MINI PROJECT - IV

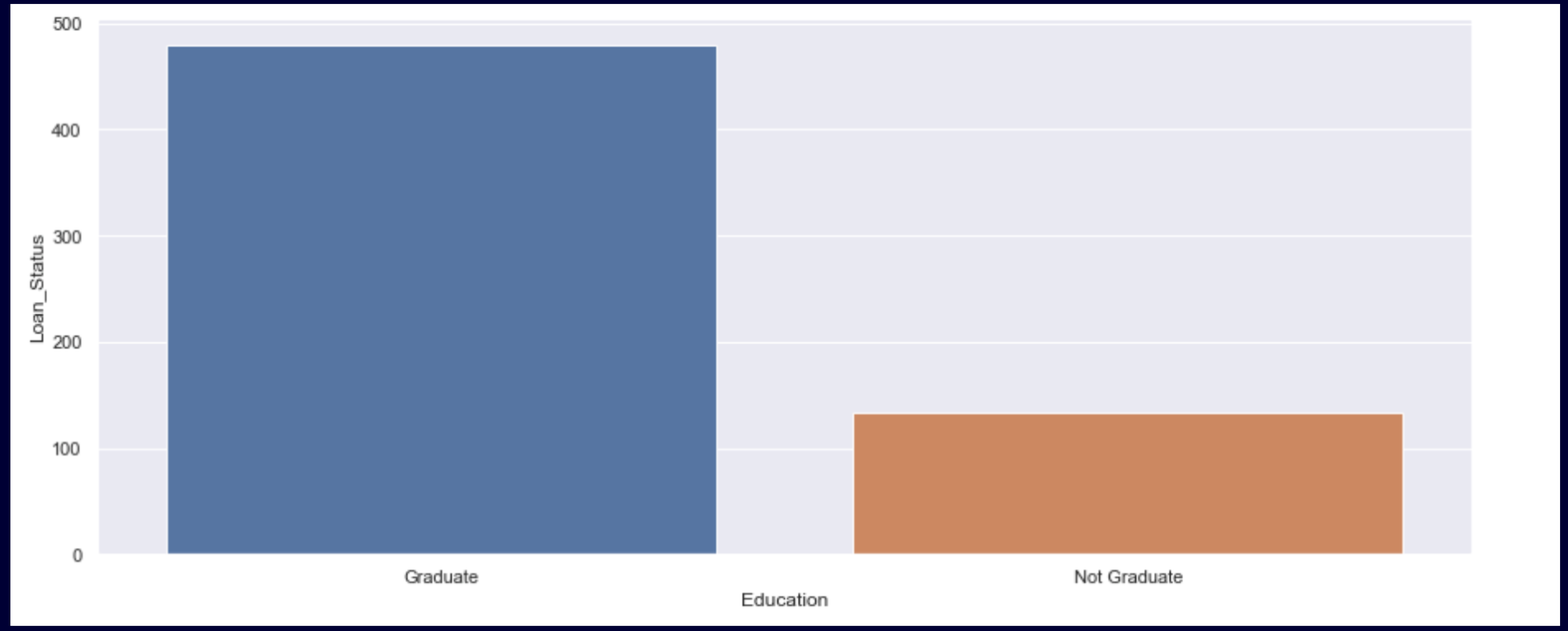
GERARD AGADA - LOAN PREDICTIONS

PROJECT FLOW STRUCTURE

- Hypothesis Generation
- Data Exploration
- Data Cleaning
- Model Building using a Pipeline
- Model deployment to the Cloud

HYPOTHESIS GENERATION - #1

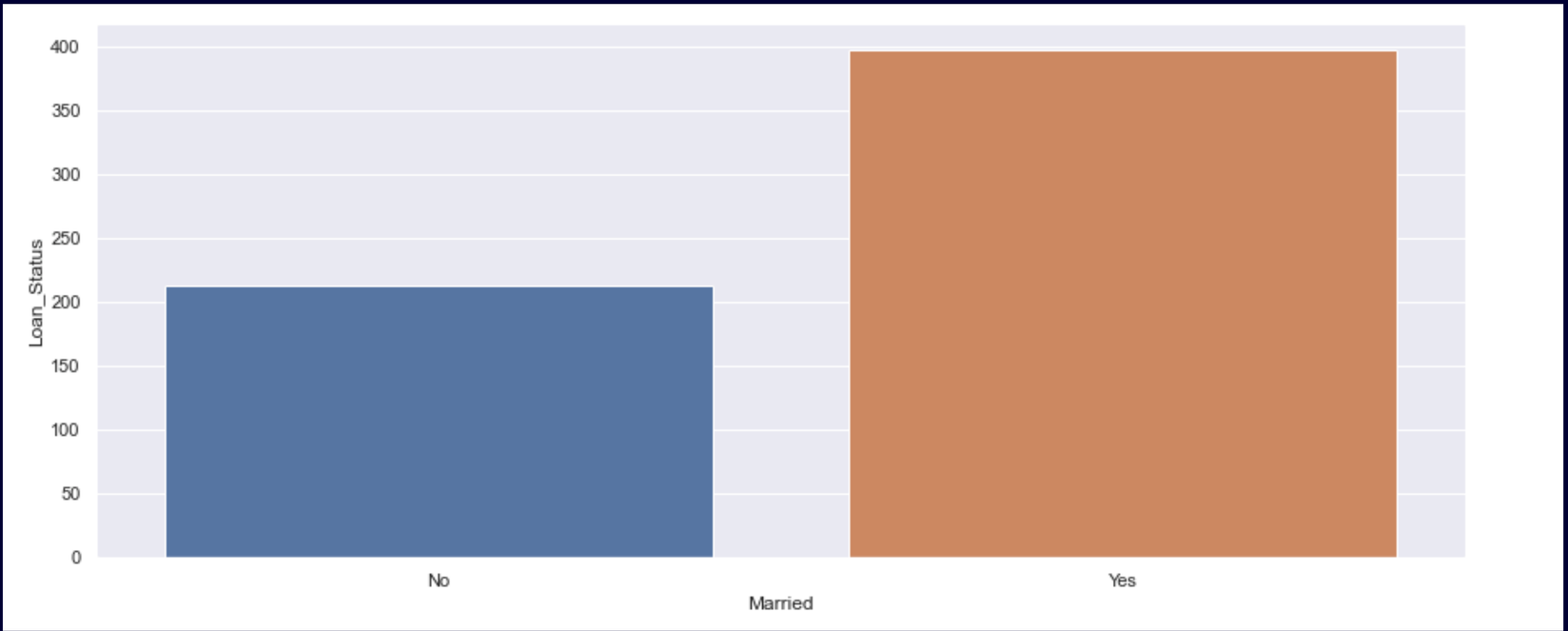
Applicants who are more educated are more likely to qualify for a loan.



	Education	Loan_Status
0	Graduate	480
1	Not Graduate	134

HYPOTHESIS GENERATION - #2

Applicants who are married are also more likely to qualify for a loan.



	Married	Loan_Status
0	No	213
1	Yes	398

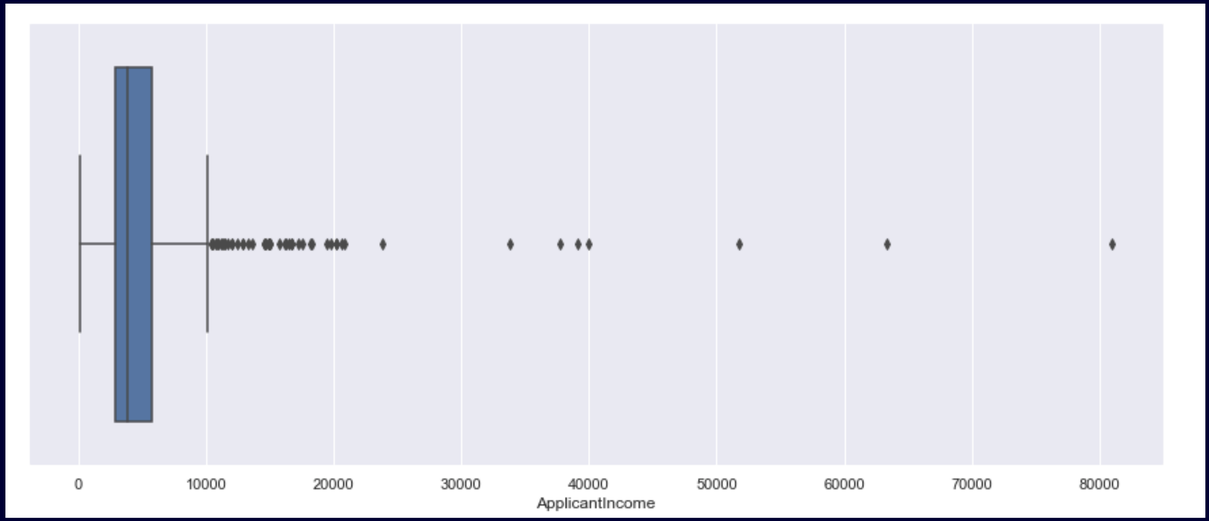
HYPOTHESIS GENERATION - #3

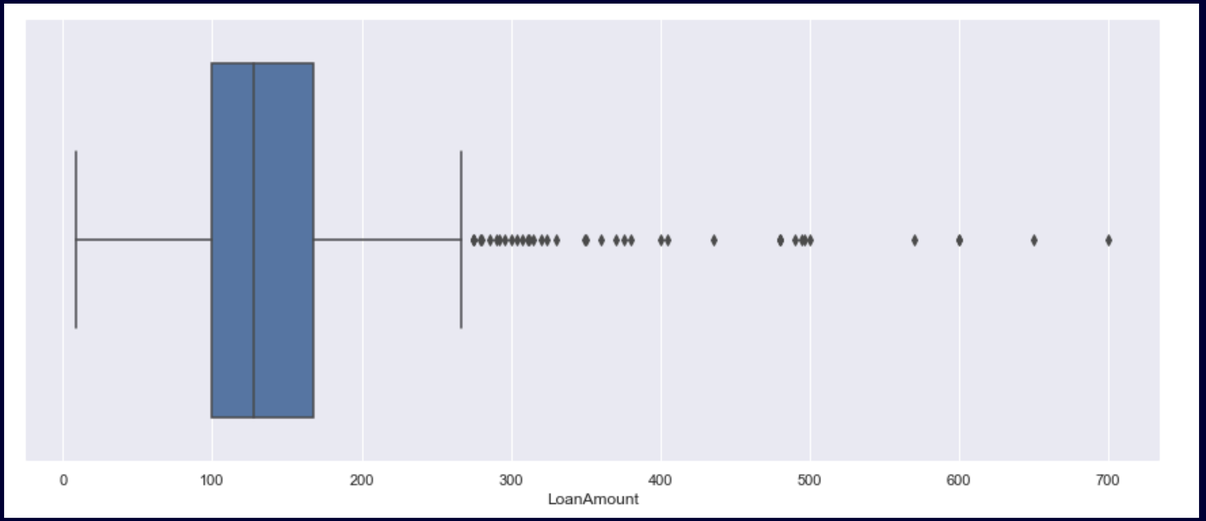
- Other Considerations
  - Creditworthiness
  - Cash Flow
  - Plan for Using Your Loan

DATA EXPLORATION

Distribution Analysis

Visualizing outliers and understanding the distribution using a box-plot plot shows that the average Applicant Income is approximately $5,000 and the average loan amount is $145.00.





DATA EXPLORATION

Categorical Variable Analysis

- Male and Female applicants who have graduated are likely to have a larger loan amount as well as more income.
- If you are self employed and have good credit history, you are more likely to have a higher income as well as a larger loan amount.

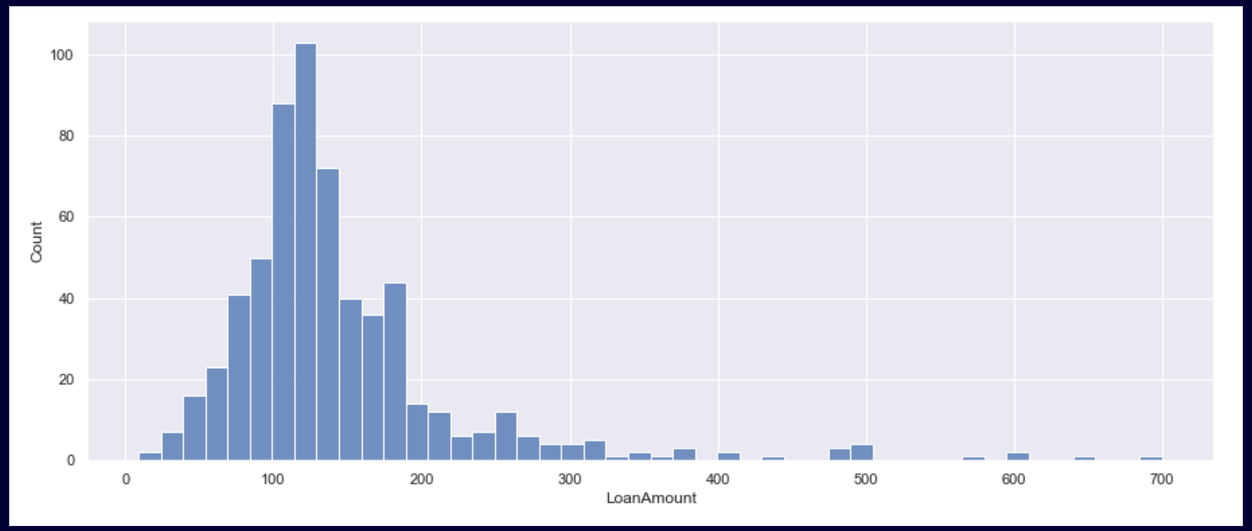
Gender	Education	ApplicantIncome	LoanAmount
Female	Graduate	4646.467391	129.855556
	Not Graduate	4629.700000	111.736842
Male	Graduate	5992.345745	157.994490
	Not Graduate	3630.061947	119.654206

Self_Employed	Credit_History	ApplicantIncome	LoanAmount
No	0.0	5458.460526	143.356164
	1.0	5019.431525	141.319892
Yes	0.0	7135.583333	175.818182
	1.0	7611.746032	166.639344

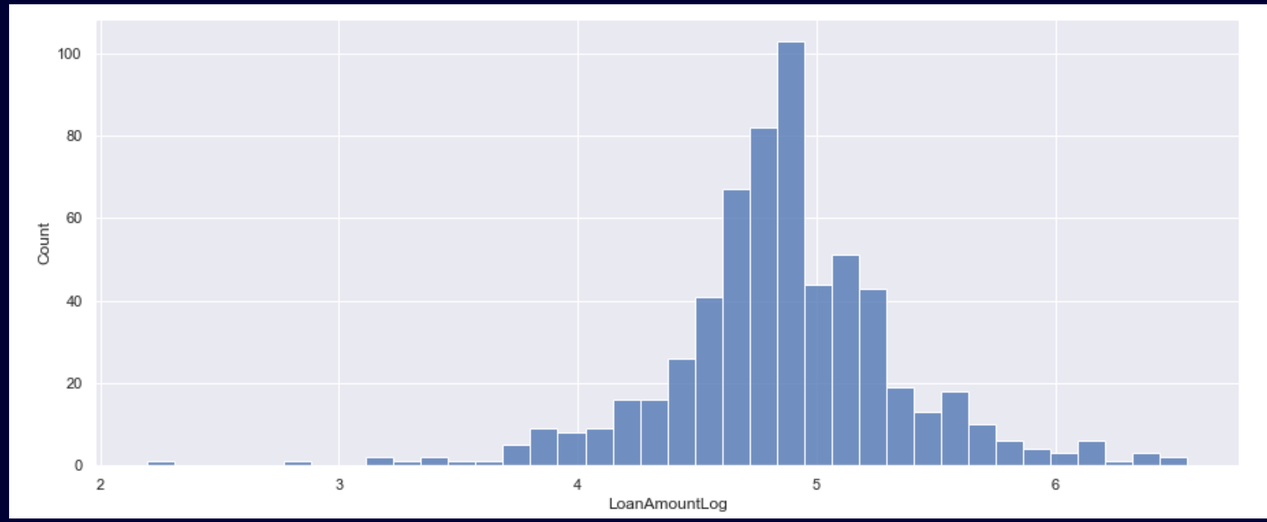
DATA CLEANING

Extreme Values - Loan Amount Log Transformation

Histogram Before Transformation



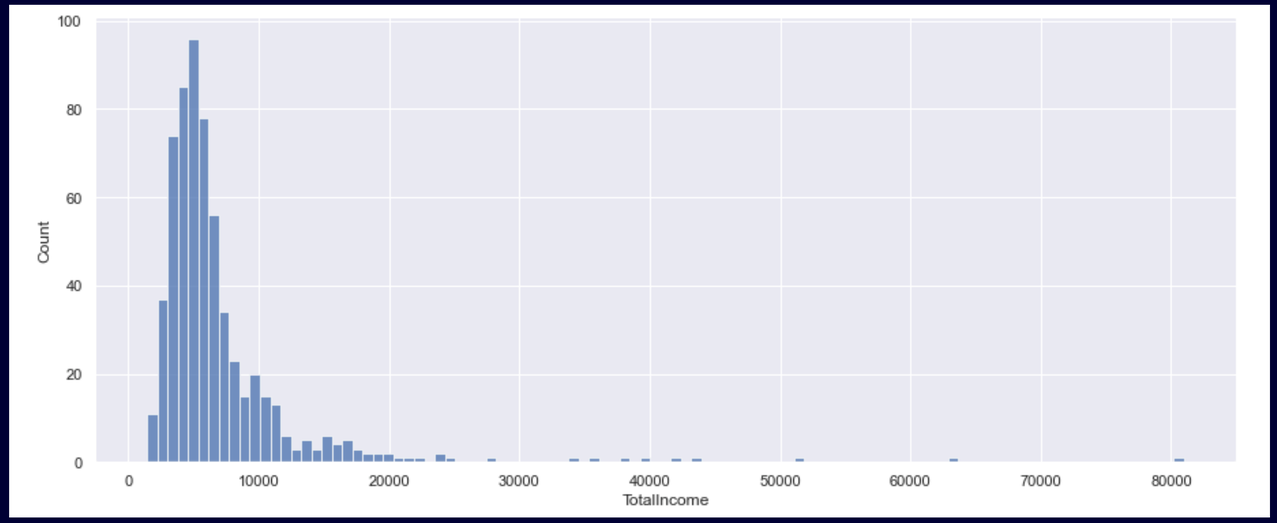
Histogram After Transformation



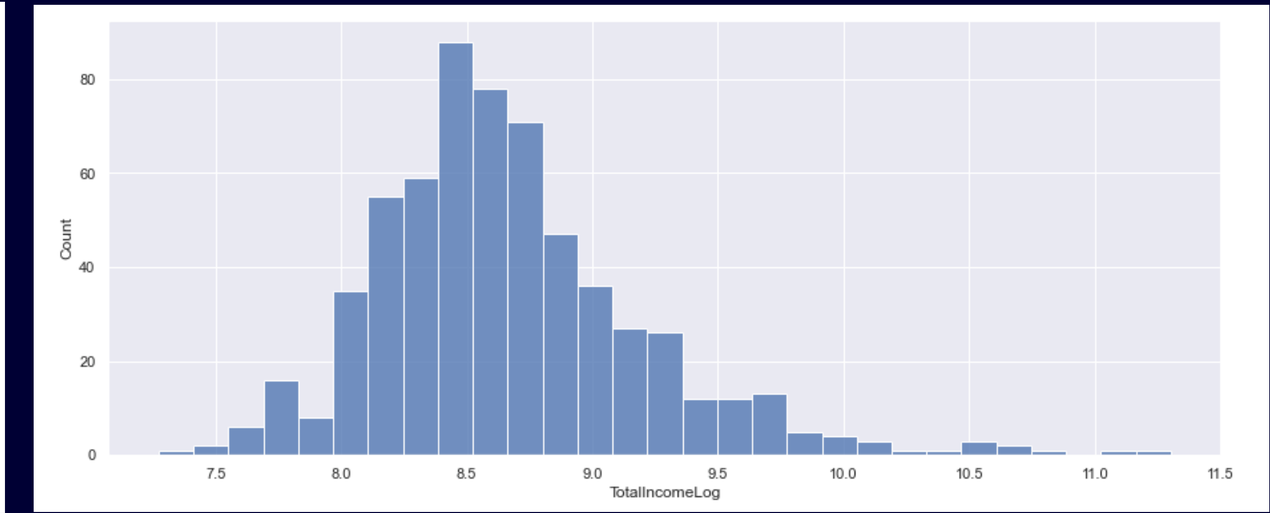
DATA CLEANING

Extreme Values - Total Income Log Transformation

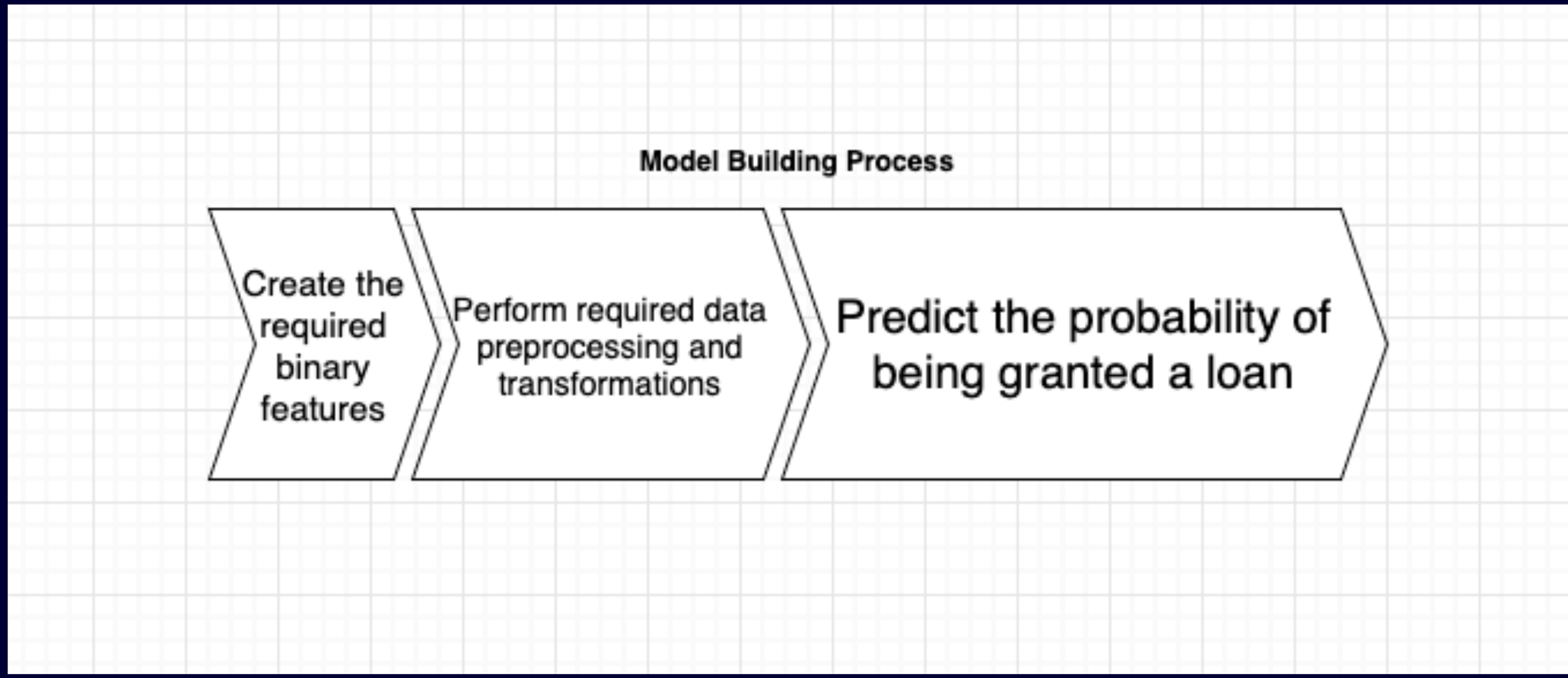
Total Income Skewed Distribution



Total Income Normal Distribution



MODEL BUILDING USING A PIPELINE



DEPLOYMENT TO THE CLOUD

```
(base_env) gcagada@Chiemekas-MacBook-Pro mini-project-IV % python mini_project_api.py
 * Serving Flask app 'mini_project_api' (lazy loading)
 * Environment: production
   WARNING: This is a development server. Do not use it in a production deployment.
   Use a production WSGI server instead.
 * Debug mode: on
 * Running on all addresses.
   WARNING: This is a development server. Do not use it in a production deployment.
 * Running on http://[redacted]5555/ (Press CTRL+C to quit)
 * Restarting with stat
 * Debugger is active!
 * Debugger PIN: [redacted]
[redacted]- - [24/Feb/2022 17:08:41] "POST / HTTP/1.1" 404 -
- - [24/Feb/2022 17:09:03] "GET / HTTP/1.1" 404 -
- - [24/Feb/2022 17:09:10] "POST /scoring HTTP/1.1" 200 -
- - [24/Feb/2022 17:25:30] "POST /scoring HTTP/1.1" 200 -
- - [24/Feb/2022 17:25:40] "POST /scoring HTTP/1.1" 200 -
- - [24/Feb/2022 17:25:45] "POST /scoring HTTP/1.1" 200 -
```

```
[ ]: json_data = {'Gender': 'Male',
      'Married': 'No',
      'Dependents': 0,
      'Education': 'Graduate',
      'Self_Employed': 'No',
      'ApplicantIncome': 500,
      'CoapplicantIncome': 5,
      'LoanAmount': 55.0,
      'Loan_Amount_Term': 360,
      'Credit_History': 1.0,
      'Property_Area': 'Urban'}

[ ]: #Check if response is 200
     print(r)

     <Response [200]>

[ ]: print(r.json())

     [0.7578125]
```

BIGGEST CHALLENGES

- EC2 Dependencies
- Adding all the components together using a Pipeline
- More time required to try different ML models.
- Binary Classification