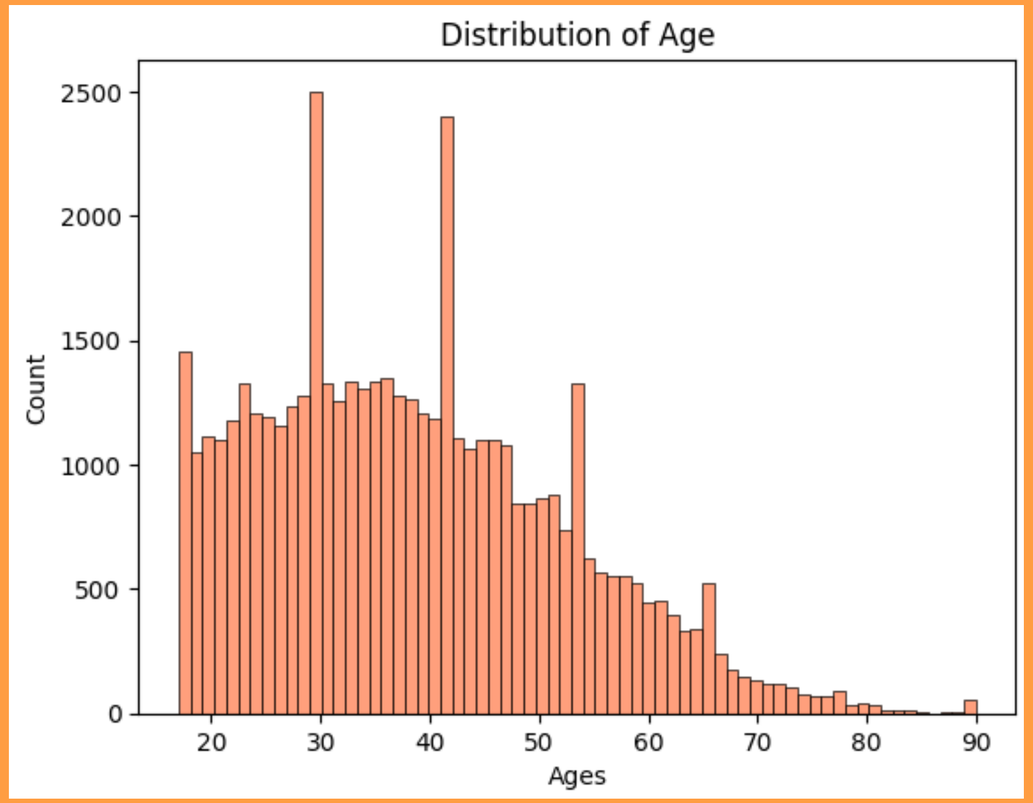
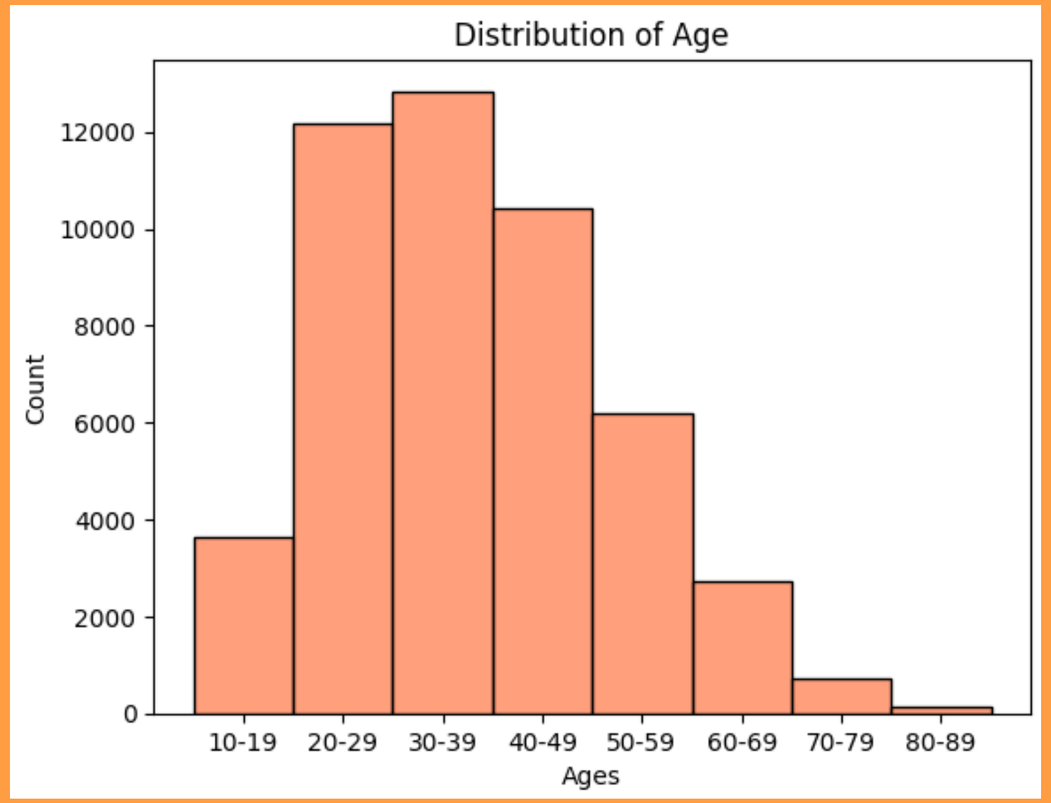
MIDTERM SKILLS EXAM: Data Wrangling and Analysis

CENSUS INCOME DATASET

LIAM CALAMBA III CPE22S3 - CPE 311

AGE





This plot shows the distribution of age of the sample.

the plot shows that the average age of the sample is starting around late 20's to early 40s.

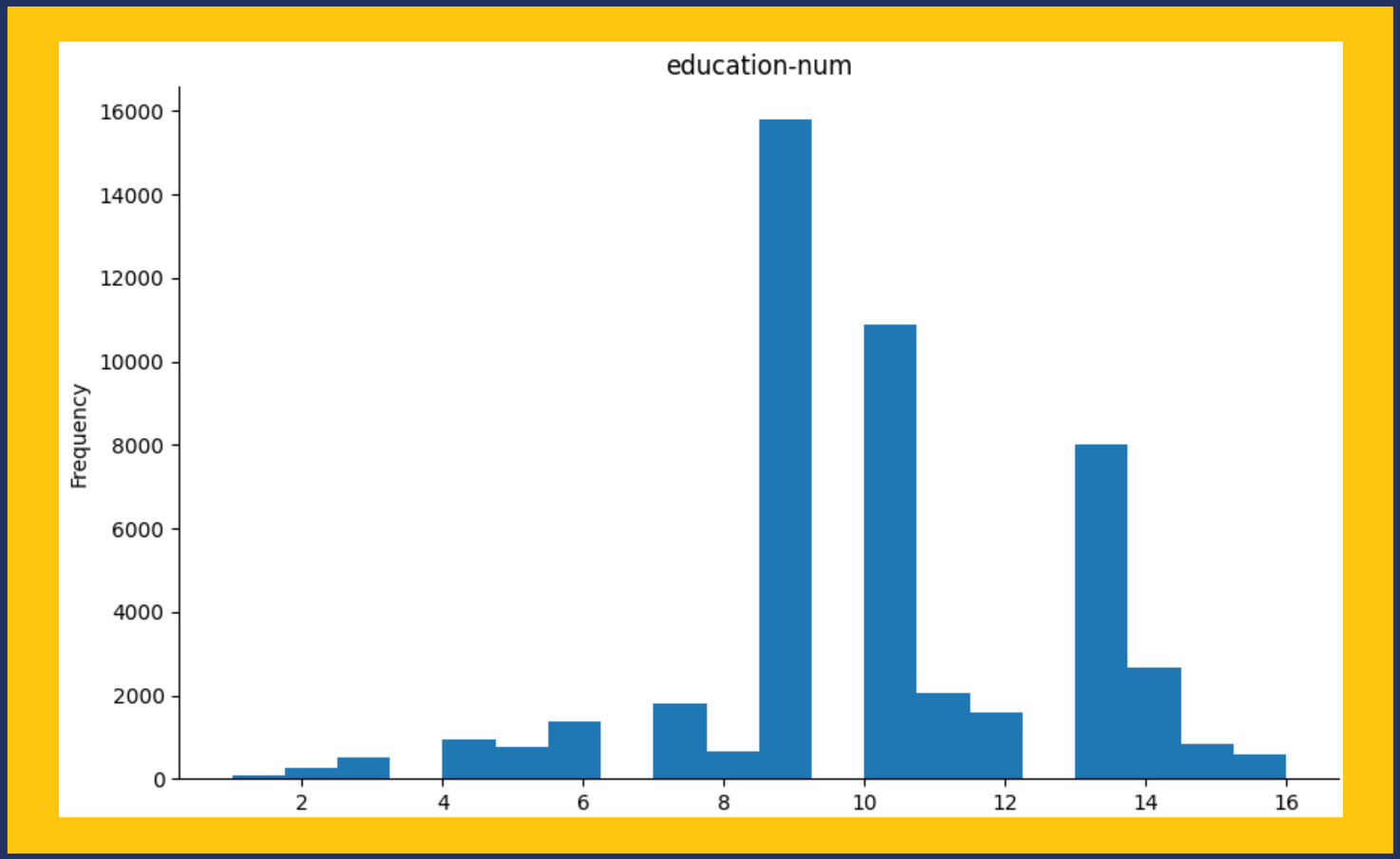
This plot shows the education attained by the sample.

the plot shows that on average the education attained by the sample is around HS Graduate to Some-Collage and there are also noticeably some who pursued to get their Bachelors.

(for reference)
education and their corresponding education number :

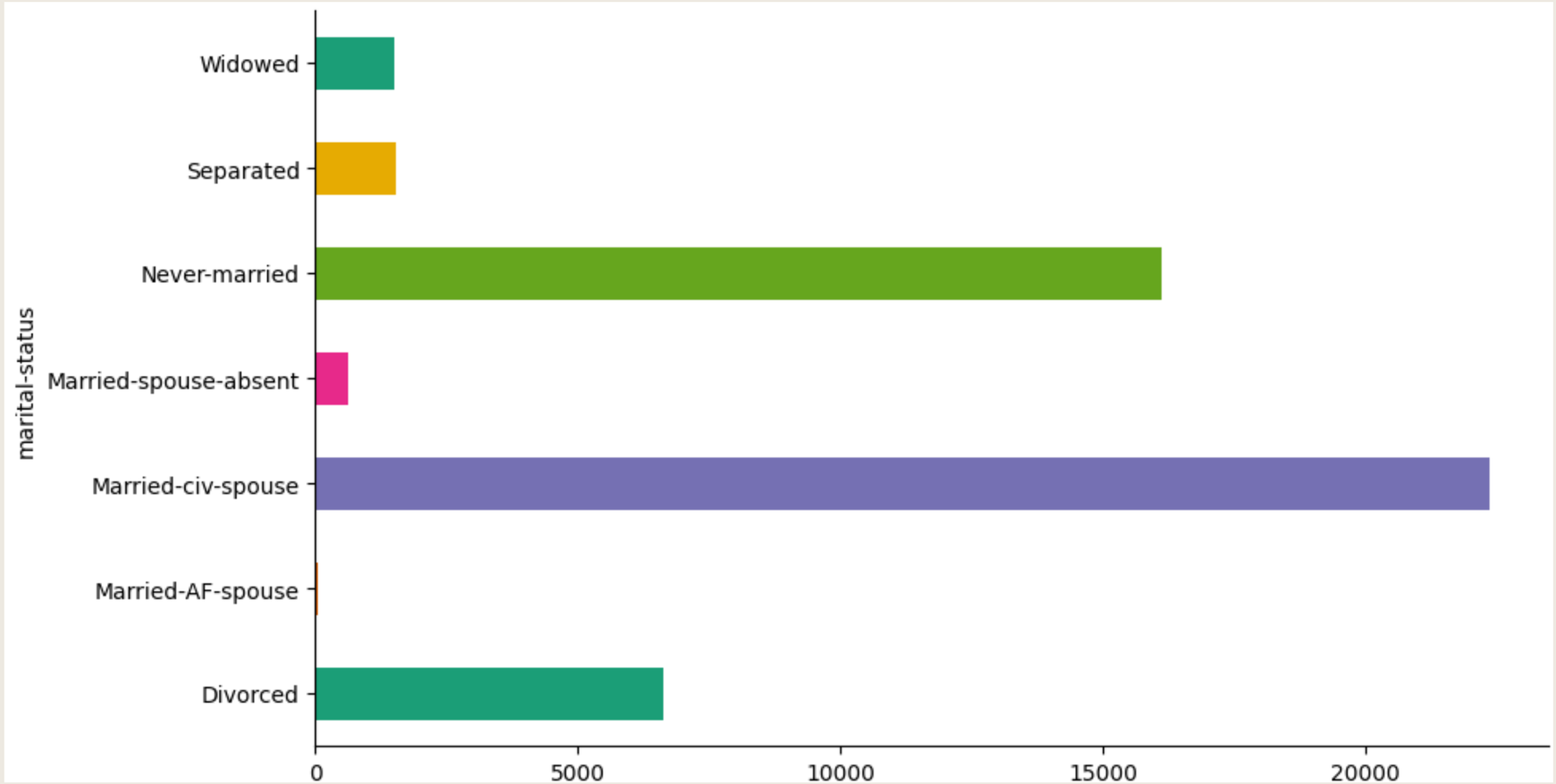
- Preschool → 1
- 1st-4th → 2
- 5th-6th → 3
- 7th-8th → 4
- 9th → 5
- 10th → 6
- 11th → 7
- 12th → 8
- HS-grad → 9
- Some-college → 10
- Assoc-voc → 11
- Assoc-acdm → 12
- Bachelors → 13
- Masters → 14
- Prof-school → 15
- Doctorate → 16

EDUCATION



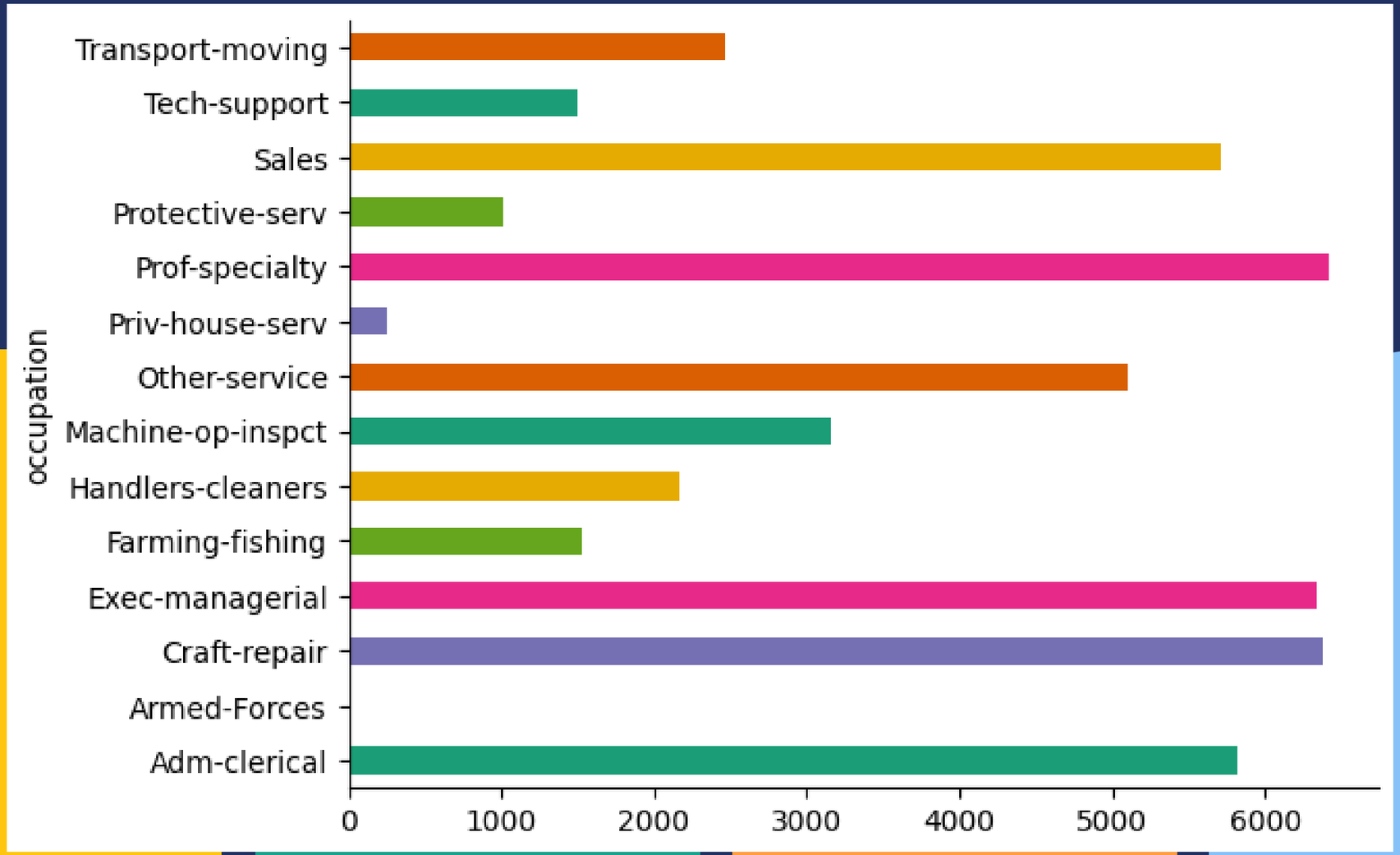
MARITAL STATUS

This plot shows the marital status of the sample.
the plot shows that the most common marital statuses in the sample are Married to a civilian spouse, followed by Never married, and lastly Divorced, with some of the sample either Widowed, Separated, or Married with absent spouse, and with a very few of the sample that is Married to a Spouse that is associated with the Armed Forces.



OCCUPATION

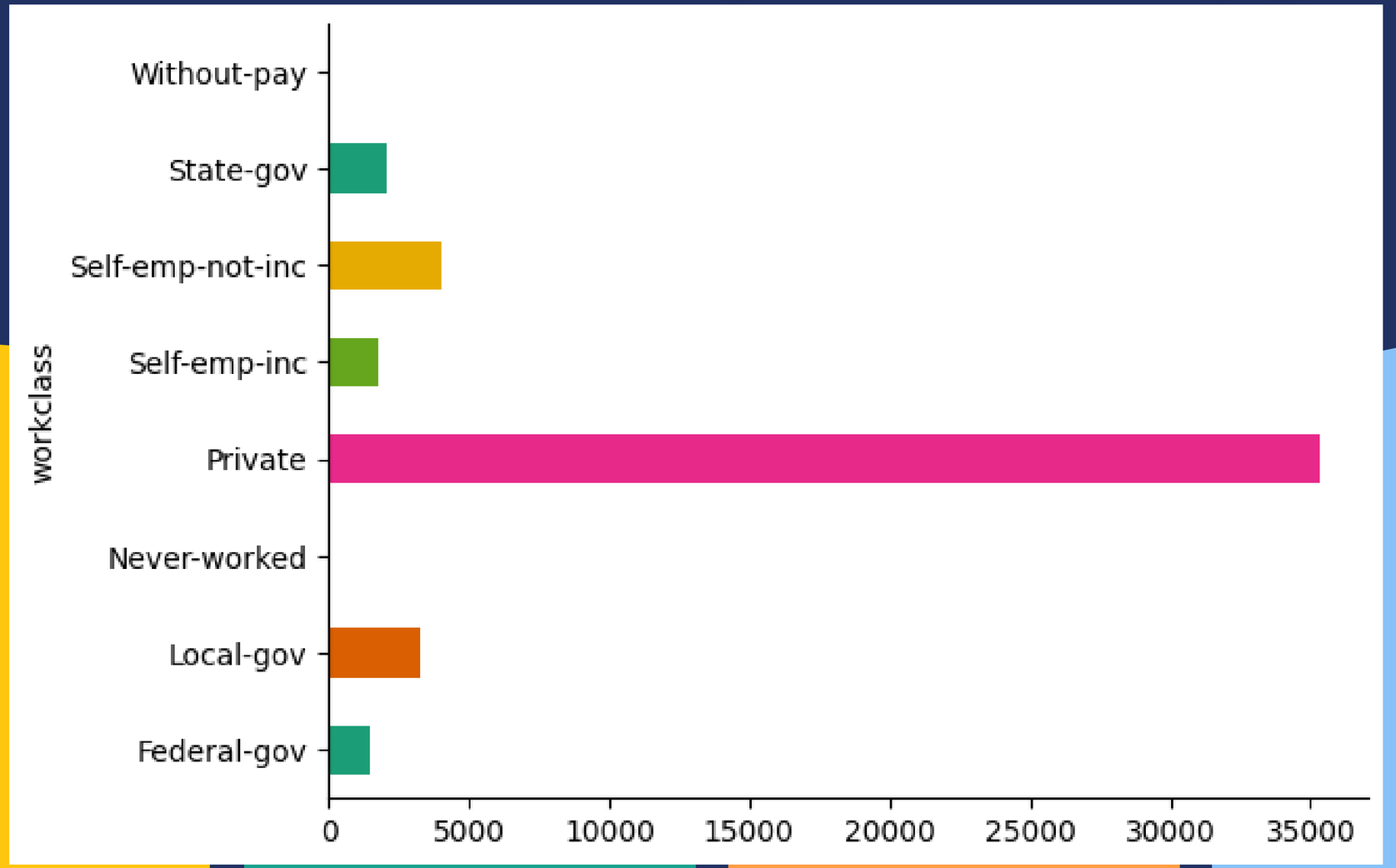
This plot shows the occupations of the sample.



WORK CLASS

This plot shows the work class of the sample.

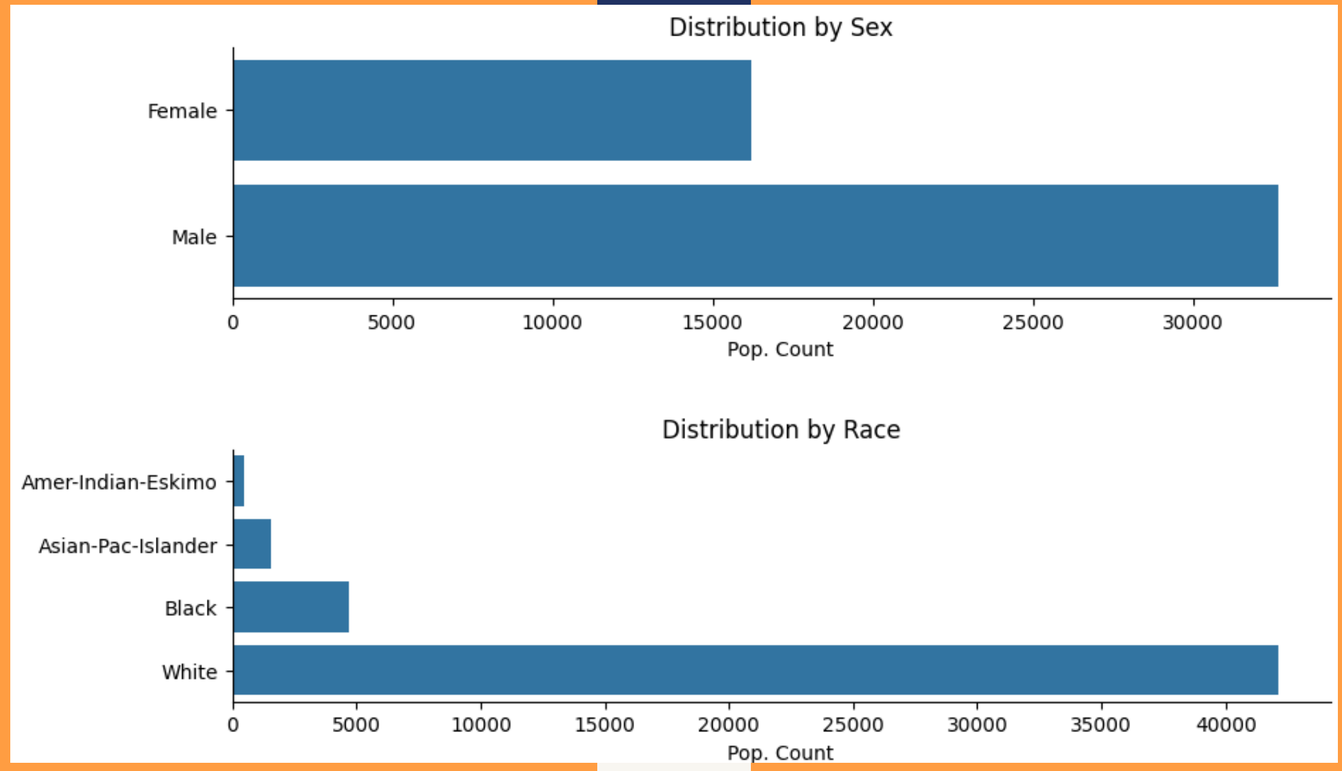
the plot shows that a lot of them works in a private company or such, while there are a few who works at the government or is self employed.



SEX & RACE

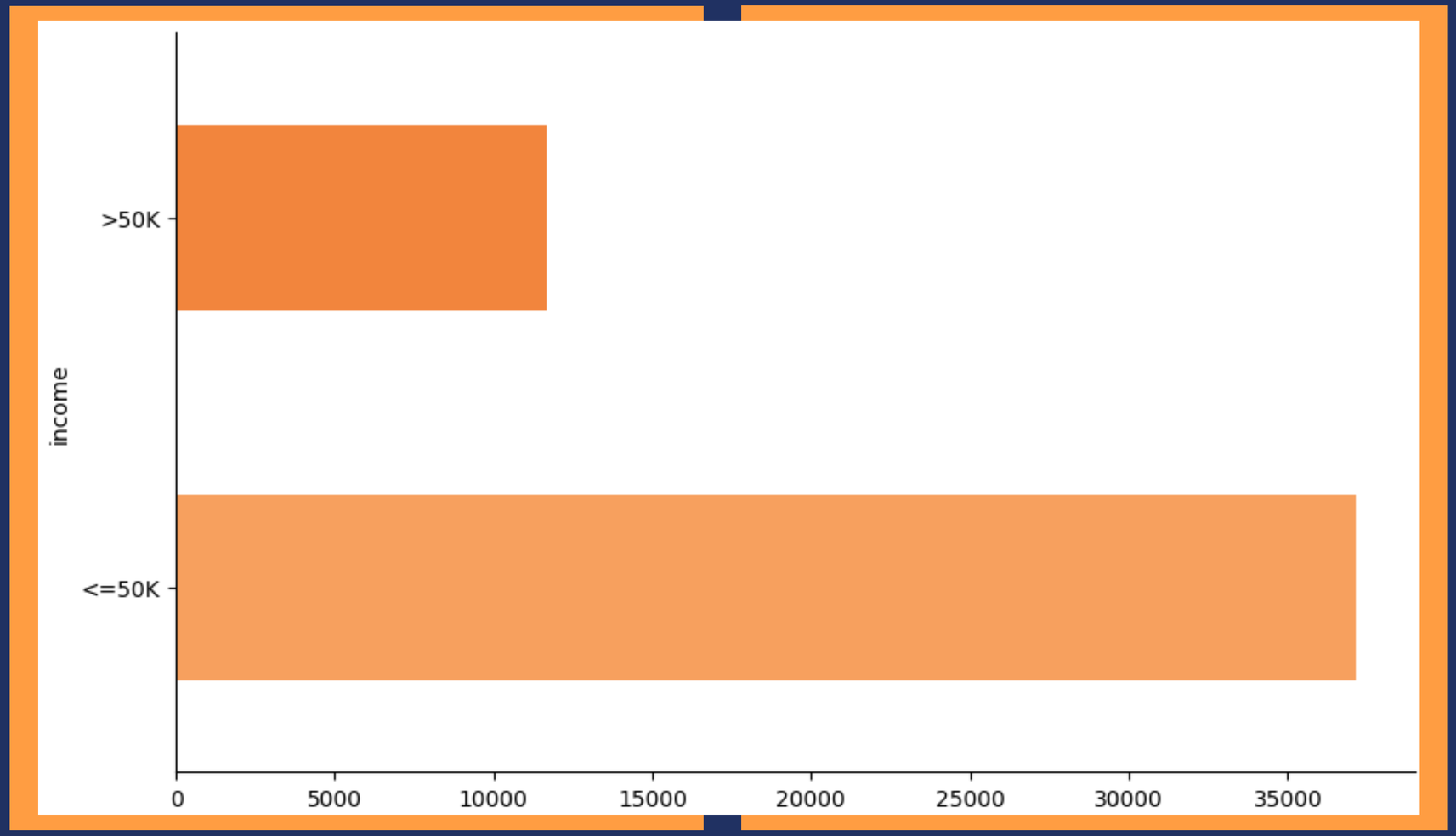
The following plots shows the Sex and Race distribution of the sample.

the top plot shows that more than 15000 identifies as Females representing one portion of the sample and double that size identifies as Males which is more than 35000 representing the other porion of the sample.



the bottom plot shows that even though the Population where the sample is taken is in the US, there are some other races that works in the US, they are possibly immigrants that seeks better opportunity abroad.

INCOME



The plot shows the income per year of the sample.

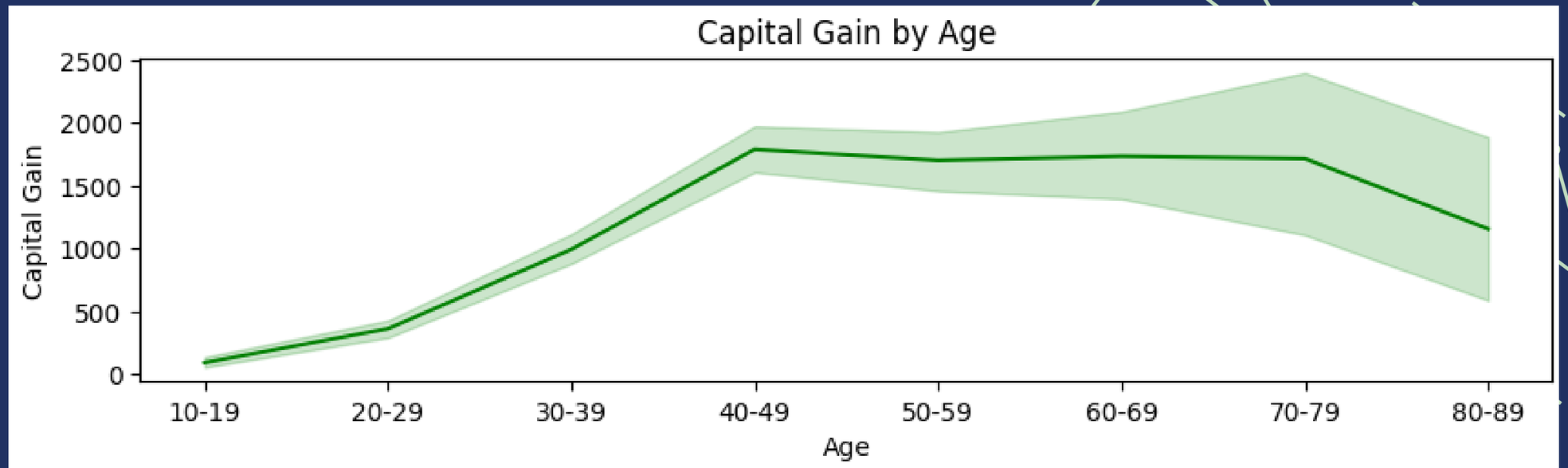
the plot shows that majority of the sample earns an income of less than 50K per year, and the others earn more than 50K per year

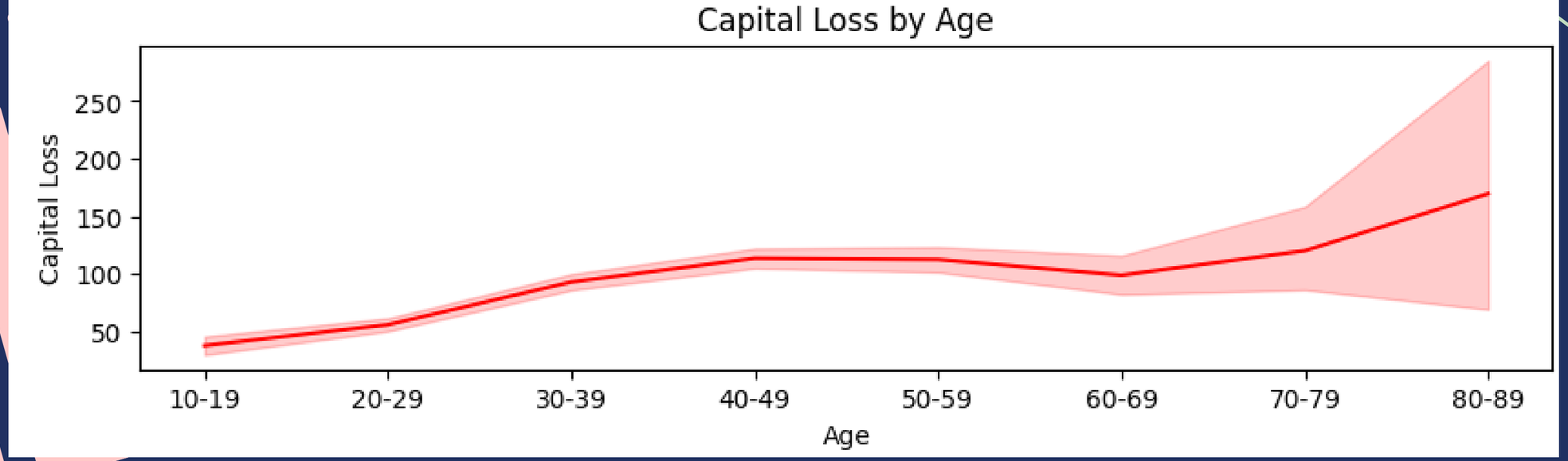
CAPITAL GAIN/LOSS AND AGE

- the graphs shows the trend of capital gain and loss according to their age, the 2 are complementary; the capital gain/loss increases as the individual grows older.

- the data sugggest that the sample tend to accumulate more capital gains (maybe from investment profits) as they grow older, specifically at around their 30s and peaks at their 40s, then it will stabilize and slowly decrease, this is probably because of having more time in their field therefore having higher income and better risk tolerance.

- capital loss seem to be noticably lower overall than capital gains, but they can always occur.

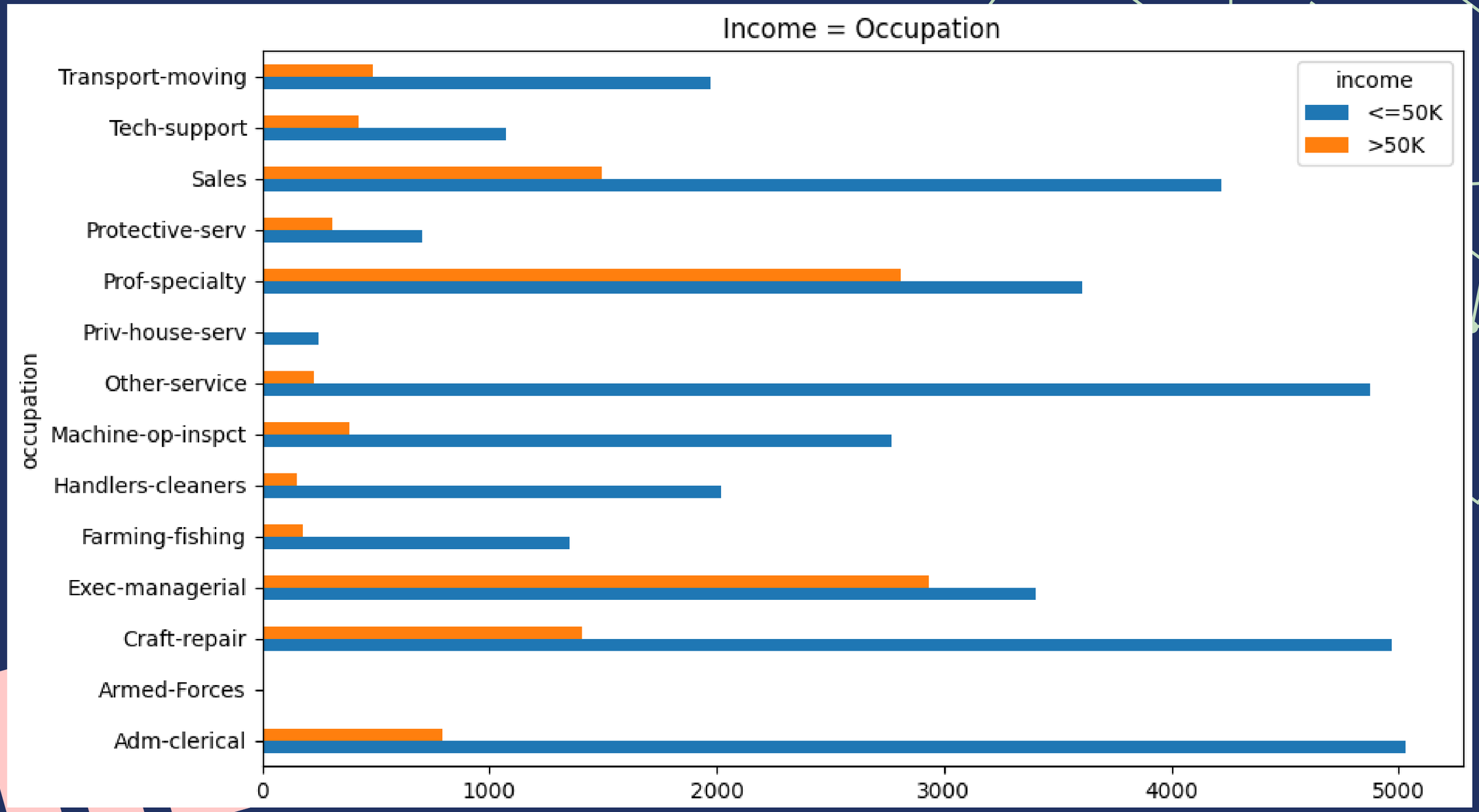

Capital Gain by Age
Capital Gain
2500
2000
1500
1000
500
0
10-19
20-29
30-39
40-49
50-59
60-69
70-79
80-89
Age


Capital Loss by Age
Capital Loss
250
200
150
100
50
10-19
20-29
30-39
40-49
50-59
60-69
70-79
80-89
Age

IS INCOME TIED TO THE OCCUPATION OF AN INDIVIDUALLOSS AND AGE

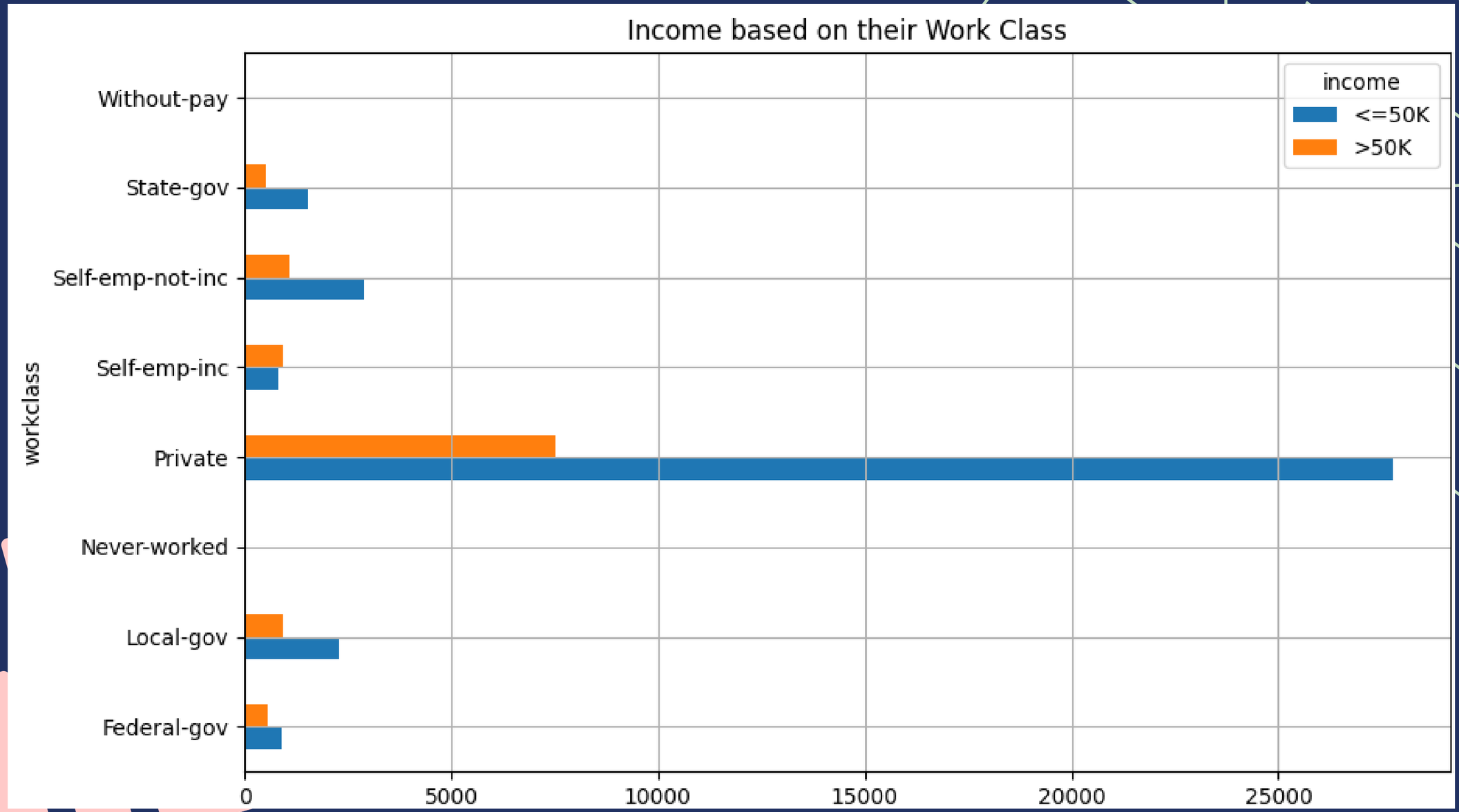
- the graph says a lot about it, with only 2 occupations that is close to each other when it comes to income (Exec-Manigerial & Prof-specialty), the rest shows that a lot of the individuals earns below 50K in these occupations and only some earns above 50K in that same occupations.

- the ones that earns above 50K in the jobs that the majority earns below 50K may be the ones that are the seasoned professionals, the ones that are in that job the longest and have a lot of experience hence the greater income.


Income = Occupation
income
<=50K
>50K
occupation
Transport-moving
Tech-support
Sales
Protective-serv
Prof-specialty
Priv-house-serv
Other-service
Machine-op-inspct
Handlers-cleaners
Farming-fishing
Exec-managerial
Craft-repair
Armed-Forces
Adm-clerical
0
1000
2000
3000
4000
5000

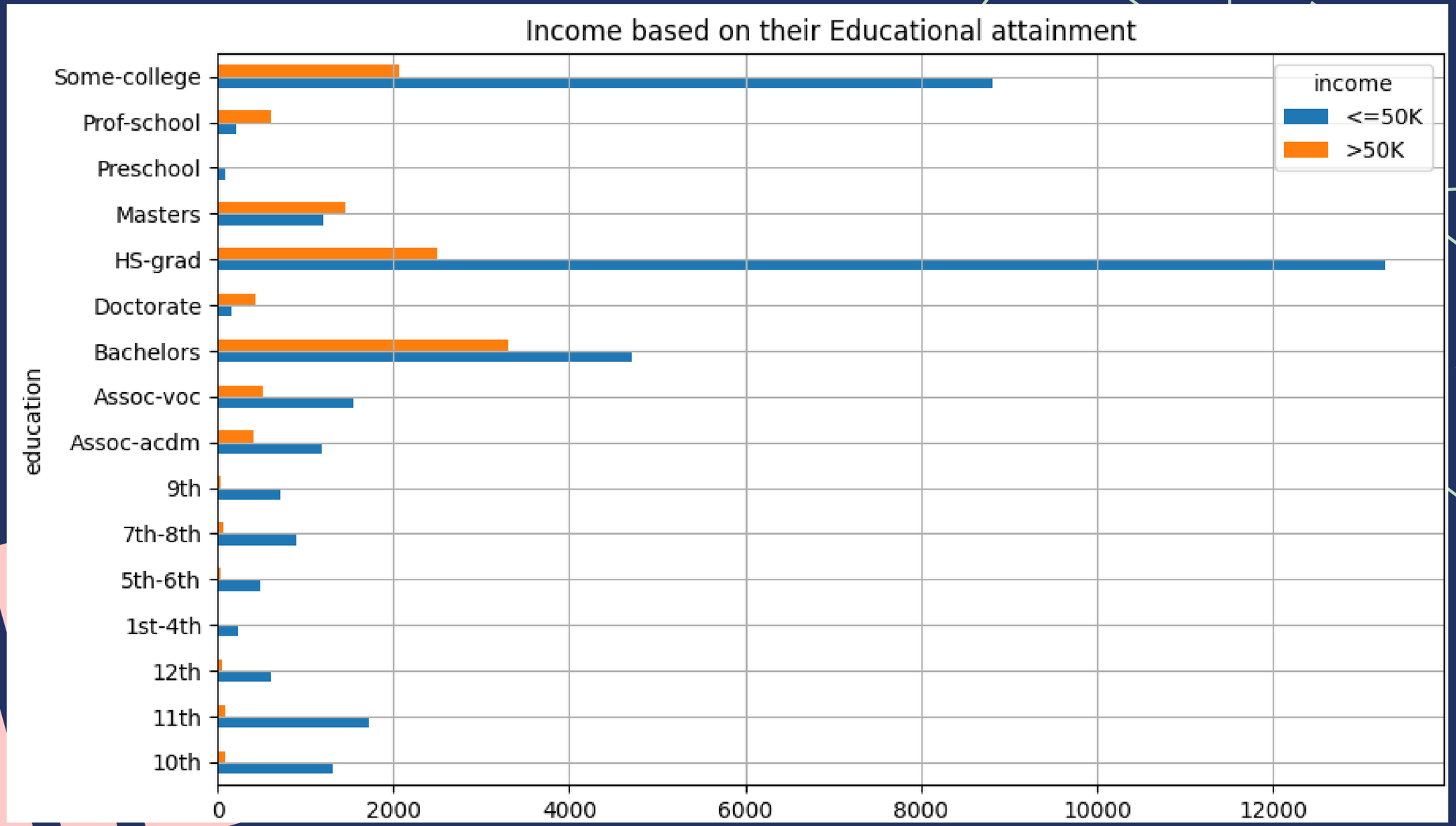
INCOME BASED ON THEIR WORK CLASS

- we can ignore without pay here because it's self explanatory.
- this shows the incomes based on Work Class, working in private seems to not have a benefit, looking at the other classes it's close to each other, there's a possibility to earn more than what they are currently earning.


Income based on their Work Class
income
<=50K
>50K
workclass
Without-pay
State-gov
Self-emp-not-inc
Self-emp-inc
Private
Never-worked
Local-gov
Federal-gov
0
5000
10000
15000
20000
25000

INCOME BASED ON THEIR EDUCATIONAL ATTAINMENT

- this shows that the ones who finished higher education tend to have greater income


Income based on their Educational attainment
education
Some-college
Prof-school
Preschool
Masters
HS-grad
Doctorate
Bachelors
Assoc-voc
Assoc-acdm
9th
7th-8th
5th-6th
1st-4th
12th
11th
10th
0
2000
4000
6000
8000
10000
12000
income
<=50K
>50K