Classifying Income
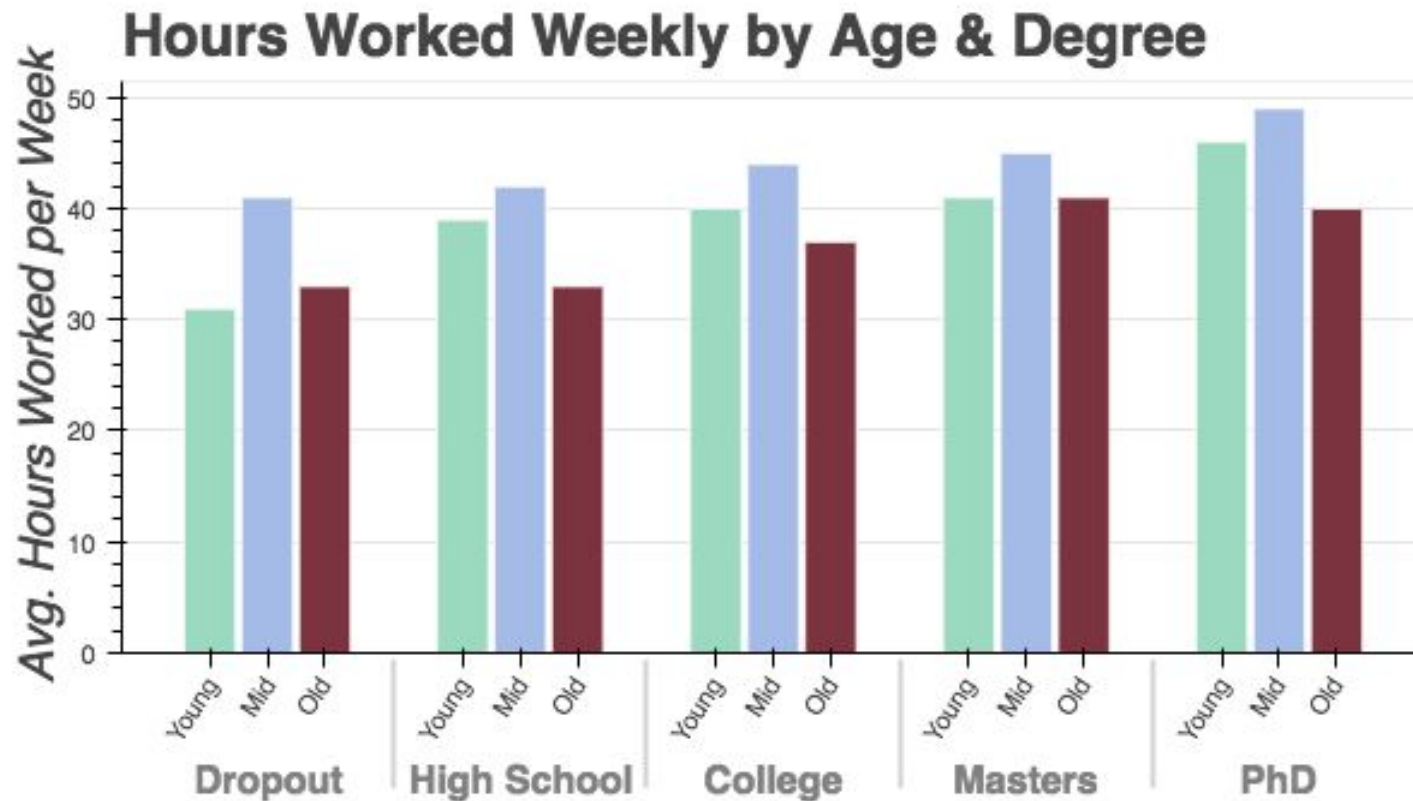
• • •

Tarek Barnes
August 2018

Question

Which demographic receives a moderate income?
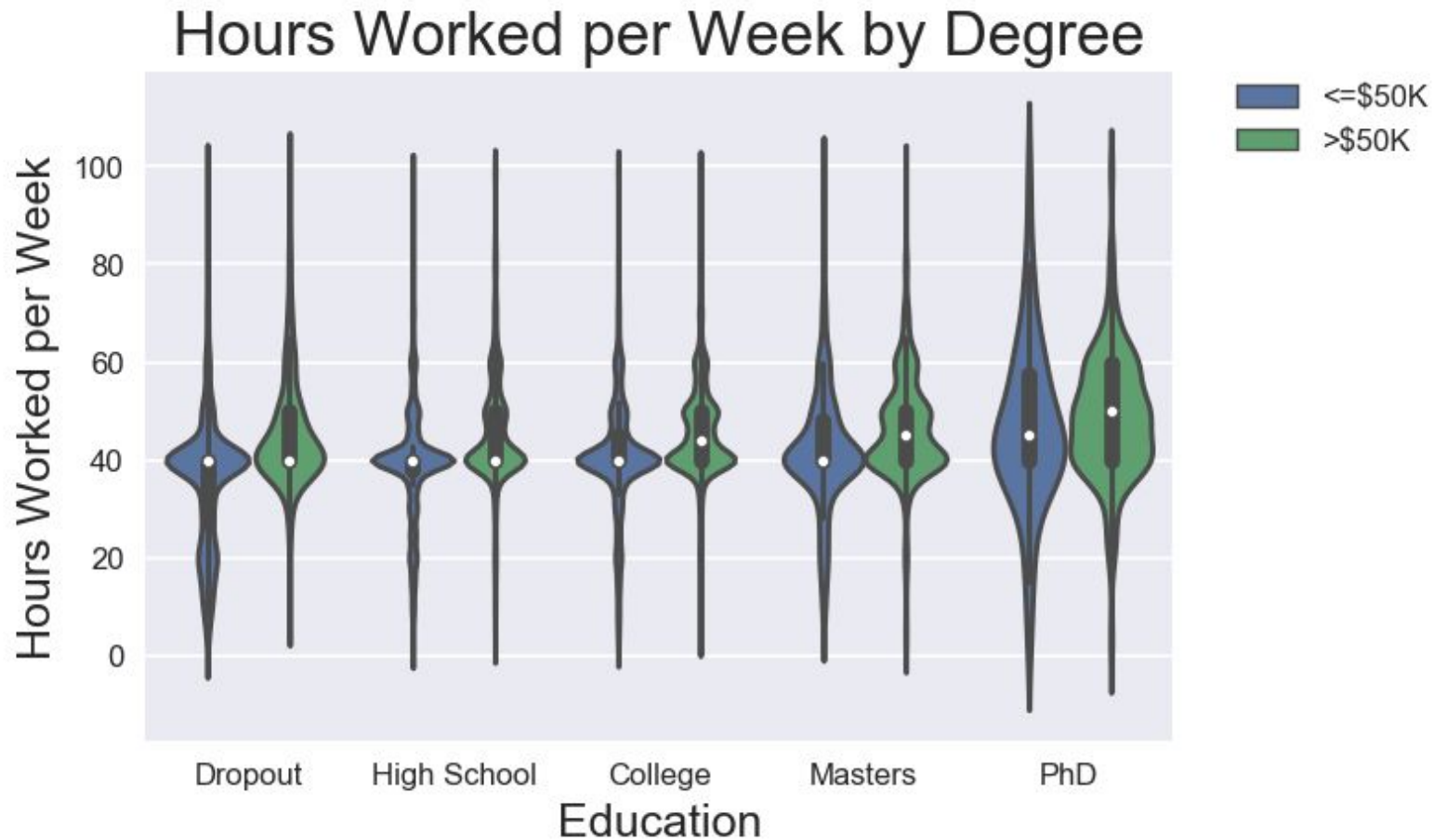
Data

- 1994 US Census

- Feature categories:
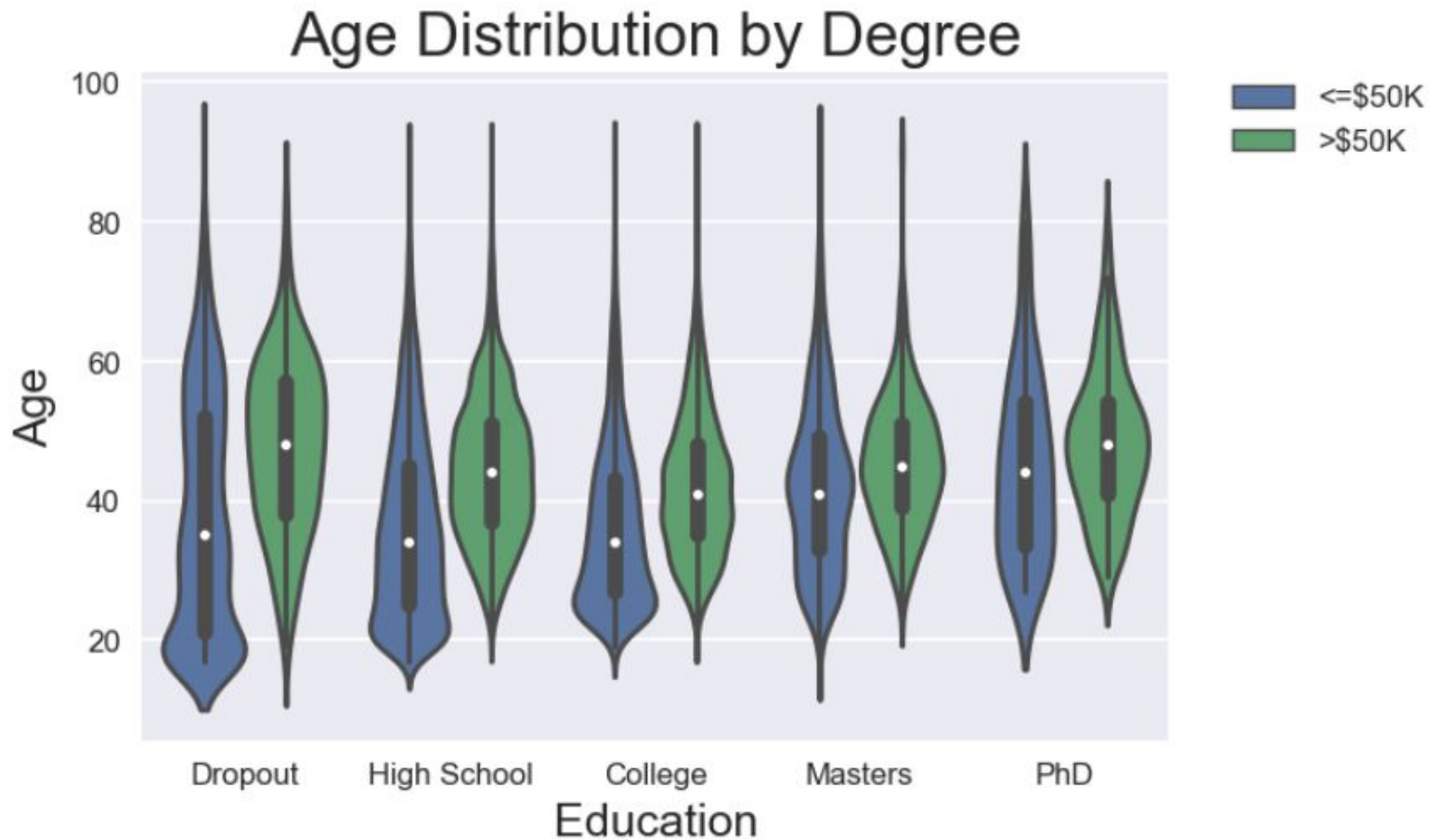 1) General
  2) Education
  3) Work history
  4) Social class

Initial Hypothesis

Hours Worked Weekly by Age & Degree

Initial Hypothesis

Hours Worked per Week by Degree

Initial Hypothesis

Age Distribution by Degree
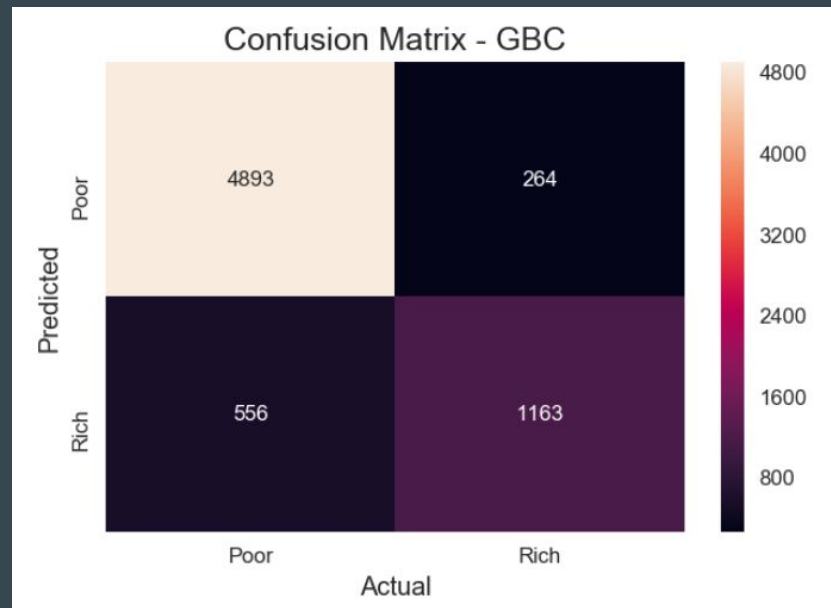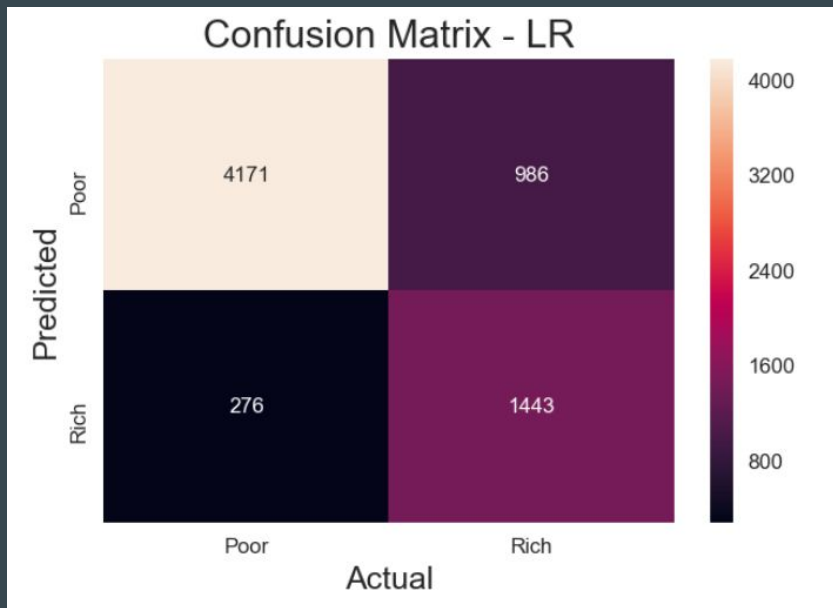
Model optimization

Other important indicators (RFE):

- Financial standing
- Marital status

Other optimization strategies:

- Clustering
- Oversampling

Optimized Models

Confusion Matrix - LR

	Poor	Rich
Poor	4171	986
Rich	276	1443

Confusion Matrix - GBC

	Poor	Rich
Poor	4893	264
Rich	556	1163

Ideal Demographic

- Married
- College graduate
- Full-time employment
- Established financial standing
- Old enough to have a midlife crisis

Questions?

Appendix - Model Performance (SMOTE)

Model	Training	Testing	Precision	Recall	F1
GBC	0.918	0.855	0.706	0.738	0.722
LR	0.831	0.806	0.581	0.83	0.684
BC	0.983	0.831	0.674	0.644	0.658
RF	0.975	0.938	0.678	0.693	0.685

Appendix - Model Performance (non-SMOTE)

Model	Training	Testing	Precision	Recall	F1
GBC	0.882	0.861	0.775	0.634	0.698
LR	0.847	0.843	0.736	0.593	0.657
BC	0.977	0.841	0.703	0.643	0.672
RF	0.863	0.85	0.788	0.559	0.654