# Predicting Income Greater and Lesser than $50k

A Study of US Census Data (1994-95)

By: Terri John

# Problem Statement and Background

- US Census Data results empower the USG to make data driven decisions regarding allocation of funding.
- Improved understanding of demographic characteristics of subpopulation.
- Income disparities.

# Methodology

## 1

### Data Cleaning and EDA

Exploring the data and preparing it for modeling

Missing Data

## 2

### Feature Engineering

Categorical Features

## 3

### Modeling

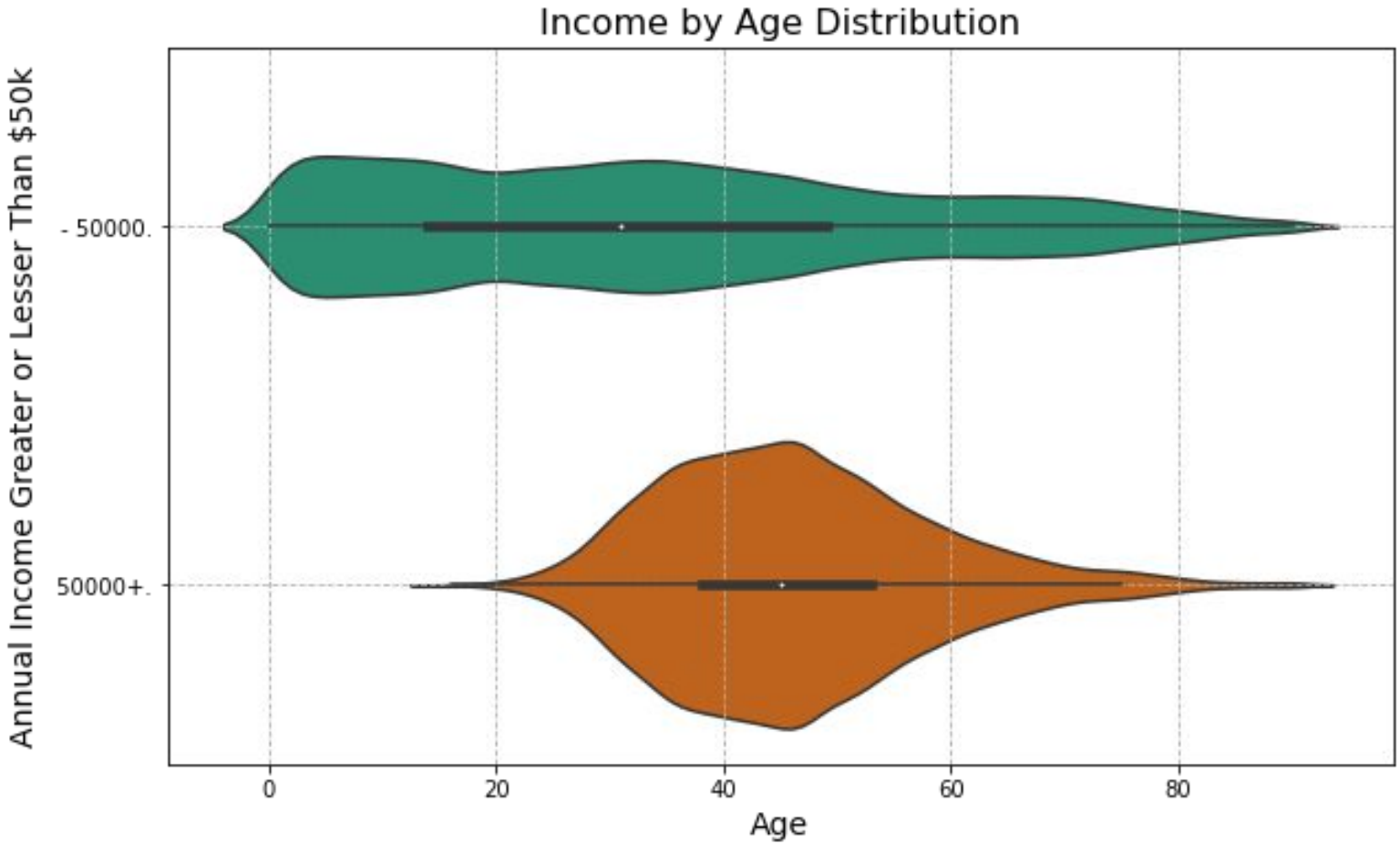
Binary Classification Model to predict income

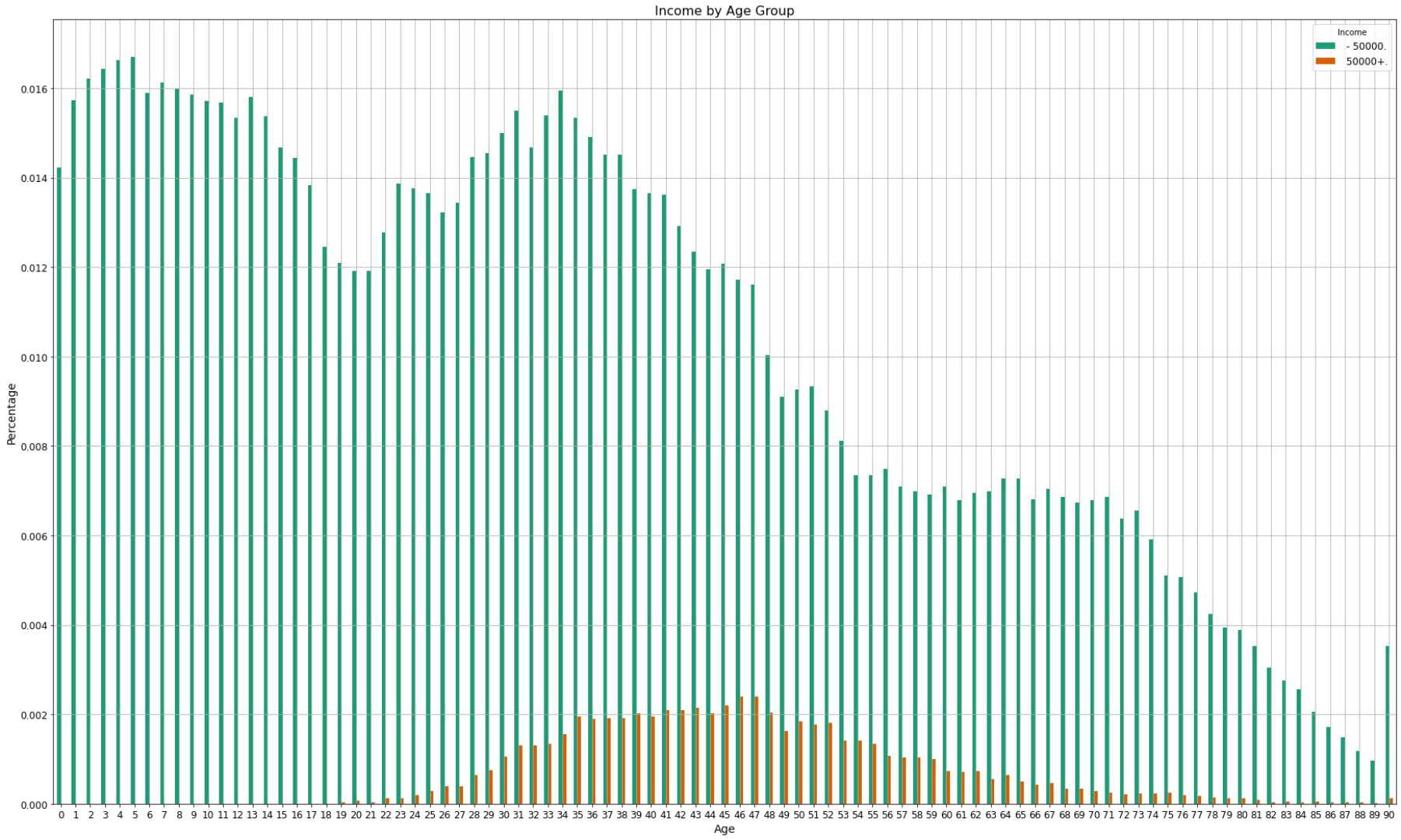
## 4

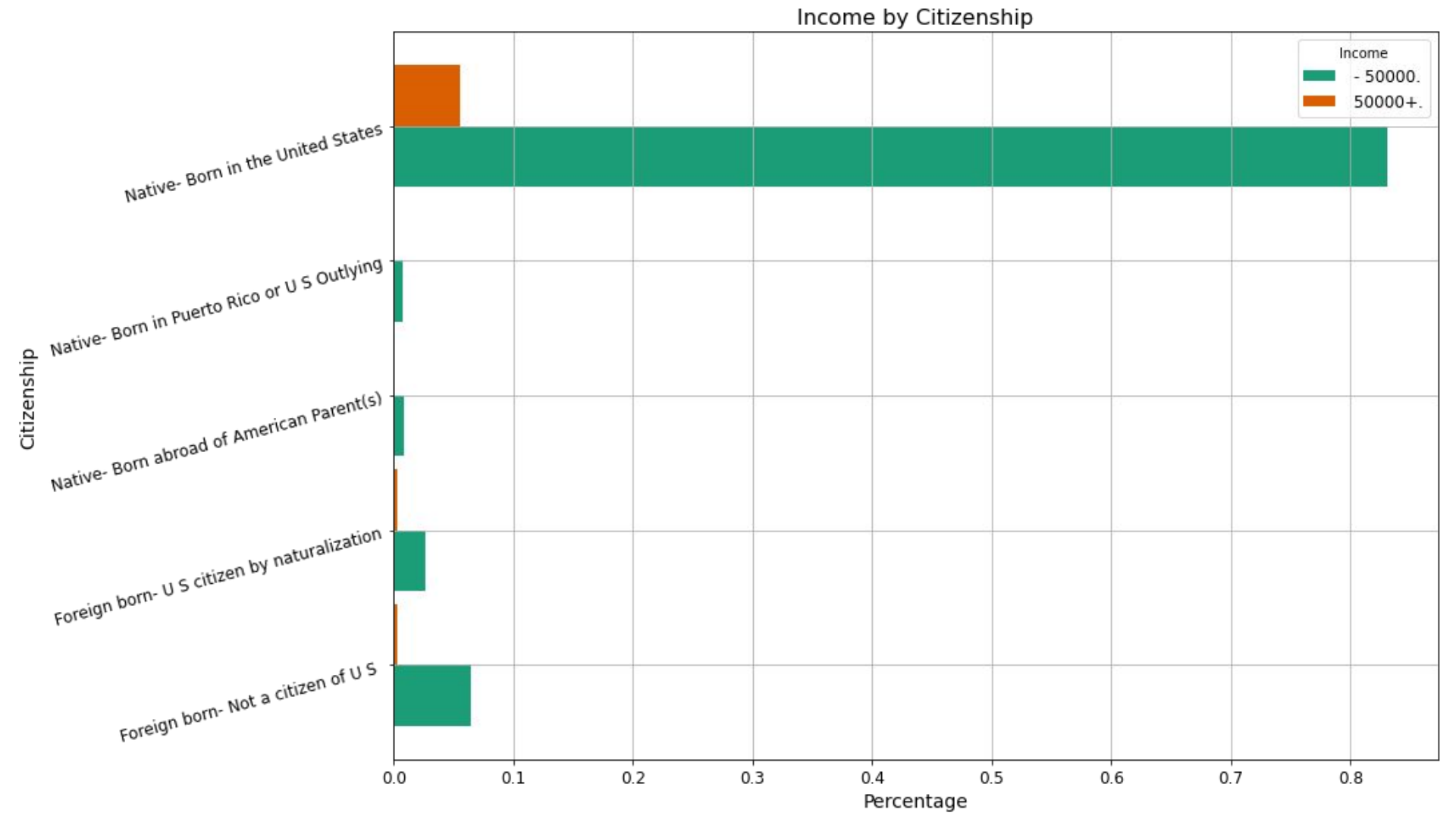
### Analysis

Model Performance

Key Takeaways

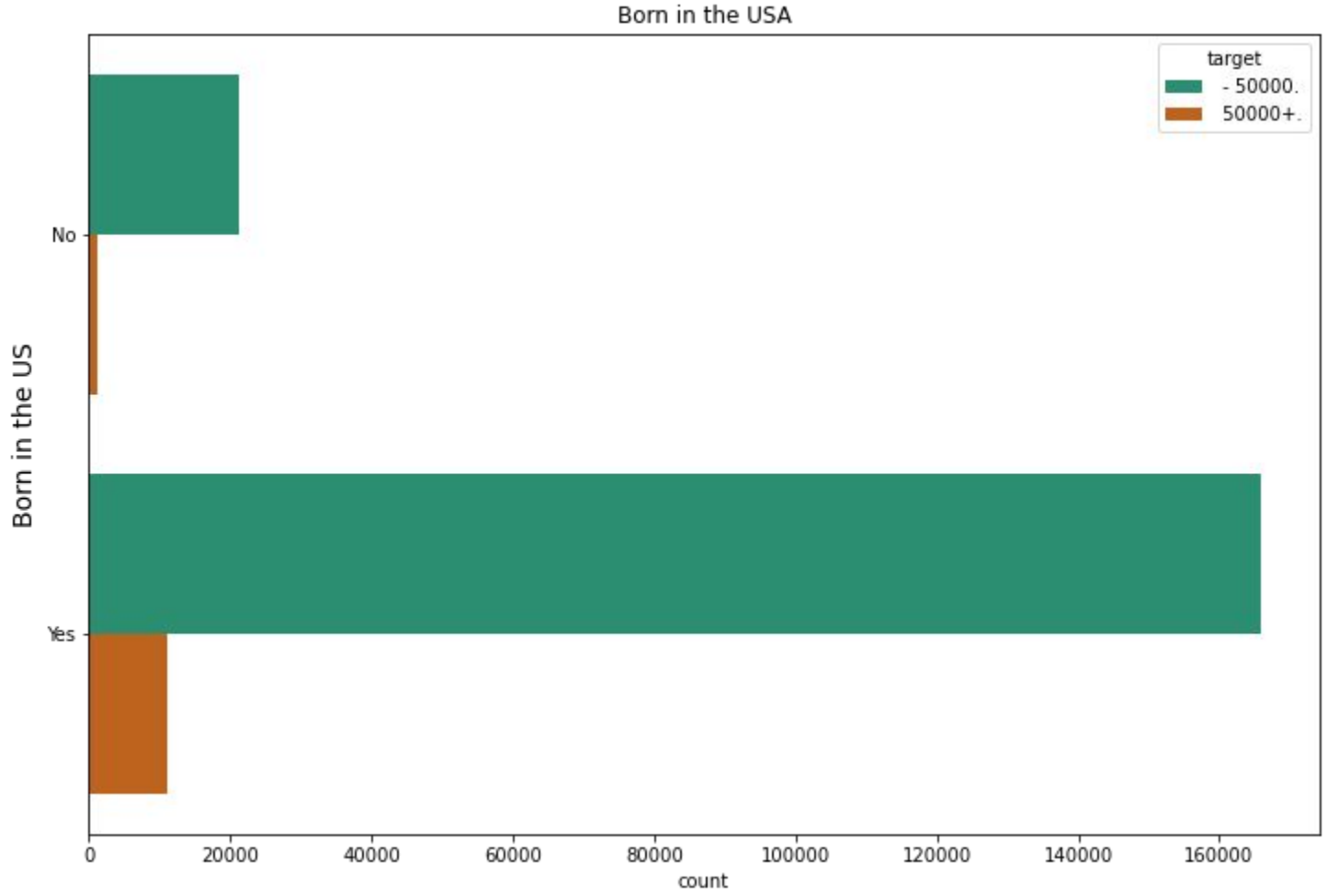

Income by Age Distribution
Annual Income Greater or Lesser Than $50k
- 50000.
50000+.
0
20
40
60
80
Age

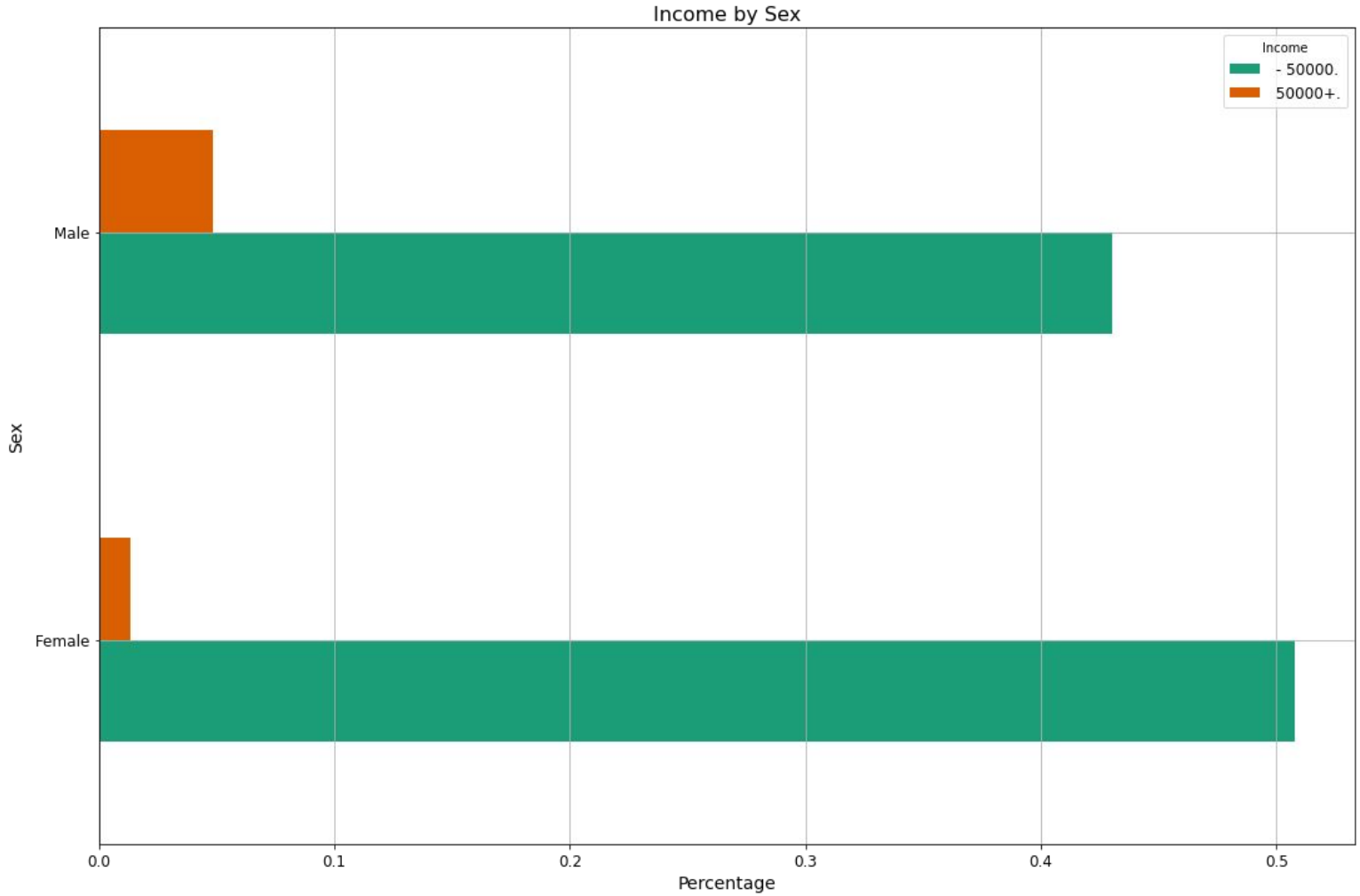

Income by Age Group
Income
- 50000.
50000+.
Percentage
0.000
0.002
0.004
0.006
0.008
0.010
0.012
0.014
0.016
Age
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90

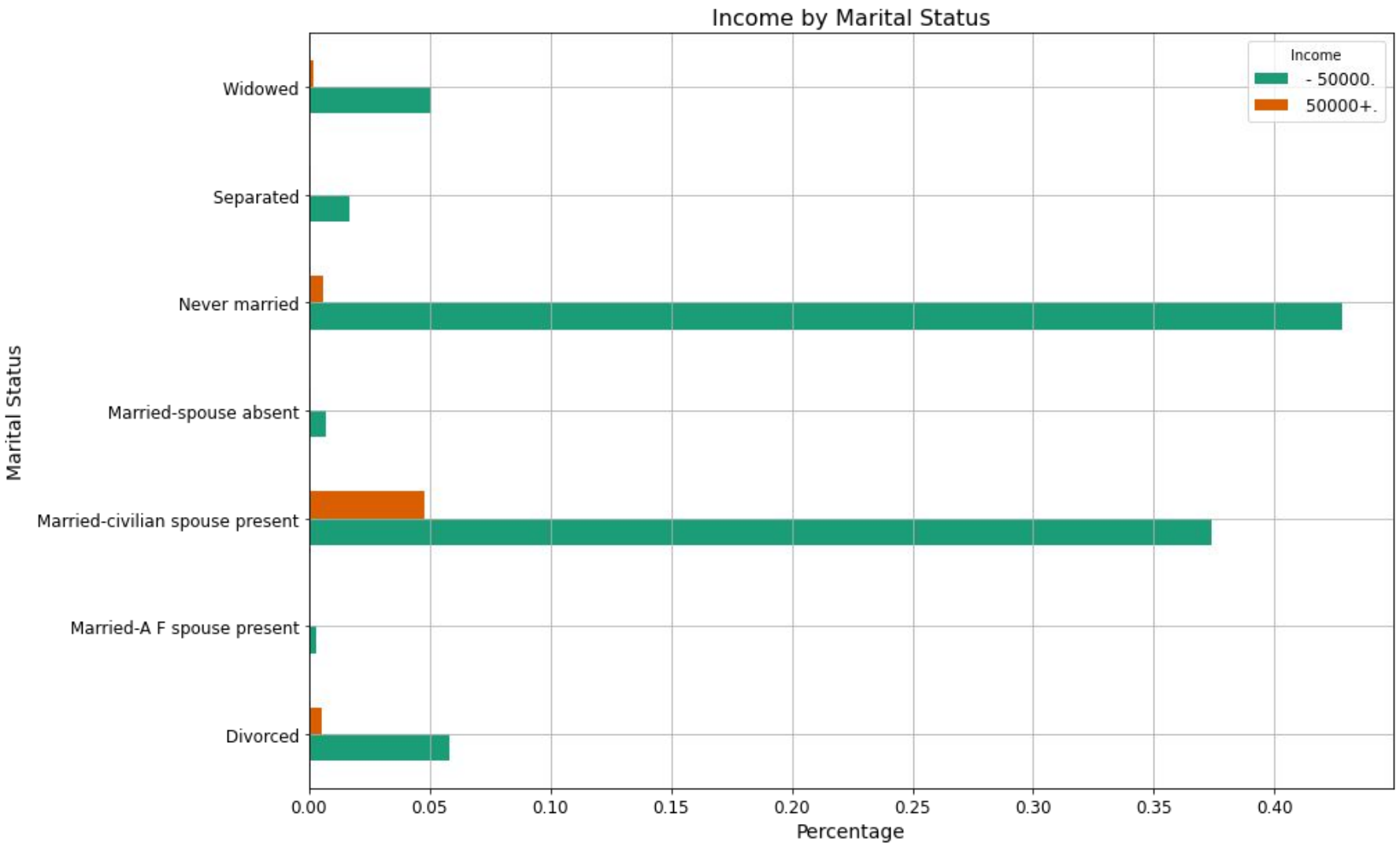

Income by Citizenship
Income
- 50000.
50000+.
Native- Born in the United States
Native- Born in Puerto Rico or U S Outlying
Native- Born abroad of American Parent(s)
Foreign born- U S citizen by naturalization
Foreign born- Not a citizen of U S
Citizenship
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
Percentage

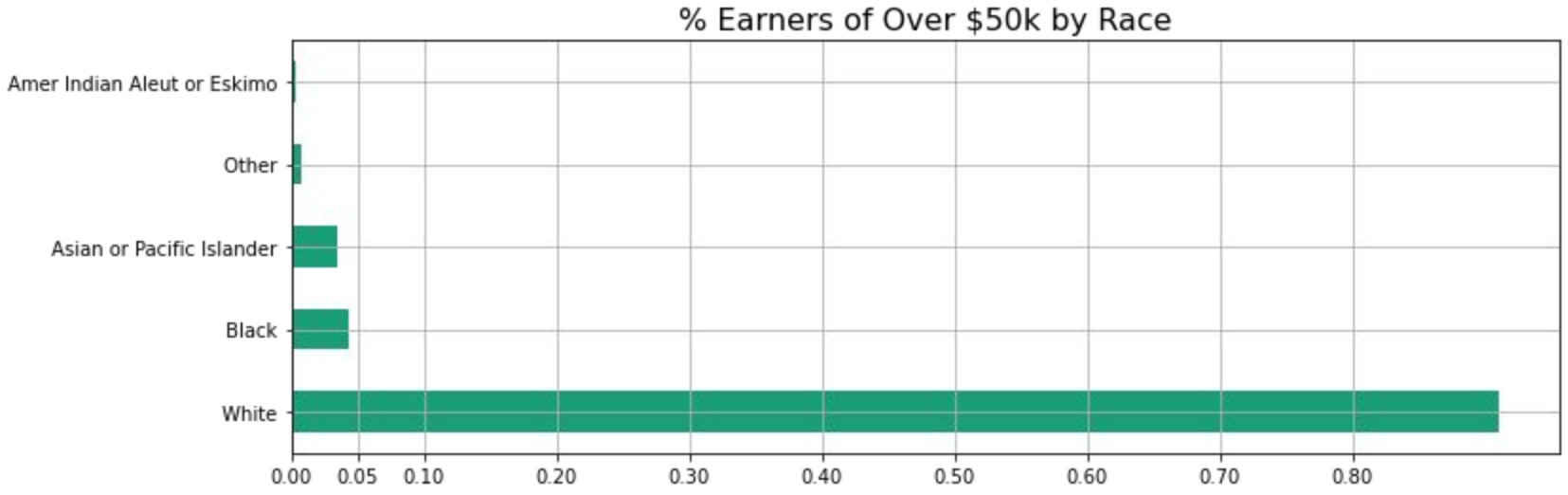
# Citizenship

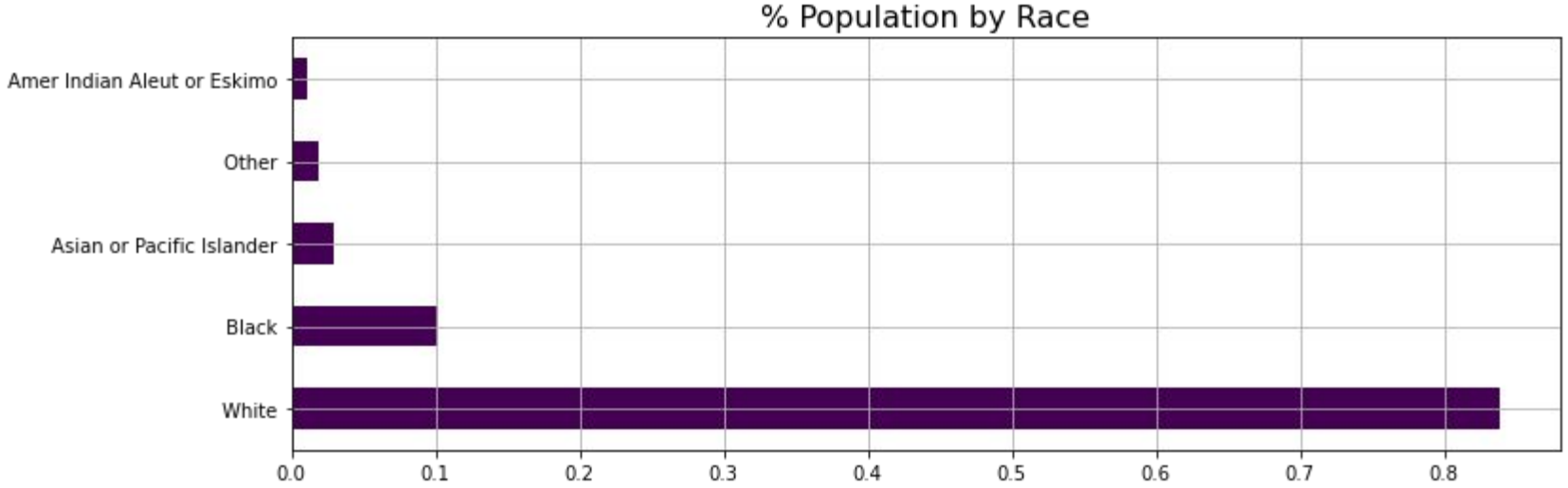
| Citizenship Status | Population % | % of Group +$50k |
|---|---|---|
| Native- Born in the United States | 88.7% | 6% |
| Foreign born- Not a citizen of U S | 6.7% | 3% |
| Foreign born- U S citizen by naturalization | 2.9% | 10% |
| Native- Born abroad of American Parent(s) | .8% | 7% |
| Native- Born in Puerto Rico or U S Outlying | .7% | 2% |


Born in the USA
target
- 50000.
50000+.
Born in the US
No
Yes
0
20000
40000
60000
80000
100000
120000
140000
160000
count


Income by Sex
Income
- 50000.
50000+.
Male
Female
Sex
0.0
0.1
0.2
0.3
0.4
0.5
Percentage


Income by Marital Status
Income
- 50000.
50000+.
Widowed
Separated
Never married
Married-spouse absent
Married-civilian spouse present
Married-A F spouse present
Divorced
Marital Status
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0.40
Percentage


% Earners of Over $50k by Race
Amer Indian Aleut or Eskimo
Other
Asian or Pacific Islander
Black
White
0.00
0.05
0.10
0.20
0.30
0.40
0.50
0.60
0.70
0.80

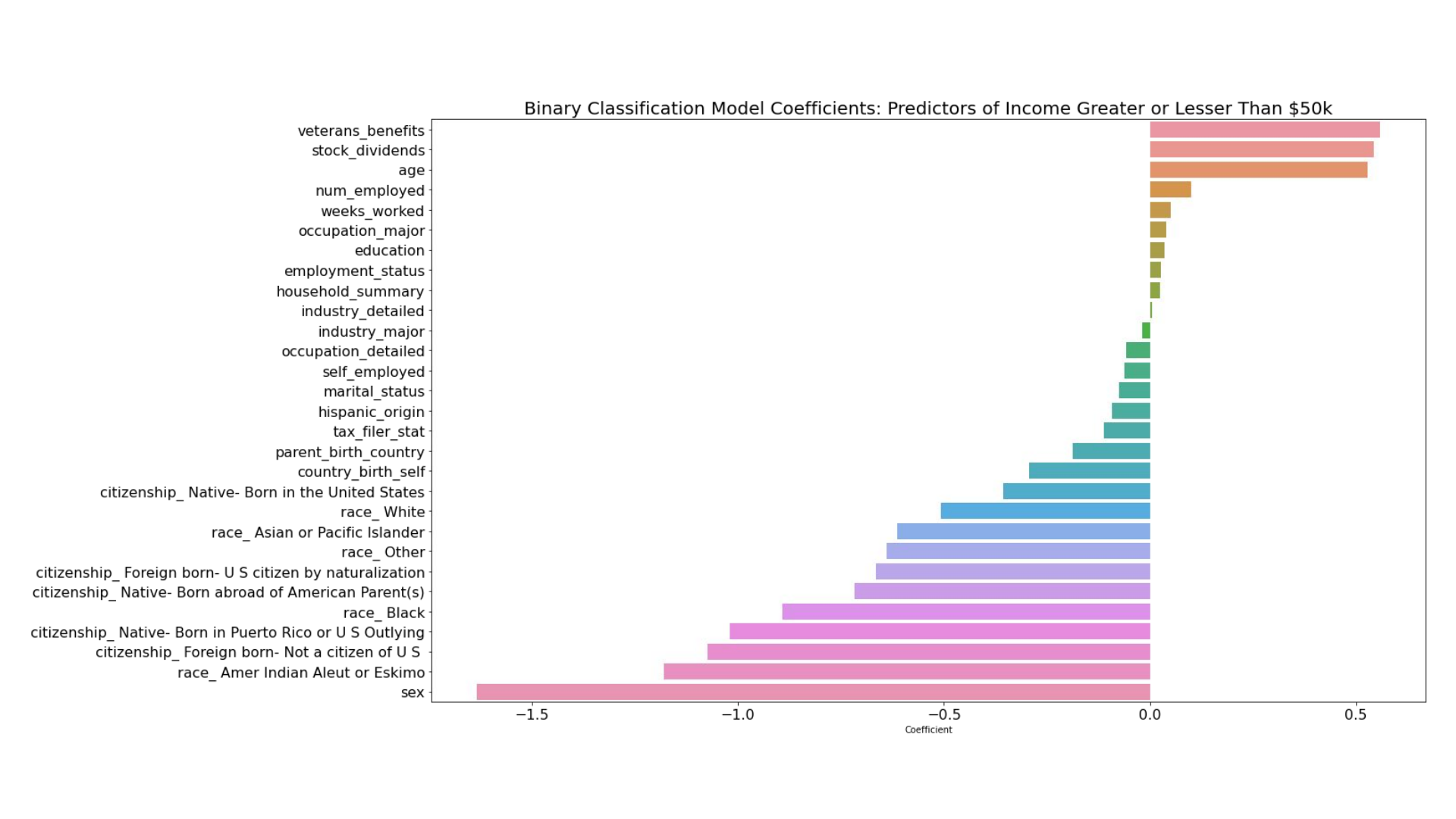

% Population by Race
Amer Indian Aleut or Eskimo
Other
Asian or Pacific Islander
Black
White
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8

# A Tale of Two Models

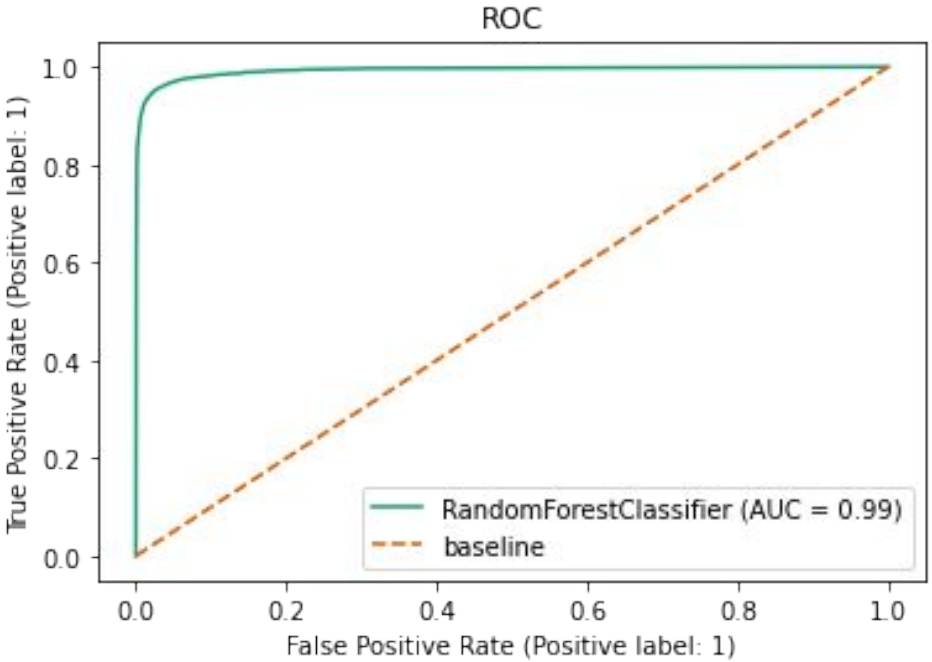
## Predicting Annual Income Greater/Lesser Than $50k

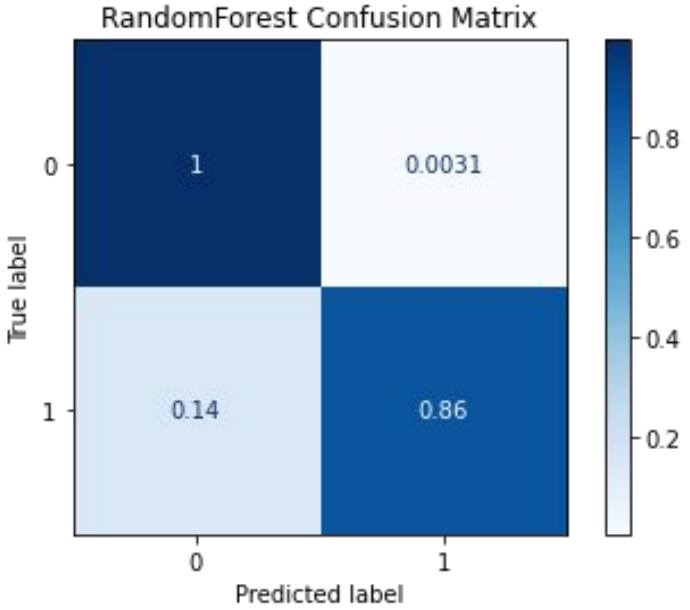
| | Train Accuracy | Test Accuracy | Recall | F1 Score |
|---|---|---|---|---|
| **Logistic Regression** | 94.5% | 94.5% | 0.24 | 0.35 |
| **RandomForest** | 99.2% | 98.9% | 0.86 | 0.90 |

*The baseline model assumes a 93% accuray score, based on the majority class.


Binary Classification Model Coefficients: Predictors of Income Greater or Lesser Than $50k
veterans_benefits
stock_dividends
age
num_employed
weeks_worked
occupation_major
education
employment_status
household_summary
industry_detailed
industry_major
occupation_detailed
self_employed
marital_status
hispanic_origin
tax_filer_stat
parent_birth_country
country_birth_self
citizenship_ Native- Born in the United States
race_ White
race_ Asian or Pacific Islander
race_ Other
citizenship_ Foreign born- U S citizen by naturalization
citizenship_ Native- Born abroad of American Parent(s)
race_ Black
citizenship_ Native- Born in Puerto Rico or U S Outlying
citizenship_ Foreign born- Not a citizen of U S
race_ Amer Indian Aleut or Eskimo
sex
−1.5
−1.0
−0.5
0.0
0.5
Coefficient

# RandomForest Model



# Conclusion and Recommendations

## Key Takeaways

- RandomForest Model
- Decisive Characteristics
- Why does it matter?

## Next Steps

- Bring in more data to fill null values.
- Cross compare features
- Pull in data for parent and self birth country

# Sources

Data for this project was provided by Dataiku.
Original data is from the US Census Bureau: http://www.census.gov/ftp/pub/DES/www/welcome.html

**Outside sources:**

- Forbes: https://www.forbes.com/sites/marisadellatto/2021/10/05/single-adults-make-less-money-than-partnered-ones-study-says/?sh=360d12ad454f

# Thanks!

**Do you have any questions?**

tjohn07@gmail.com

https://www.linkedin.com/in/terri-john/

https://www.terrijohn.com

**CREDITS:** This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik**

Please keep this slide for attribution