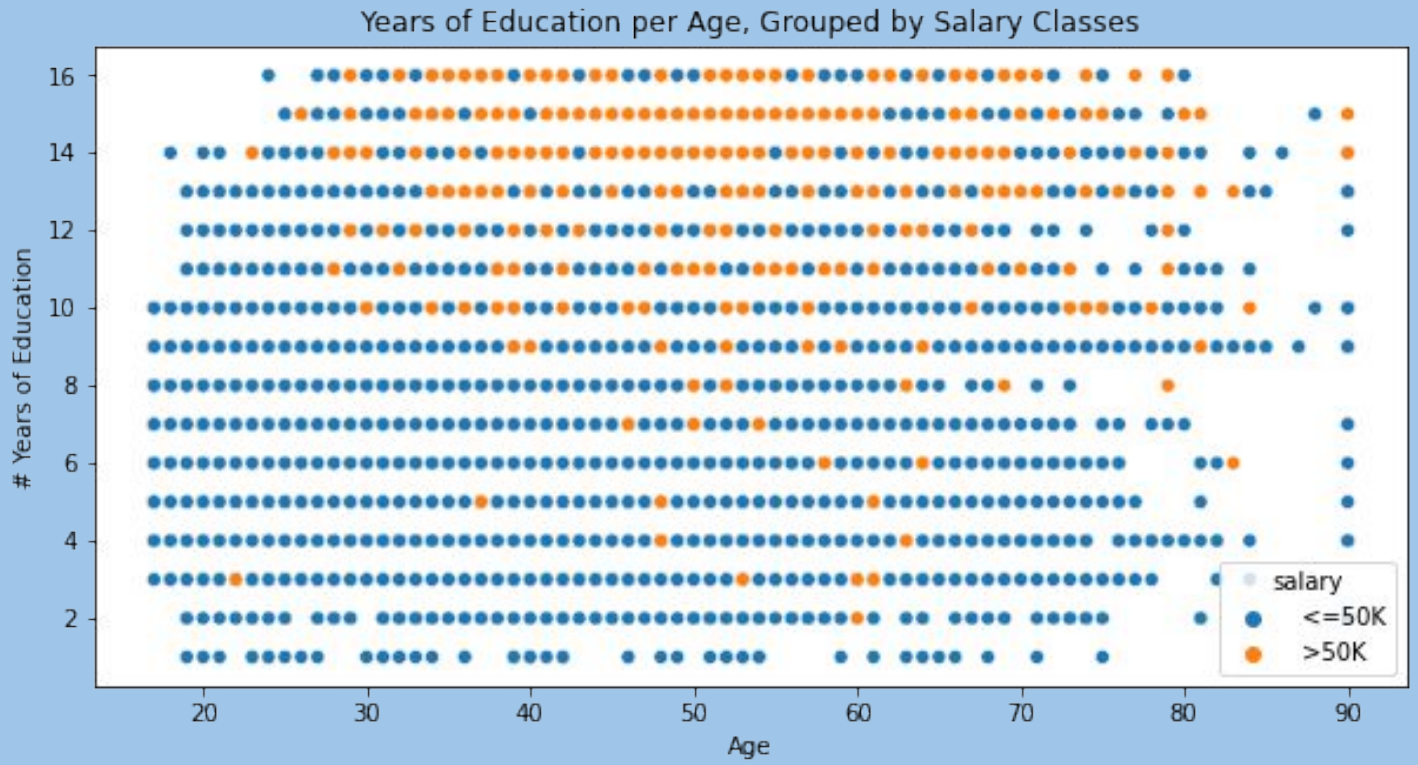
Salary Prediction

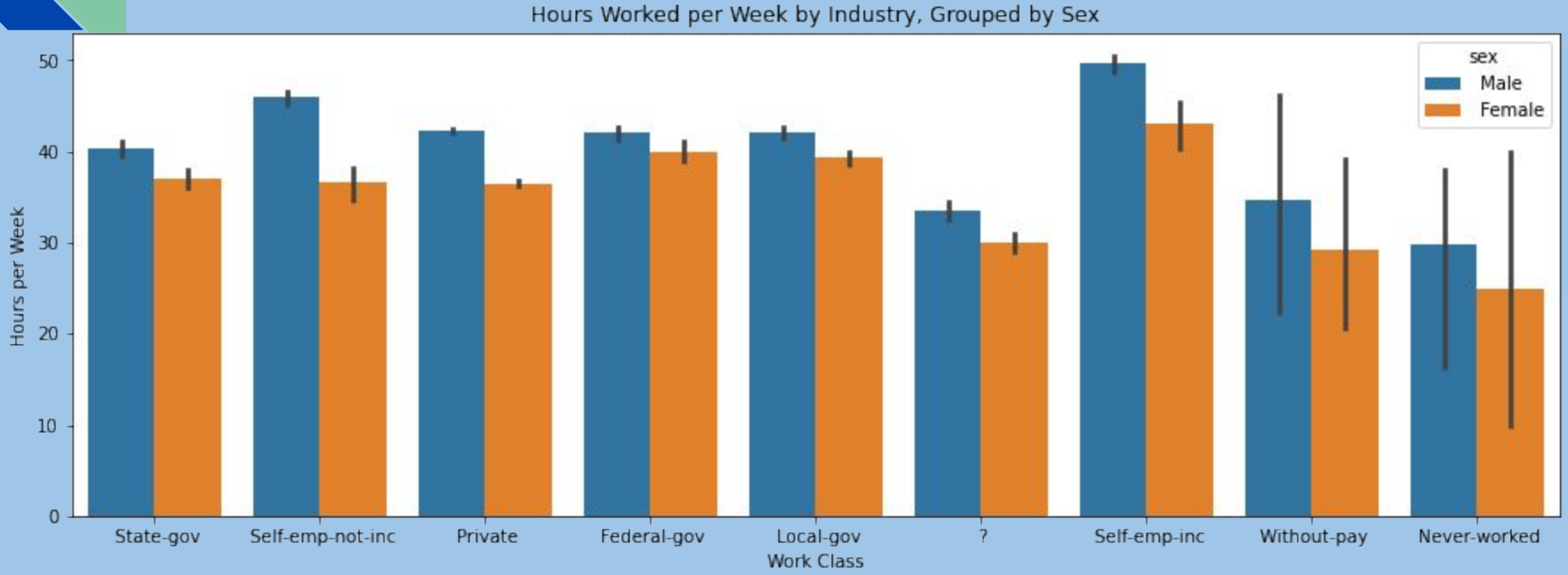
Presented by:
Hadassah Port

Data Introduction & Stakeholder

- The data is from the 1994 US Census
- Includes features such as someone's education level, marital status, and professional sector
- Target is predicting whether someone makes more or less than $50,000 annually
- Potential stakeholders: Credit Card companies, banks, Marketing agencies



Between ages 32-61, people with at least 13 years of education are likely to have greater than 50K salaries.



Across all categories, men work more hours than women. The previous graph showed that more hours worked lead to more pay. Based on that observation, men are expected to make more than women.

Limitations

	precision	recall
<=50K	0.87	0.92
>50K	0.69	0.58
accuracy		
macro avg	0.78	0.75
weighted avg	0.83	0.83

- Initial target balance was 75/25
- Final accuracy was 84%
- Accuracy couldn't be taken at face value for determining model's performance
- Precision: 0 .87 for <=50K; 0 .69 for >50K
- Recall: 0.92 for <=50K; 0.58 for >50K
- Increased false positives and negatives could potentially lead to credit card companies either lending too much credit to someone who won't be able to pay it all back or miss out on lending more money to someone who has the ability to repay it

Strengths

- SMOTE
- Updated target balance: 50/50
- Increased, more balanced Precision and Recall
    - Precision: 0 .89 for <=50K; 0 .80 for >50K
    - Recall: 0.77for <=50K; 0.90 for >50K
- Accuracy still 84% but can be taken at face value
- Less false positives and negatives allows companies to more accurately set credit limits based on salaries

Final Recommendations

- KNN vs Logistic Regression
- KNN with SMOTE had high Precision, Recall, and Accuracy