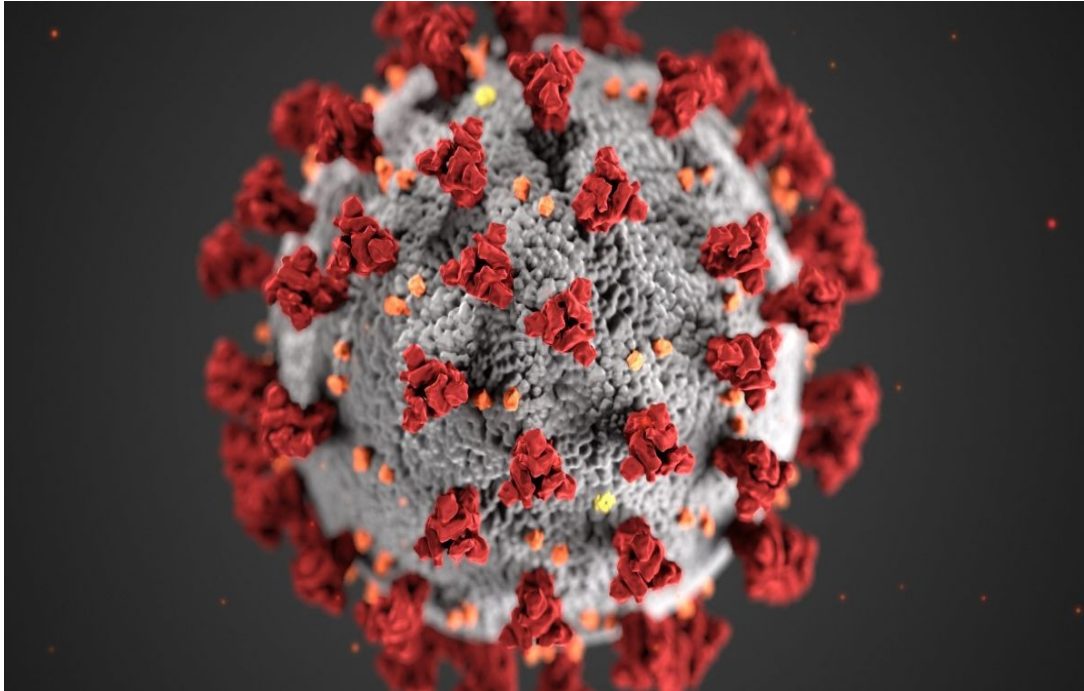


Census Project

Problème :



- Covid-19 :
 - plus de 1/3 de la pop Mondiale
 - plus de 1/4 de la pop US

Solution



United States[®]
Census
Bureau

DataSet

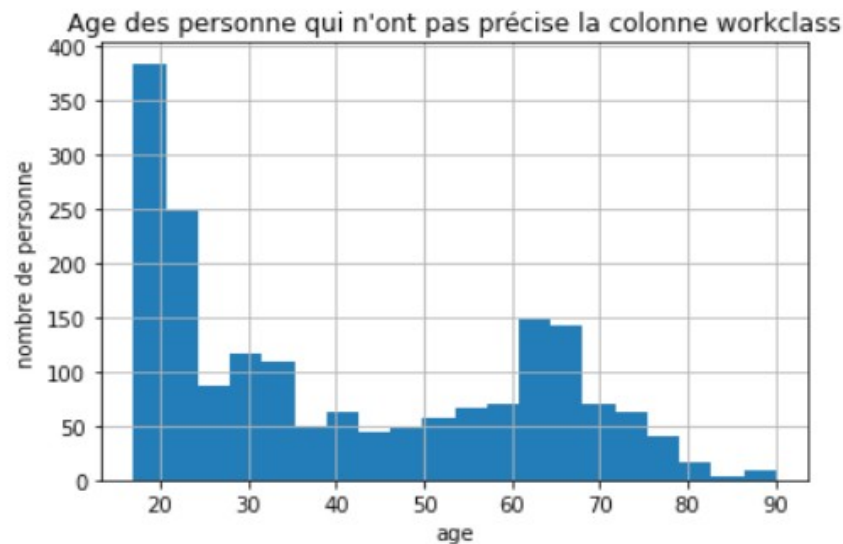
	age	workclass	fnlwgt	education	education.num	marital.status	occupation	relationship	race	sex	capital.gain	capital.loss	hours.per.week	native.country	income
0	90	?	77053	HS-grad	9	Widowed	?	Not-in-family	White	Female	0	4356	40	United-States	<=50K
1	82	Private	132870	HS-grad	9	Widowed	Exec-managerial	Not-in-family	White	Female	0	4356	18	United-States	<=50K
2	66	?	186061	Some-college	10	Widowed	?	Unmarried	Black	Female	0	4356	40	United-States	<=50K
3	54	Private	140359	7th-8th	4	Divorced	Machine-op-inspct	Unmarried	White	Female	0	3900	40	United-States	<=50K
4	41	Private	264663	Some-college	10	Separated	Prof-specialty	Own-child	White	Female	0	3900	40	United-States	<=50K

- '?' = NaN
- Transformer en classe la colonne **Income**

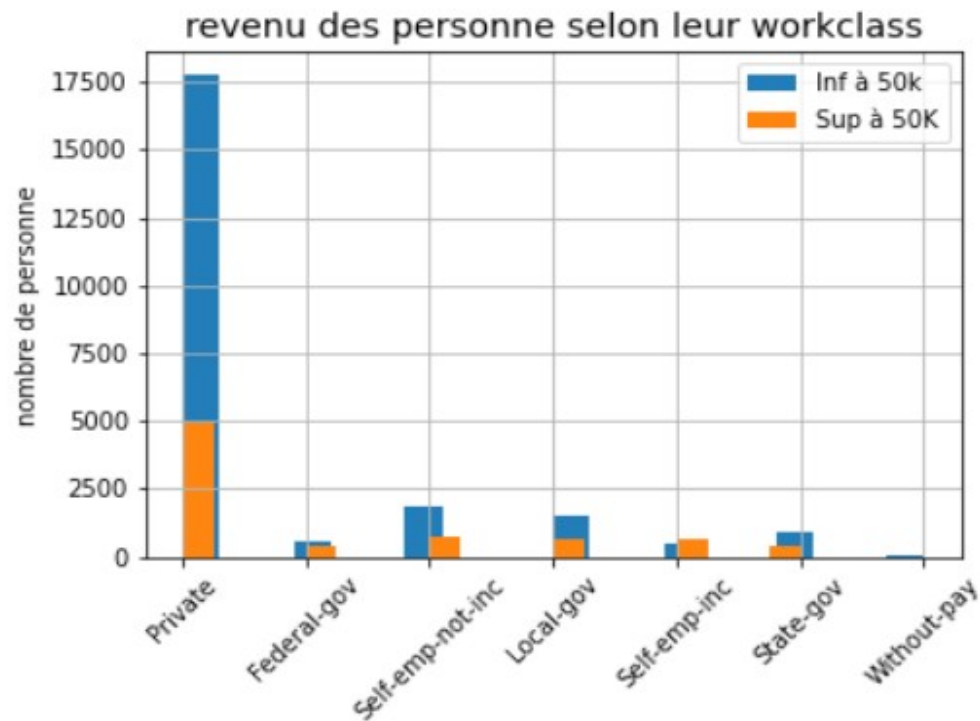
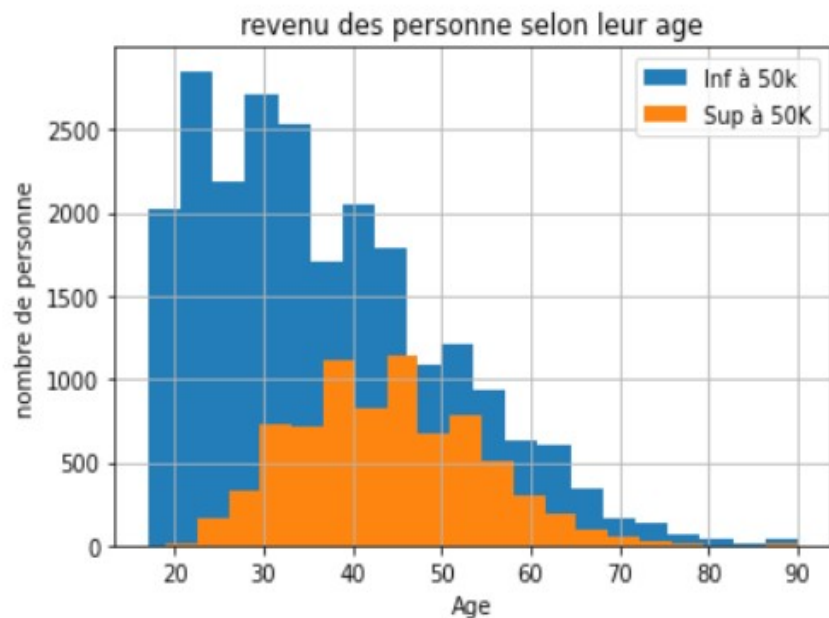
Cleaning

	age	workclass	fnlwgt	education	education.num	marital.status	occupation
8874	18	Never-worked	206359	10th	6	Never-married	NaN
13675	23	Never-worked	188535	7th-8th	4	Divorced	NaN
17089	17	Never-worked	237272	10th	6	Never-married	NaN
21934	18	Never-worked	157131	11th	7	Never-married	NaN

- Workclass : 1836 NaN
- Occupation : 18443 NaN
- Native.country : 583 NaN

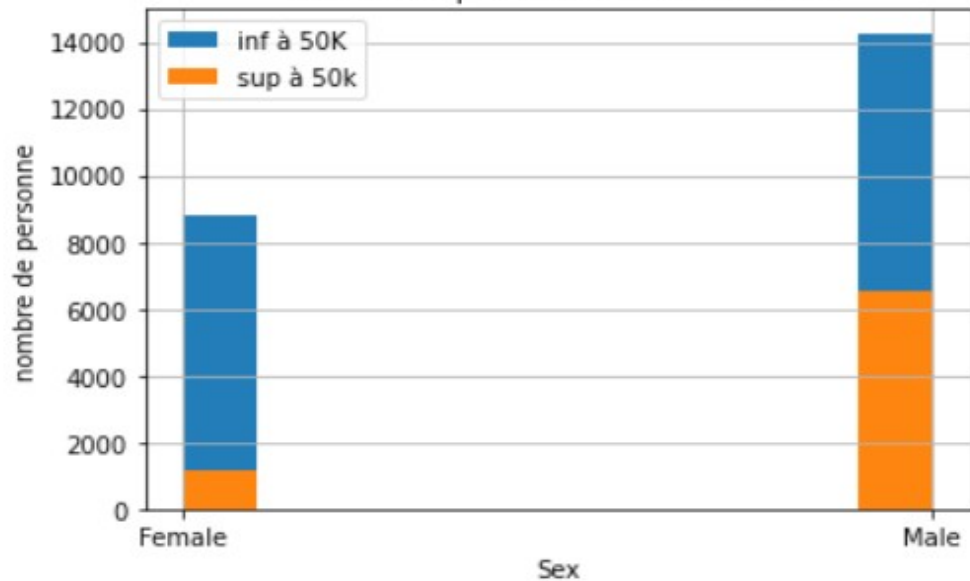


Exploratory data analysis

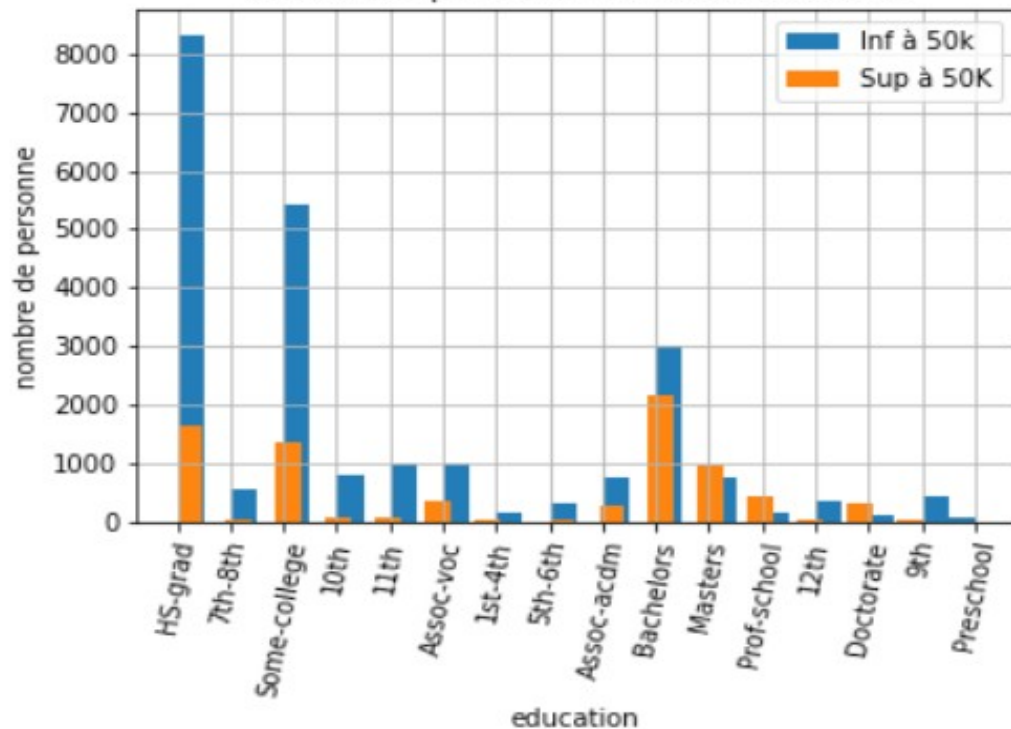


Exploratory data analysis

Revenu de personne selon leur sexe



revenu des personne selon leur éducation



Préprocessing

	age	workclass	fnlwgt	education.num	marital.status	occupation	relationship	race	sex	capital.gain	capital.loss	hours.per.week	income
1	82	5	132870	9	4	0	2	0	1	0	4356	18	0
3	54	5	140359	4	3	8	3	0	1	0	3900	40	0
4	41	5	264663	10	5	1	4	0	1	0	3900	40	0
5	34	5	216864	9	3	11	3	0	1	0	3770	45	0
6	38	5	150601	6	5	4	3	0	0	0	3770	40	0

Self-emp-inc (0) < ... < Private(5) < Without-pay(6)

education : 1 < 2 < ... < 15 < 16

...

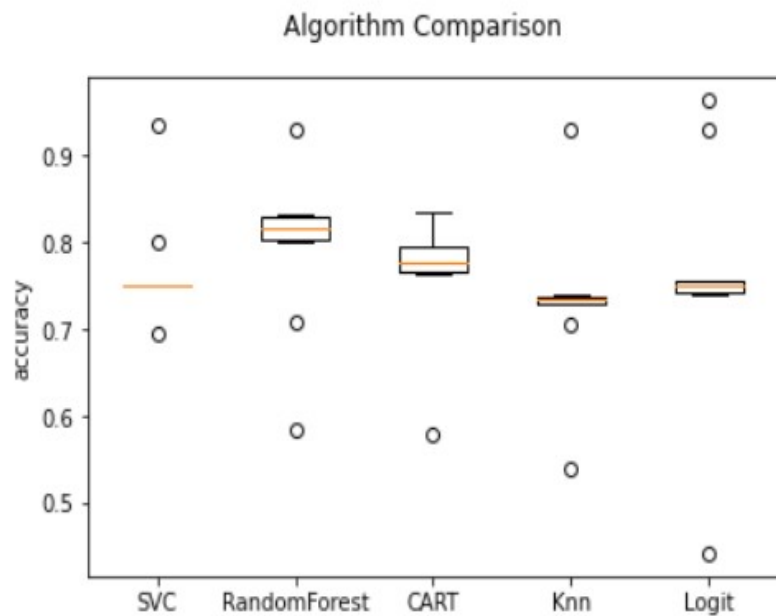
Préprocessing

- Relationship & Marital Status
- Occupation & education

	age	workclass	fnlwgt	education.num	marital.status	occupation	relationship	race	sex	capital.gain	capital.loss	hours.per.week	income
age	1	-0.205066	-0.0765397	0.0435671	-0.110281	-0.133124	-0.383336	-0.0303278	-0.0821167	0.0803922	0.0604089	0.101879	0.242431
workclass	-0.205066	1	0.0410379	-0.154787	0.0953975	0.148184	0.137665	0.0107877	0.0716317	-0.0697357	-0.0424757	-0.120952	-0.151066
fnlwgt	-0.0765397	0.0410379	1	-0.043509	0.0216077	0.0175113	0.0344762	0.089015	-0.0265686	-0.000188977	-0.0100105	-0.0228105	-0.00944568
education.num	0.0435671	-0.154787	-0.043509	1	-0.0987776	-0.435761	-0.14718	-0.0785989	-0.0074431	0.124247	0.0793208	0.151241	0.33464
marital.status	-0.110281	0.0953975	0.0216077	-0.0987776	1	0.139518	0.637997	0.13236	0.434936	-0.0703064	-0.0675779	-0.172369	-0.378875
occupation	-0.133124	0.148184	0.0175113	-0.435761	0.139518	1	0.203716	0.119004	0.0463892	-0.094863	-0.0758415	-0.167142	-0.301335
relationship	-0.383336	0.137665	0.0344762	-0.14718	0.637997	0.203716	1	0.120541	0.275989	-0.0828104	-0.0812867	-0.258945	-0.398294
race	-0.0303278	0.0107877	0.089015	-0.0785989	0.13236	0.119004	0.120541	1	0.112538	-0.018831	-0.0263973	-0.057256	-0.0975671
sex	-0.0821167	0.0716317	-0.0265686	-0.0074431	0.434936	0.0463892	0.275989	0.112538	1	-0.0477575	-0.0470098	-0.230321	-0.216626
capital.gain	0.0803922	-0.0697357	-0.000188977	0.124247	-0.0703064	-0.094863	-0.0828104	-0.018831	-0.0477575	1	-0.0323325	0.0795476	0.221871
capital.loss	0.0604089	-0.0424757	-0.0100105	0.0793208	-0.0675779	-0.0758415	-0.0812867	-0.0263973	-0.0470098	-0.0323325	1	0.0539609	0.151461
hours.per.week	0.101879	-0.120952	-0.0228105	0.151241	-0.172369	-0.167142	-0.258945	-0.057256	-0.230321	0.0795476	0.0539609	1	0.228547
income	0.242431	-0.151066	-0.00944568	0.33464	-0.378875	-0.301335	-0.398294	-0.0975671	-0.216626	0.221871	0.151461	0.228547	1

Model

Kfold = 10



Kfold = 5

