

The background is a dark navy blue. It features several thin, gold-colored lines that form abstract, angular shapes. These lines radiate from the central text box, extending towards the corners and edges of the frame, creating a sense of dynamic movement and modern design.

# PREDICTING INCOME

A Classification Project - By Kyle Turner

The background features several thin, light-colored lines that intersect to form various geometric shapes, including triangles and polygons, scattered across the dark blue field.

\$50,000

# THE CLIENT



**GROW SMART**

- Non-profit
- Future Planning
- US Roll out

# DATA SCIENCE SOLUTION

- Classification Model
- Use Census Data
  - $>50K$  /  $\leq 50K$



# THE DATA



## INCOME CLASSIFICATION

- Census data from Kaggle
- 32,561 rows/ 15 columns



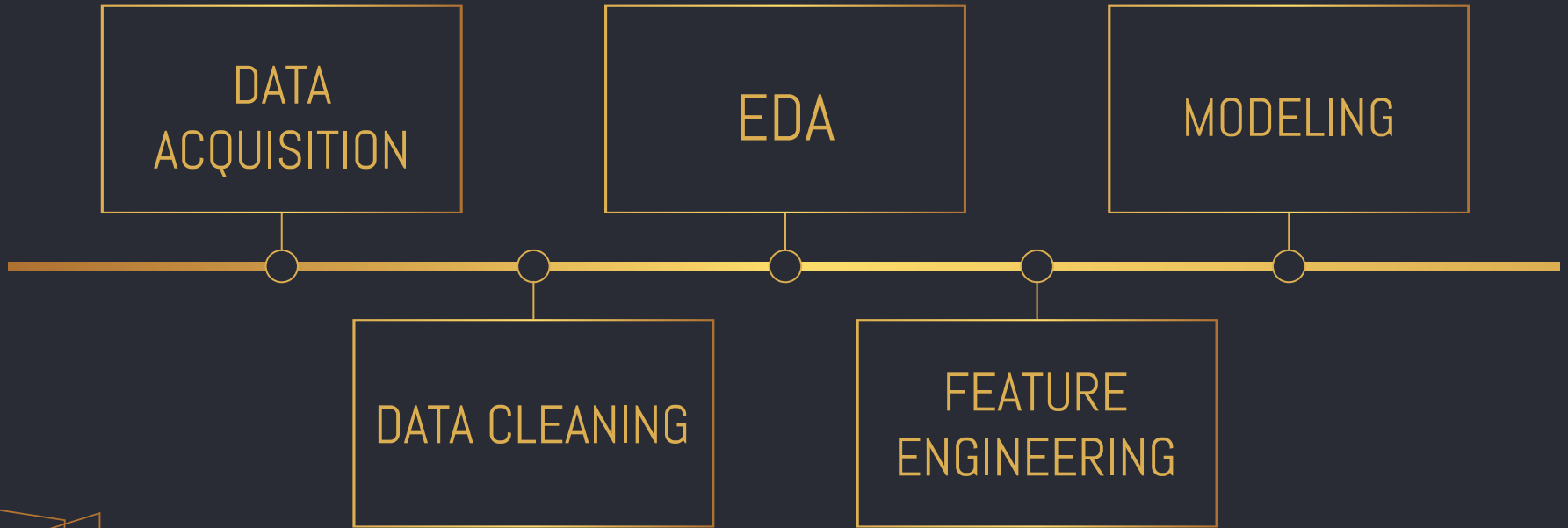
## TARGET

Income:

0= ( $\leq 50K$ )

1= ( $> 50K$ )

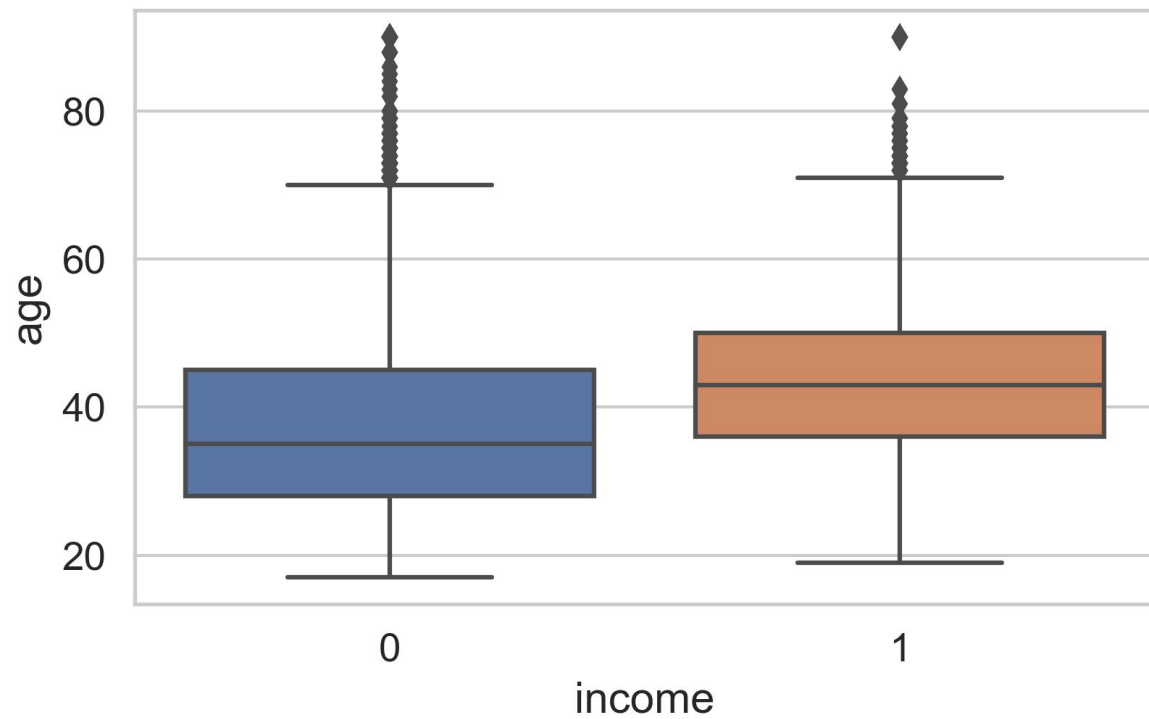
# METHODOLOGY



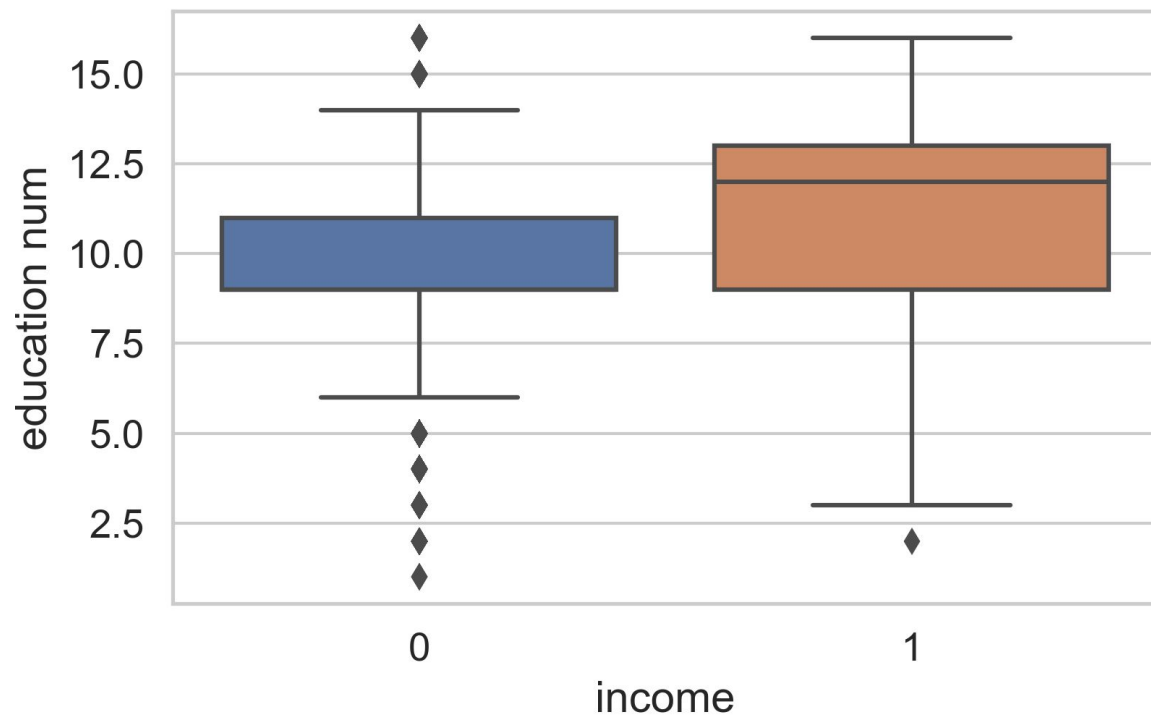
# FEATURES

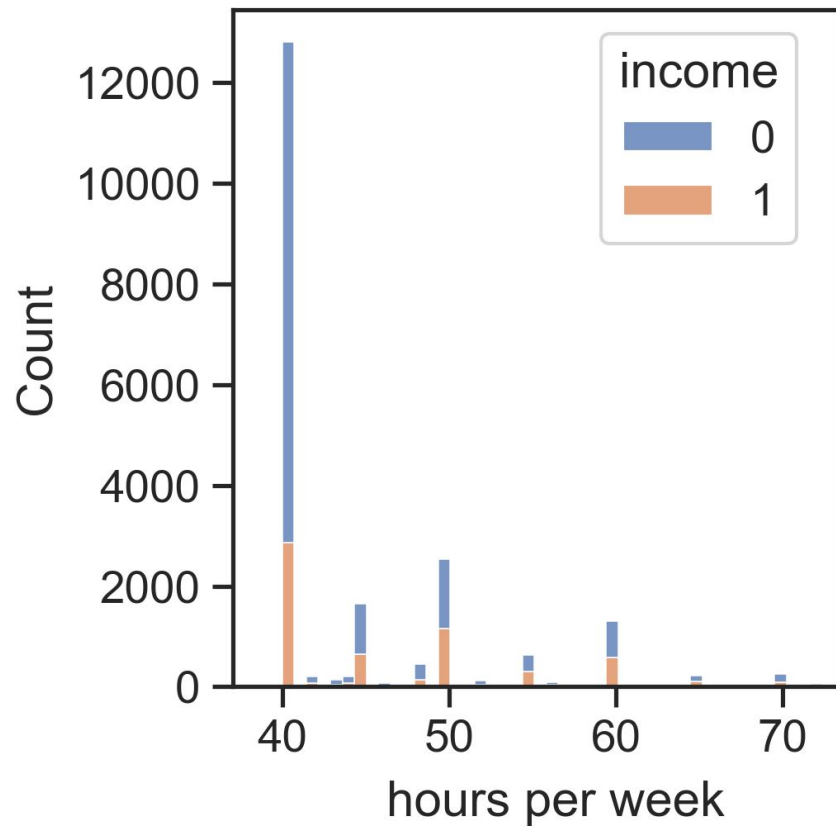
- Age
- Sex
- Race
- Education
- Occupation
- Marital Status
- Hours Worked



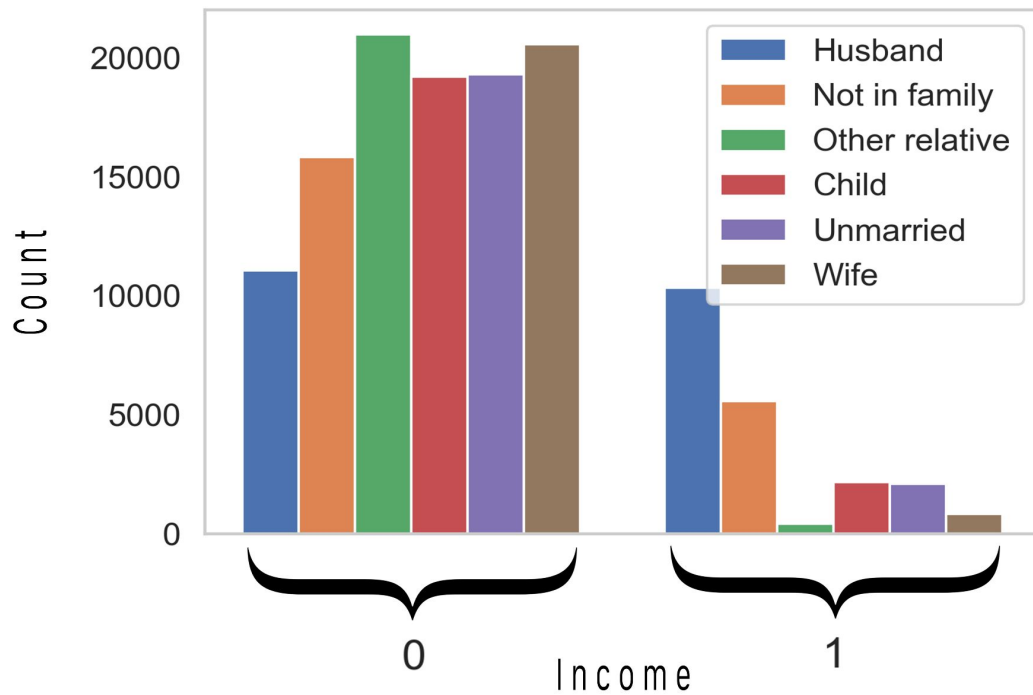




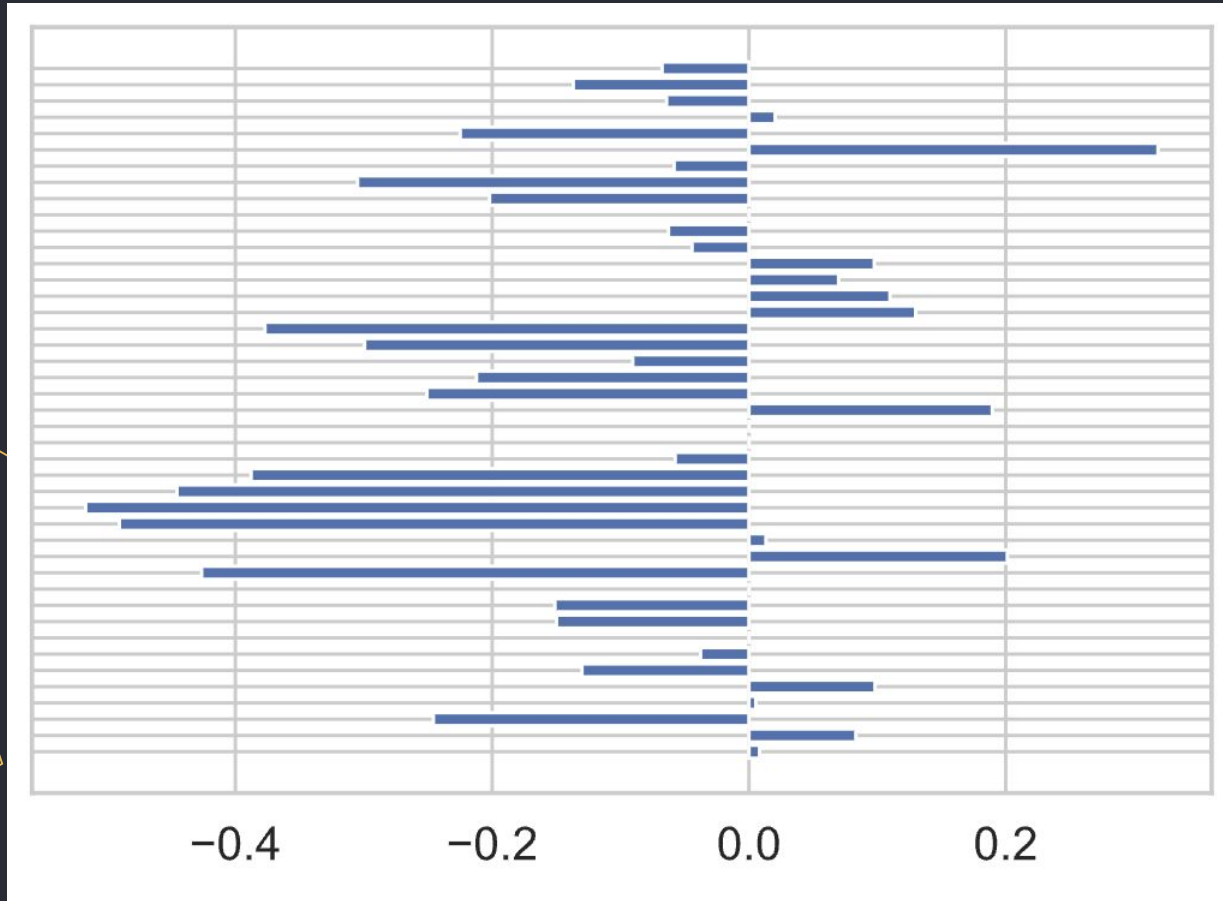




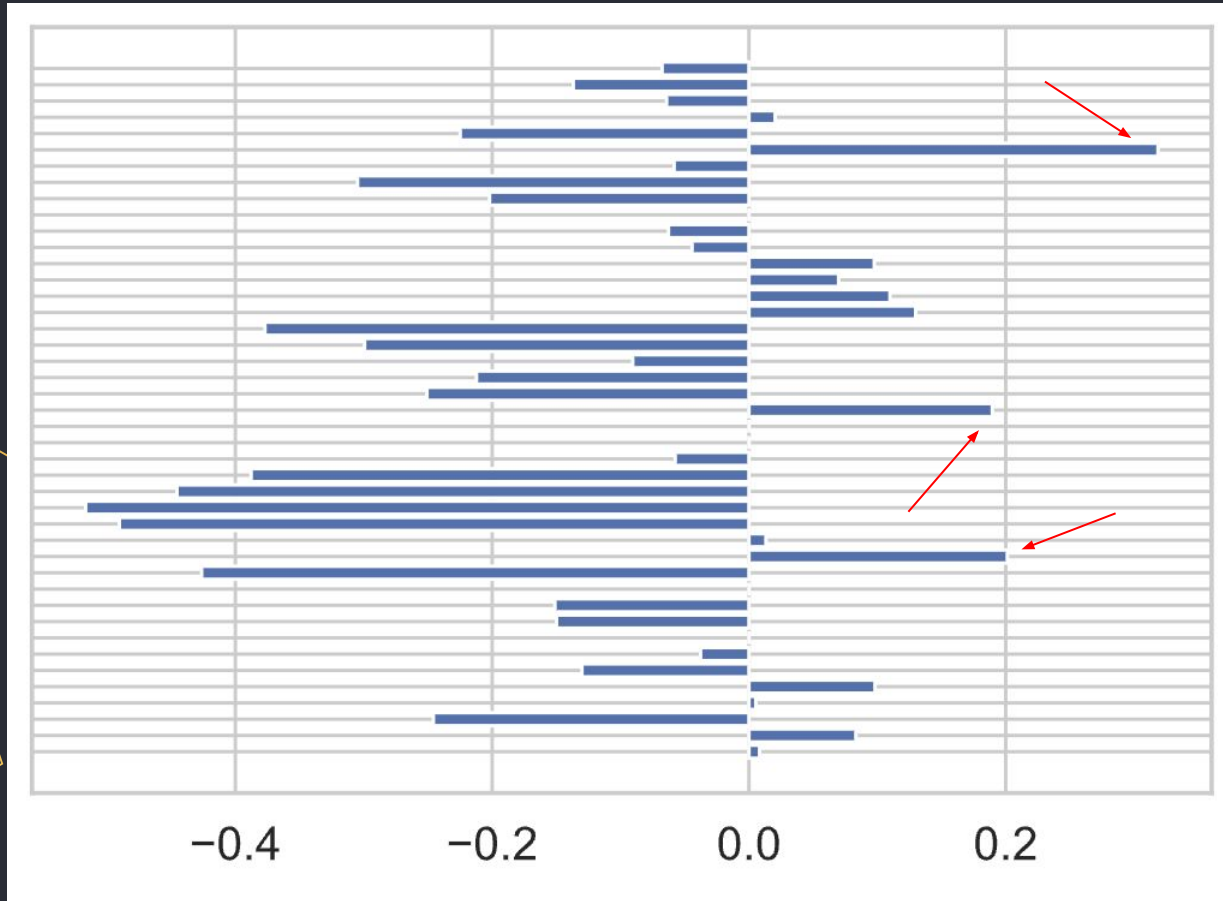
# Family Relationship



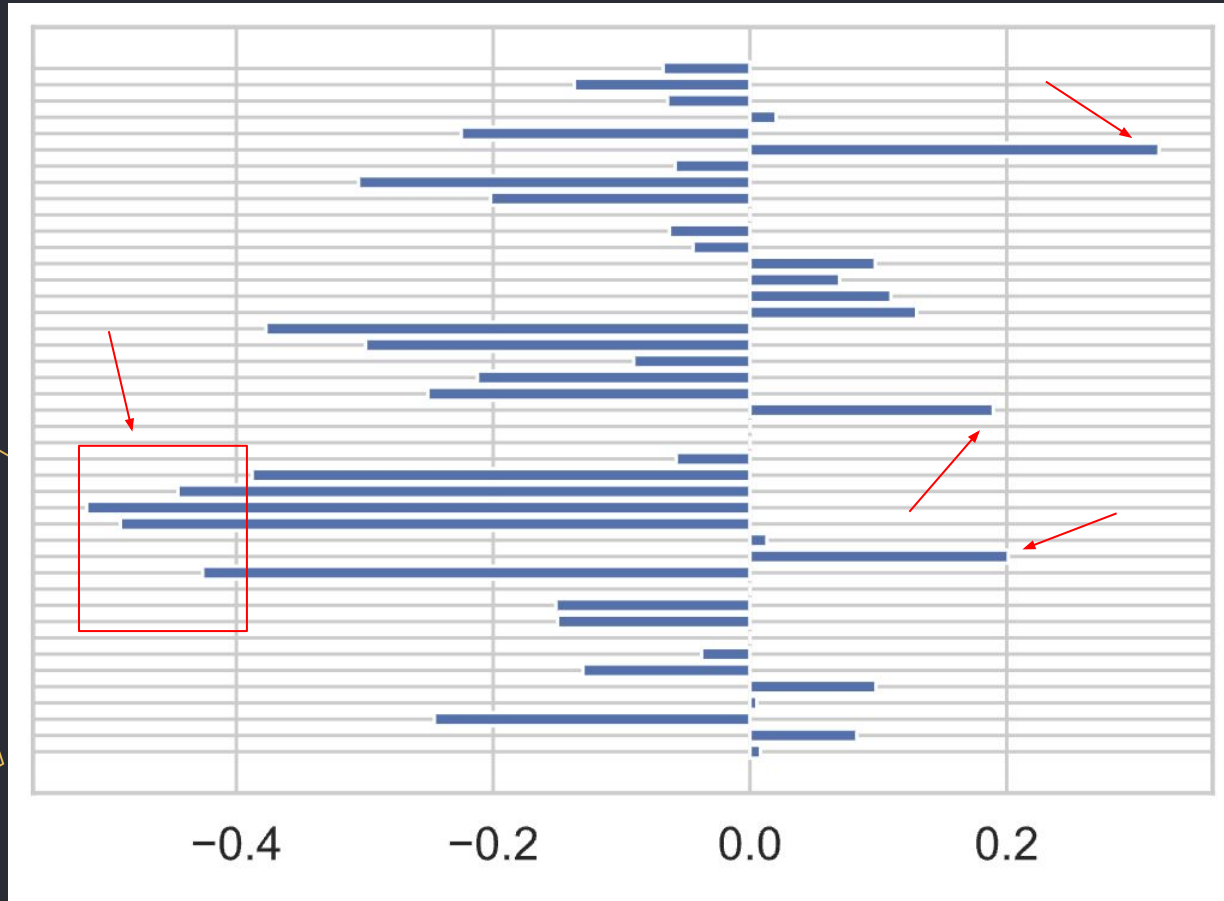
# COEFFICIENTS- using divide by 4 rule



# COEFFICIENTS- using divide by 4 rule



# COEFFICIENTS- using divide by 4 rule



# STANDOUT COEFFICIENTS

|                             |       |   |        |
|-----------------------------|-------|---|--------|
| Relationship-<br>Wife       | 0.322 | Marital Status-<br>Separated                | -0.438 |
| Occupation-<br>Exec Manager | 0.189 | Marital Status-<br>Never Married            | -0.509 |
| Workclass-<br>Without Pay   | 0.206 | Marital Status-<br>Divorced                 | -0.419 |
|                             |       | Marital Status-<br>Married<br>Spouse Absent | -0.483 |

# MODELING

|    | LOGISTIC REGRESSION |    | LOGISTIC REGRESSION CV |    | RIDGE CLASSIFIER |    | LOGISTIC REGRESSION OVERSAMPLE |    | LOGISTIC REGRESSION CV OVERSAMPLE |    | RIDGE CLASSIFIER OVERSAMPLE |    |
|----|---------------------|----|------------------------|----|------------------|----|--------------------------------|----|-----------------------------------|----|-----------------------------|----|
|    | ACCURACY            | F1 | ACCURACY               | F1 | ACCURACY         | F1 | ACCURACY                       | F1 | ACCURACY                          | F1 | ACCURACY                    | F1 |
| F1 | 0.641               |    | 0.638                  |    | 0.632            |    | 0.671                          |    | 0.638                             |    | 0.632                       |    |
|    | 0.802               |    | 0.801                  |    | 0.800            |    | 0.776                          |    | 0.774                             |    | 0.773                       |    |



# FINAL MODEL

LOGISTIC REGRESSION TEST SCORES:

Accuracy 0.796

F1 0.614

# FUTURE WORK

- Add More Features
- Different Feature Engineering



A grayscale photograph of architectural blueprints being worked on. Two hands are visible: one on the left holding a ruler, and one on the right holding a pencil. The blueprints show a cross-section of a building with a gabled roof. Various technical labels are present, including 'Roof-- 407ap', 'Garagr roof', 'Over.', 'AP', '7456 Garage', 'FL', '330a Door open', '17D', 'NGa7', and '2.60'. The image is overlaid with a semi-transparent dark blue layer and the text 'THANK YOU' in large, bold, yellow capital letters. Thin yellow geometric lines are also scattered across the image.

# THANK YOU

