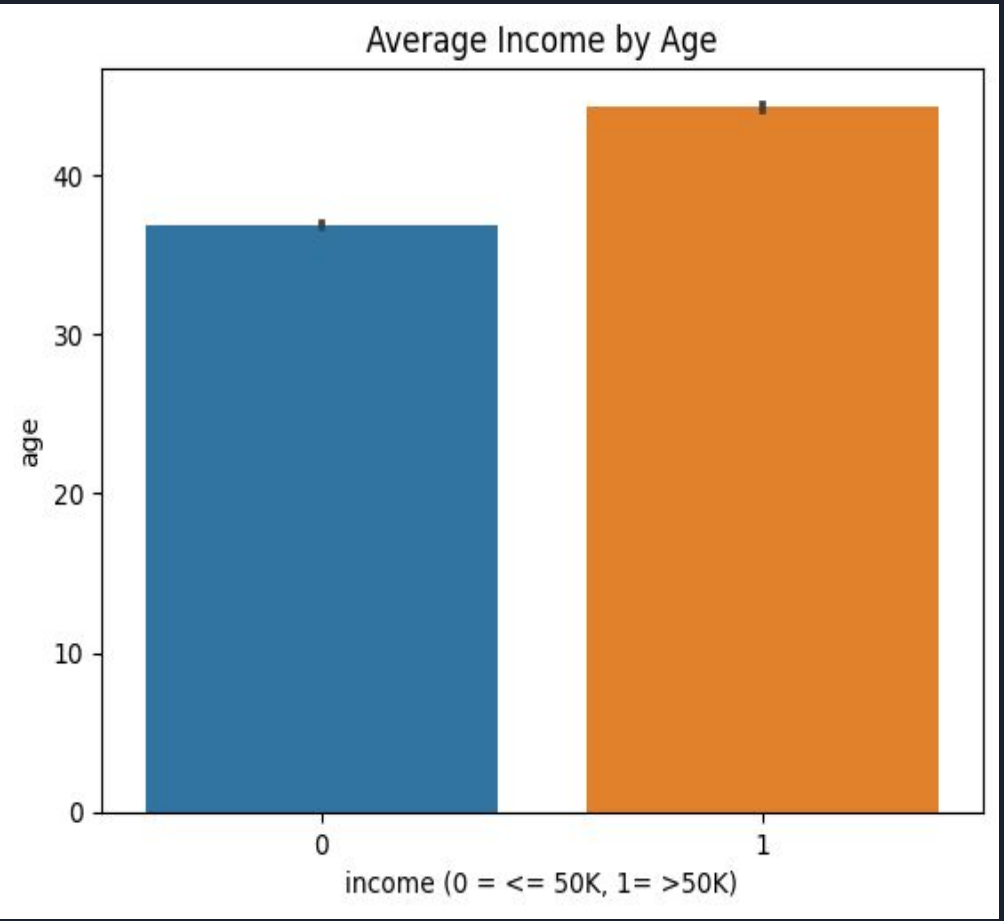
Adult Income

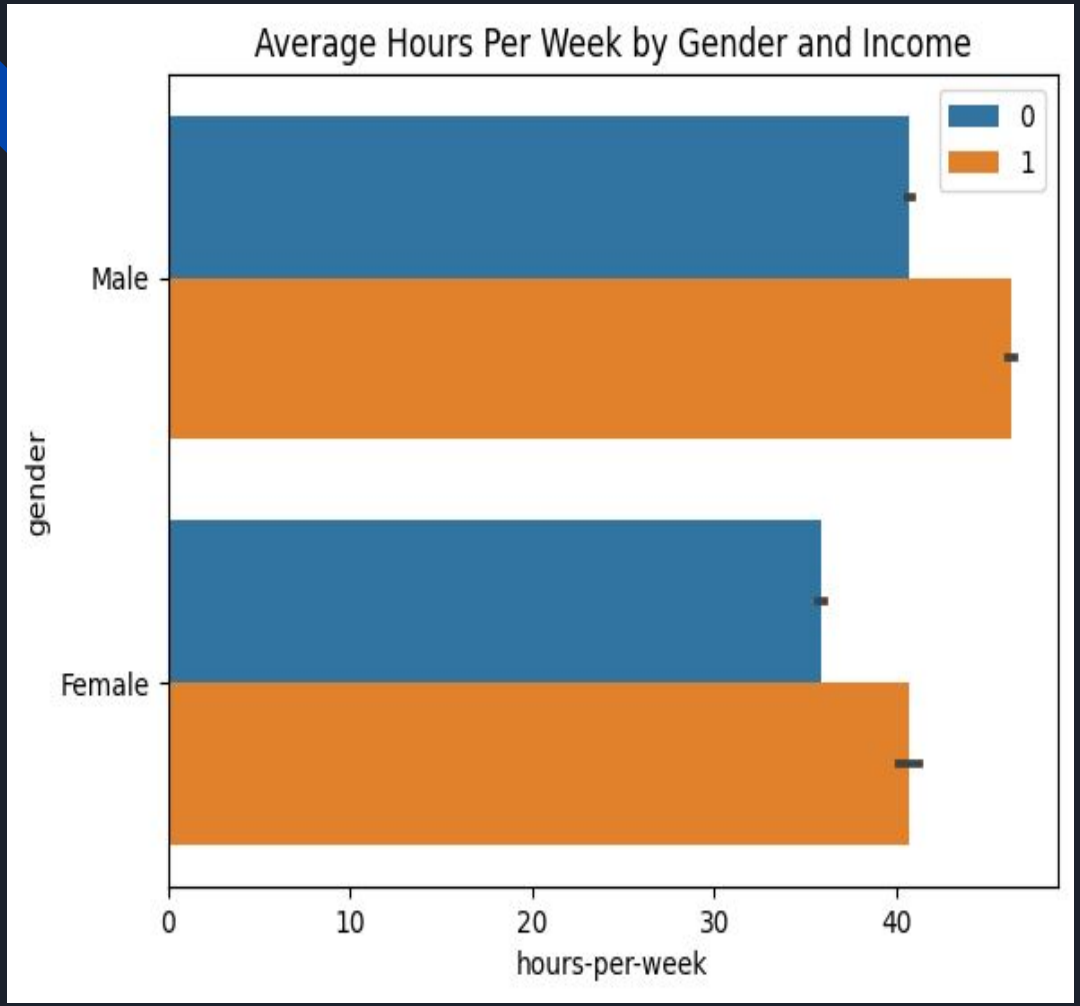
William Rodemoyer

Project Objective

- Banking Companies are our Stakeholders
- We are exploring the possibility of predicting an individual's income level, based on some the individuals personal information for our stakeholders.
- The income is divided into two classes: <=50K and >50K
- A person's annual income results from various factors. Some of the factors that we used are education level, age, gender, occupation, and etc.



- On average, those who make over 50K annually, are older in age.



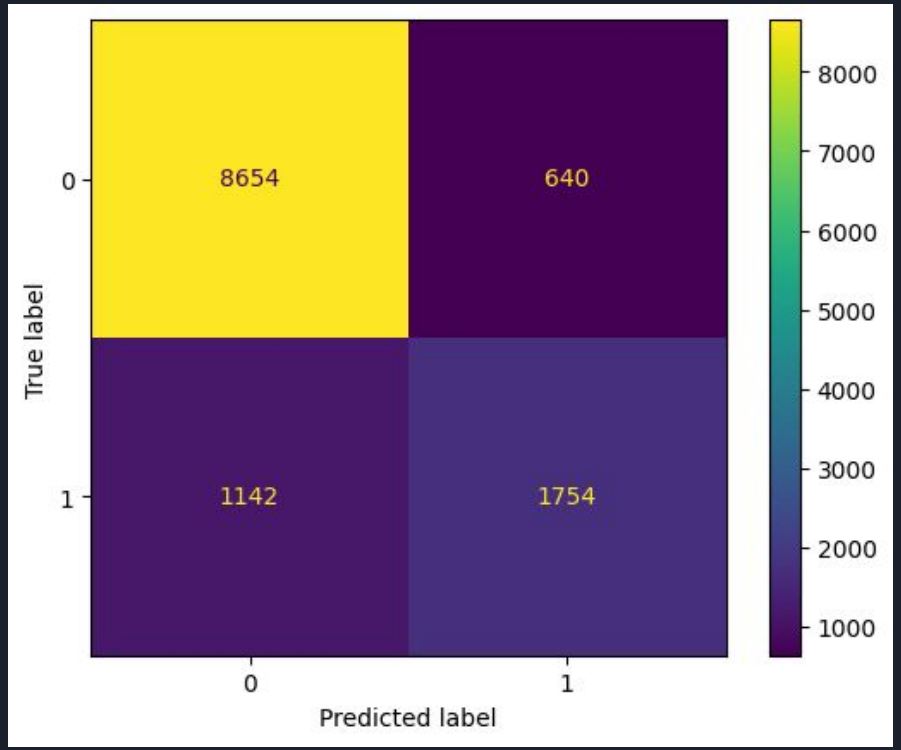
- Both Genders who make more than 50k, work more hours per week than those who make 50k or less.

- On average, males work more hours per week compared to females when comparing their respectable income classes.

Starting Model

- Our starting model show how unbalanced it is predicting if someone makes over 50k.
  - 1142 people were predicted to make 50k or less, that actually made over 50k.
  - While only 1754 people making over 50k were predicted accurately.
  - That is too much imbalance.

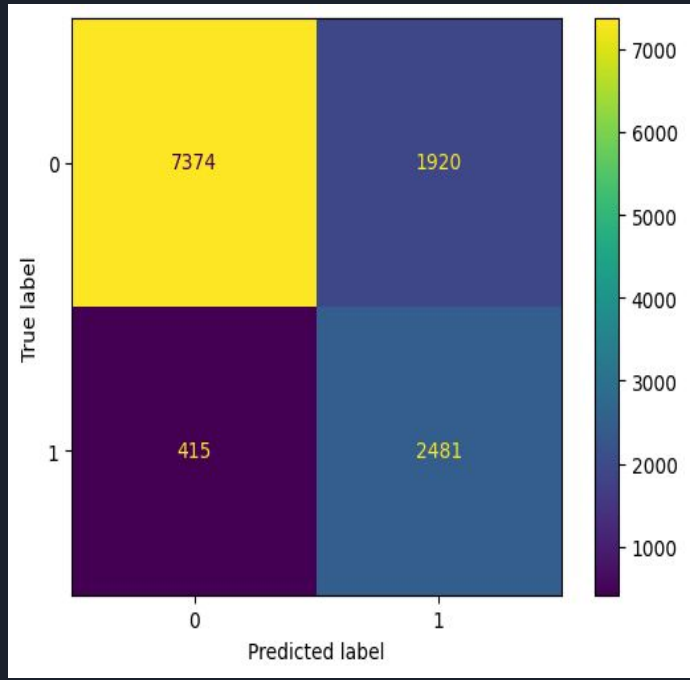
- I needed to create a new create model with more balance.



Final Model

- **Pros:**
  - Accurately predicting 81% of incomes. (Good, Not Great Percentage)

- **Cons: The other 19%**
  - 415 people are predicted to make 50k or less, but actually make more than 50k
  - 1920 people are predicted to make over 50k, but make equal to or less than 50k



Final Recommendations

- Add another column or two.
  - Ex. pay (salary, hourly, commission, etc.), overtime hours, etc.
- Generalize some of the values within the columns, less complex can often time be easier on our models.