

Visualization

Definition:

- 1) **Correlation Matrix** : Correlation is a statistical technique that shows how two variables are related. [Pandas dataframe.corr\(\)](#) method is used for creating the correlation matrix. It is used to find the pairwise correlation of all columns in the dataframe. Any null values are automatically excluded. For any non-numeric data type columns in the dataframe it is ignored.

Returns: DataFrame : Correlation matrix.

- 2) **Vertical bar plot** : A bar plot is a plot that presents categorical data with rectangular bars with lengths proportional to the values that they represent. A bar plot shows comparisons among discrete categories. One axis of the plot shows the specific categories being compared, and the other axis represents a measured value.

Returns : matplotlib.axes.Axes or np.ndarray of them :

An ndarray is returned with one matplotlib.axes.Axes per column when subplots=True.

3) **A box plot** : is a method for graphically depicting groups of numerical data through their quartiles. The box extends from the Q1 to Q3 quartile values of the data, with a line at the median (Q2). The whiskers extend from the edges of box to show the range of the data. By default, they extend no more than $1.5 * \text{IQR}$ ($\text{IQR} = \text{Q3} - \text{Q1}$) from the edges of the box, ending at the farthest data point within that interval. Outliers are plotted as separate dots.

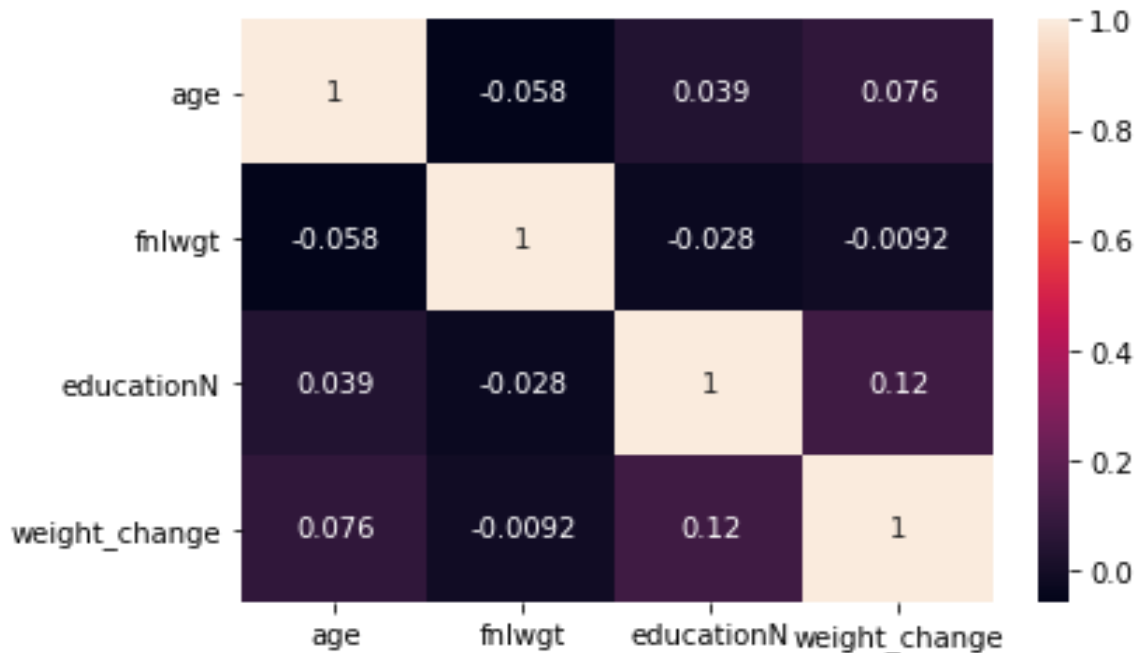
return_type : {'axes', 'dict', 'both'} or None, default 'axes'
The kind of object to return. The default is axes.

- 'axes' returns the matplotlib axes the boxplot is drawn on.
- 'dict' returns a dictionary whose values are the matplotlib Lines of the boxplot.
- 'both' returns a namedtuple with the axes and dict.
- when grouping with by, a Series mapping columns to return_type is returned.

If return_type is None, a NumPy array of axes with the same shape as layout is returned.

Returns: result of data.

1)Correlation Matrix:

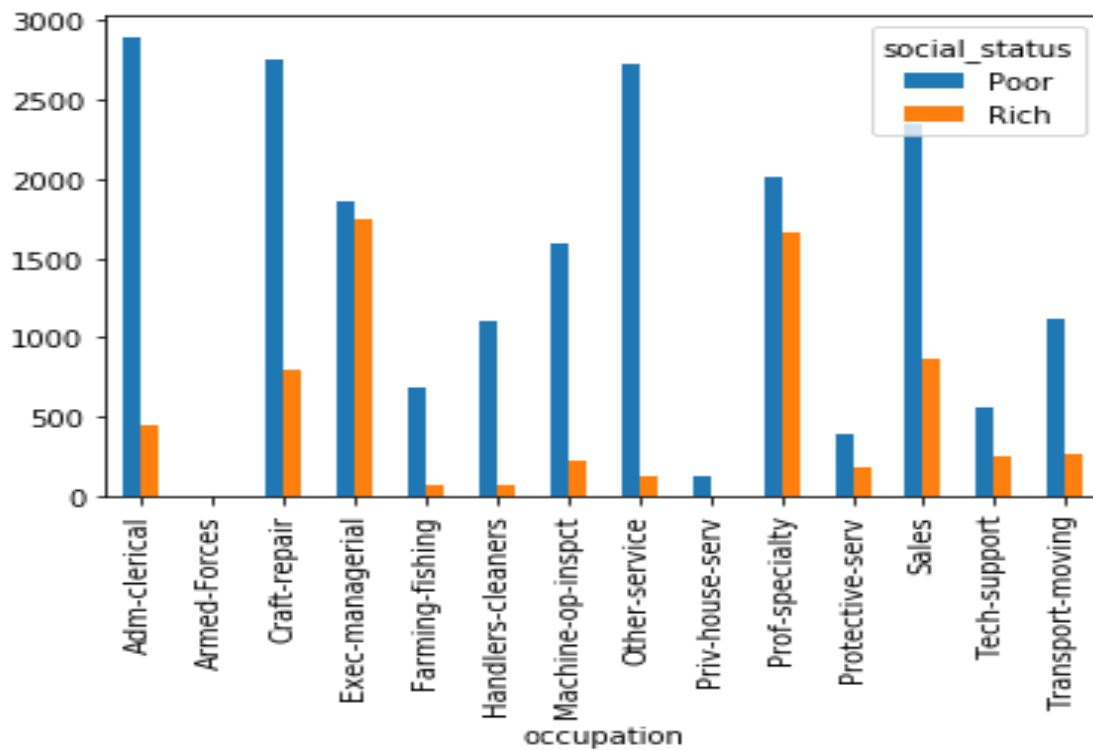


- We can note the following:

-White means that the correlation between the data is strong.

-everything that is away from white means that the data correlation is decreasing and with the black color it means that the data link does not exist.

2)Relation between occupation and social status :

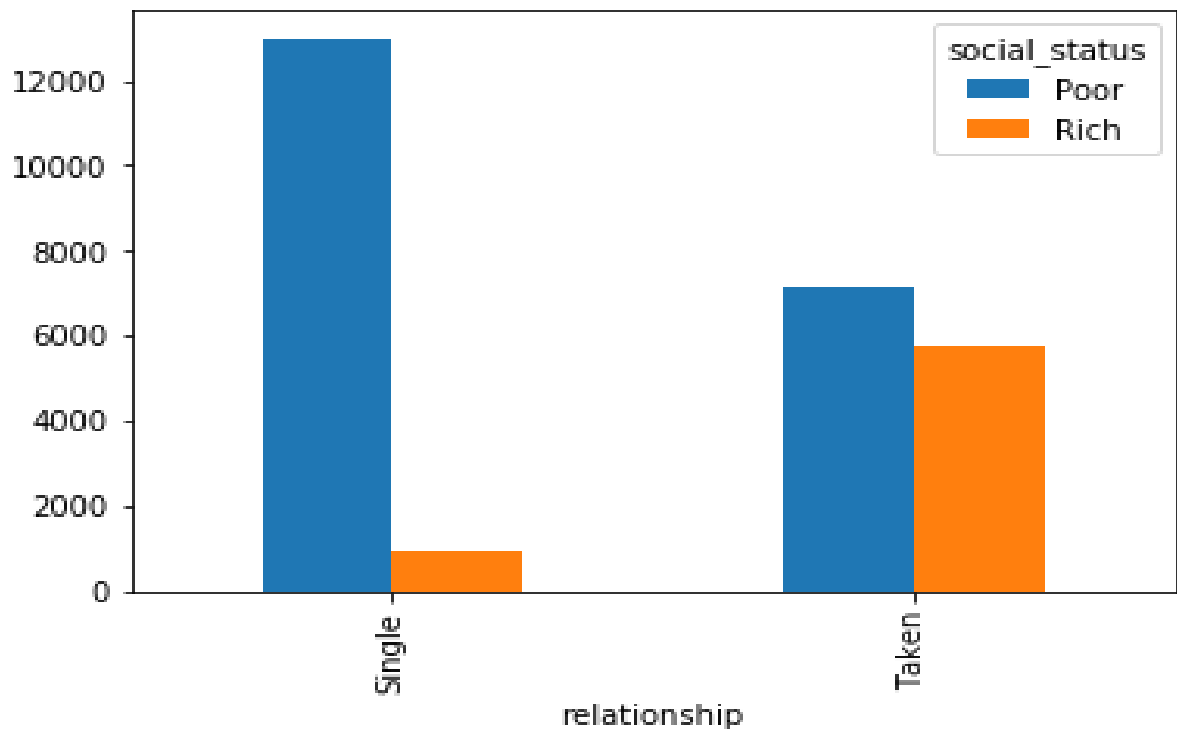


- We can note the following:

-That's show us the relation between occupation and social status.

- 1) The percentage of the poor is greater than that of the rich in all occupation.
- 2) The largest proportion of the poor in "Adm-clerical".
- 3) Largest percentage of the rich in "executives-managers".
- 4) The smallest proportion of the poor in " Armed-Forces".
- 5)The smallest proportion of the rich in "Armed-Forces".

3)Relation between relationship and social_status :



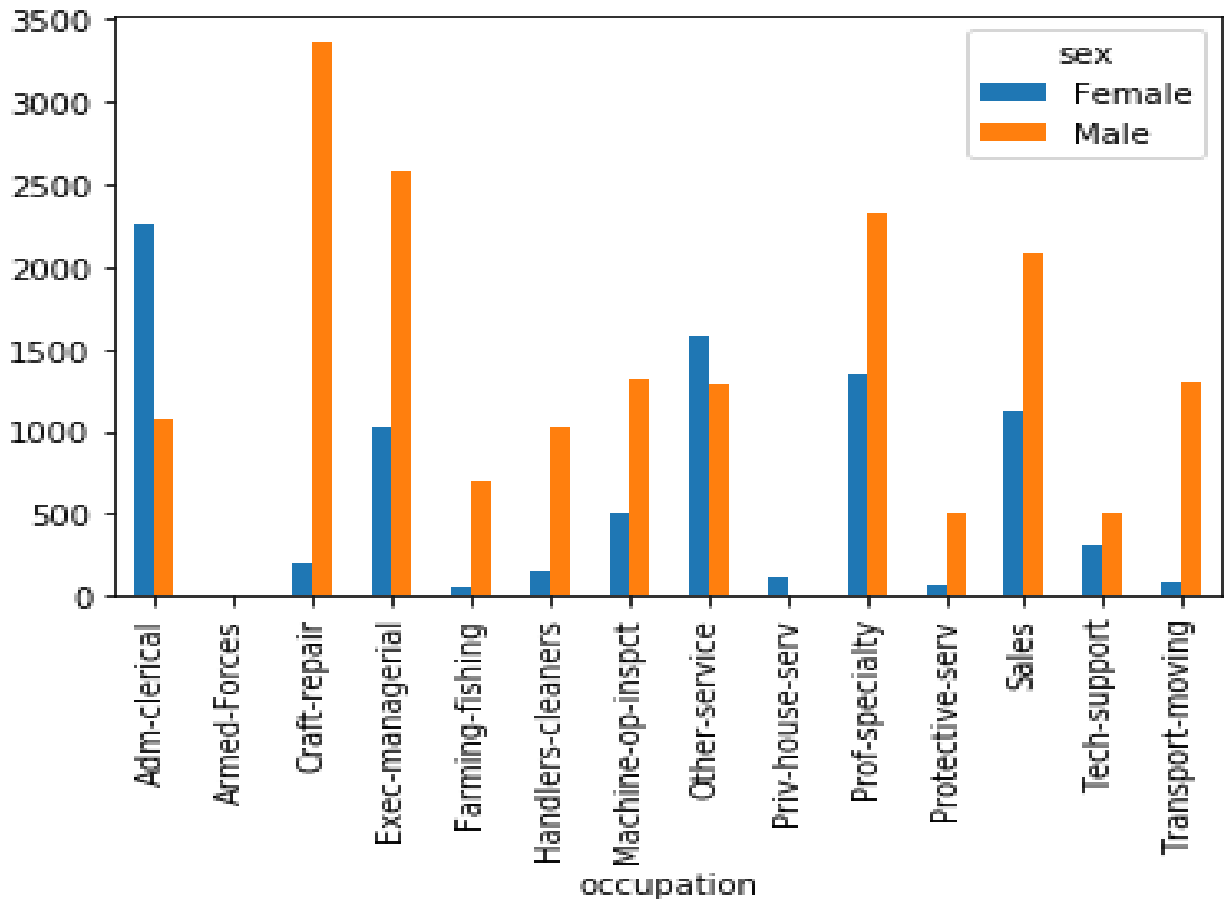
- We can note the following:

-That's show us the relation between relationship and social status.

1) That in "single" the percentage of the poor people more than rich people.

2) That in "Taken" the percentage of the poor people more than rich people.

4)Relation between occupation and sex :

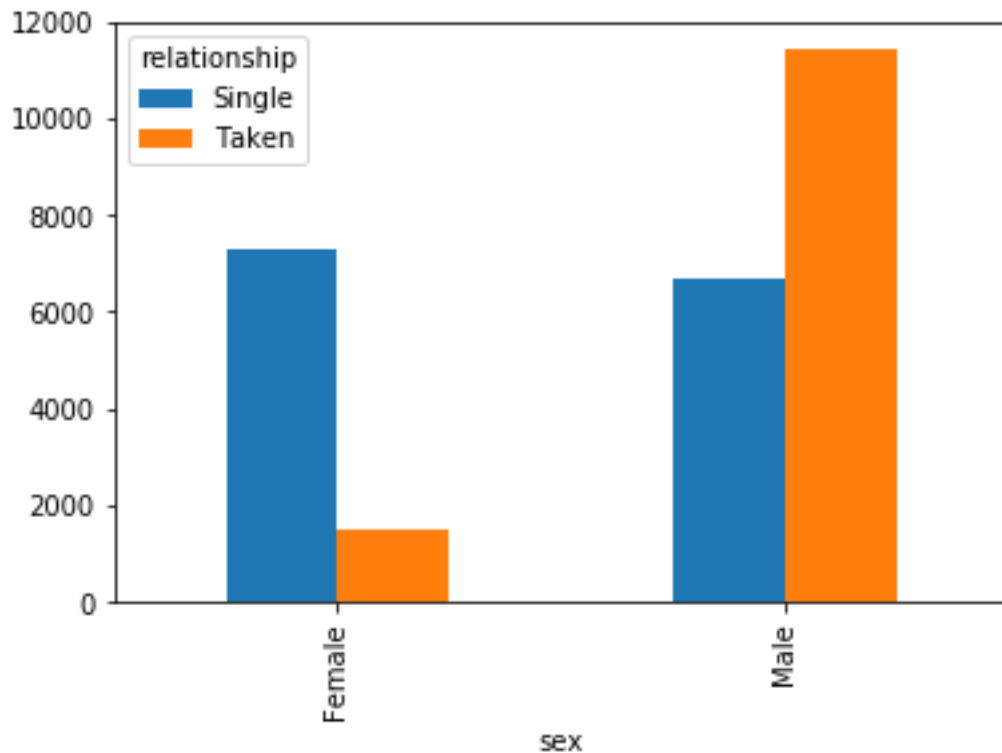


-We can note the following:

-That's show us the relation between occupation and sex.

- 1) The largest proportion of the Male in " Craft-repair ".
- 2) The largest proportion of the Female in " Adm-clerical ".
- 3) The smallest proportion of the Male in " Armed-Forces".
- 4)The smallest proportion of the Female in "Armed-Forces".

5)Relation between sex and relationship :



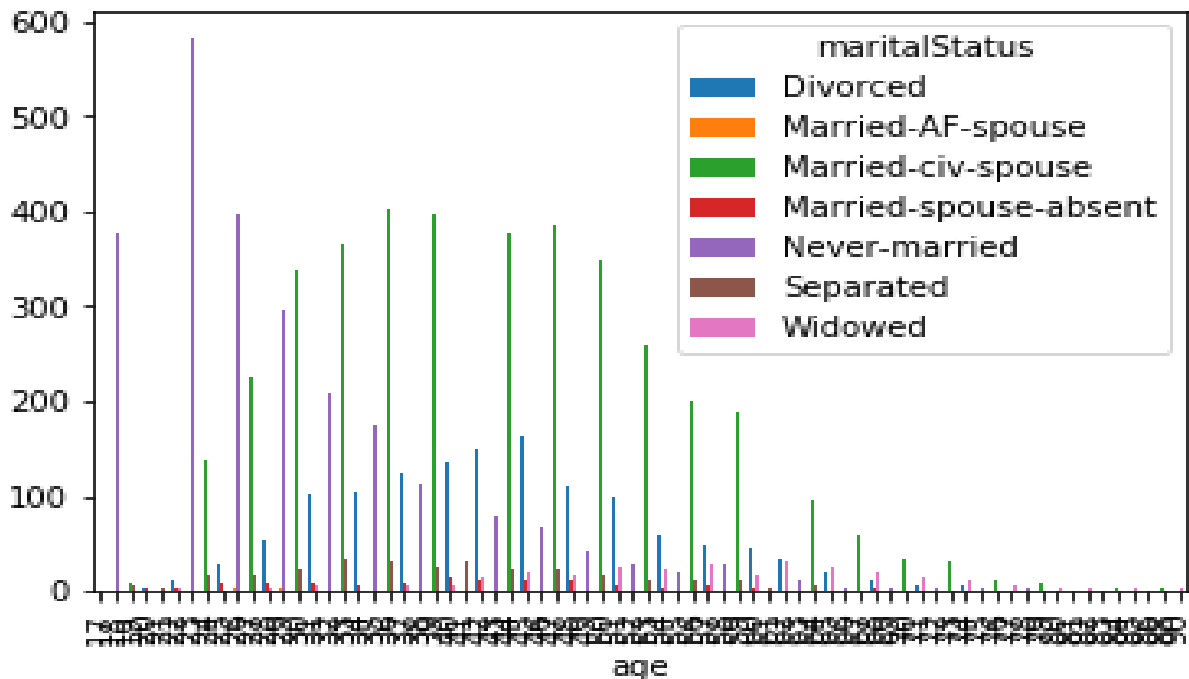
- We can note the following:

-That's show us the relation between sex and relationship.

1) The largest proportion of female are not in relationship.

2) The largest proportion of male are in relationship.

6)Relation between age and maritalStatus :



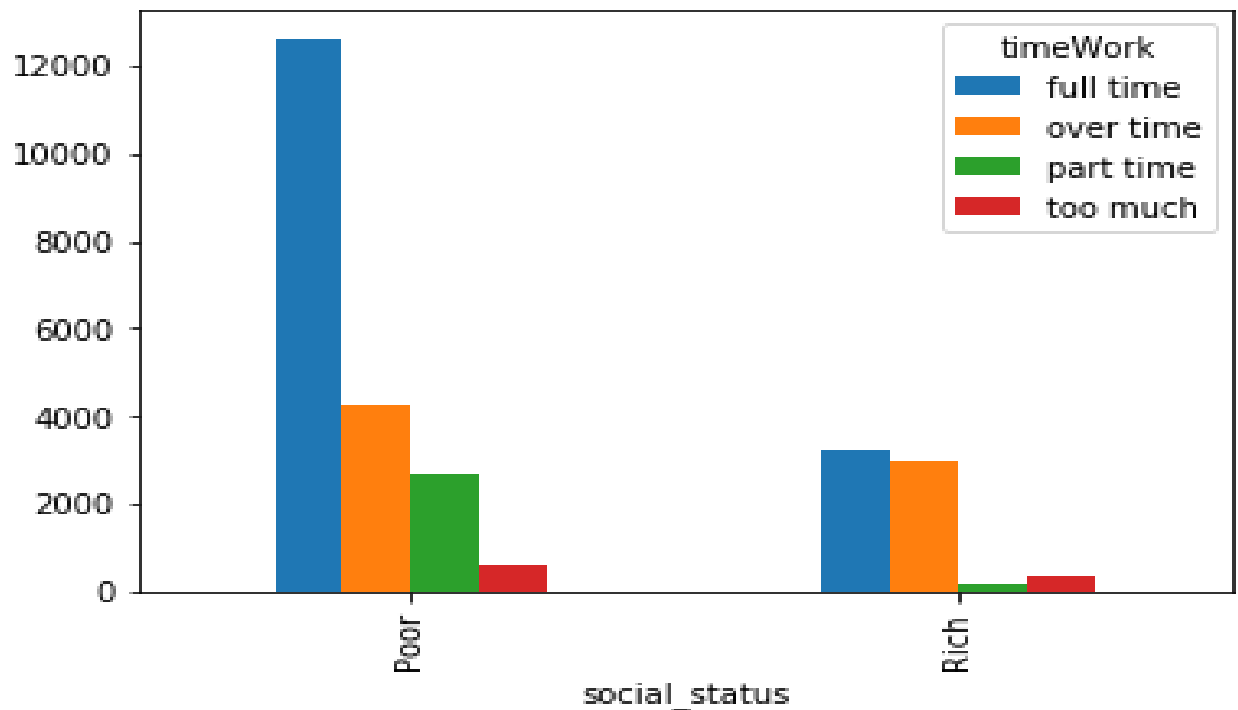
- We can note the following:

-That's show us the relation between age and marital status

Example:

- 1)The largest percentage of "Married-civ-spouse" in age 37.
- 2)The largest percentage of "Never-married" in age 23 .
- 3)The largest percentage of "Divorced" in age 45.
- 4)The largest percentage of "Separated" in age 34.
- 5)The largest percentage of "Widowed" in age 55.
- 6)The largest percentage of "Married-spouse-absent" in age 47 .
- 7)The largest percentage of "Married-AF-spouse" in age 29.

7)Relation between social_status and timeWork :



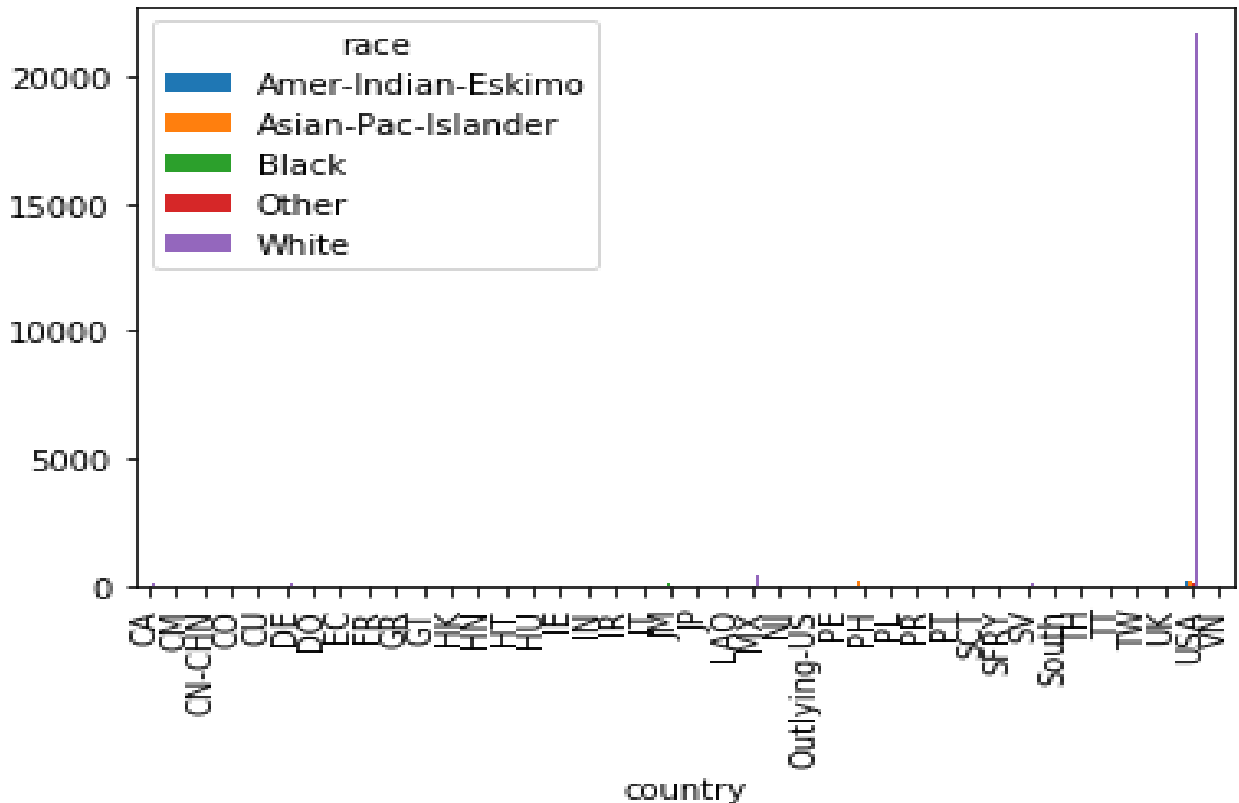
- We can note the following:

-That's show us the relation between social status and time work.

1) That the largest percentage of the poor work full time.

2) That the largest percentage of the rich work full time.

8)Relation between country and race :



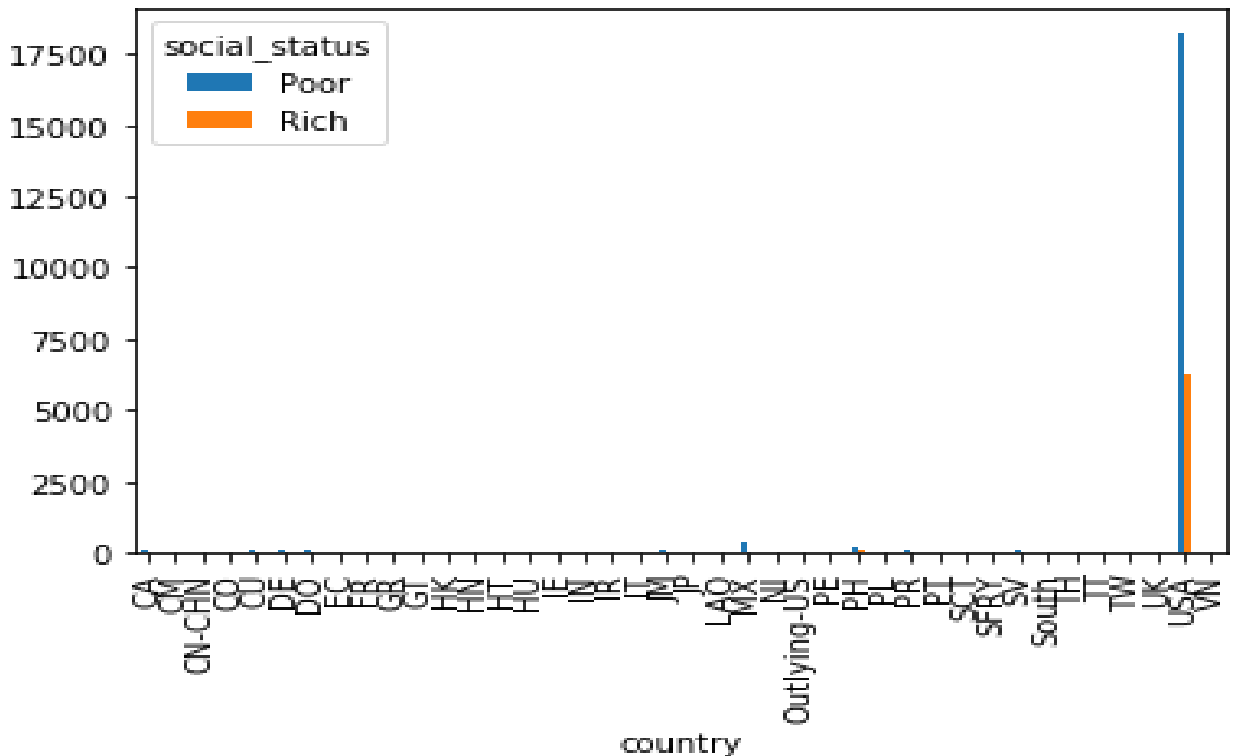
- We can note the following:

-That's show us the relation between country and race

Example:

- 1) The largest percentage of "Other" in USA.
- 2) The largest percentage of "Amer-Indian-Eskimo" in USA.
- 3) The largest percentage in MX is "White".
- 4) The largest percentage in PH is "Asian-Pac-Islander".
- 5) The largest percentage in TT is "Black".

9)Relation between country and social status :



- We can note the following:

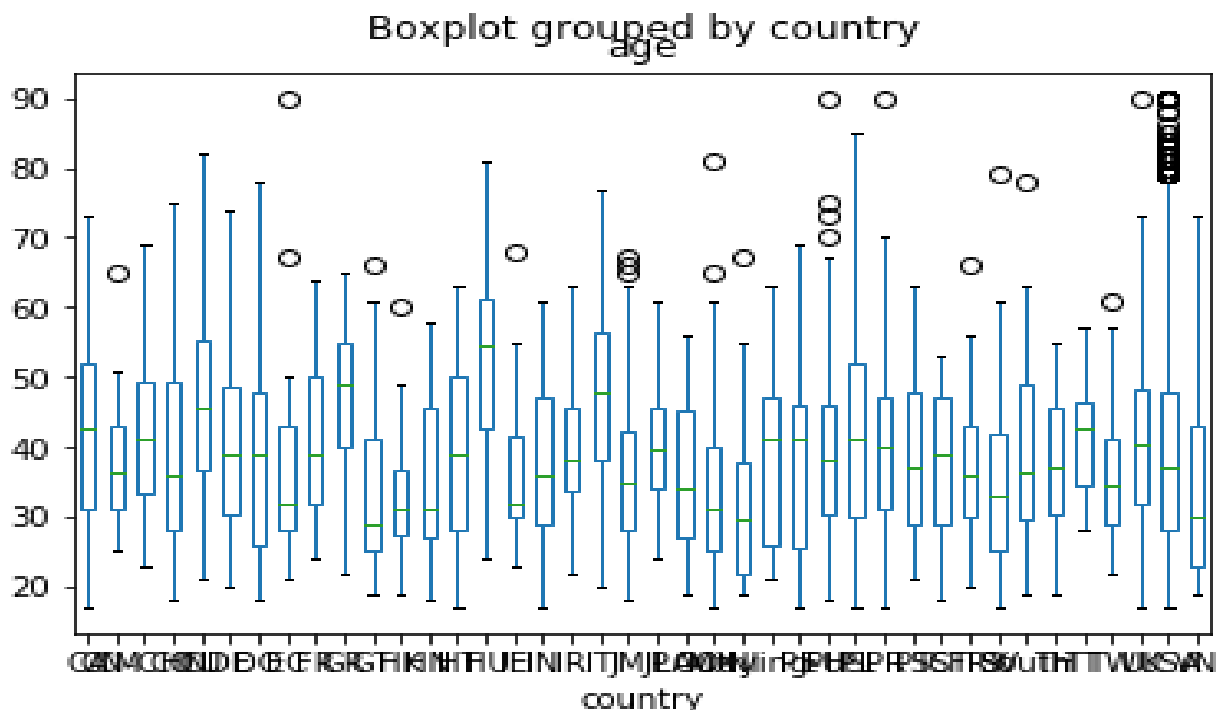
-That's show us the relation between country and social status.

-We conclude that the poor are the most in all countries.

Example :

- 1) The largest percentage in USA is the poor.
- 2) The largest percentage in MX is the poor.
- 3) The largest percentage in HH is the poor.
- 4) The largest percentage in TT is the poor.

10) Relation between country and age by boxplot:



- We can note the following:

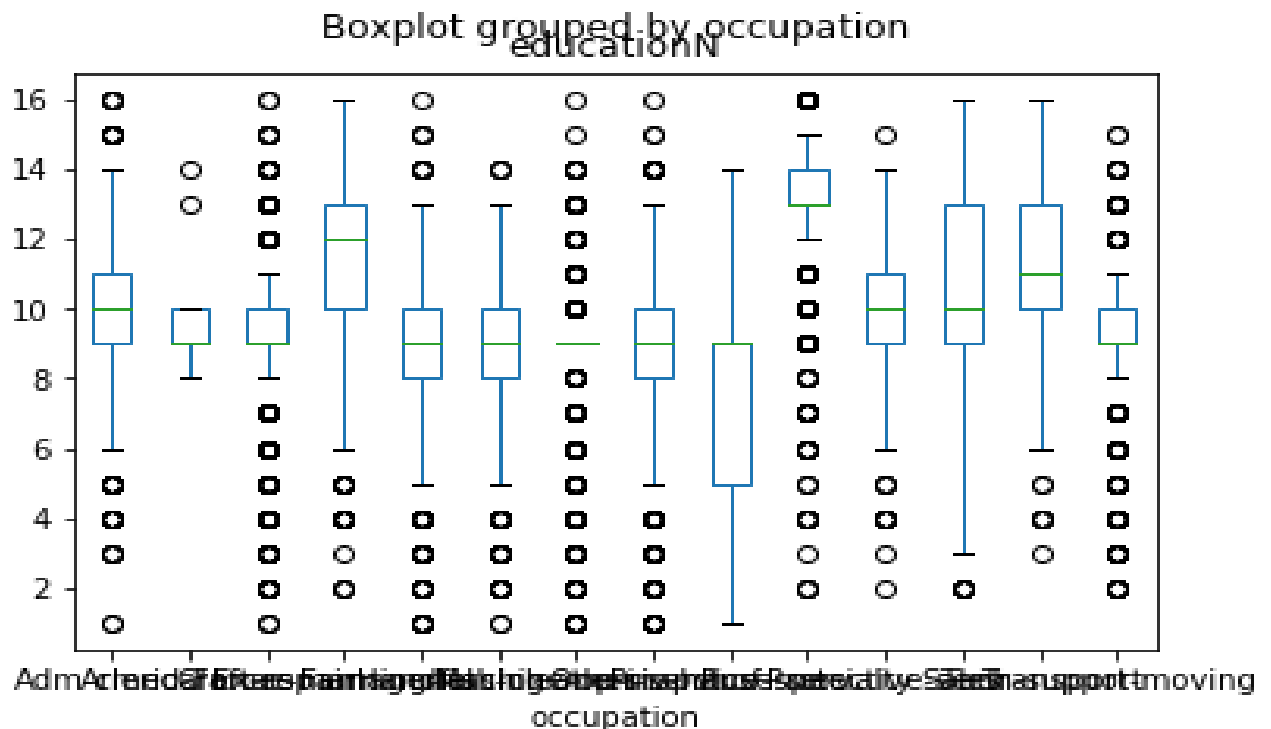
-That's show us the relation between country and age.

-we can see in this boxplot there is many outliers and this due to the variety in the ages and countries.

Example :

- 1) Age groups in FR in range from 24 to 64.
- 2) Age groups in UK in range from 17 to 90.
- 3) Age groups in MX in range from 17 to 81.
- 4) Age groups in TT in range from 28 to 57.

11)Relation between occupation and educationN by boxplot:



- We can note the following:

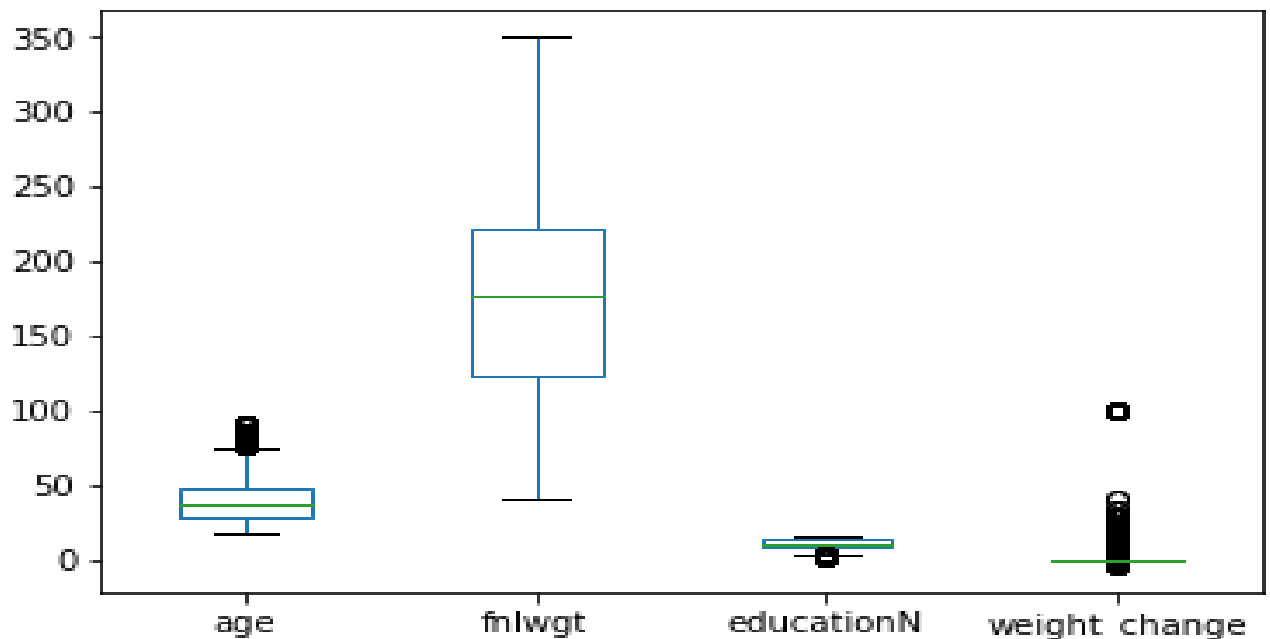
-That's show us the relation between occupation and the years of education.

-we can see in this boxplot there is many outliers and this due to the variety in the occupation and the years of education.

Example :

- 1) School years range in " Armed-Forces " from 8 to 14
- 2) School years range in " Protective-serv " from 2 to 14
- 3) School years range in " Prof-specialty " from 2 to 16
- 4) School years range in " Transport-moving " from 2 to 15

12)adult_clean boxplot:



- **We can note the following:**

- 1) we can see in this boxplot there is many outliers in age and this due to the variety in ages.
- 2) We can see in this box there are many outliers in the final weight “fnlwgt” and this is due to the variety in weights.
- 3) We can see in this box there are many outliers in the educationN and this is due to the variety in the years of education
- 4) We can see in this box that there are many outliers in the change of weight and this is due to the variation in how much weights change from one to one