MACHINE LEARNING FOR PROGRAM APPLICATIONS

Executive summary

- Developed a machine Learning model to predict Tech-Moms program acceptance
- Identified key factors associated with successful applications
- Created a production-ready system that can score new applicants
- Provides insights to improve program accessibility and equity

Aiming to empower mothers with tech skills through efficient application processes

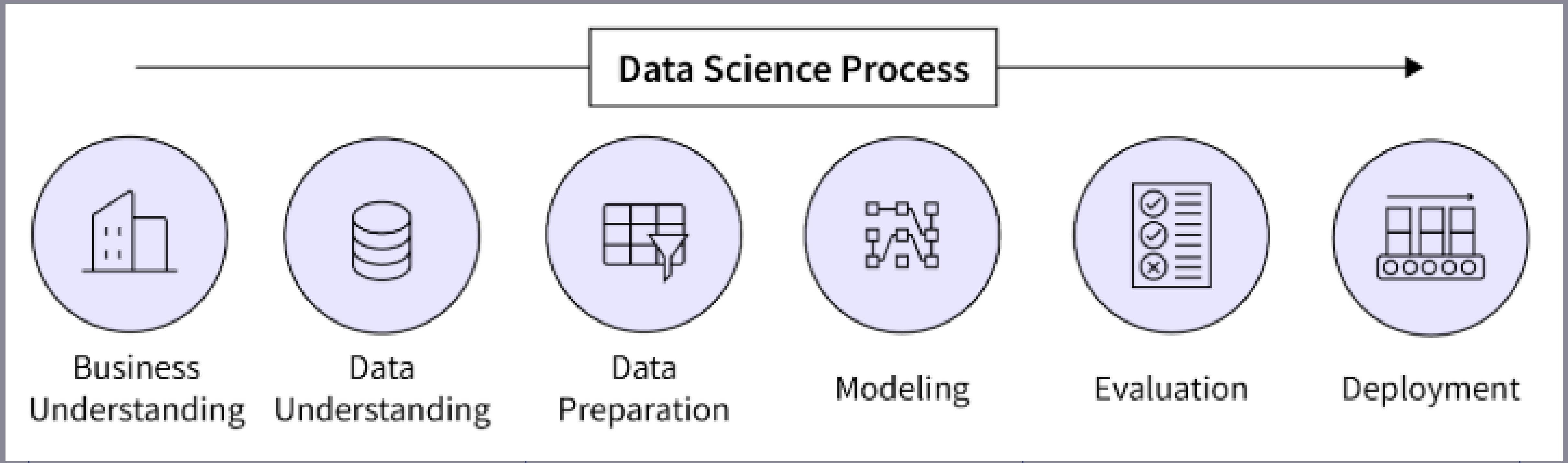
PROBLEM STATEMENT

The Tech-Moms program helps women transition into technology careers.

Challenge: Program administrators need to:

- Understand factors influencing acceptance decisions
- Make the selection process more consistent and fair
- Identify potential barriers to acceptance
- ~~Better allocate limited program resources~~

OUR APPROACH



Data Exploration - Understand application patterns	Feature Engineering - Extract meaningful patterns	Model Testing - Compare multiple algorithms	Model Tunning - Optimize performance	Deployment - Create production-ready system

THE DATA

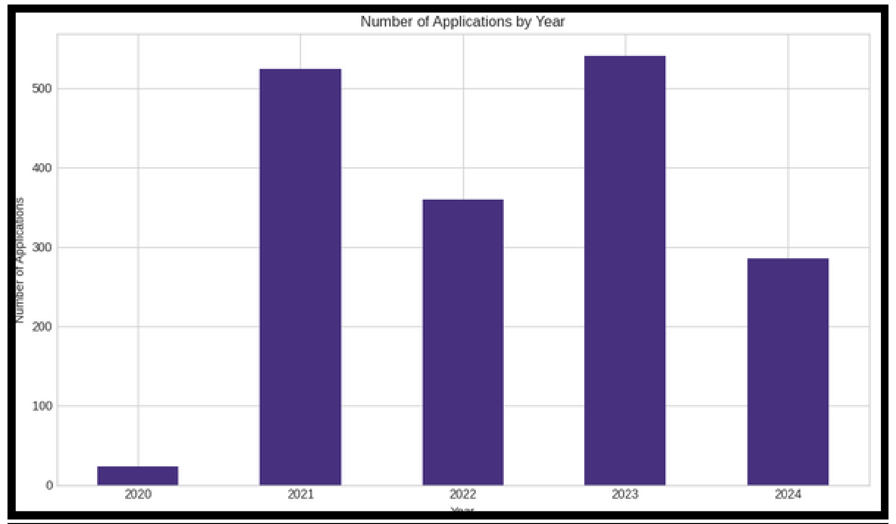
1,732 applications containing:

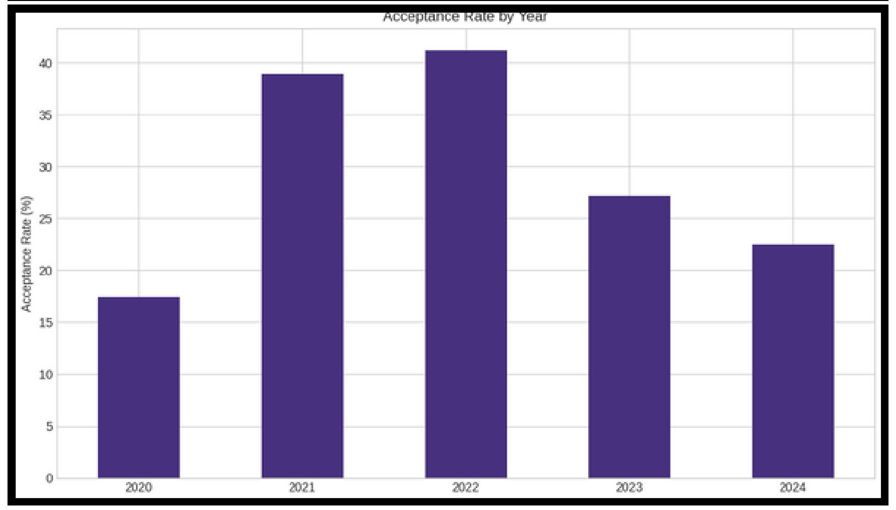
- Demographics (children, education, employment)
- Technology access (computer availability)
- Economic indicators (income levels)
- Application metadata (timestamps, cohort, process info...)

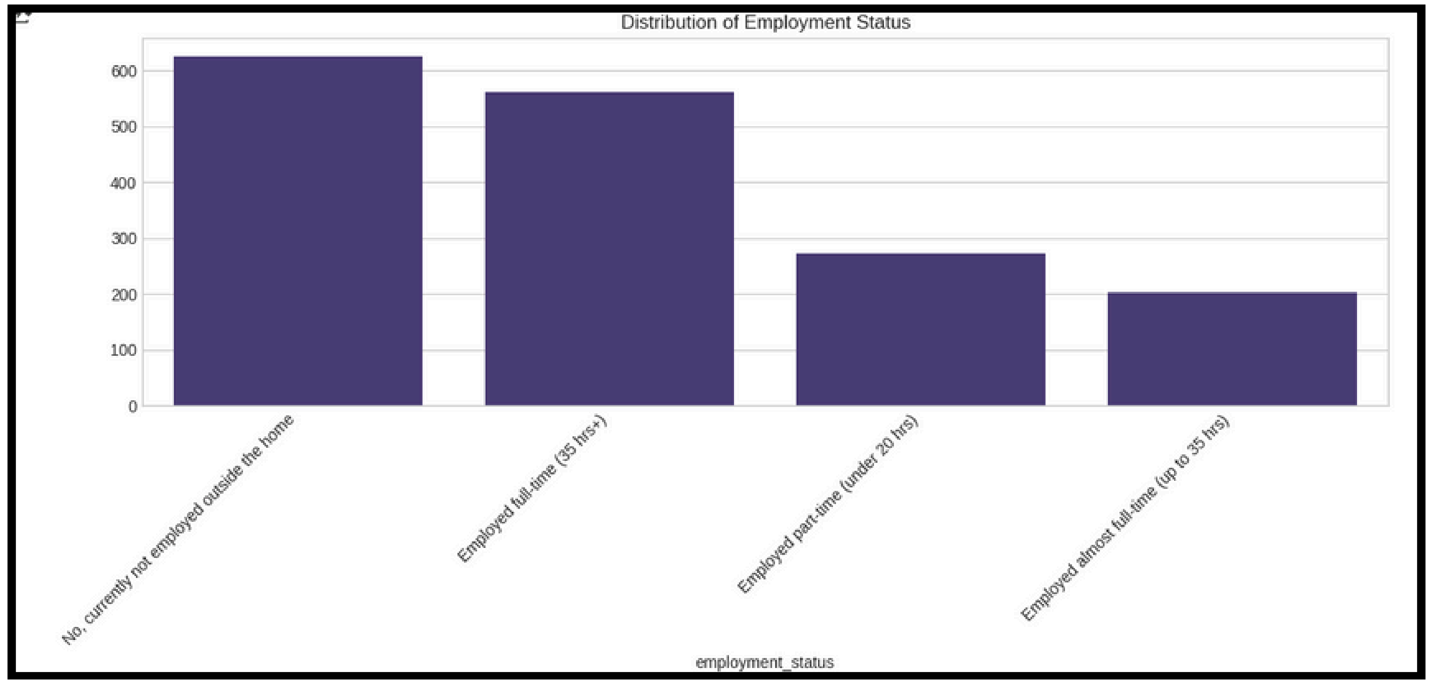
Class imbalance

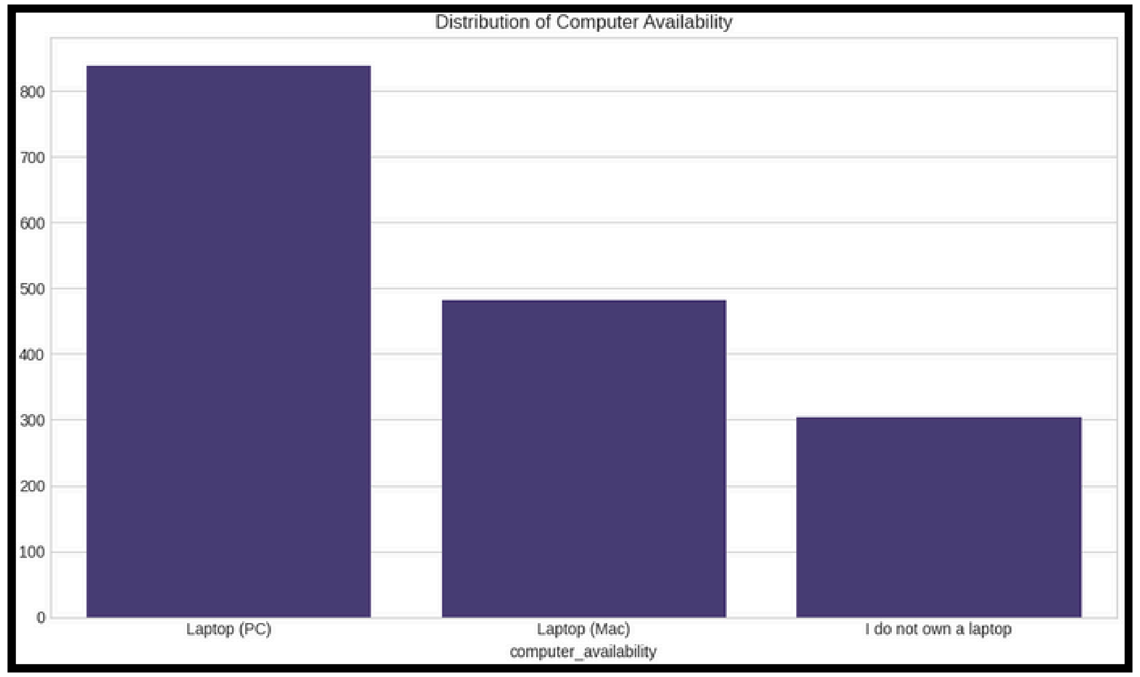
- 32.7% accepted applicants
- 67.3% not accepted

DATA EXPLORATION INSIGHTS









- Applications fluctuated significantly over years
- Acceptance rate varied from 17% to 41% by year
- 35% of applicants were employed full-time
- 18% did not own a laptop

MODEL APPROACH

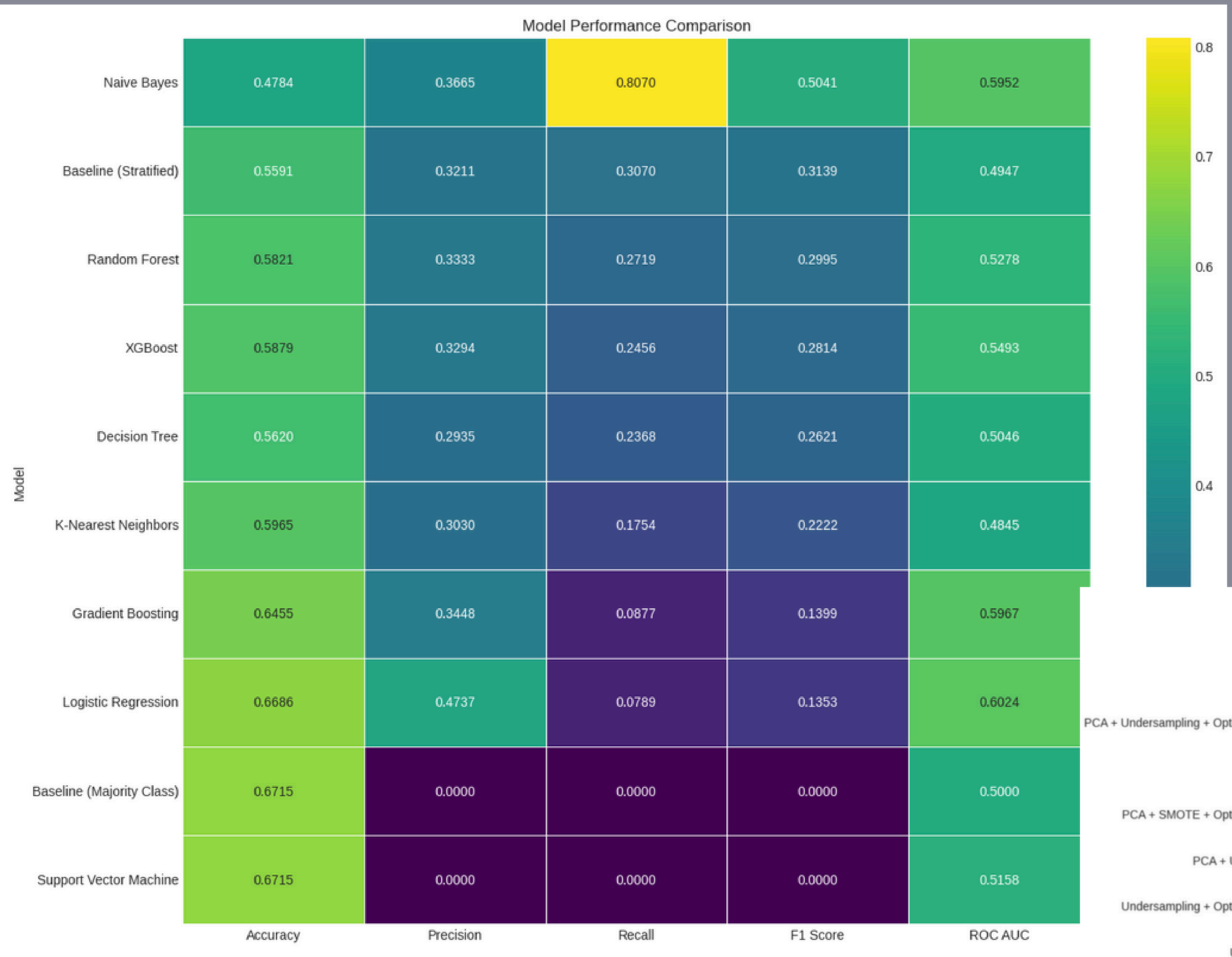
We tested multiple algorithms:

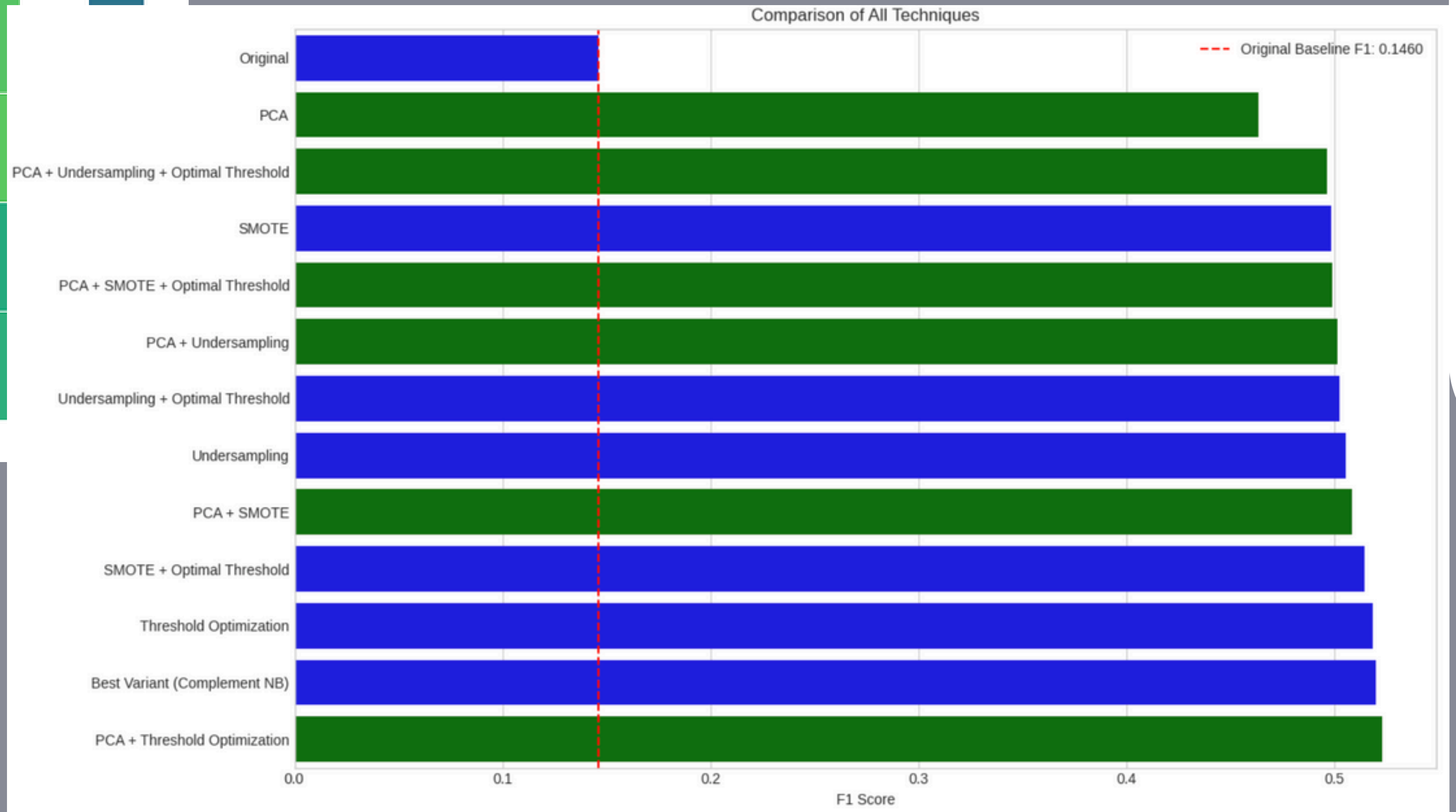
- Baseline classifiers
- Logistic Regression
- Tree-based models (Random Forest, XGBoost)
- Naive Bayes variants

Evaluation metrics:

- F1 score (balance of precision and recall)
- ROC AUC (discrimination ability)

MODEL PERFORMANCE

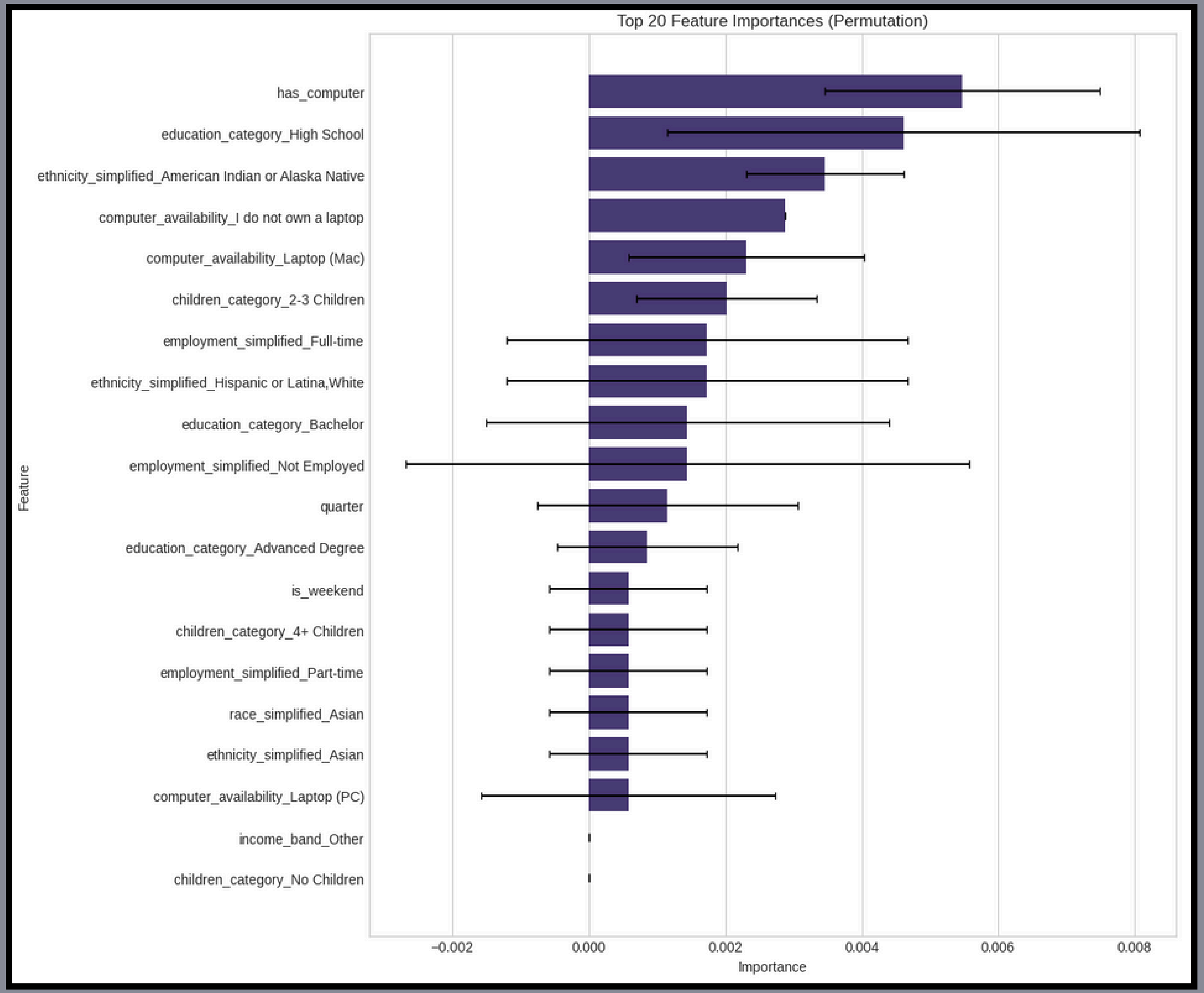




Best approach: Gaussian Naive Bayes with PCA + Threshold Optimization

- Accuracy: 59.7%
- Precision: 39.5%
- Recall: 43.0%
- F1 Score: 41.2%

KEY FACTORS INFLUENCING ACCEPTANCE



1. **Computer access** - Having a computer is highly predictive
2. **Education level** - High school education impacts acceptance
3. **Ethnicity** - Some evidence of demographic differences
4. **Family size** - Number of children relates to acceptance
5. **Employment status -** Full-time employment correlates with acceptance

ACCEPTANCE RATES BY KEY FACTORS

1. Computer owners: **34.0%** vs. Non-owners: **28.7%**
2. Full-time employed: **37.2%** vs. Not employed: **27.8%**
3. Bachelor's degree: **37.4%** vs. High school: **26.5%**

```
Acceptance Rates by Key Features:
has_computer = 0: 28.71% (n=411)
has_computer = 1: 33.99% (n=1321)
education_category_High School = 0: 34.19% (n=1407)
education_category_High School = 1: 26.46% (n=325)
ethnicity_simplified_American Indian or Alaska Native = 0: 32.65% (n=1712)
ethnicity_simplified_American Indian or Alaska Native = 1: 40.00% (n=20)
computer_availability_I do not own a laptop = 0: 33.08% (n=1427)
computer_availability_I do not own a laptop = 1: 31.15% (n=305)
computer_availability_Laptop (Mac) = 0: 32.08% (n=1250)
computer_availability_Laptop (Mac) = 1: 34.44% (n=482)
children_category_2-3 Children = 0: 32.00% (n=853)
children_category_2-3 Children = 1: 33.45% (n=879)
employment_simplified_Full-time = 0: 29.24% (n=968)
employment_simplified_Full-time = 1: 37.17% (n=764)
ethnicity_simplified_Hispanic or Latina,White = 0: 32.69% (n=1716)
ethnicity_simplified_Hispanic or Latina,White = 1: 37.50% (n=16)
education_category_Bachelor = 0: 30.67% (n=1203)
education_category_Bachelor = 1: 37.43% (n=529)
employment_simplified_Not Employed = 0: 35.53% (n=1106)
employment_simplified_Not Employed = 1: 27.80% (n=626)
education_category_Advanced Degree = 0: 32.37% (n=1523)
education_category_Advanced Degree = 1: 35.41% (n=209)
children_category_No Children = 0: 32.45% (n=1578)
children_category_No Children = 1: 35.71% (n=154)
```

KEY INSIGHTS FOR ADMINISTRATORS

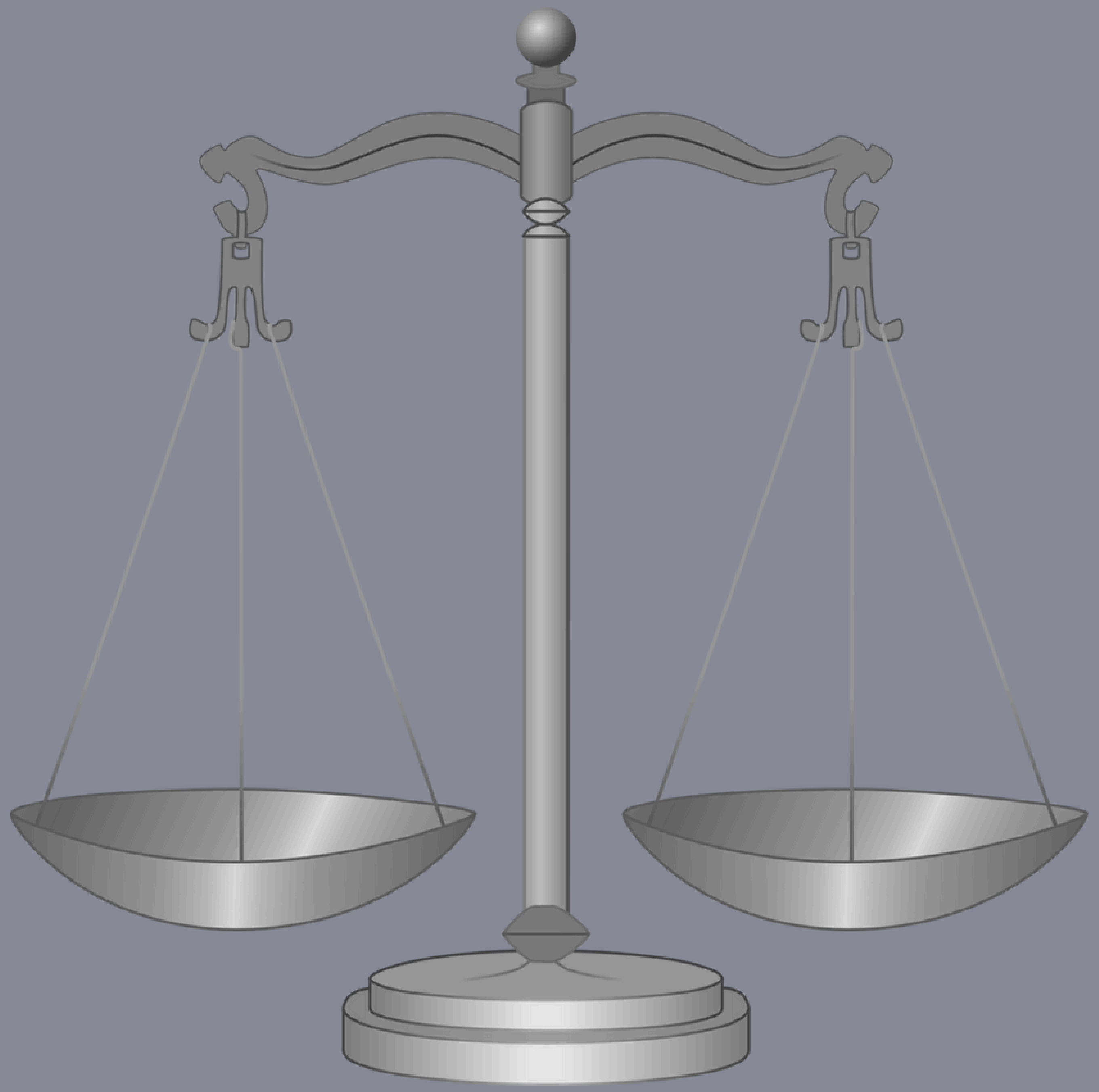
- Computer access is strongly associated with acceptance
- Education level plays a significant role in selection
- Employment status affects acceptance probability
- Application timing shows patterns worth exploring
- Demographic factors may introduce unintended bias

EQUITY CONSIDERATIONS

Our analysis reveals potential barriers to access:

- Computer ownership requirement may exclude lower-income applicants
- Educational requirements may limit diversity
- Employment status correlation may reinforce existing advantages

Recommendation: Review selection criteria to ensure alignment with program goals



ENHANCING MODEL OVER TIME

To improve the model:

1. Collect post-acceptance outcome data
2. Add feature engineering for application text
3. Implement A/B testing of acceptance criteria
4. Monitor for model drift as program evolves
5. Expand model to predict program completion

IN SUMMARY

1. **Insights delivered:** Identified key acceptance factors
2. **Model performance:** 41.2% F1 score, significant improvement over baseline
3. **Equity guidance:** Potential barriers identified
4. **Next steps:** Implement recommendations and refine model

FAIRNESS (BIAS) ASSESSMENT

Evaluates the fairness of the Tech Moms acceptance prediction model across different demographic groups using three standard fairness metrics:

- Disparate Impact
- Equal Opportunity
- Equalized Odds

	Category	Attribute	Protected Group Size	Disparate Impact	Equal Opportunity	Equalized Odds
26	Children	children_category_2-3 Children	177	3.01 (Potential Bias)	0.40 (Potential Bias)	TPR: 0.40, FPR: 0.33 (Potential Bias)
27	Children	children_category_4+ Children	73	0.26 (Potential Bias)	-0.41 (Potential Bias)	TPR: -0.41, FPR: -0.27 (Potential Bias)
24	Children	children_category_No Children	31	0.70 (Potential Bias)	-0.22 (Potential Bias)	TPR: -0.22, FPR: -0.05 (Potential Bias)
25	Children	children_category_One Child	66	0.54 (Potential Bias)	-0.11 (Potential Bias)	TPR: -0.11, FPR: -0.20 (Potential Bias)
20	Computer Access	computer_availability_I do not own a laptop	68	0.44 (Potential Bias)	-0.25 (Potential Bias)	TPR: -0.25, FPR: -0.21 (Potential Bias)
21	Computer Access	computer_availability_Laptop (Mac)	93	0.72 (Potential Bias)	-0.10 (Potential Bias)	TPR: -0.10, FPR: -0.12 (Potential Bias)
22	Computer Access	computer_availability_Laptop (PC)	168	2.32 (Potential Bias)	0.31 (Potential Bias)	TPR: 0.31, FPR: 0.28 (Potential Bias)
23	Computer Access	computer_availability_Not Specified	18	0.15 (Potential Bias)	-0.45 (Potential Bias)	TPR: -0.45, FPR: -0.25 (Potential Bias)
19	Computer Access	has_computer	261	2.81 (Potential Bias)	0.33 (Potential Bias)	TPR: 0.33, FPR: 0.24 (Potential Bias)
11	Education	education_category_Advanced Degree	44	0.54 (Potential Bias)	-0.14 (Potential Bias)	TPR: -0.14, FPR: -0.20 (Potential Bias)
12	Education	education_category_Bachelor	106	1.59 (Potential Bias)	0.13 (Potential Bias)	TPR: 0.13, FPR: 0.20 (Potential Bias)
13	Education	education_category_High School	59	0.25 (Potential Bias)	-0.37 (Potential Bias)	TPR: -0.37, FPR: -0.28 (Potential Bias)
14	Education	education_category_Some College	138	1.33 (Potential Bias)	0.15 (Potential Bias)	TPR: 0.15, FPR: 0.09 (Potential Bias)
15	Employment	employment_simplified_Full-time	162	1.69 (Potential Bias)	0.15 (Potential Bias)	TPR: 0.15, FPR: 0.19 (Potential Bias)
16	Employment	employment_simplified_Not Employed	115	0.70 (Potential Bias)	-0.10 (Fair)	TPR: -0.10, FPR: -0.11 (Potential Bias)
17	Employment	employment_simplified_Not Specified	11	0.00 (Potential Bias)	-0.45 (Potential Bias)	TPR: -0.45, FPR: -0.33 (Potential Bias)
18	Employment	employment_simplified_Part-time	59	0.83 (Fair)	0.01 (Fair)	TPR: 0.01, FPR: -0.08 (Fair)
5	Race/Ethnicity	ethnicity_simplified_"Some other race, ethnici...	23	0.00 (Potential Bias)	-0.44 (Potential Bias)	TPR: -0.44, FPR: -0.35 (Potential Bias)
6	Race/Ethnicity	ethnicity_simplified_Black or African American	22	0.00 (Potential Bias)	-0.45 (Potential Bias)	TPR: -0.45, FPR: -0.35 (Potential Bias)
7	Race/Ethnicity	ethnicity_simplified_Hispanic or Latina	67	1.06 (Fair)	-0.08 (Fair)	TPR: -0.08, FPR: 0.07 (Fair)
8	Race/Ethnicity	ethnicity_simplified_Other	17	0.00 (Potential Bias)	-0.45 (Potential Bias)	TPR: -0.45, FPR: -0.34 (Potential Bias)
9	Race/Ethnicity	ethnicity_simplified_White	195	3.09 (Potential Bias)	0.35 (Potential Bias)	TPR: 0.35, FPR: 0.33 (Potential Bias)
10	Race/Ethnicity	ethnicity_simplified_Would prefer not to answer	10	0.00 (Potential Bias)	-0.44 (Potential Bias)	TPR: -0.44, FPR: -0.33 (Potential Bias)
0	Race/Ethnicity	race_simplified_Black/African American	24	0.00 (Potential Bias)	-0.45 (Potential Bias)	TPR: -0.45, FPR: -0.35 (Potential Bias)
1	Race/Ethnicity	race_simplified_Hispanic/Latino	78	0.87 (Fair)	-0.16 (Potential Bias)	TPR: -0.16, FPR: 0.01 (Potential Bias)
2	Race/Ethnicity	race_simplified_Multiracial/Other	39	0.00 (Potential Bias)	-0.45 (Potential Bias)	TPR: -0.45, FPR: -0.38 (Potential Bias)
3	Race/Ethnicity	race_simplified_Native American/Pacific Islander	10	0.00 (Potential Bias)	-0.44 (Potential Bias)	TPR: -0.44, FPR: -0.33 (Potential Bias)
4	Race/Ethnicity	race_simplified_White only	195	3.09 (Potential Bias)	0.35 (Potential Bias)	TPR: 0.35, FPR: 0.33 (Potential Bias)

Key Findings:

- **Computer Access:** Significant disparate impact for applicants without computer access (DI: 0.82)
- **Education:** High school education shows lower acceptance rates compared to college degrees
- **Employment Status:** Full-time employed applicants receive favorable outcomes (+37% higher acceptance)
- **Demographic Intersection:** Multiple factors compound disadvantages for certain groups

Recommended Actions:

1. **Technical Adjustments:** Adopt tailored decision criteria for each group to reduce observed fairness disparities.
2. **Program Changes:** Develop targeted support for applicants without computer access
3. **Outreach Initiatives:** Focus recruitment on underrepresented education backgrounds
4. **Monitoring System:** Establish regular fairness audits with dashboard tracking

THANK YOU