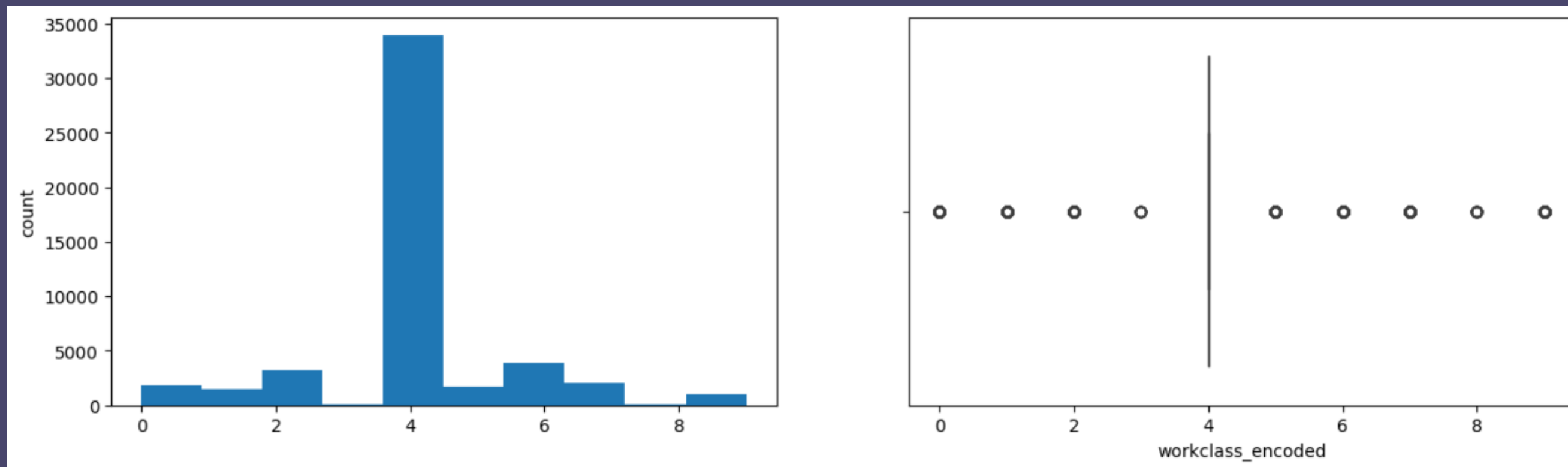
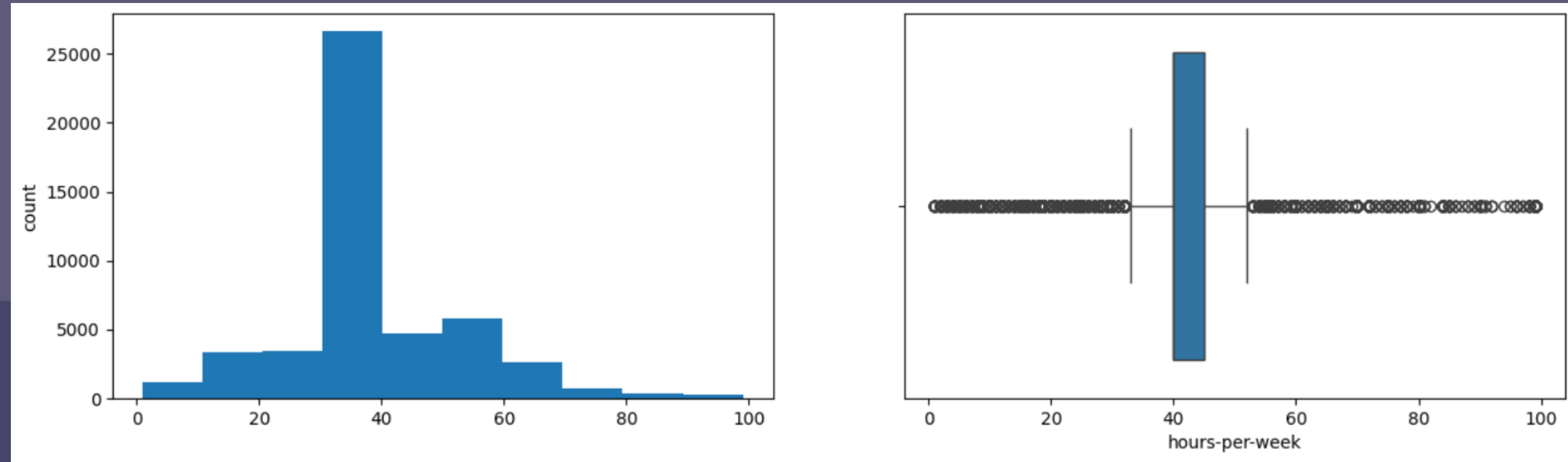


EXPLORATORY DATA ANALYSIS



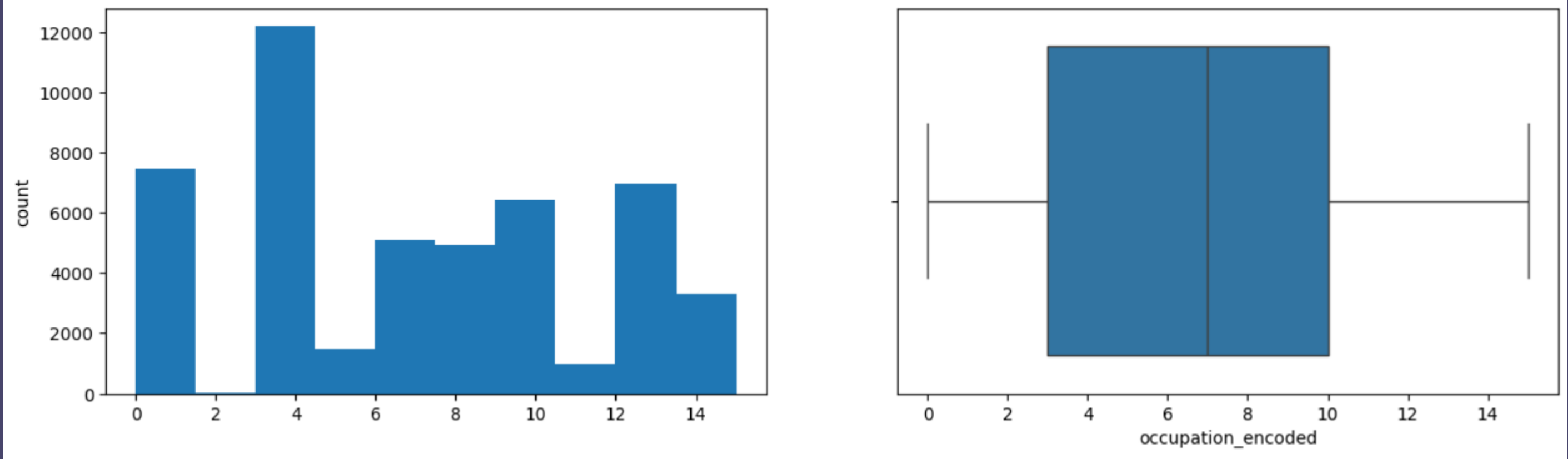
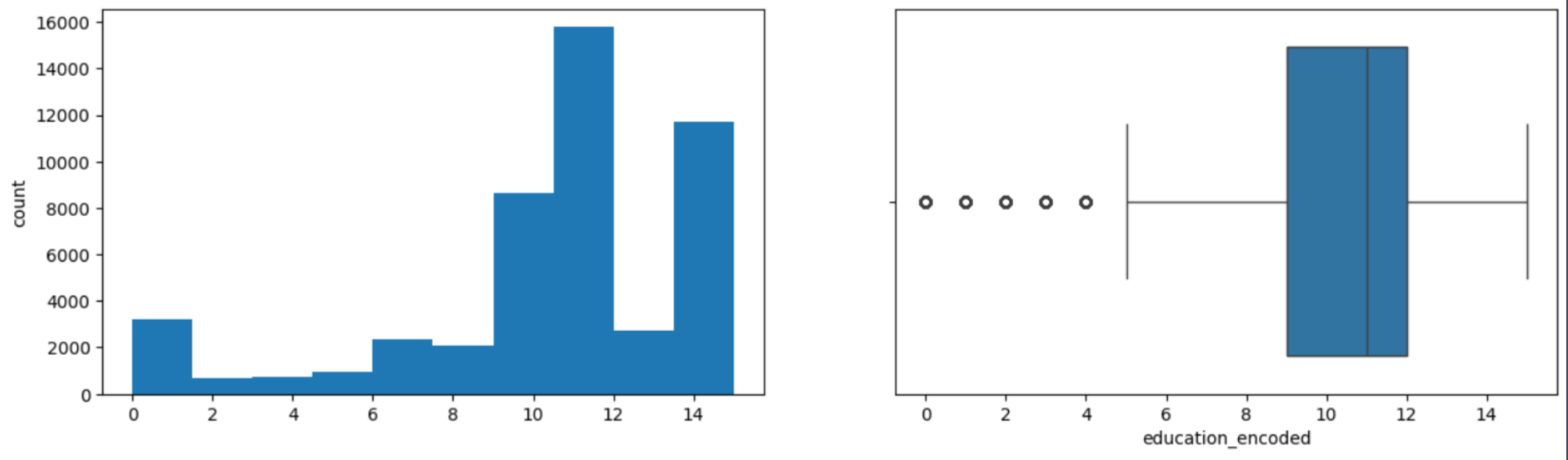
A HISTOGRAM AND BOX PLOT IS USED TO SHOW THE PATTERN OF THE VARIABLES, AS SOME VARIABLES HAVE SKEWNESS AND OUTLIERS.

THE BARPLOT OF HOURS-PER-WEEK INDICATES A MODERATELY SKEW OF 0.24



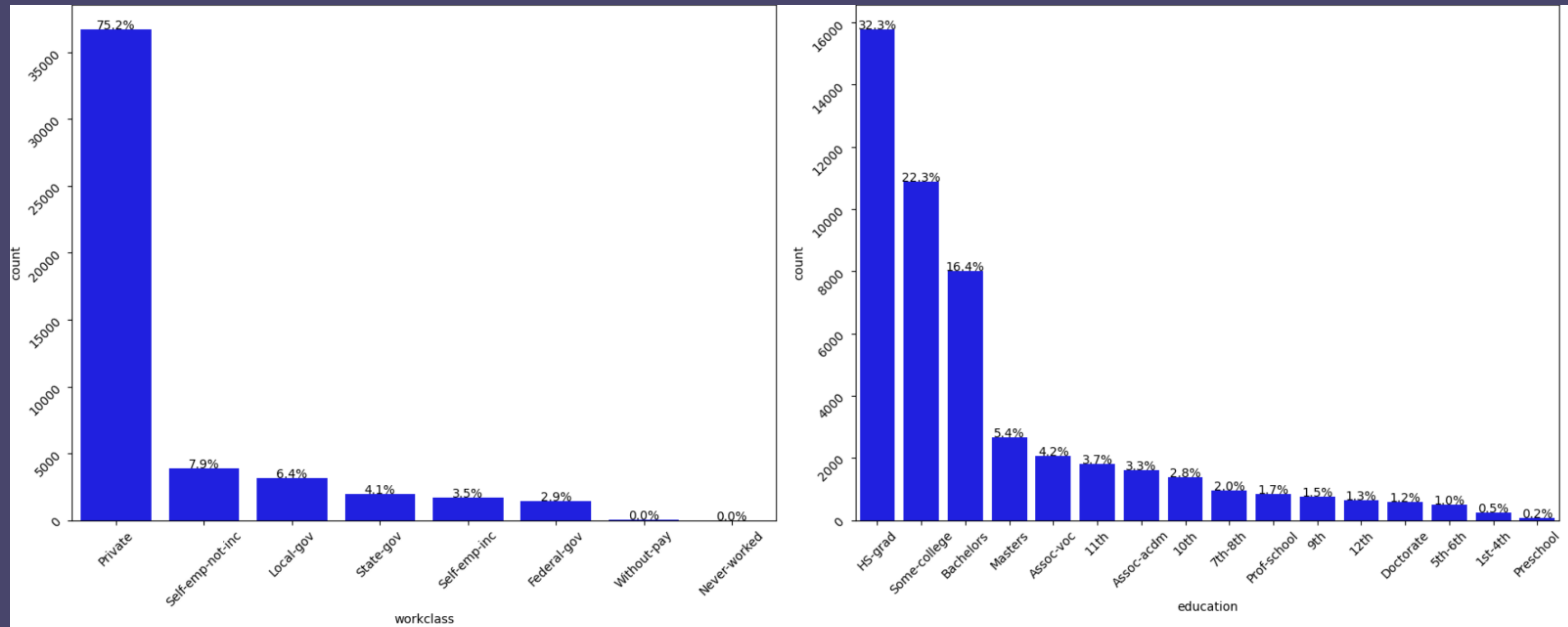
FOR THE WORKCLASS, IT IS ALMOST SYMMETRIC AND CLOSE TO ZERO WITH A SKEW OF 0.6

**THE BARPLOT FOR
EDUCATION HAS A SKEW OF
-0.94, AND FOR THE
OCCUPATION IT HAS 0.12
SKEWNESS**

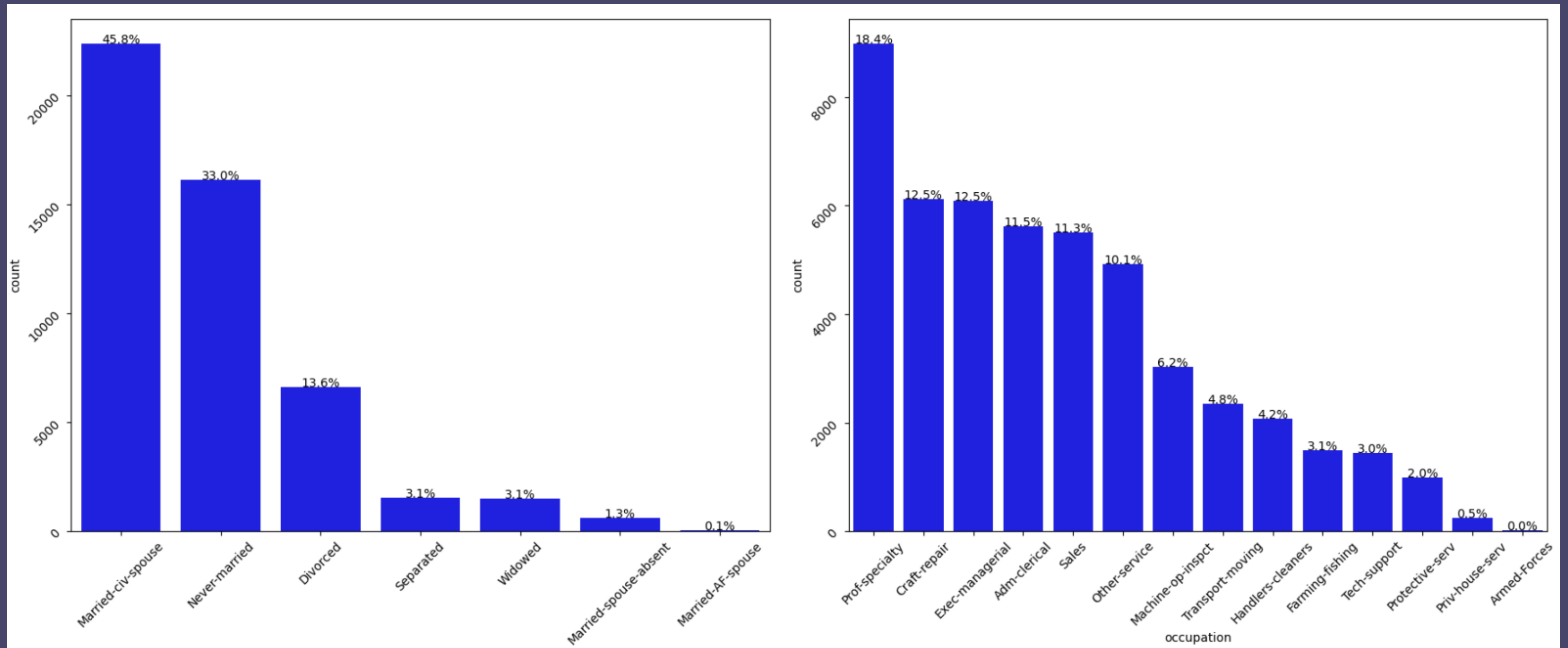


**I USED HISTOGRAM AND
BARPLOTS TO SEE IF THE
VARIABLES ARE DISTRIBUTED
NORMALLY AND PERFORM
BETTER DESCRIPTIVE
ANALYSIS**

CATEGORICAL VARIABLES ARE BEING VISUALIZED USING A COUNT PLOT

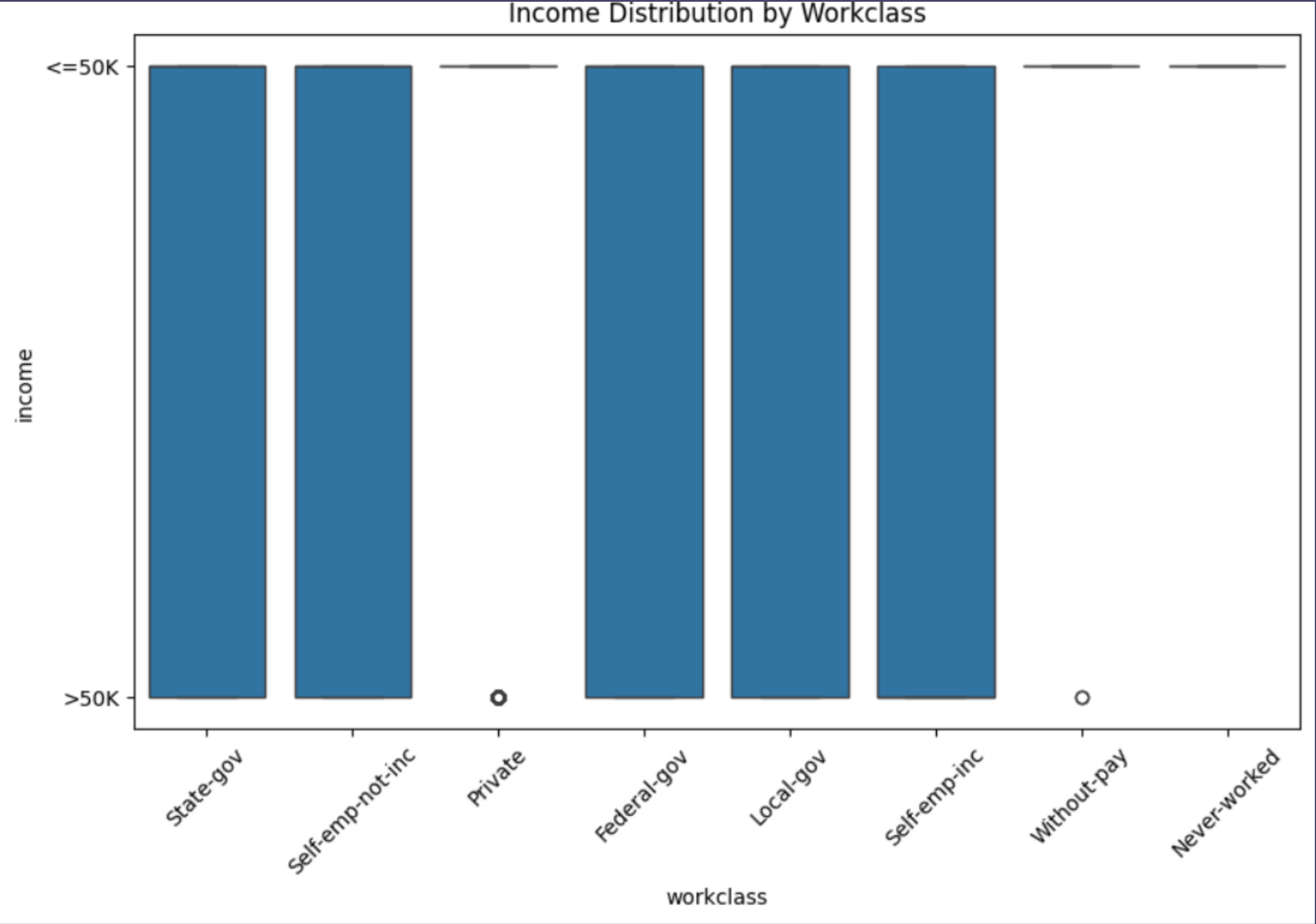


CATEGORICAL VALUES PROVIDE THE PATTERN OF FACTORS INFLUENCING THE INCOME AS WE CAN SEE IN THIS GRAPH 75.2% ARE IN PRIVATE SECTOR, 32.3% ARE THOSE WHO GRADUATED ONLY IN HIGH SCHOOL

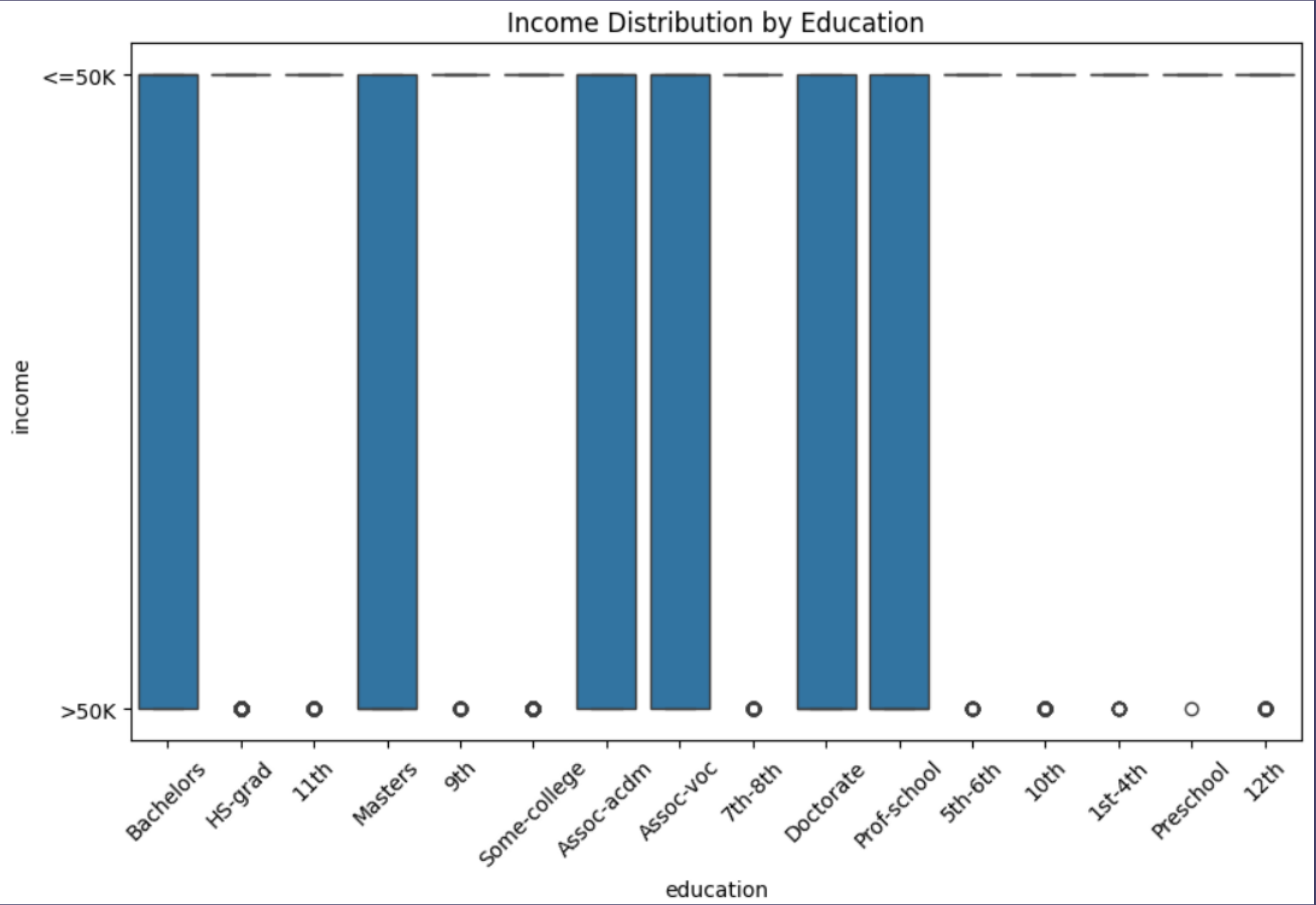


THERE ARE 45.8% THAT ARE MARRIED AND AN 18.4% WITH PROFFESIONAL SPECIALTY. THIS PERCENTAGES ARE FACTORS THAT MAY IMPACT INCOME AND HOW THEY CORRELATE TO EACH OTHER AND FINDING OUT THE SIGNIFICANCE OF THESE VARIABLES TO INCOME

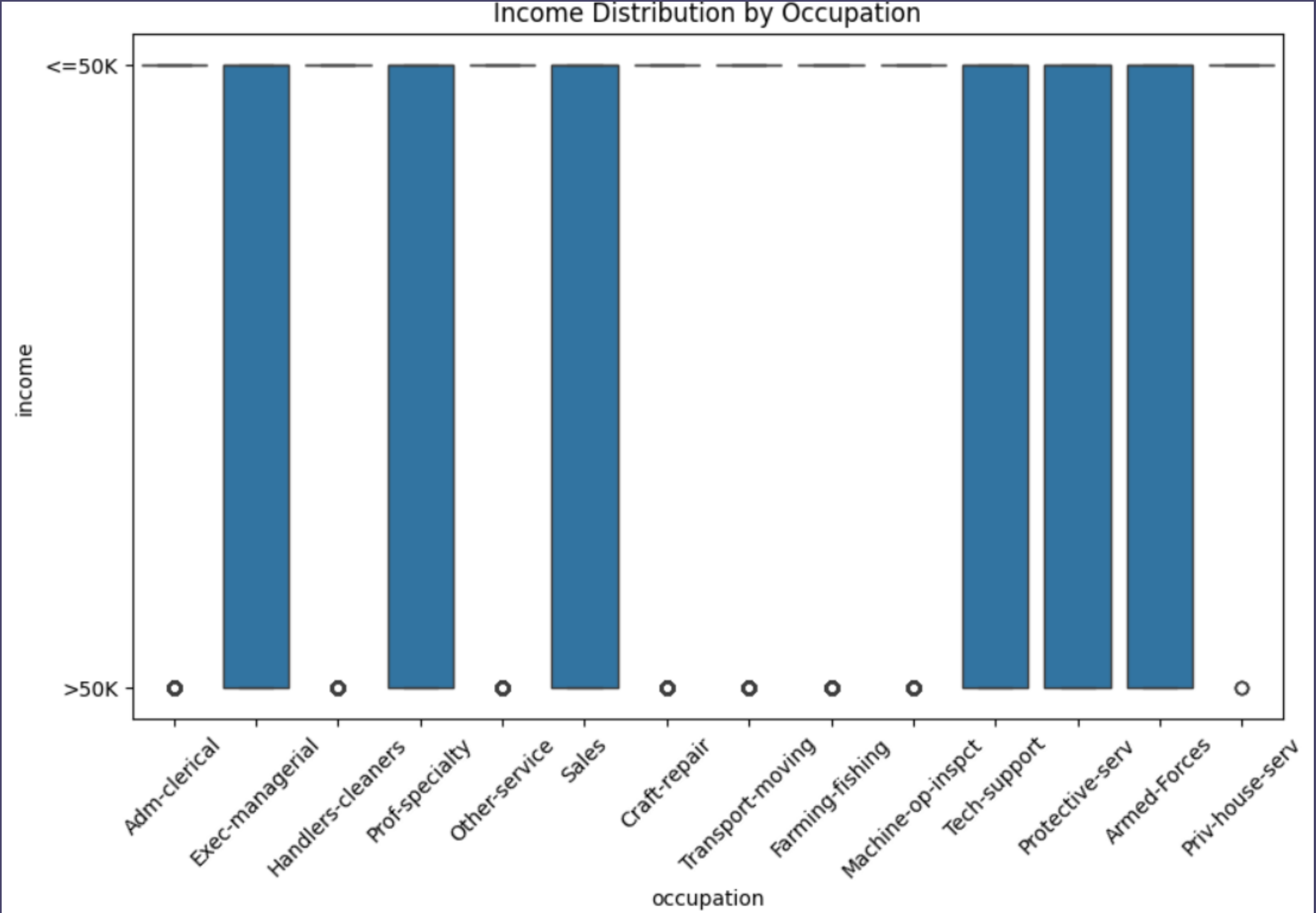
**THIS GRAPH SHOWS THE
DISTRIBUTION OF INCOME
BY WORKCLASS, MOST OF
THE PEOPLE THAT ARE
WORKING UNDER A
GOVERNMENT ARE LIKELY
TO EARN MORE THAN 50K
COMPARED TO PRIVATE
SECTOR**



**AS WE CAN SEE IN THIS GRAPH,
MAJORITY OF THE PEOPLE WHO EARNS
GREATER THAN 50K ARE THOSE WHO
FINISHED A COLLEGE DEGREE**



IN THIS GRAPH, MOST OCCUPATIONS
THAT PAY MORE THAN 50K INCLUDES
PROFESSIONAL SPECIALTY, WHICH IS
IDENTIFIED IN THE COUNT PLOT SO
THERE IS RELEVANCY



THE CORRELATION HEATMAP SHOWS HOW SIGNIFICANT IS THE RELATIONSHIP OF EACH VARIABLES

THE INCOME HAS A STRONG POSITIVE CORRELATION TO AGE, CAPITAL GAIN, CAPITAL LOSS, HOURS PER WEEK, AND EDUCATION

