# Adult Income Prediction

Nena Esaw

# Project Description

- **The purpose of this project is predicting if an adult income will be greater than $50,000 based on occupation, education, age and other attributes.**

- **The Adult Income dataset was extracted from the 1994 US Census database. The goal is to determine if several factors contribute to individual making a salary over $50,000 with classification techniques.**

# Adult Income Dataset

The dataset was sourced from Kaggle - Adult Income Dataset.

For this dataset, there were 48,852 rows and 15 columns.

1. age: continuous.
2. workclass: Private, Self-emp-not-inc, Self-emp-inc, Federal-gov, Local-gov, State-gov, Without-pay, Never-worked.
3. fnlwgt: continuous.
4. education: Bachelors, Some-college, 11th, HS-grad, Prof-school, Assoc-acdm, Assoc-voc, 9th, 7th-8th, 12th, Masters, 1st-4th, 10th, Doctorate, 5th-6th, Preschool.
5. education-num: continuous.
6. marital-status: Married-civ-spouse, Divorced, Never-married, Separated, Widowed, Married-spouse-absent, Married-AF-spouse.
7. occupation: Tech-support, Craft-repair, Other-service, Sales, Exec-managerial, Prof-specialty, Handlers-cleaners, Machine-op-inspct, Adm-clerical, Farming-fishing, Transport-moving, Priv-house-serv, Protective-serv, Armed-Forces.
8. relationship: Wife, Own-child, Husband, Not-in-family, Other-relative, Unmarried.
9. race: White, Asian-Pac-Islander, Amer-Indian-Eskimo, Other, Black.
10. sex: Female, Male.
11. capital-gain: continuous.
12. capital-loss: continuous.
13. hours-per-week: continuous.
14. native-country: United-States, Cambodia, England, Puerto-Rico, Canada, Germany, Outlying-US(Guam-USVI-etc), India, Japan, Greece, South, China, Cuba, Iran, Honduras, Philippines, Italy, Poland, Jamaica, Vietnam, Mexico, Portugal, Ireland, France, Dominican-Republic, Laos, Ecuador, Taiwan, Haiti, Columbia, Hungary, Guatemala, Nicaragua, Scotland, Thailand, Yugoslavia, El-Salvador, Trinadad&Tobago, Peru, Hong, Holand-Netherlands.

class: >50K, <=50K

# Stakeholders

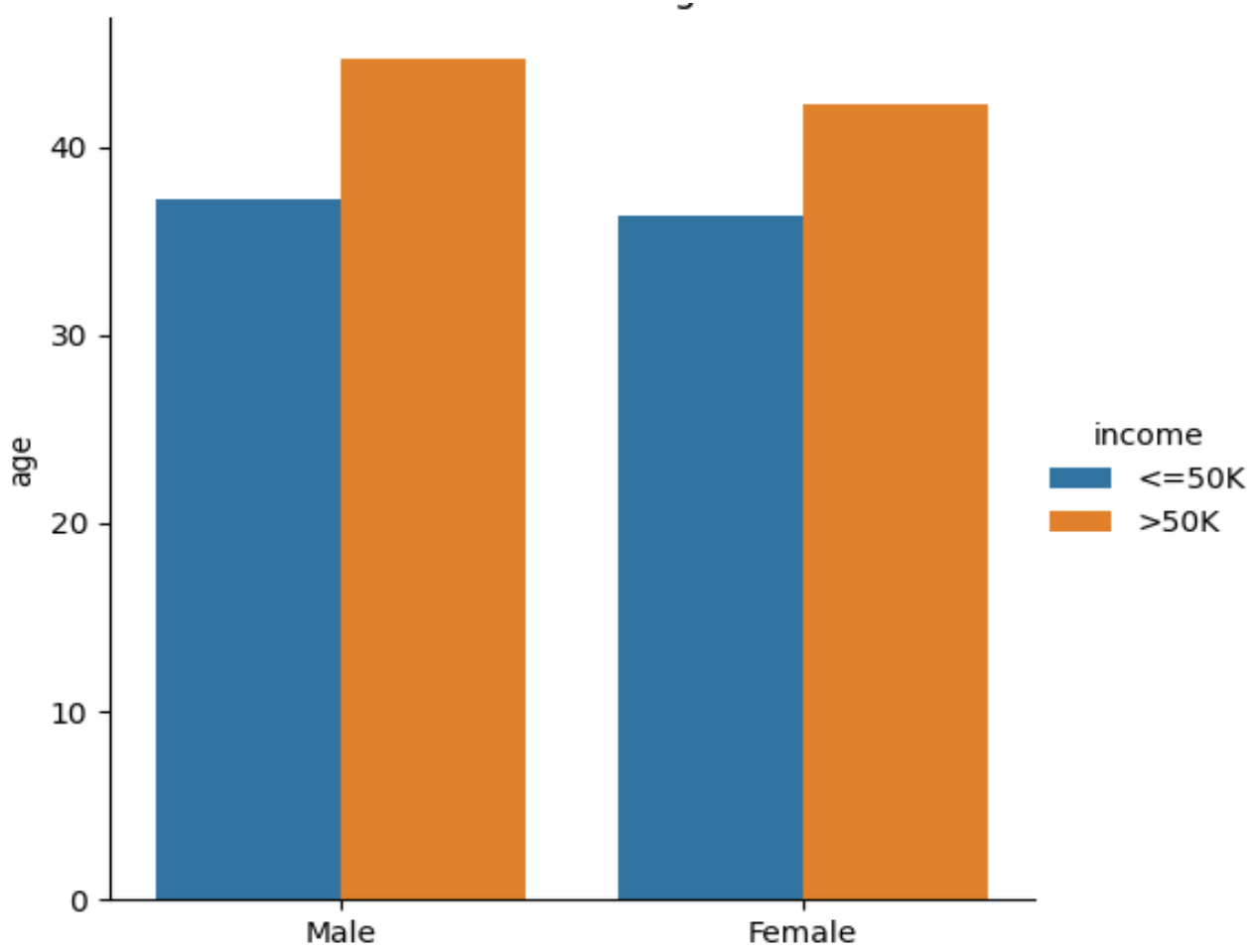
This analysis will be used by individuals who want to determine what factors contribute to making a salary over $50,000.

# Key Findings

•We can see that more male over the age of 40 generate a higher income over 50k than compared to females.

•Females over the age 40 make over 50k as well but not as much as males.

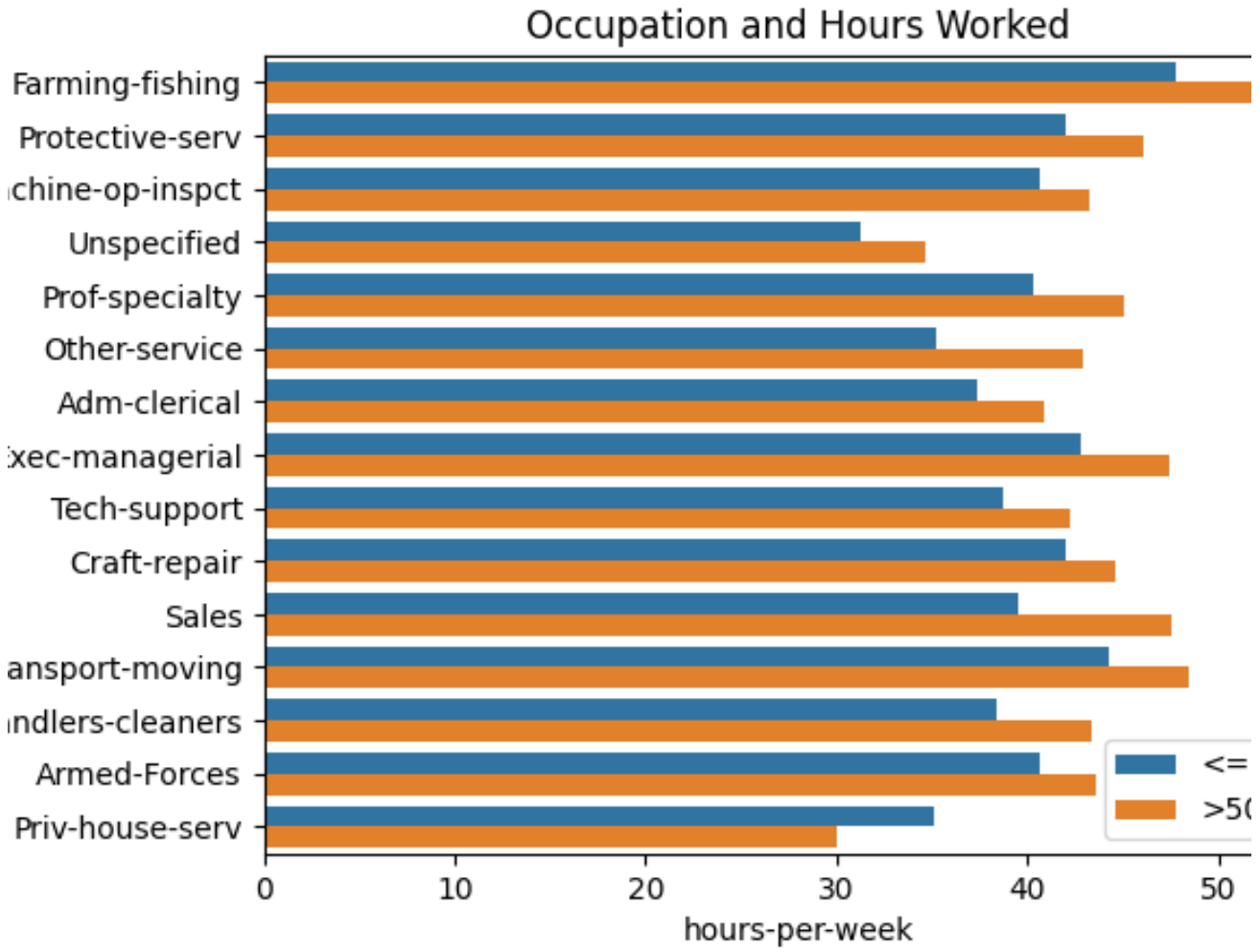
•We can determine that over the over of 40 for both male and females they generate a higher income.



# Key Findings

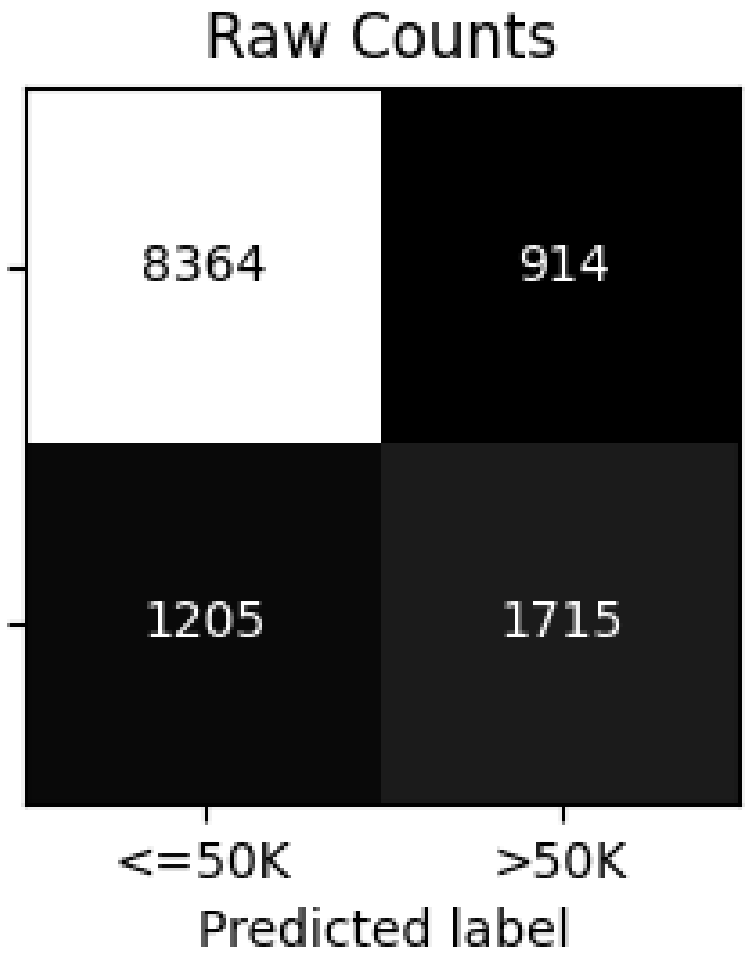
•This graph shows up that Farming-fishing occupation generates the highest income and work more than 50 hours per week.

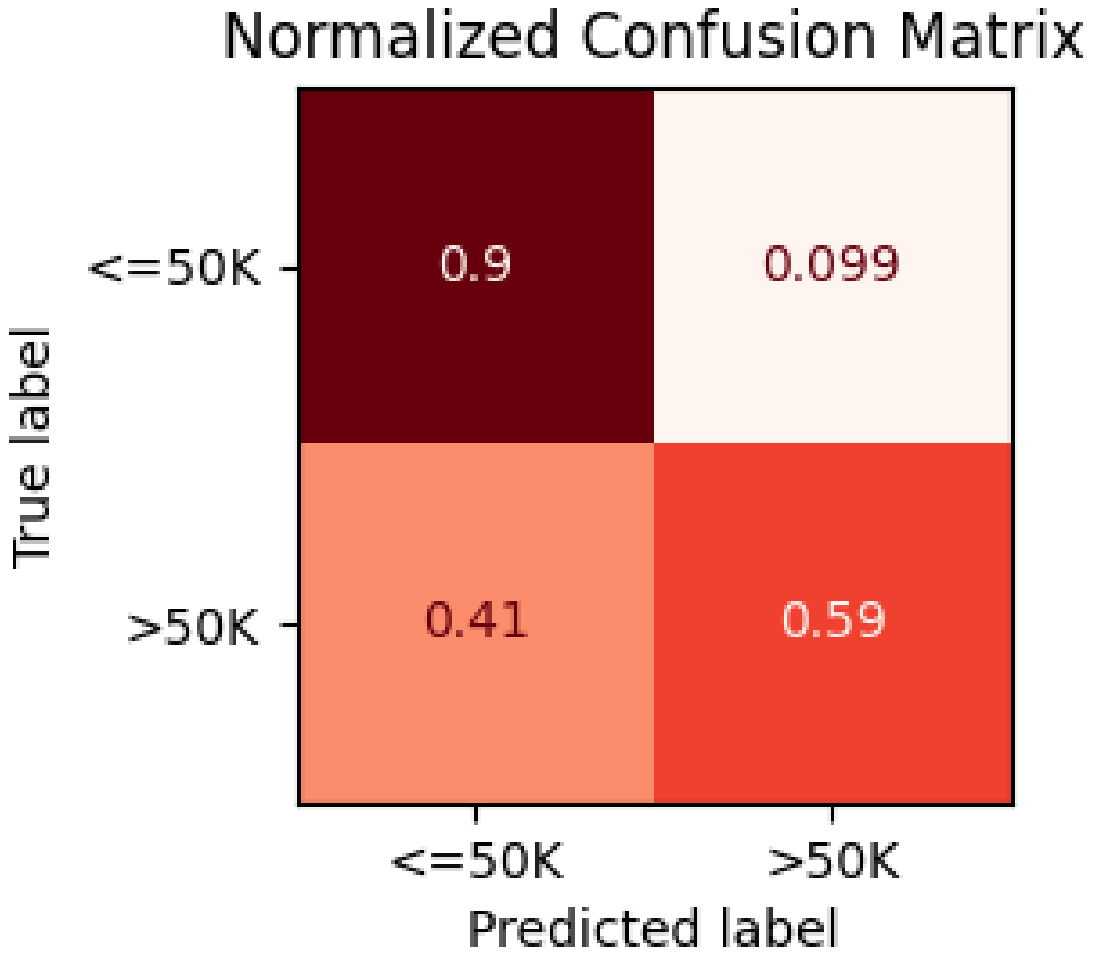
•We can also see that most of the work class that work over 40 hours per week generate over 50k income.



# Key Findings

- The KNN model performed the best with an 83% accuracy. The recall for incomes under 50k was predicting at 90%. However, the over 50k income was only predicting at 59%.
- The false negatives were at a 41% while the false positive were 9%. Which means that our model does a great job of predicting incomes that are less than 50k.
- The model is not great at predicting an individual's income over 50k.



# Recommendations

- The KNN model might predict better if the data were more balanced. There were about 4330 values that were unknown or missing data. Having a more accurate dataset would improve the predictions for this model.
- Incorrectly estimating a person's income to be over $50,000 should not be used with this model due to the high false negatives.