**Shape of the dataset: (48842, 16)**

**First few rows of the dataset:**

**Unnamed: 0 age workclass fnlwgt education education-num \**

**0 0 39 State-gov 77516 Bachelors 13**

**1 1 50 Self-emp-not-inc 83311 Bachelors 13**

**2 2 38 Private 215646 HS-grad 9**

**3 3 53 Private 234721 11th 7**

**4 4 28 Private 338409 Bachelors 13**

**marital-status occupation relationship race sex \**

**0 Never-married Adm-clerical Not-in-family White Male**

**1 Married-civ-spouse Exec-managerial Husband White Male**

**2 Divorced Handlers-cleaners Not-in-family White Male**

**3 Married-civ-spouse Handlers-cleaners Husband Black Male**

**4 Married-civ-spouse Prof-specialty Wife Black Female**

**capital-gain capital-loss hours-per-week native-country income**

**0 2174 0 40 United-States <=50K**

**1 0 0 13 United-States <=50K**

**2 0 0 40 United-States <=50K**

**3 0 0 40 United-States <=50K**

**4 0 0 40 Cuba <=50K**

**Dataset Information:**

**<class 'pandas.core.frame.DataFrame'>**

**RangeIndex: 48842 entries, 0 to 48841**

**Data columns (total 16 columns):**

**# Column Non-Null Count Dtype**

**--- ------ -------------- -----**

**0 Unnamed: 0 48842 non-null int64**

**1 age 48842 non-null int64**

**2 workclass 47879 non-null object**

**3 fnlwgt 48842 non-null int64**

**4 education 48842 non-null object**

**5 education-num 48842 non-null int64**

**6 marital-status 48842 non-null object**

**7 occupation 47876 non-null object**

**8 relationship 48842 non-null object**

**9 race 48842 non-null object**

**10 sex 48842 non-null object**

**11 capital-gain 48842 non-null int64**

**12 capital-loss 48842 non-null int64**

**13 hours-per-week 48842 non-null int64**

**14 native-country 48568 non-null object**

**15 income 48842 non-null object**

**dtypes: int64(7), object(9)**

**memory usage: 6.0+ MB**

**None**

**Summary Statistics:**

**Unnamed: 0 age fnlwgt education-num capital-gain \**

**count 48842.000000 48842.000000 4.884200e+04 48842.000000 48842.000000**

**mean 24420.500000 38.643585 1.896641e+05 10.078089 1079.067626**

**std 14099.615261 13.710510 1.056040e+05 2.570973 7452.019058**

**min 0.000000 17.000000 1.228500e+04 1.000000 0.000000**

**25% 12210.250000 28.000000 1.175505e+05 9.000000 0.000000**

**50% 24420.500000 37.000000 1.781445e+05 10.000000 0.000000**

**75% 36630.750000 48.000000 2.376420e+05 12.000000 0.000000**

**max 48841.000000 90.000000 1.490400e+06 16.000000 99999.000000**

**capital-loss hours-per-week**

**count 48842.000000 48842.000000**

**mean 87.502314 40.422382**

**std 403.004552 12.391444**

**min 0.000000 1.000000**

**25% 0.000000 40.000000**

**50% 0.000000 40.000000**

**75% 0.000000 45.000000**

**max 4356.000000 99.000000**

**Missing Values (before handling):**

**Unnamed: 0 0**

**age 0**

**workclass 963**

**fnlwgt 0**

**education 0**

**education-num 0**

**marital-status 0**

**occupation 966**

**relationship 0**

**race 0**

**sex 0**

**capital-gain 0**

**capital-loss 0**

**hours-per-week 0**

**native-country 274**

**income 0**

**dtype: int64**

**Missing Values (after handling):**

**Unnamed: 0 0**

**age 0**

**workclass 0**

**fnlwgt 0**

**education 0**

**education-num 0**

**marital-status 0**

**occupation 0**

**relationship 0**

**race 0**

**sex 0**

**capital-gain 0**

**capital-loss 0**

**hours-per-week 0**

**native-country 0**

**income 0**

**dtype: int64**

**Accuracy Scores:**

**Random Forest Accuracy: 0.9616**

**XGBoost Accuracy: 0.9150**

**Naive Bayes Accuracy: 0.7858**

**Logistic Regression Accuracy: 0.7793**

**Decision Tree Accuracy: 0.9559**

**K-Nearest Neighbors Accuracy: 0.8764**

**Support Vector Machine Accuracy: 0.8709**





