

# Adult Dataset Profiling Report



## Overview

Brought to you by [YData](#)

Overview

Alerts **9**

Reproduction

### Dataset statistics

Number of variables	15
Number of observations	32561
Missing cells	0
Missing cells (%)	0.0%
Duplicate rows	23
Duplicate rows (%)	0.1%
Total size in memory	17.6 MiB
Average record size in memory	568.3 B

### Variable types

Numeric	6
Categorical	9

## Variables

Select Columns

**age**

Real number ( $\mathbb{R}$ )

Distinct73

Distinct (%)0.2%

Missing0

Missing (%)0.0%

Infinite0

Infinite (%)0.0%

Mean38.581647

Minimum17

Maximum90

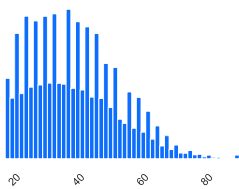
Zeros0

Zeros (%)0.0%

Negative0

Negative (%)0.0%

Memory size254.5 I



More details

workclass

Categorical

Distinct9

Distinct (%)< 0.1%

Missing0

Missing (%)0.0%

Memory size1.8 MiB

Private22696

Self-em...25

Local-gov2093

?1836

State-gov1298

Other va...2097

More details

fnlwgt

Real number ( $\mathbb{R}$ )

Distinct21648

Distinct (%)66.5%

Missing0

Missing (%)0.0%

Infinite0

Infinite (%)0.0%

Mean189778.37

Minimum12285

Maximum148470

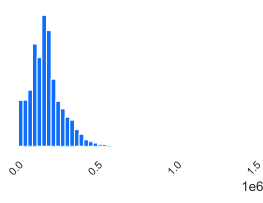
Zeros0

Zeros (%)0.0%

Negative0

Negative (%)0.0%

Memory size254.5 MB



More details

education

Categorical

High correlation

Distinct16

Distinct (< 0.1%)< 0.1%

Missing0

Missing (%)0.0%

Memory size1.8 MiB

HS-grad10501

Some-co...7291

Bachelors5355

Masters172

Assoc-voc13

Other va...6309

More details

education-num

Real number ( $\mathbb{R}$ )

High correlation

Distinct16

Distinct (%)< 0.1%

Missing0

Missing (%)0.0%

Infinite0

Infinite (%)0.0%

Mean10.080679

Minimum1

Maximum16

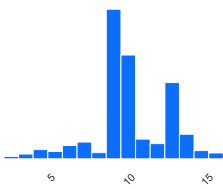
Zeros0

Zeros (%)0.0%

Negative0

Negative (%)0.0%

Memory size254.5 I



More details

marital-status

Categorical

Distinct7

Distinct (%)< 0.1%

Missing0

Missing (%)0.0%

Memory size2.0 MiB

Married-...14976

Never-m...10683

Divorced4443

Separated1025

Widowed993

Other va...441

More details

occupation

Categorical

Distinct15

Distinct (%)< 0.1%

Prof-spe...4140

Craft-re...4099

Exec-ma...4066

Missing	0
Missing (%)	0.0%
Memory size	1.9 MiB

Adm-cle...3770

Sales3650

Other va...12836

More details

relationship

Categorical

High correlation

Distinct	6
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	1.8 MiB

Husband13193

Not-in-f...8305

Own-child5068

Unmarried3446

Wife15

More details

race

Categorical

Imbalance

Distinct	5
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	1.7 MiB

White27816

Black31

Asian-Pa...1039

Amer-In...311

Other271

More details

sex

Categorical

High correlation

Distinct	2	Male	21790
Distinct (%)	< 0.1%	Female	10771
Missing	0		
Missing (%)	0.0%		
Memory size	1.7 MiB		

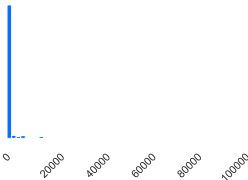
More details

capital-gain

Real number ( $\mathbb{R}$ )

Zeros

Distinct	119	Minimum	0
Distinct (%)	0.4%	Maximum	99999
Missing	0	Zeros	29849
Missing (%)	0.0%	Zeros (%)	91.7%
Infinite	0	Negative	0
Infinite (%)	0.0%	Negative (%)	0.0%
Mean	1077.6488	Memory size	254.5 I



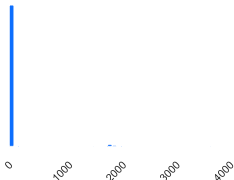
More details

capital-loss

Real number ( $\mathbb{R}$ )

Zeros

Distinct	92	Minimum	0
Distinct (%)	0.3%	Maximum	4356
Missing	0	Zeros	31042
Missing (%)	0.0%	Zeros (%)	95.3%
Infinite	0	Negative	0
Infinite (%)	0.0%	Negative (%)	0.0%
Mean	87.30383	Memory size	254.5 MB

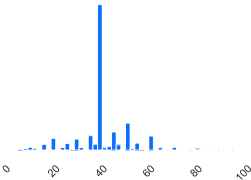


More details

hours-per-week

Real number ( $\mathbb{R}$ )

Distinct	94	Minimum	1
Distinct (%)	0.3%	Maximum	99
Missing	0	Zeros	0
Missing (%)	0.0%	Zeros (%)	0.0%
Infinite	0	Negative	0
Infinite (%)	0.0%	Negative (%)	0.0%
		Memory size	254.5 MB



Mean 40.437456

More details

native-country

Categorical

Imbalance

Distinct 42

Distinct (%) 0.1%

Missing 0

Missing (%) 0.0%

Memory size 1.9 MiB

United-S... 29170

Mexico 643

? 583

Philippin... 198

Germany 137

Other va... 1830

More details

income

Categorical

Distinct 2

Distinct (%) < 0.1%

Missing 0

Missing (%) 0.0%

Memory size 1.7 MiB

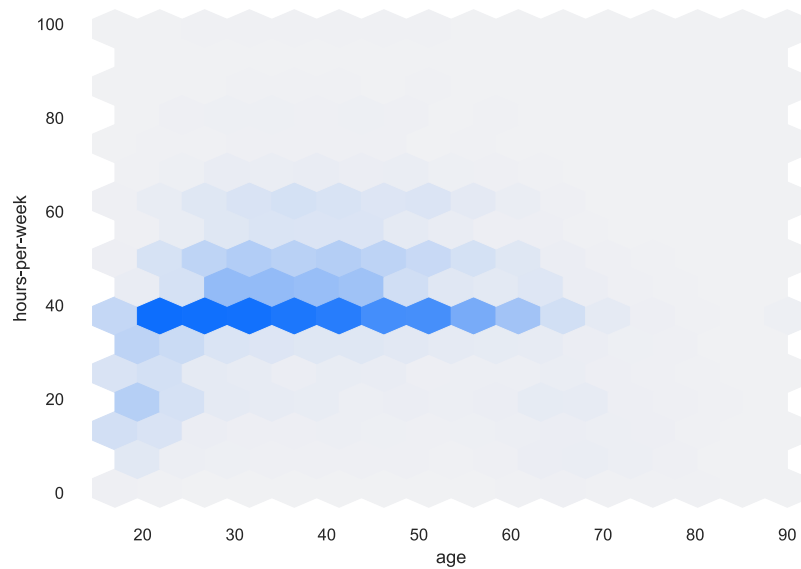
<=50K 24720

>50K 7841

More details

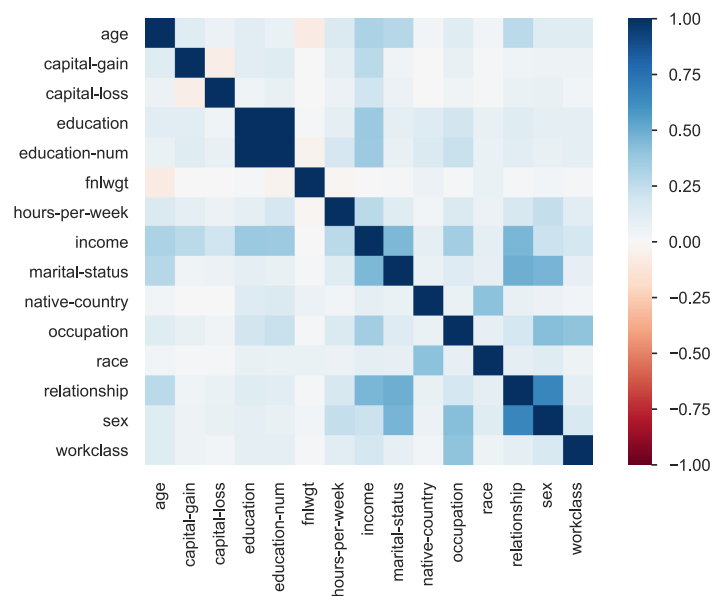
Interactions



[age](#)[fnlwgt](#)[education-num](#)[capital-gain](#)[capital-loss](#)[hours-per-week](#)[hours-per-week](#)[age](#)[fnlwgt](#)[education-num](#)[capital-gain](#)[capital-loss](#)

## Correlations

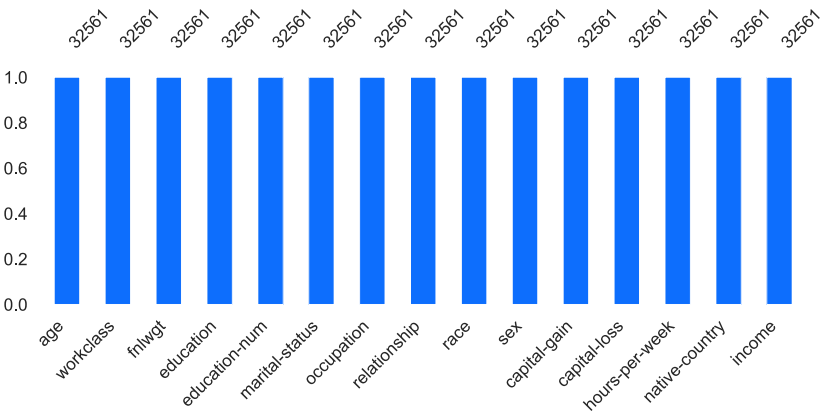
[Auto](#)[Heatmap](#)[Table](#)



## Missing values

Count

[Matrix](#)



A simple visualization of nullity by column.

# Sample

First rows [Last rows](#)

	age	workclass	fnlwgt	education	education-num
0	39	State-gov	77516	Bachelors	13
1	50	Self-emp-not-inc	83311	Bachelors	13
2	38	Private	215646	HS-grad	9
3	53	Private	234721	11th	7
4	28	Private	338409	Bachelors	13
5	37	Private	284582	Masters	14
6	49	Private	160187	9th	5
7	52	Self-emp-not-inc	209642	HS-grad	9

	age	workclass	fnlwgt	education	education-num
8	31	Private	45781	Masters	14
9	42	Private	159449	Bachelors	13

## Duplicate rows

### Most frequently occurring

	age	workclass	fnlwgt	education	education-num	marital-status
8	25	Private	195994	1st-4th	2	Never-married
0	19	Private	97261	HS-grad	9	Never-married
1	19	Private	138153	Some-college	10	Never-married
2	19	Private	146679	Some-college	10	Never-married
3	19	Private	251579	Some-college	10	Never-married
4	20	Private	107658	Some-college	10	Never-married
5	21	Private	243368	Preschool	1	Never-married
6	21	Private	250051	Some-college	10	Never-married
7	23	Private	240137	5th-6th	3	Never-married

	age	workclass	fnlwgt	education	education- num	marital- status
9	25	Private	308144	Bachelors	13	Never-married

Report generated by [YData](#).

