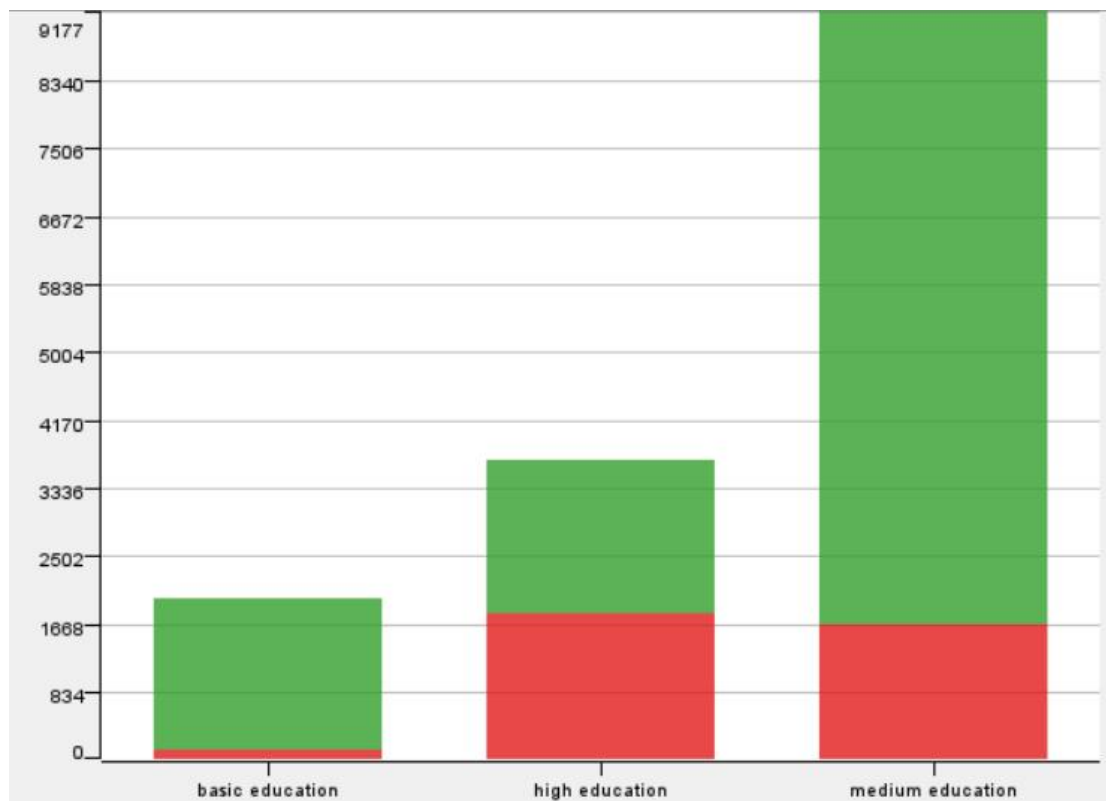
Znormalizowany histogram dla zmiennej *education* z wartościami zmiennej *income*



Również po znormalizowaniu przy latach edukacji respondentów powyżej 8 lat można zauważyć większą ilość respondentów mających zarobki powyżej 50 tys niż przy latach edukacji poniżej 8 lat. Przy latach edukacji (14-16) respondentów mających wynagrodzenie wyższe niż 50 tys. rocznie jest prawie 75%.

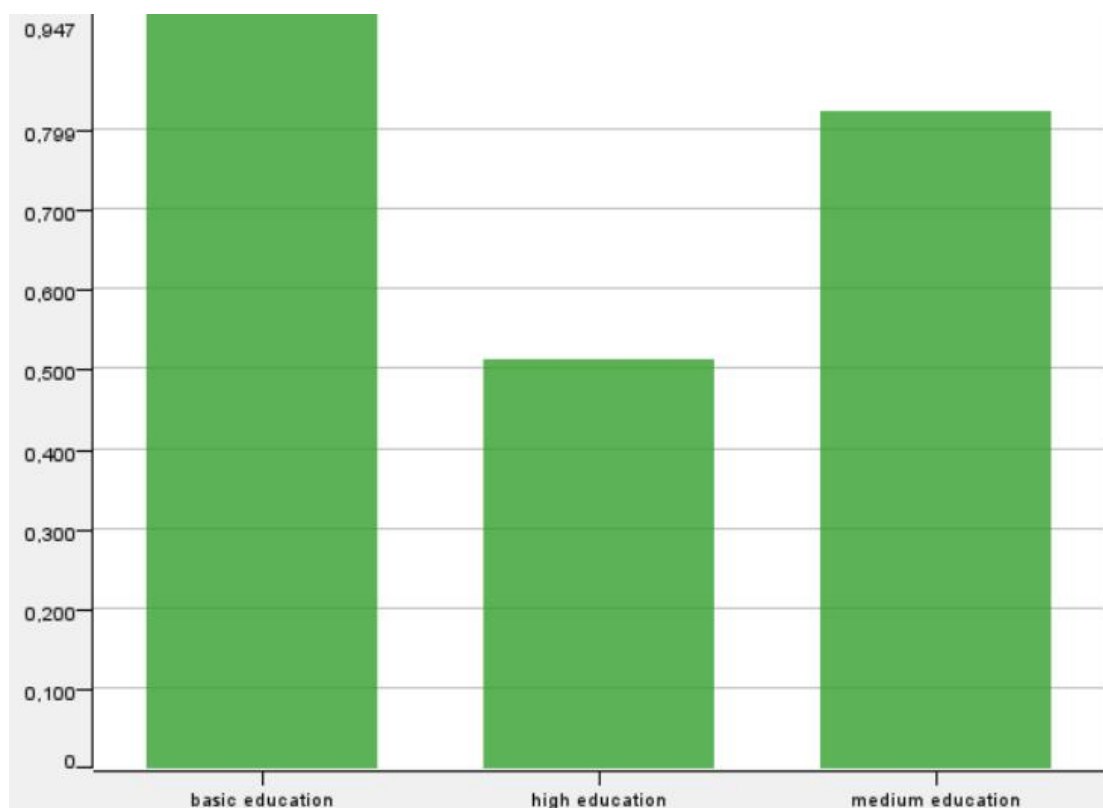
Wnioskować można że im więcej lat edukacji tym większa możliwość na zarobki powyżej 50 tys. rocznie.

Zestawiony wykres słupkowy dla zmiennej *education_binned* z wartościami zmiennej *income*



Najwięcej respondentów jest z grupy *medium education*, a najmniej z grupy *basic education*. W przypadku *basic education* bardzo mało respondentów zarabia więcej niż 50 tys. rocznie. W kolejnych dwóch grupach liczba respondentów zarabiających powyżej 50 tys. rocznie wzrasta.

Znormalizowany zestawiony wykres słupkowy dla zmiennej *education_binned* z wartościami zmiennej *income*



Po znormalizowaniu można zauważyć że w grupie *basic education* prawie 95% respondentów zarabia nie więcej 50 tys. rocznie. W grupie *high education* prawie 50% respondentów ma zarobki większe niż 50 tys. rocznie.

Wnioskować można że im wyższe wykształcenie tym większa możliwość zarobków powyżej 50 tys. rocznie.