# Adult Income

By: Rosa R

Adult Income

This data describes the features related to adult incomes. We compare the difference between someone who pursued higher education vs someone who didn't and what that outcome was.

Who is your stakeholder, and what problem are you solving for them? The Department of Education to show areas of focus for improvement of outcomes.
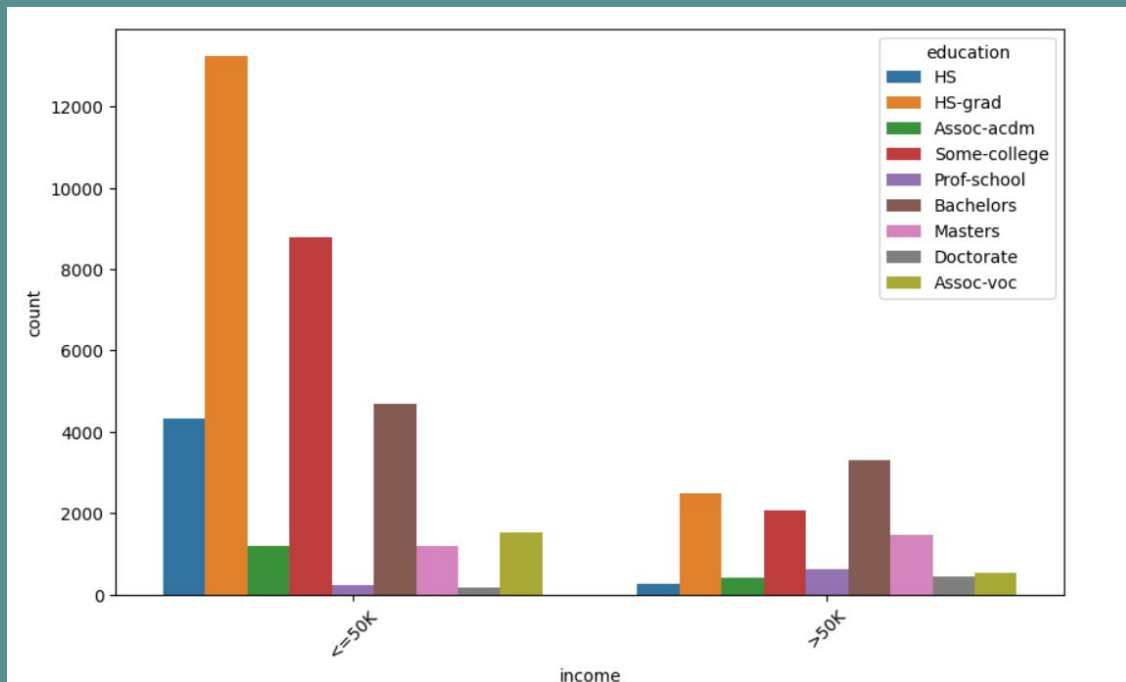
Visualizations

- A seaborn countplot and histogram visualized for numeric datatype column.

- Also, a barplot was visualized for categorical column.

- This gave a good baseline for the numeric and categorical columns.
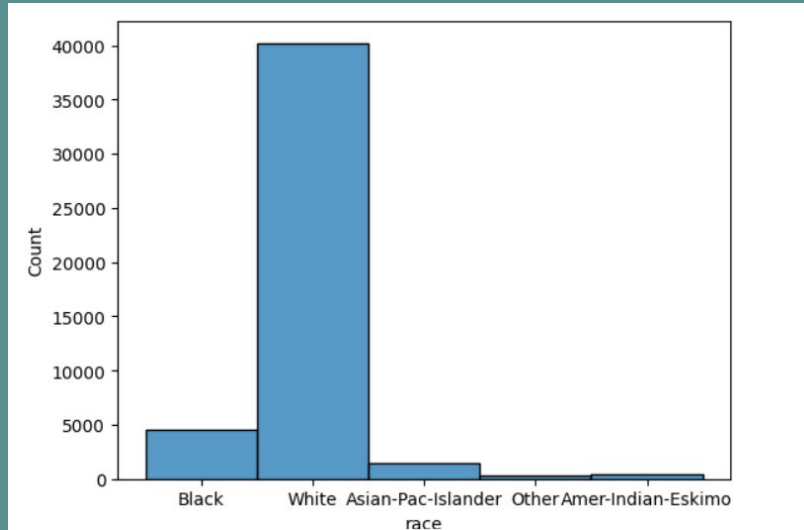
The majority of employees are white compared to any other race.

Income and Education

Here you can see a clear correlation between income and education. Very few people with a high school education (not grad) can still make over 50k.

Machine Learning Using the Following Models:

- Logistic Regression Model

- Random Forest Classifier Model

- KNN Classifier Model

Models Evaluated & Results

The Final Model Chosen was a Logistic Regression PCA Tuned Model.

	Model	Accuracy Score	Precision Score	Recall Score	F1 Score	ROC	Best Parameters	Execution Time
0	Random Forest	0.815565	0.647258	0.584068	0.614041	0.738645	N/A	16.70
1	Logistic Regression	0.794108	0.562208	0.814915	0.665375	0.801021	N/A	5.08
2	KNeighbors	0.689629	0.438200	0.835254	0.574828	0.738016	N/A	13.83
3	Logistic Regression Tuned	0.795300	0.563964	0.815932	0.666944	0.802155	{'logisticregression__C': 0.5, 'logisticregres...	633.42
4	Logistic Regression PCA Tuned	0.793767	0.560998	0.823051	0.667216	0.803497	{'logisticregression__C': 0.5, 'logisticregres...	531.10
5	Random Forest PCA Tuned	0.790446	0.556794	0.812542	0.660786	0.797788	{'randomforestclassifier__max_depth': 10, 'ran...	3511.21
6	KN Tuned	0.743784	0.493762	0.791525	0.608152	0.759647	{'kneighborsclassifier__n_neighbors': 10}	121.06

Recommendations

Model Performance

- Given the results of the models after running with PCA I think Logistic Regression PCA Tuned performed slightly better than Random Forest therefore I would use this as my prediction model and suggest to my stakeholder. KNeighbors actually scored lower before using PCA but still lower than both other models.