

US Census bureau Adults income: Machine Learning

Title: Census Income

Members:

Eldan Abdrashim
Milka Wafula
Pablo dela Fuente

Introduction

The objective of this project is to determine people income,based on age and number of hours.

What question is your data answering?

From the dataset for 1994 US Census bureau database,Does age matter when it comes to incomes

Group's hypothesis?

Do older people working on average number of hours in a week earn more than young people

What was your processes for evaluating the data?

- Excel, IPython/Pandas
- Tableau

Data cleaning

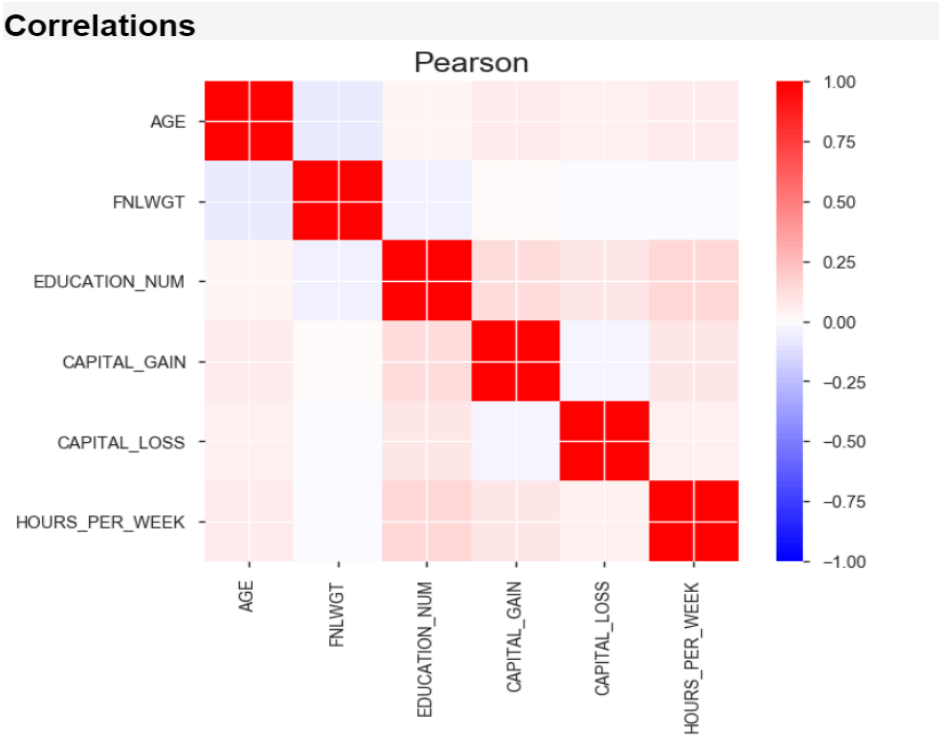
- Replacing null values and removing outliers

Data analysis

- Grouping data and correlation



Correlation



We noted

- The relationship between age and hours per week is negative linear relationship
- Strong negative linear relationship

Data head structure

```
data.head()
```

	AGE	WORKCLASS	FNLWGT	EDUCATION	EDUCATION_NUM	MARITAL_STATUS	OCCUPATION	RELATIONSHIP	RAC
0	39	State-gov	77516	Bachelors	13	Never-married	Adm-clerical	Not-in-family	Whi
1	50	Self-emp-not-inc	83311	Bachelors	13	Married-civ-spouse	Exec-managerial	Husband	Whi
2	38	Private	215646	HS-grad	9	Divorced	Handlers-cleaners	Not-in-family	Whi
3	53	Private	234721	11th	7	Married-civ-spouse	Handlers-cleaners	Husband	Blac
4	28	Private	338409	Bachelors	13	Married-civ-spouse	Prof-specialty	Wife	Blac



Overview

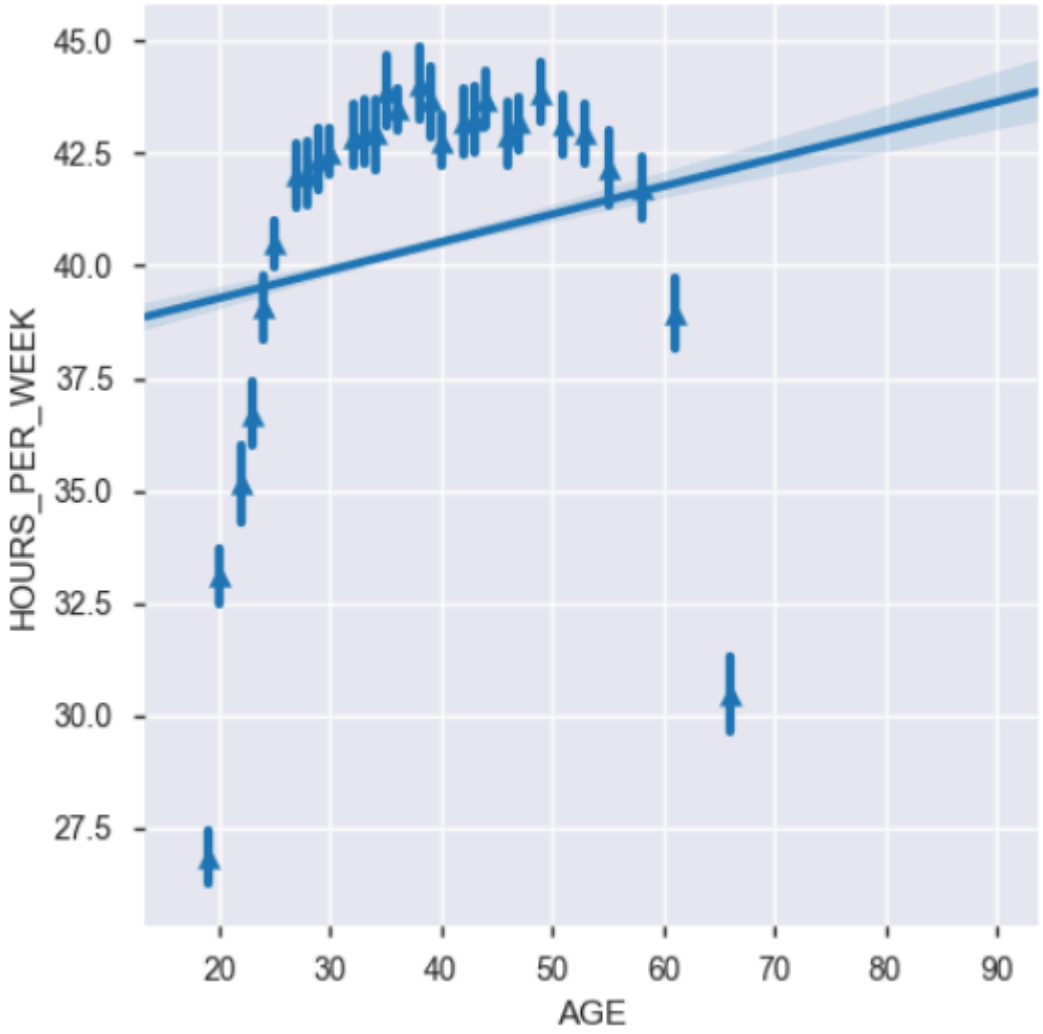
Dataset info

Number of variables	15
Number of observations	32561
Total Missing (%)	0.0%
Total size in memory	3.7 MiB
Average record size in memory	120.0 B

Variables types

Numeric	6
Categorical	9
Boolean	0
Date	0
Text (Unique)	0
Rejected	0
Unsupported	0

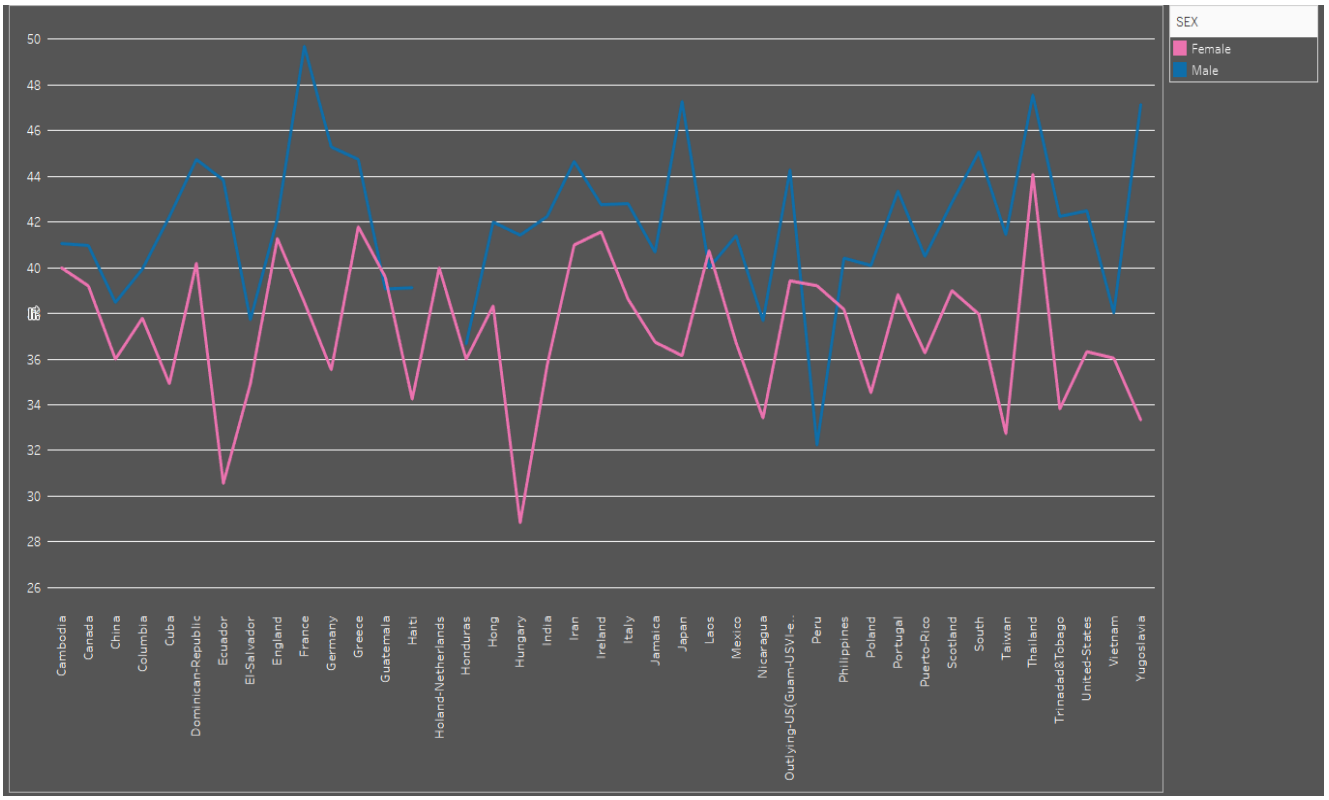




- People between age 25-55 work for a maximum of 45 hours per week and 33 hours per week minimum



Average Hours Per Week - GENDER

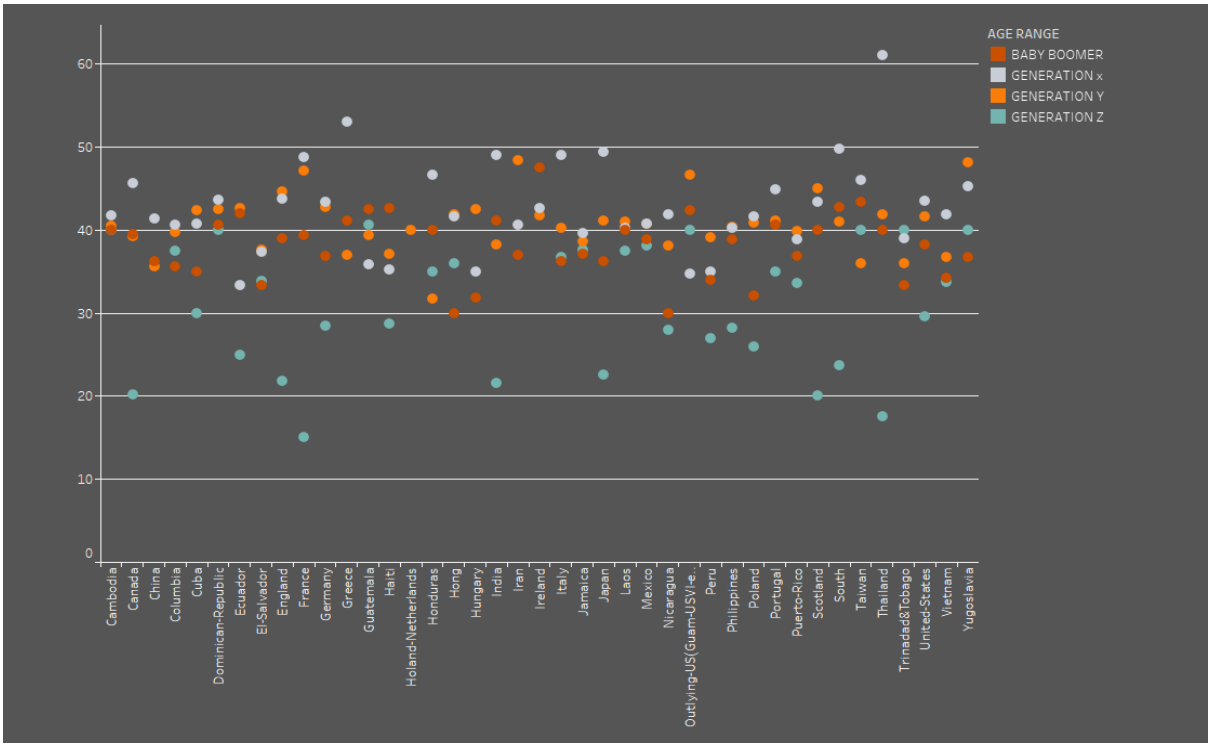


We noted:

- Men work more than women
- Women in Peru work more hours than men



AGE PROFILE: Average Hours Per Week



We noted:

- There is higher polynomiaL linear relationship where generation z working below 30 hours per week
- Generation age x y work more than 30 hours per week



AGE RANGE: Education by Race Profile



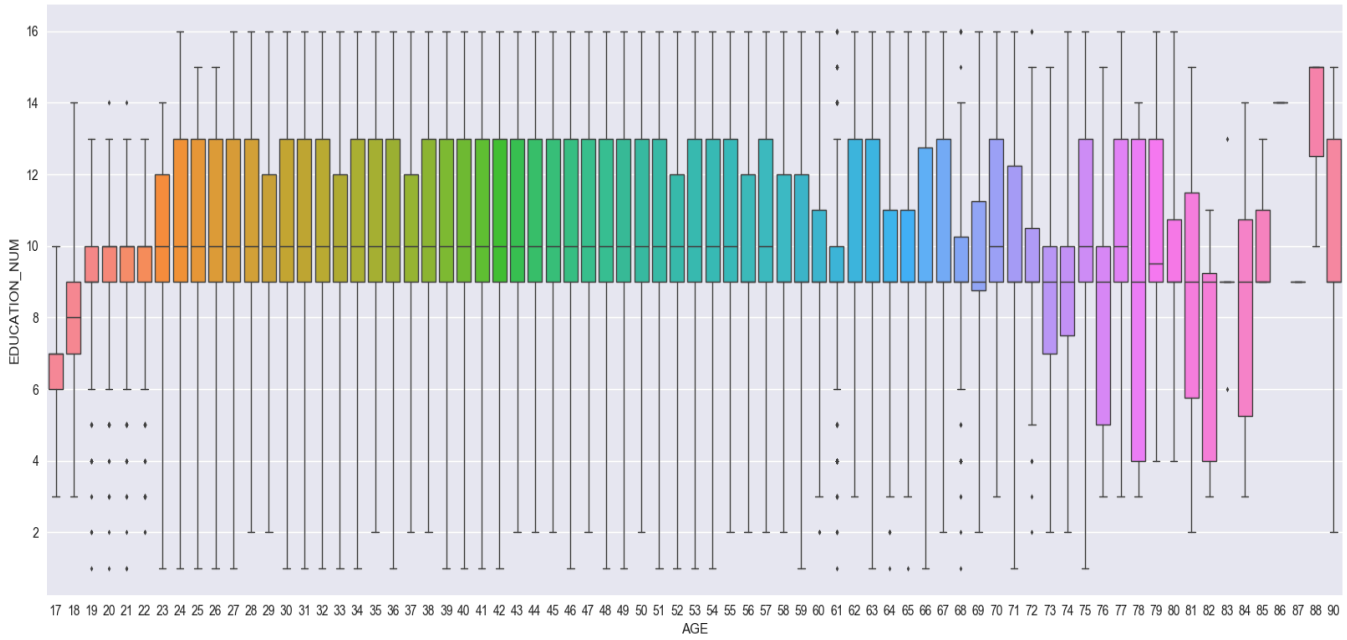
We noted;

- Black,white and Asia generation z
  have the same bachelors probability rate.

- Blacks have the highest probability rate in Doctorate level on all generation



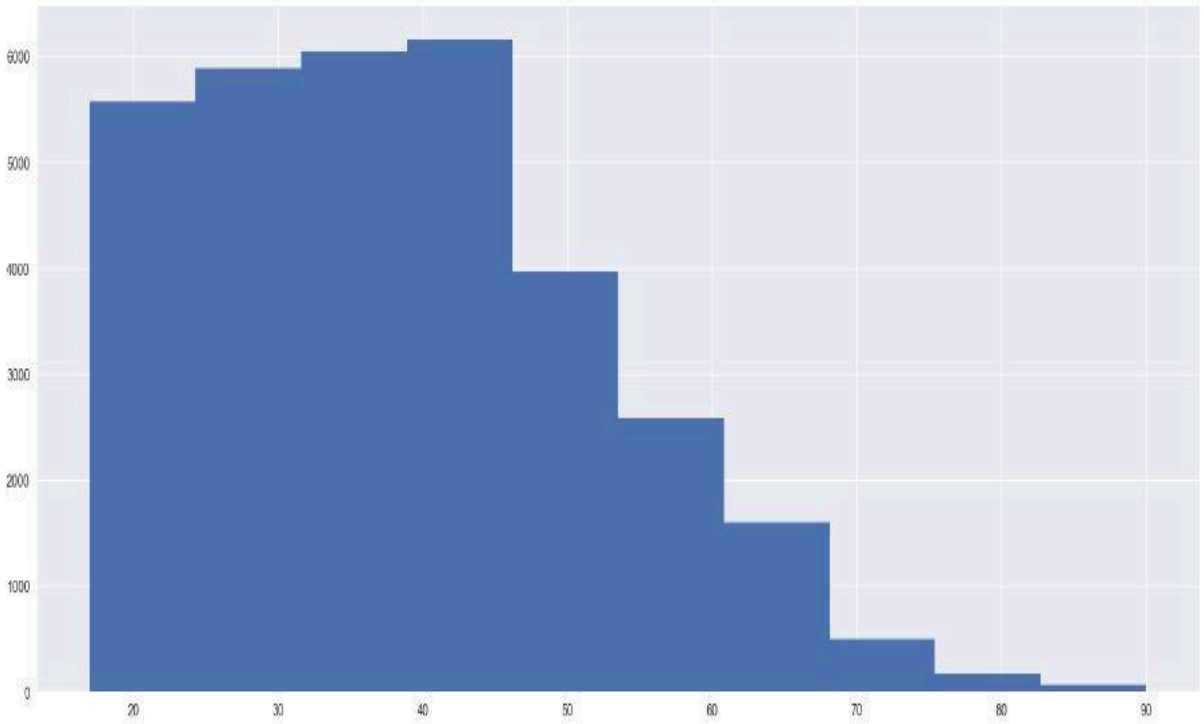
Education by age



We noted:

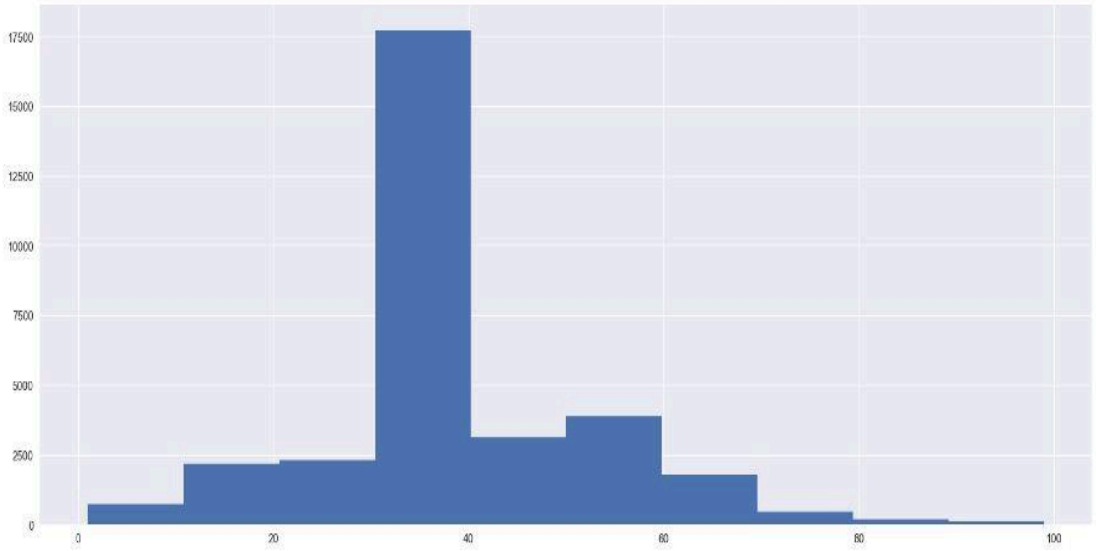
- Age between 19-73 had a higher education of collage school to doctorate






```
In [29]: plt.hist(data.HOURS_PER_WEEK)
```

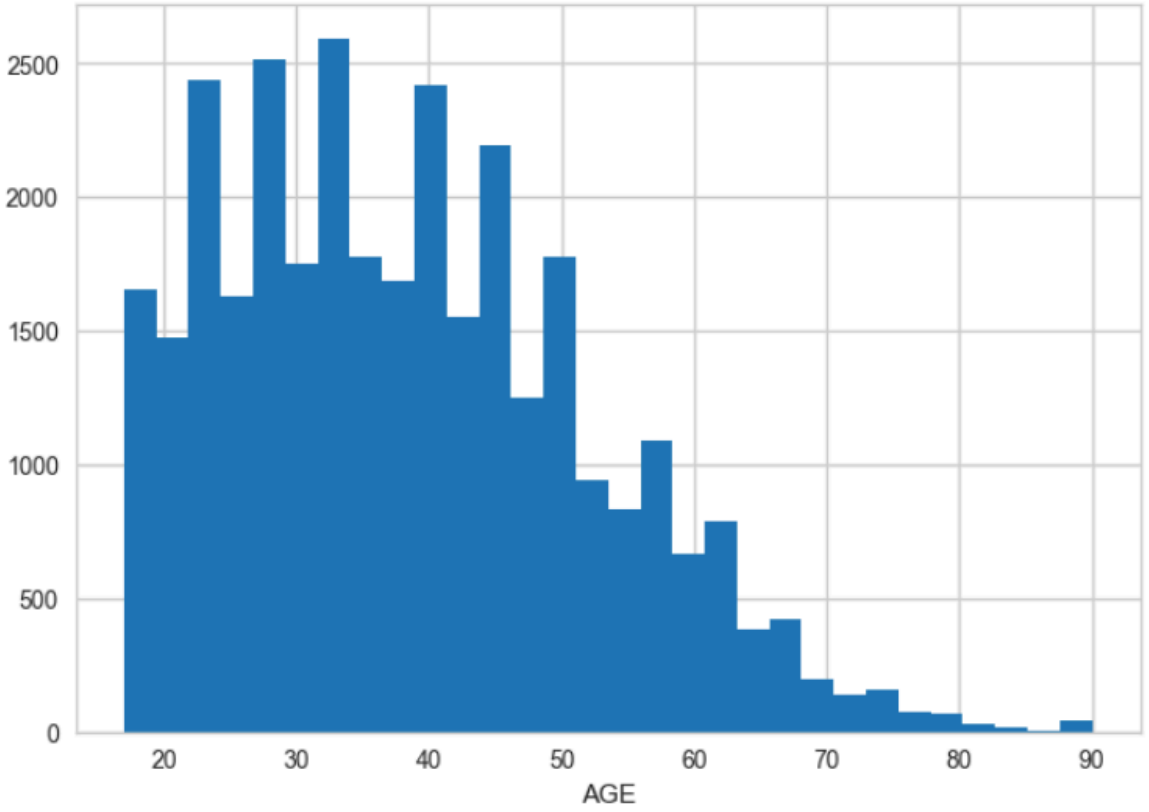
```
Out[29]: (array([  736.,  2192.,  2317., 17735.,  3119.,  3877.,  1796.,   448.,
                   202.,   139.]),
          array([ 1. , 10.8, 20.6, 30.4, 40.2, 50. , 59.8, 69.6, 79.4, 89.2, 99. ]),
          <a list of 10 Patch objects>)
```





We noted;

- Many people work for 30-40 hours per week.
- Less people work for more than 60 hours per week.



Conclusion

- The negative correlation linear relationship indicates older people with higher education, work for more hours per week.
- People age between 25-50 working on average number of hours in a week earn more than young people below age 25.
- Blacks have the highest rate in Doctorate level on all generation
- There is higher probability of generation z working below 30 hours per week



- Generation age x, y work more than 30 hours per week

Resource

1. Ronny Kohavi and Barry Becker ,UC Irvine Machine Learning Repository; 16/07/2018 ;https://archive.ics.uci.edu/ml/datasets/Adult

Acknowledgements:

we would like to thank the entire Udacity and scholarship team for making this possible!