



Tax Data Prediction

Christopher Hyek

7/21/2019

1994 Census Data Set

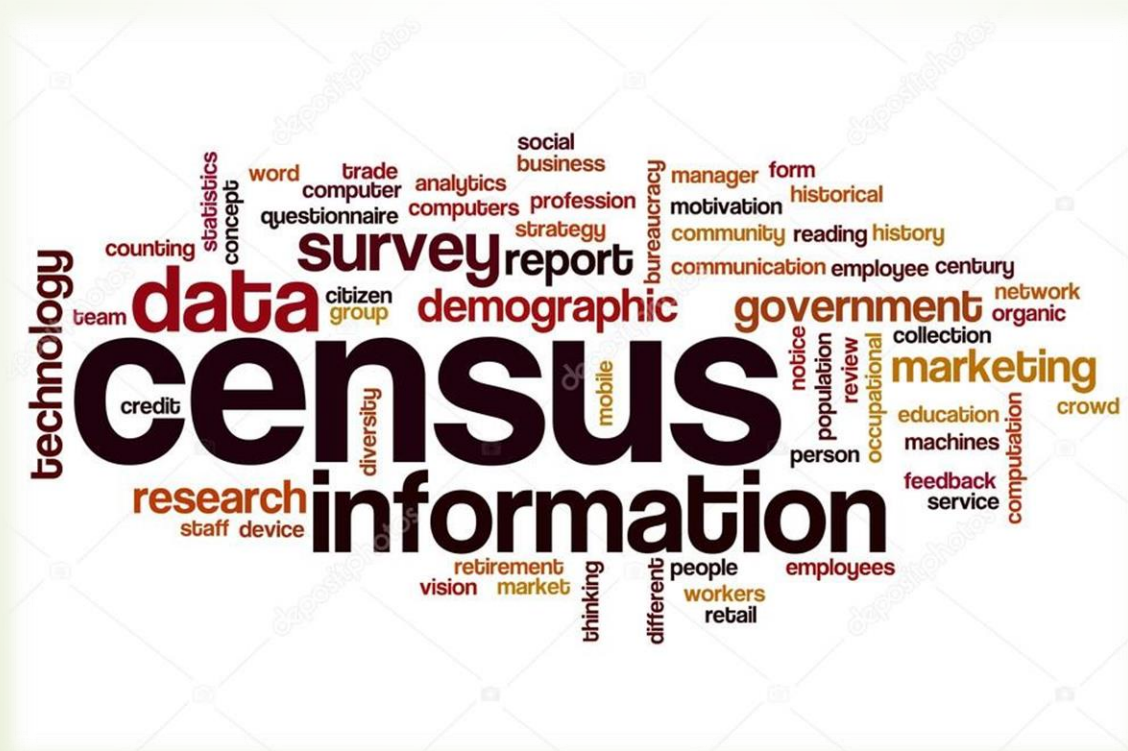
Goal

Predict annual salary based on data



Data Acquisition

To predict tax data we obtain
census bureau information

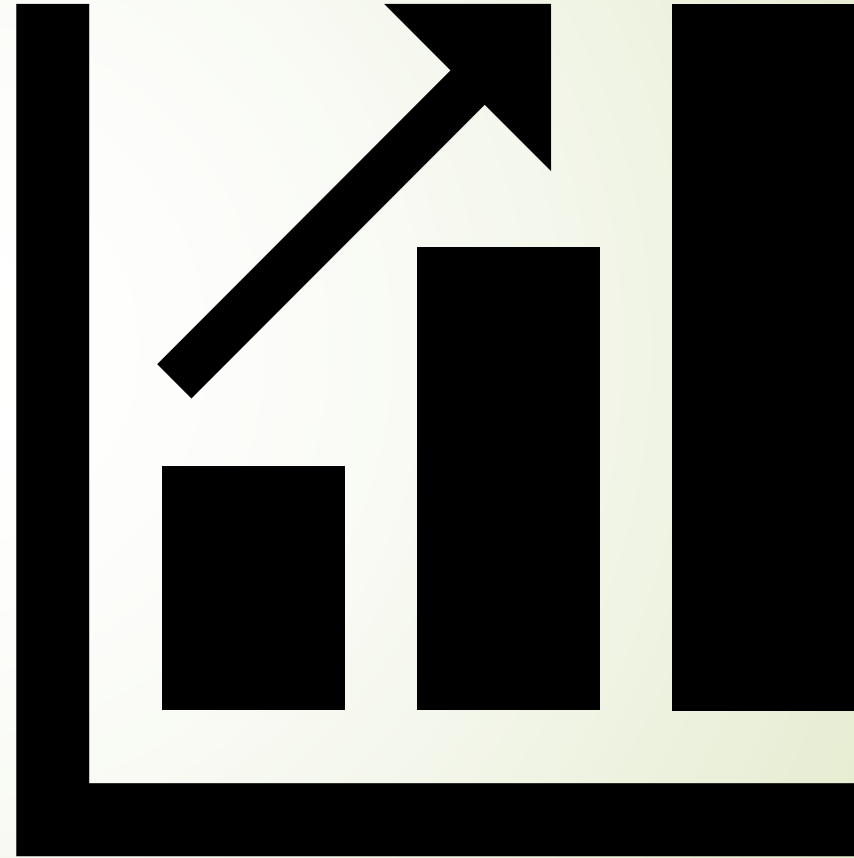


Scrubbing

Removing serious outliers

Using proper labels

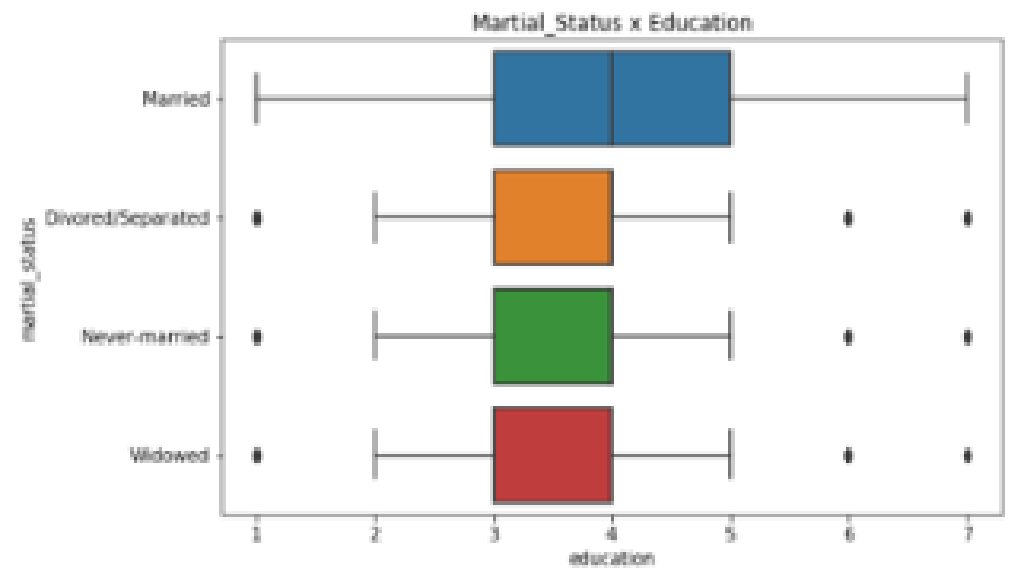
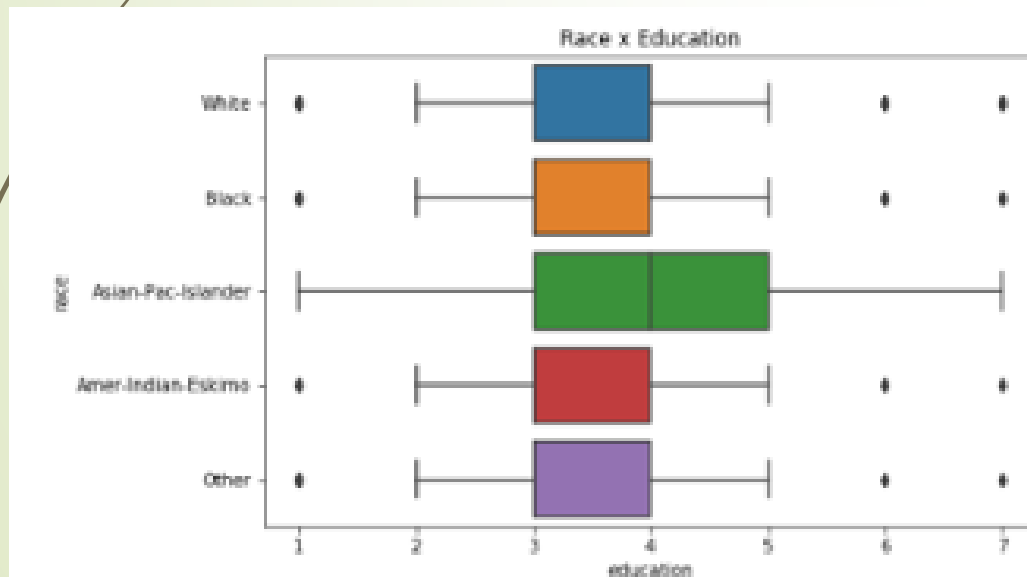
Removing typos

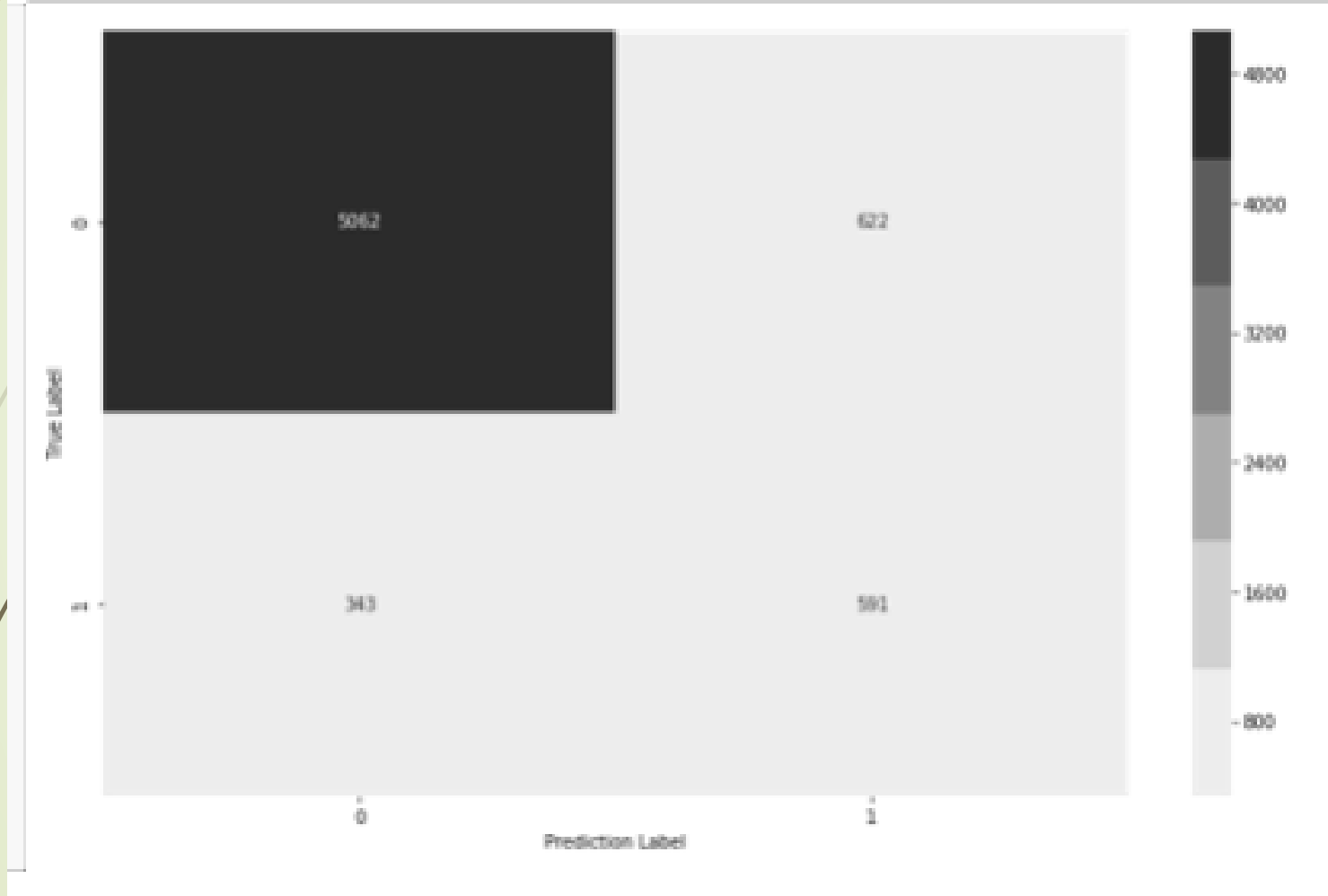


Exploring

There are very peculiar patterns from the data

Most of it is categorical so very few hard connections





Modeling

Logistic Regression

Focus on precision

Validate Data



Interpreting

85% accuracy

48% precision

67% accuracy





Recommendations

- For better Precision we need:
 - To focus on either marital status over relationship
 - To redo Capital gains and capital loss for clarity
 - Properly label for unemployed and never-worked people
- Why?
 - Through testing they improved our model by removing these.



Future Recommendations

We should:

- Attempt a new model to verify

- Try different constraints

- Refine the survey data and try again



Thank you