



Adult Income Model

Matthew Kwak



Stakeholders

Politicians - This data was provided by the US Census. The government is able to predict how well their employees income stacks up against other occupations. Also, the government will be able to predict

Executives - Executives are able to see if there is inequity regarding income in regards to age, race, gender, or native country. If there is such a high inequity, Executives should try to help provide more equity.

Guidance Counselors - Guidance Counselors will be able to use this model and dataset to predict future students and their decisions on education.

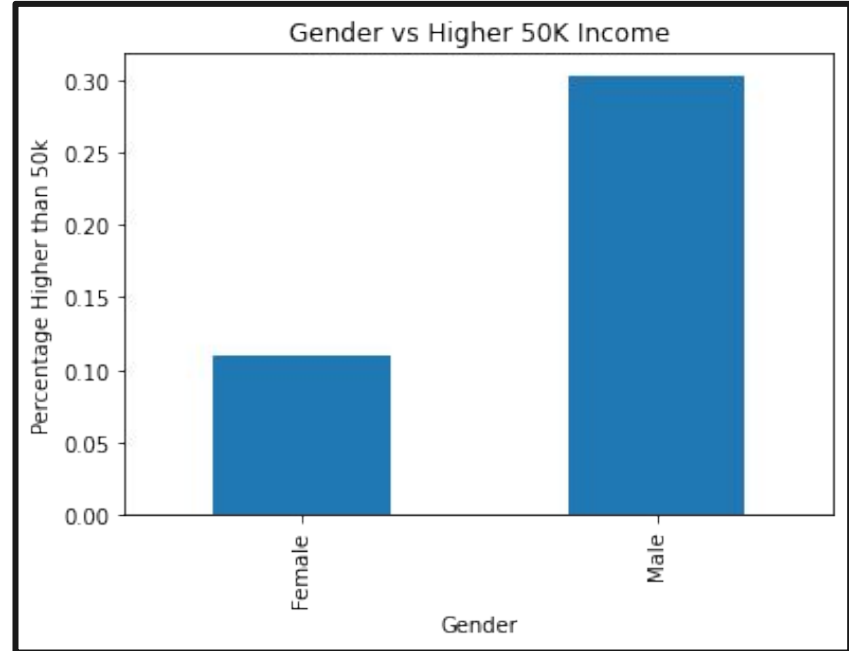


Adult Income Dataset

- An adult's income is based off many different factors such as gender, race, education, occupation, and relationship status.
- Target Value is Income
 - Does an adult make above or less than \$50k
- The total dataset has around 45,000 instances
- Dataset provided is from 5/19/1996
- References:
 - <http://www.cs.toronto.edu/~delve/data/adult/desc.html>
 - <http://www.cs.toronto.edu/~delve/data/adult/adultDetail.html>

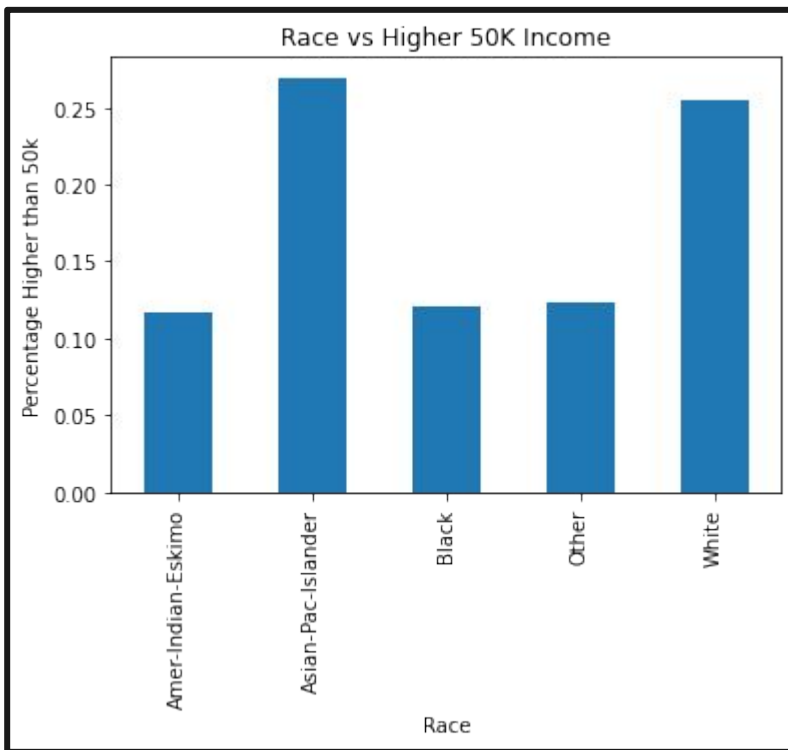
Gender

- It was found that out of the total females, only around 11% made over 50k while around 30% of males made over 50k. This is a 1:3 difference of the amount of males making over 50k than females. We do have twice as many males than females which may contribute to the percentage disparity for income.



Race

- The Asian-Pac-Islander and White races had around 25% of their population make over 50k, while the Amer-Indian-Eskimo, Black, Other races only had around 11% of their population make above 50k. The dataset is very heavily White race dominated.





Strengths

- Accuracy of the predictions of the model is 85%
- There is a substantial number of data points for the model to learn
- The class balance is fairly balance with 76% make less than or equal to \$50k and 24% make above \$50k
- Able to predict the less than or equal to \$50k group very well (92%)

Limitations

- This is a heavily dominated male and White set.
- There is not enough representation from all people
- This dataset is from 1996
- The occupation description could be more specific rather than “private” as the biggest group
- Unable to predict make above \$50k as there are many false negatives (36%)



Conclusion

- In conclusion, we would be able to use this model to help predict people's income based off the features provided in the US Census.
- Politicians would be able to use this model to predict how well their district is performing on census to census
- Executives would be able to use this model to predict how well any changes to their equity initiatives to help gender and race have more salary
- Guidance Counselors would be able to predict how well a graduating class performs in regards to salary as they receive higher education