

# Predicting Student's Income

**By: AUDREY EMERIBE**



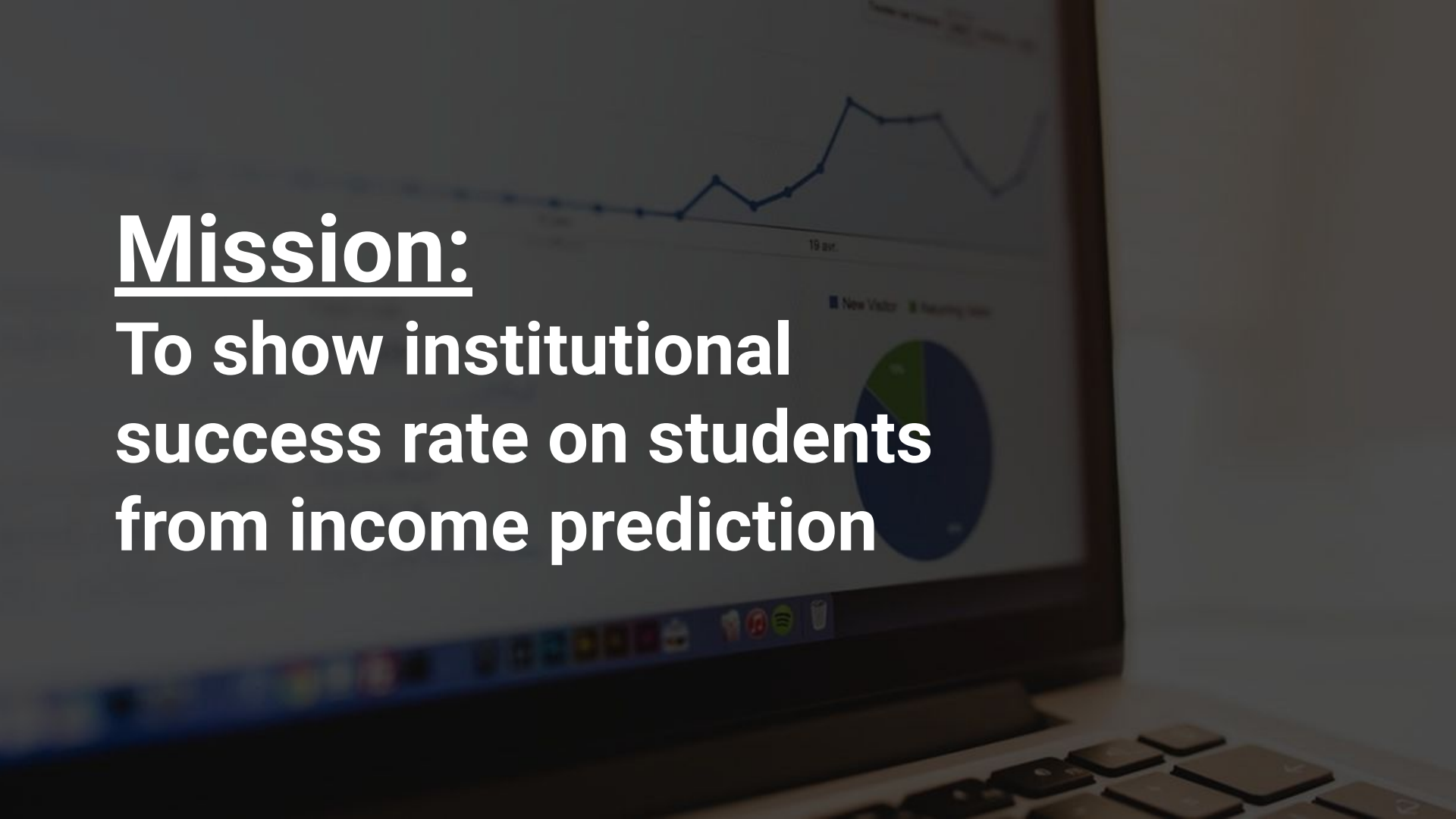
# BUSINESS QUESTION:

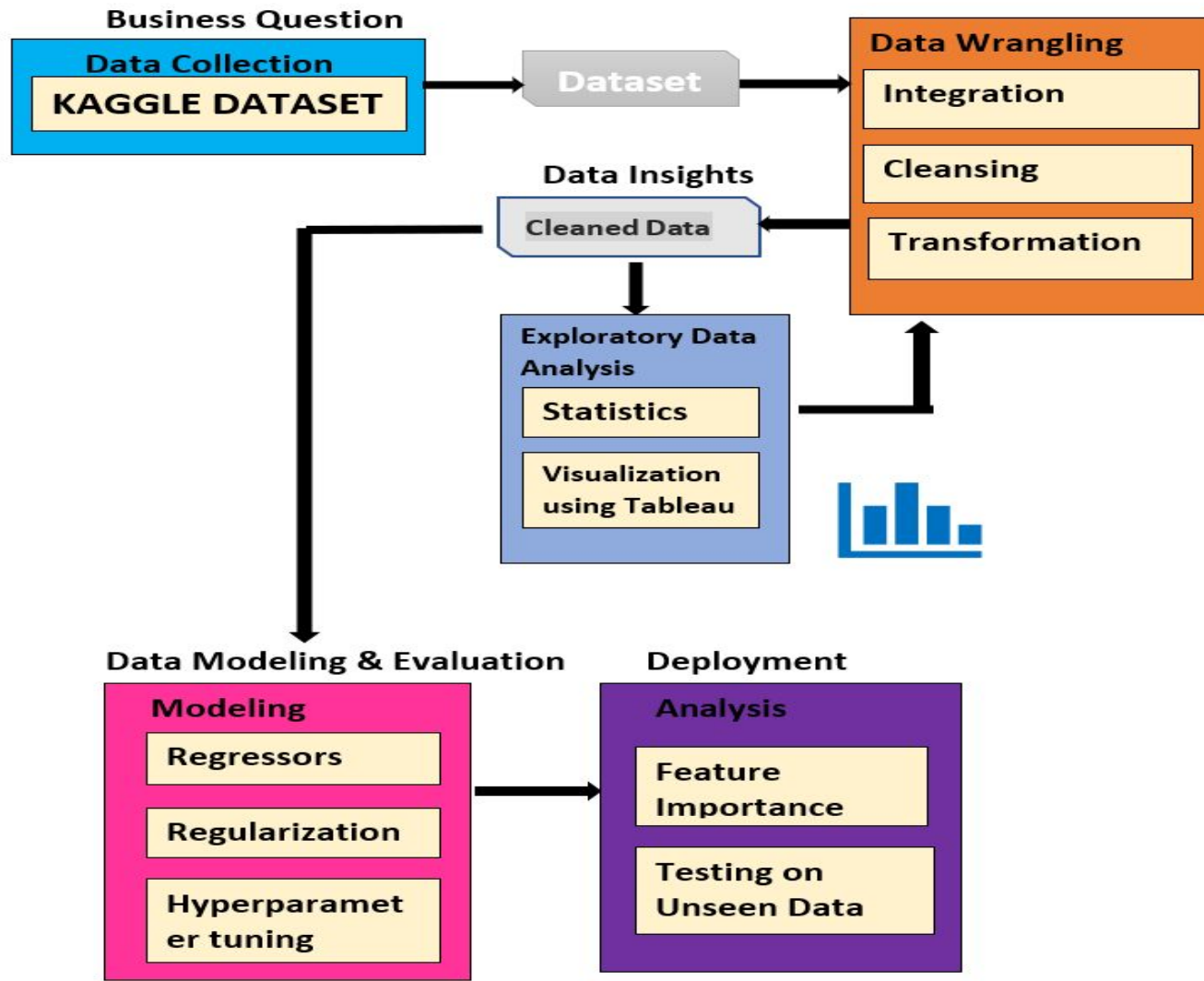
**How can I predict student's earning after their enrollment in United States institutions of higher education using machine learning?**



# Mission:

To show institutional  
success rate on students  
from income prediction





## METHODOLOGY FLOWCHART

# MICROSOFT PROFESSIONAL CAPSTONE DATASET

- The data was compiled from various sources and made publicly available by the United States Department of Education



- The compiled dataset was made available by Harsh Sharma on Kaggle platform



# DATA CONTENTS:

## BEFORE EDA:

- 26, 299 rows
- 297 columns of programs, degrees, SAT score, school characteristics

## AFTER EDA:

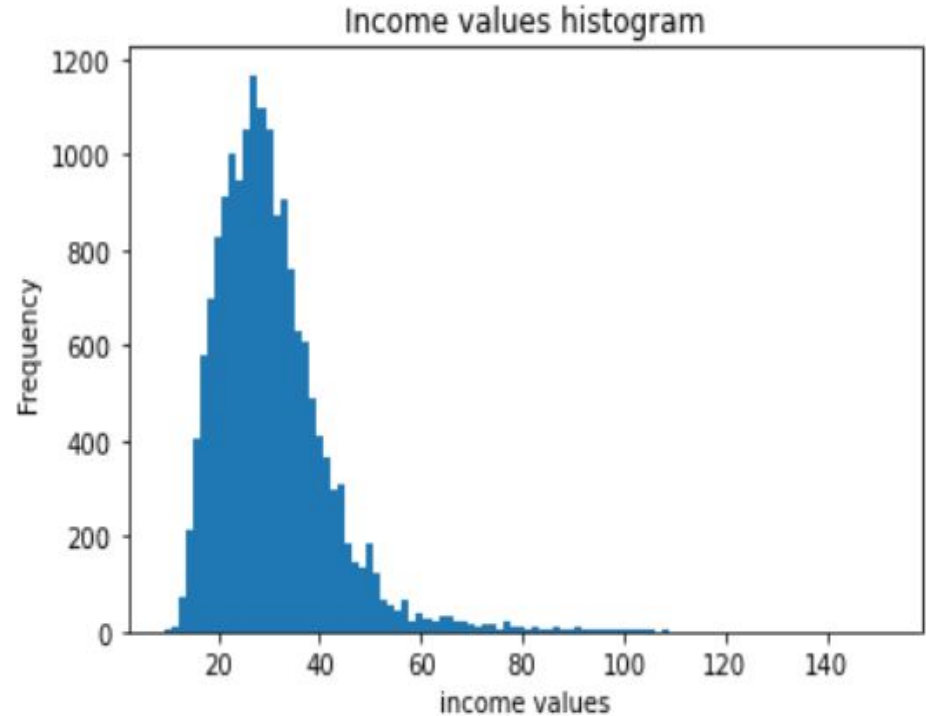
- 26,186 rows
- 73 columns




## Target Variable:

**INCOME  
DISTRIBUTION ->**

\*Target is  
continuous, so it's  
a regression  
problem



# HYPERPARAMETER TUNING USING GRID SEARCH



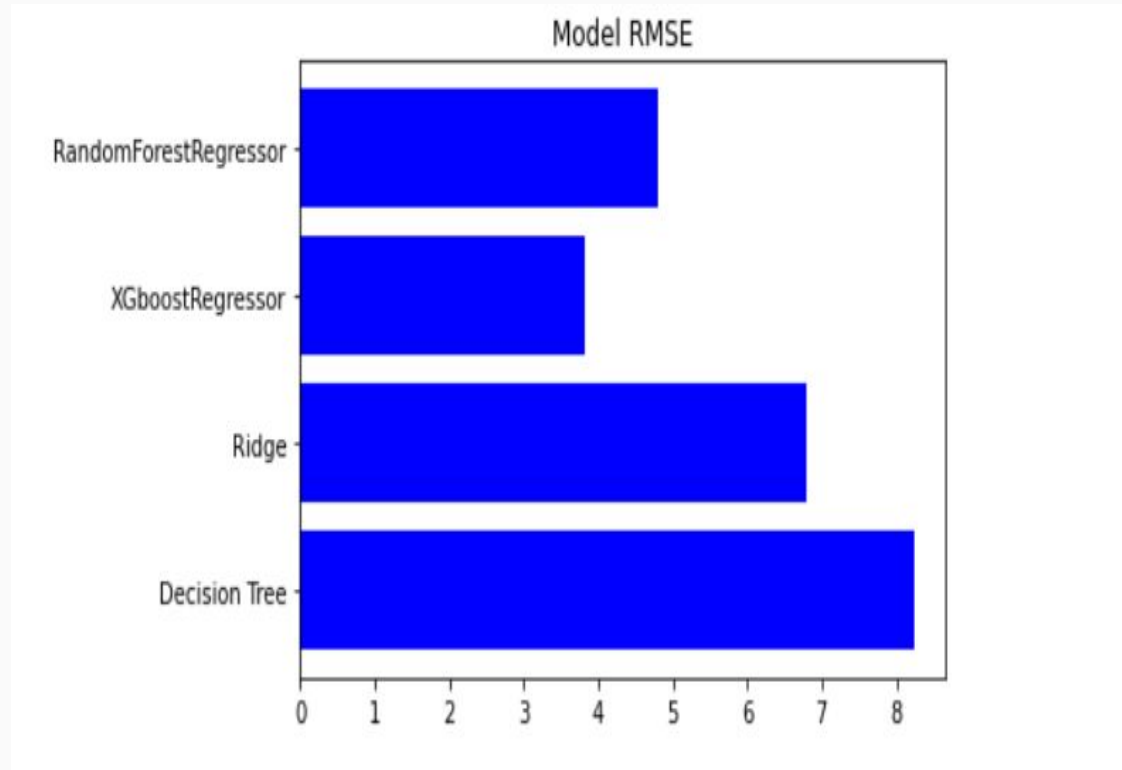
## R-SQUARED SCORE

- Decision Tree  
Regressor: 43%
- Ridge regression: 61%
- XGBoost Regressor:  
88%
- Random Forest  
Regressor: 81%



## Loss Function

USED THE ROOT  
MEAN SQUARED  
ERROR (RMSE)



The background of the slide features a stylized wireframe head on the left, composed of a mesh of triangles with glowing blue nodes. To the right of the head is a complex network of glowing blue lines and nodes, resembling a circuit board or a neural network diagram. The entire scene is set against a dark blue background.

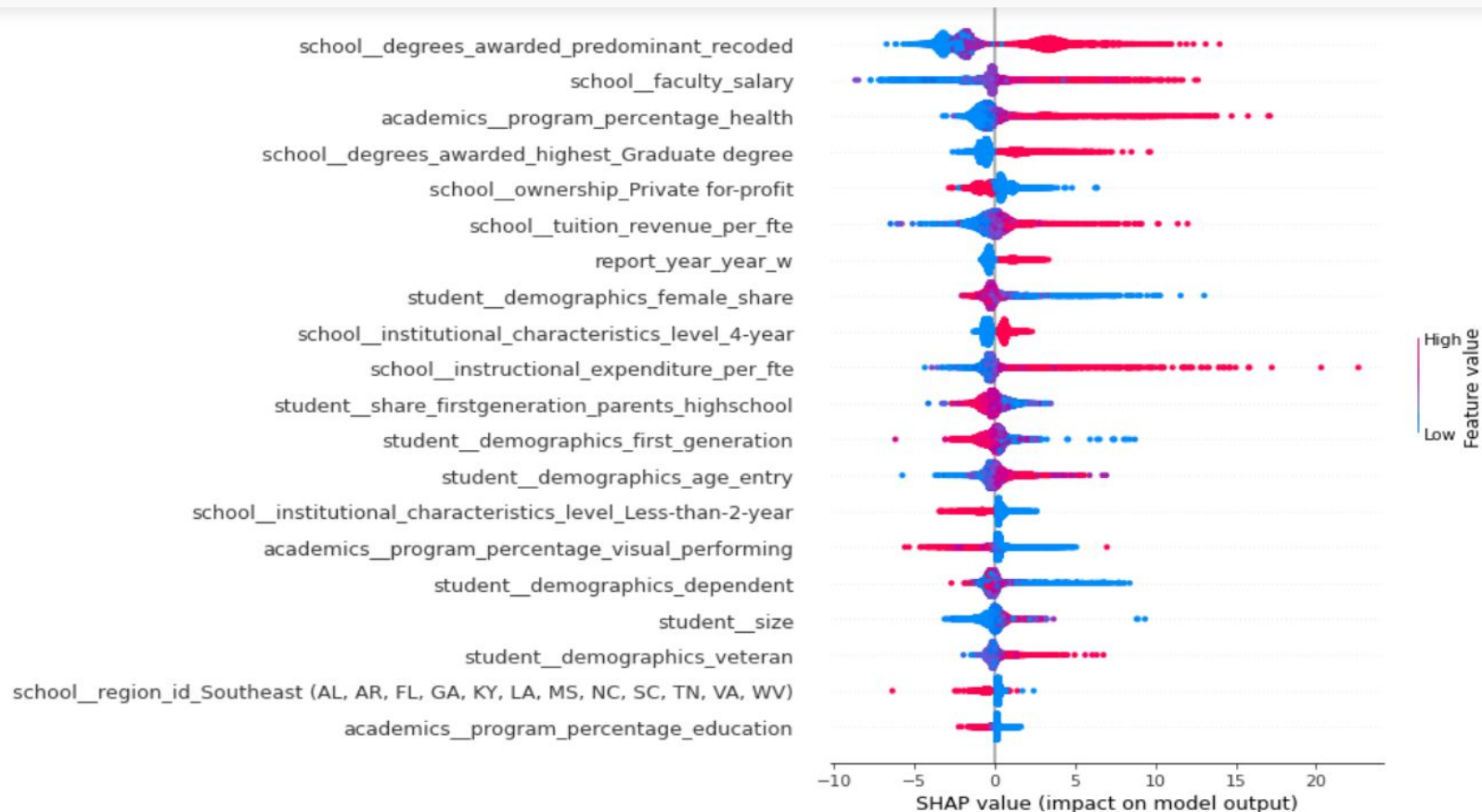
## MODEL USED FOR PREDICTION:

- XGBOOST Regressor
- R-Squared score: 88%
- RMSE: 3.78

# Interpreting XGBoost Regressor Model Result Using Shapley Values



# Shapley value is the average marginal contribution of a feature value across all possible coalition

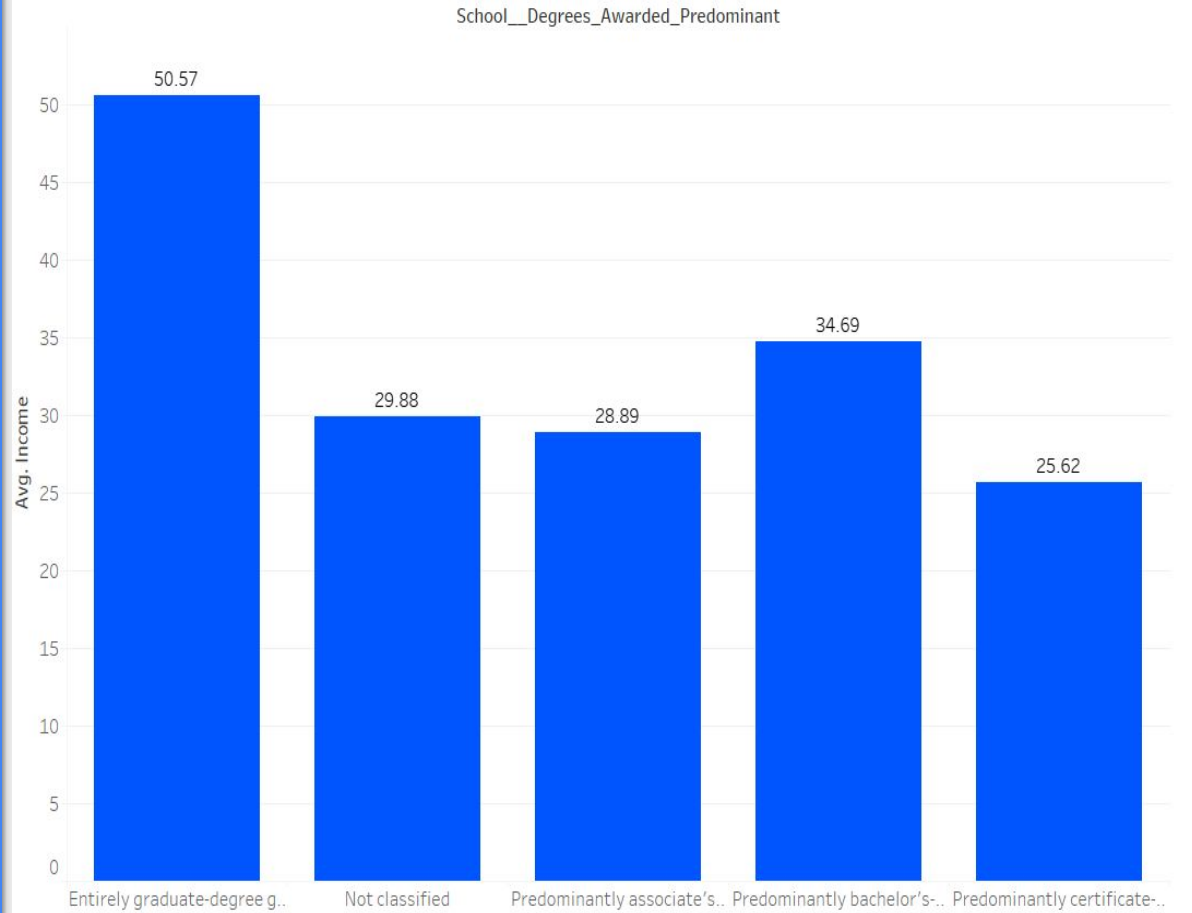


# IMPORTANT FEATURES NOTED

- Degrees awarded, most importantly Graduate Degrees
- School ownership (Private for profit)
- Health Programs
- School instructional expenditure

# Most Important Feature:

School degree awarded : Graduate degree has the highest income



# CONCLUSION AND NEXT STEPS

- Degrees, school characteristics, instructional expenditure allocated and health programs all influence income predictions
- With the Time data encoded as categorical (year w, f..) there was no way to evaluate the change in income predicted over the years and also the change of features that influenced income prediction



**[linkedin.com/in/audreyemeribe](https://linkedin.com/in/audreyemeribe)**

