

## Topic

Analysis of Socioeconomic Factors Influencing Income in the U.S.

## Data

1. Y:

Income ( $\leq 50K$  - 0,  $> 50K$  - 1)(category)

2. X:

Age(continue),

Workclass(category),

Education(category),

Education-num(continue),

Marital-status(category),

Occupation(category),

Gender(category),

Race(category),

Hours-per-week(continue)

(We can delete a few var)

<https://www.kaggle.com/datasets/mastmustu/income>

## Data Cleaning

1. Delete NULL

## Descriptive Statistics

1. (Mean, Median, SD, Min, Max, IQR)

2. Density:

(Age, Working hours) VS Income(0/1), (X = Independent ) (Y = dependent)

3. Boxplot:

(Age, Working hours) VS Income(0/1), (X = Independent ) (Y = dependent)

4. Bar:

(Gender, Race, Education) VS Income(0/1) (X = independent) (Col = dependent)

## **Analysis**

### 1. Chi-square:

- (1) Gender VS Income
- (2) Marital-status VS Income
- (3) Education VS Income

### 2. T-test/ Z-test

- (1) Age VS Income
- (2) Hours-per-week VS Income
- (3) Education-num VS Income

### 3. ANOVA

- (1) Race VS Income
- (2) Workclass VS Income
- (3) Educaiton VS Income

(We can delete a few tests)