Predictive Modeling and Fairness in Higher Education

IDC 6940: Capstone Project | Fall 2025

Mentor: Dr. Agoritsa Polyzou

Osmel Cereijo
Daniela Zaccardi

Computational Data Analytics Track

Project Overview

Problem Area

- Universities increasingly use **predictive models** for admissions decisions
- These systems may **unintentionally** amplify historical inequities
- Underrepresented groups may face **unequal** admission outcomes
- Ensuring **fairness** is critical in high-stakes academic decisions

(Bird et al., 2024; Mehrabi et al., 2019; Van Busum & Fang, 2025)

Background

What is Bias?

- ML Bias: when patterns in historical data encode human or systemic biases, influencing model predictions.
- Statistical Bias: systematic inaccuracies in estimating outcomes or group representation.

Why Fairness Matters?

- ML models shape admissions decisions and student opportunities.
- Bias can disadvantage underrepresented applicants.
- Fairness evaluation ensures transparent, accountable, and ethical decision-making.

(Barocas et al., 2023; Mehrabi et al., 2019; Hardt et al., 2016)

Goals

1 **Understand Bias in Predictive Models** — Explore how machine learning models can develop bias across different groups

2 **Mitigate Bias** — Investigate methods to alleviate bias while maintaining predictive accuracy

Objectives

1. Predict admissions outcomes using student admission data
2. Evaluate fairness across demographic groups
3. Determine the best methods to remove the most amount of bias possible from models

- How accurately can we predict admissions decisions?
- What features most strongly influence predictions?
- Does the model have bias, and if yes, which groups are the most effected?
- If bias exists, is it harmful or justifiable?
- What methods can detect, measure, and mitigate bias?
- How much bias can be effectively reduced without sacrificing accuracy?

Motivation

Bias in Admissions Practices

Legacy admissions and historical practices favor certain groups and reinforce equality.

Policy & Oversight

In August 2025, the U.S. Department of Education ordered a federal audit of admissions data to investigate potential racial bias.

Legal Context

Students for Fair Admissions v. Harvard (2023):

Supreme Court ruled that race-based admission practices are unconstitutional.

(Binkley, 2025; Lawyers' Committee for Civil Rights Under Law, 2023; U.S. Department of Education, 2025)

Prior Art: Quantifying Bias

- Explicitly evaluated group-level algorithmic fairness.

- Common fairness metrics included:
  - Statistical parity
  - Disparate impact
  - Equal opportunity
  - Equalized odds
  - Subgroup accuracy and error-rate gaps

- Prior work consistently stresses the need to examine baseline group disparities before building predictive models.

(Bird et al., 2024; Gandara et al., 2024;
Van Busum & Fang, 2025; Raftopoulos, et al., 2024)

Prior Art: Models and Findings

Models Used

- Decision Trees
- Random Forests
- Logistic Regression

- Why these models?
  - Interpretable
  - Common in education data
  - Easy to audit for fairness

Similar Findings

- High accuracy across most models
- But worse performance for minority/underrepresented groups
- Reinforced the need for fairness evaluation and mitigation

(Bird et al., 2024; Gandara et al., 2024;
Van Busum & Fang, 2025; Raftopoulos, et al., 2024)

Challenges

Data Quality & Preprocessing

Handling missing values, inconsistent features and sensitive attributes while preserving privacy and data integrity.

Bias Detection & Measurement

Selecting appropriate fairness metrics and handling situations where different metrics give conflicting signals.

Bias Mitigation Trade-Offs

Reducing bias without sacrificing too much model performance.

Model Transparency & Explainability

Ensuring models are explainable, important for evaluating whether sensitive features are influencing predictions.

(Bird et al., 2024; Gandara et al., 2024;
Van Busum & Fang, 2025; Raftopoulos, et al., 2024)

Data Sources & Description

Source & Size

- **Source:** FIU Admissions
- **Population:** Fall 2024 Undergraduate applicants
- **Size:** 30K applicants
- **Features:** 42

Feature Categories

- **6 Demographic Information Features:** e.g., gender, age, ethnicity, etc.
- **6 High-School and Academic Background Features:** e.g., type and state of high-school, highest level of education, etc.
- **22 Standardized Test Scores Features:** e.g., ACT and/or SAT scores
- **8 College Application Details Features:** e.g., intended major, admission type and outcome

Sample Applicant

Demographic Information

- Gender: Female
- Age: 18
- Ethnicity: Hispanic/Latino
- Country of Birth: United States
- Military Status: None
- Florida Residency: In-State

High-School & Academic Background

- High-School State: Florida
- High-School Type: Public
- Highest Edu. Level: High-School Diploma
- High-School GPA: 4.0
- Undergrad/Grad GPA: not available

Standardized Test Scores

- ACT Score/Sub-Scores: not available
- ACT to SAT Conversion: 1,330
  - English/Reading/Writing: 740
- SAT Score: 1,250
- SAT Subs-Scores:
  - 720 Reading & Writing, 530 Math

Application Details

- Admitted: Yes
- Admission Type: FTIC
- Intended Major: Nursing
- Admitted to Honors College: No
- Enrolled: Yes
- Matriculated: Yes

Data Quality Issues & Challenges

	Description	Action	Result
Data Quality	Applicants neither admitted nor denied	Treat as not admitted	Avoid losing 12% of our data for analysis/modeling
	Non-resident aliens category in Ethnicity	Unable to use other features as a proxy	Leave as is
High Cardinality Features	Intended major (134 values)	Map to FIU college	Intended college (9 values)
	Country of birth (196 values)	Map to continent	Continent of birth (7 values)
Missing Values	Features with 90% or more missing values	Exclude these features	Improve missing values and decrease data complexity
	Features missing values between 50% to 90%	Exclude ACT/SAT sub-scores and combine totals into a single SAT-based score	

Lits of Features

Demographic Information

- Gender
- Age
- Ethnicity
- Continent of Birth
- US Military Status
- Florida Residency

High-School & Academic Background

- Highest Level of Education
- High-School GPA
- High-School GPA Missing (Flag)

Standardized Test Scores

- SAT Total Score
- SAT Total Score Missing (Flag)

College Application Details

- Admitted
- Admission Type
- Intended College

Data Overview

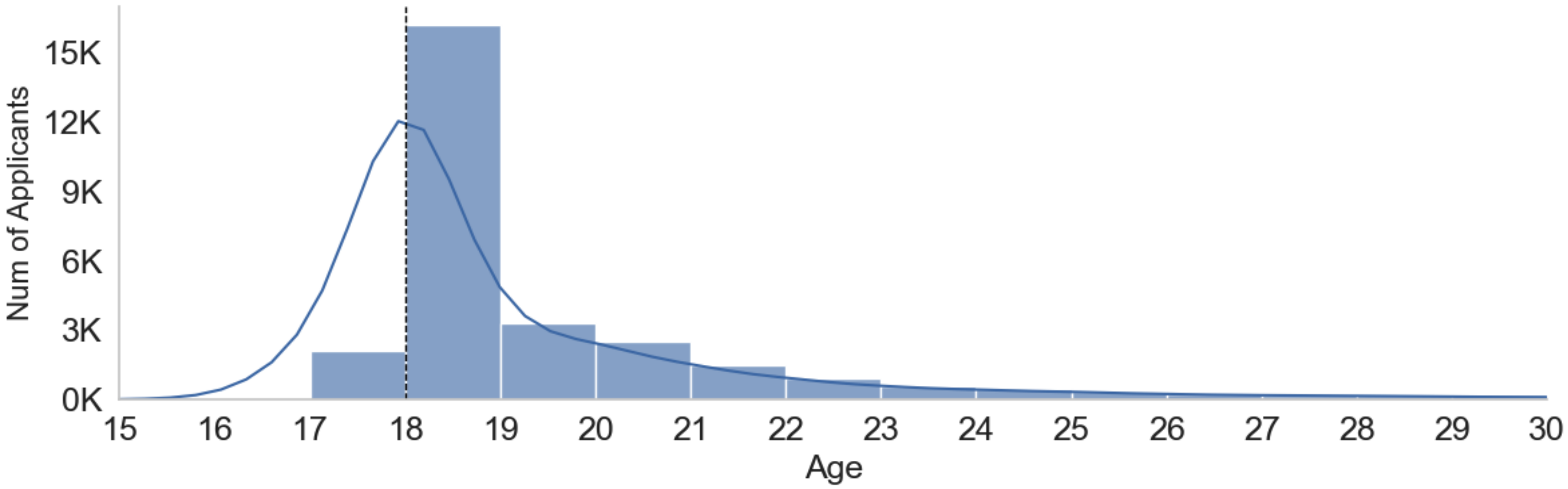
48%
Admission Rate

71%
First-Time in College

58%
Female Applicants

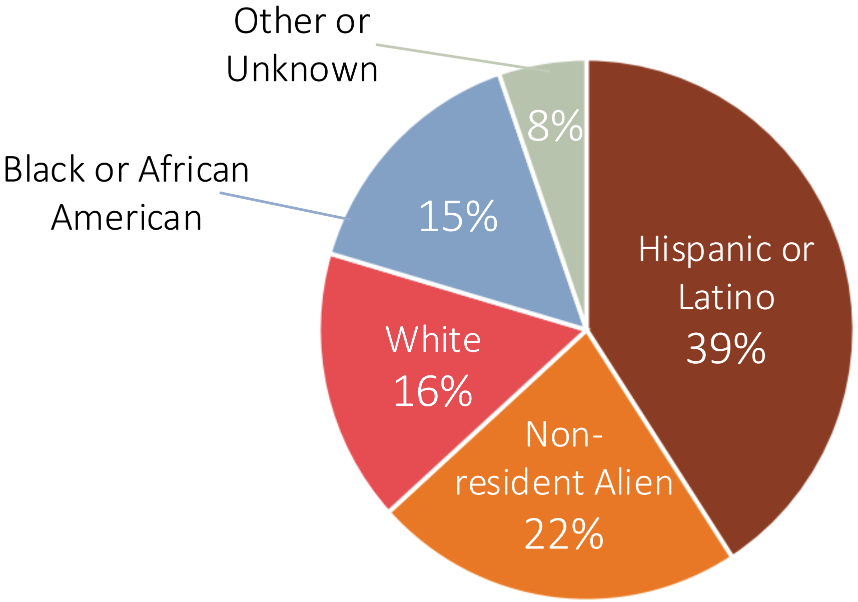
57%
Florida Residency

Age Distribution

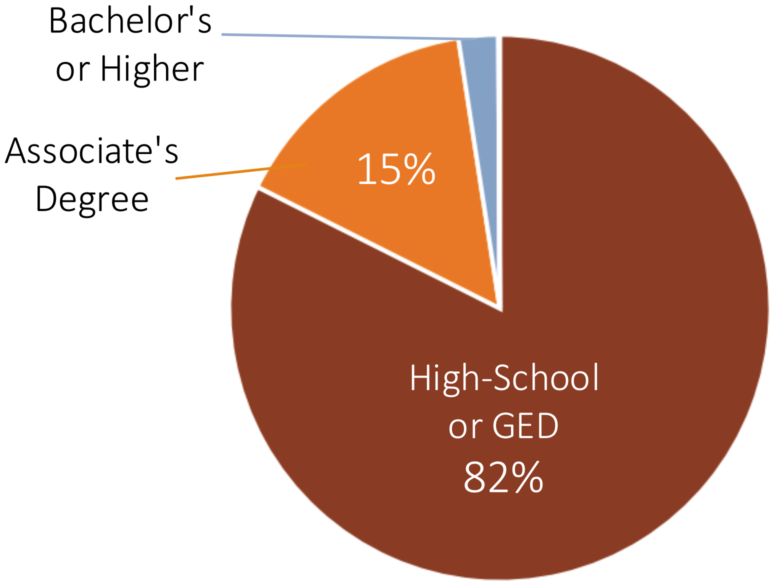


75%
20 years or younger
-
min 15
max 68
std 4.3
median 18
mean 20

Ethnicity



Highest Level of Education

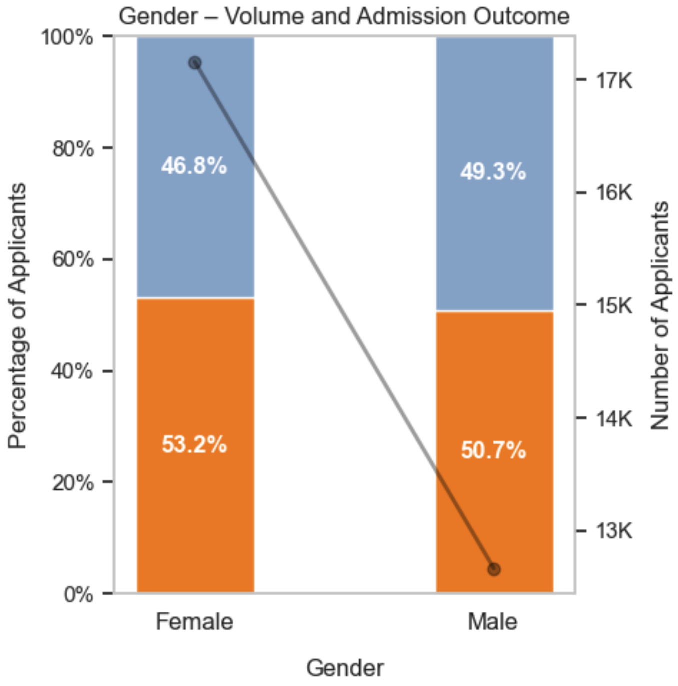


EDA and Bias Investigation

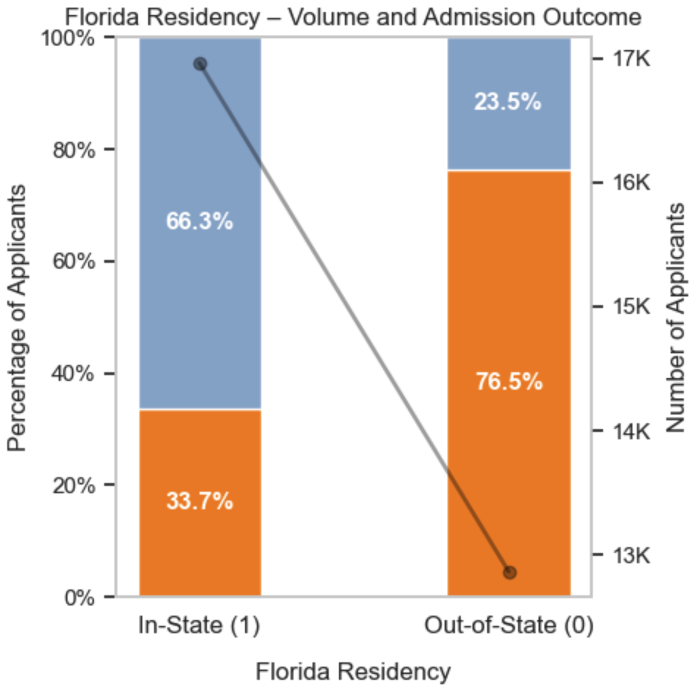
Understanding imbalance and baseline disparities

Binary Features

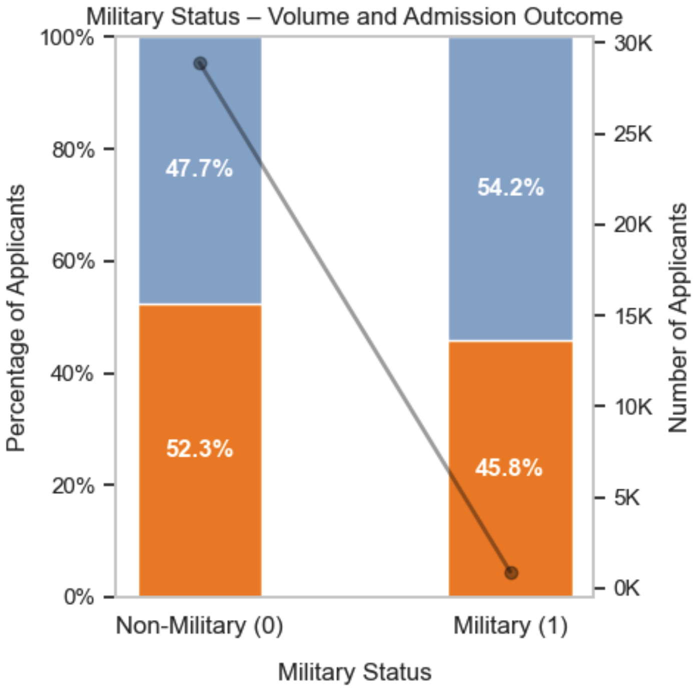
Gender



Florida Residency

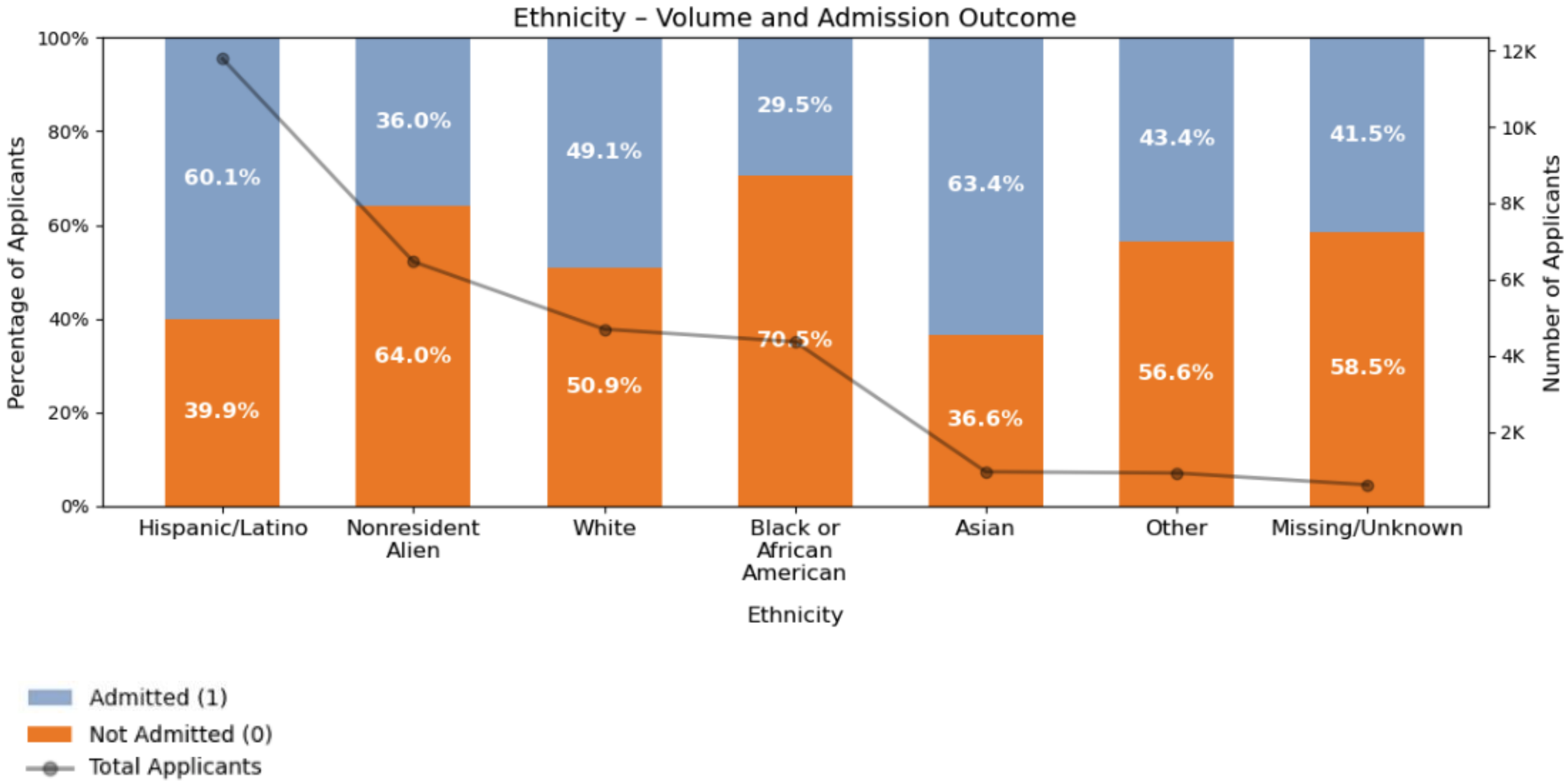


Military Status



Admitted (1)
Not Admitted (0)
Total Applicants

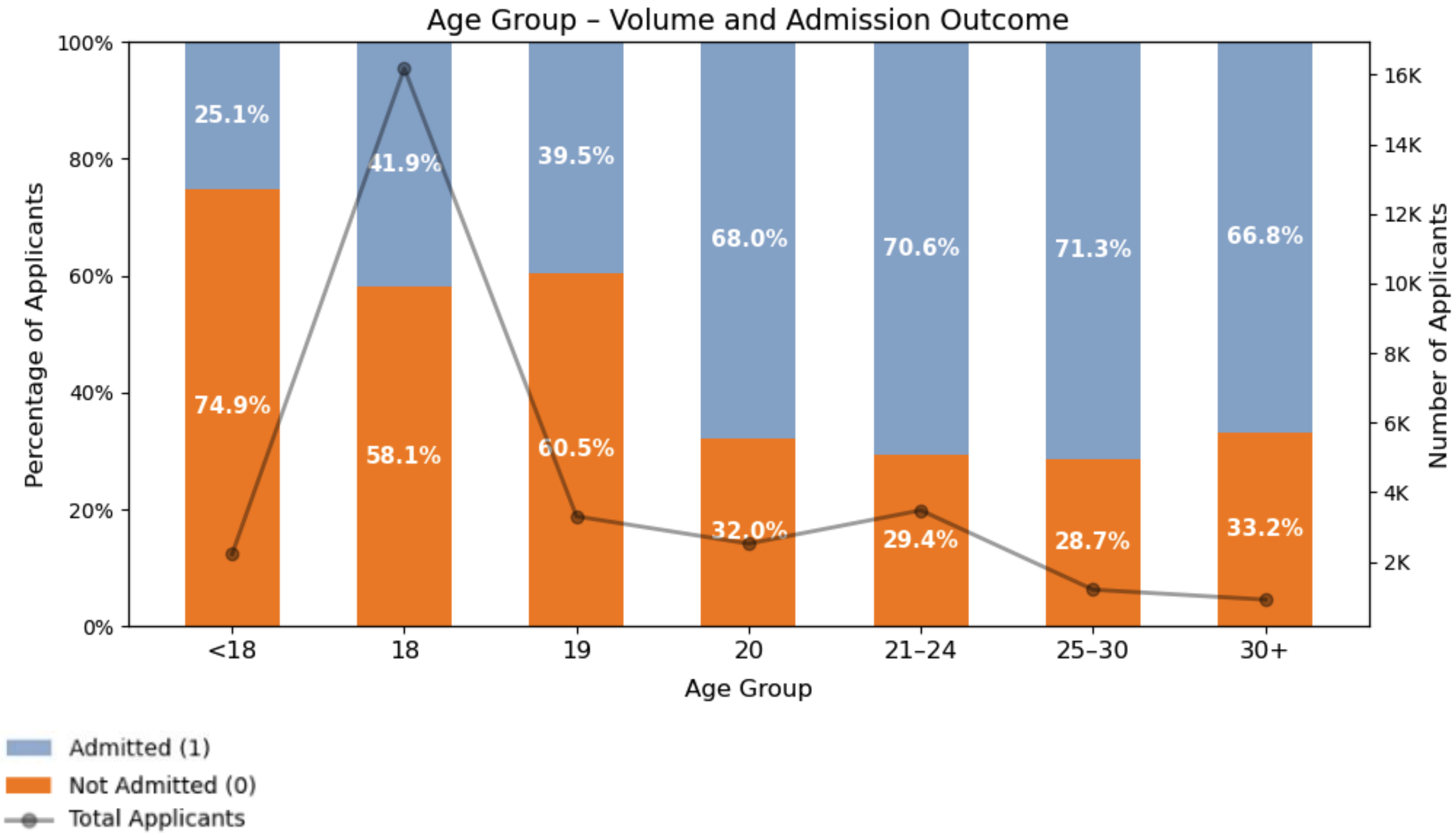
Ethnicity



Key Insights

- Majority of Hispanic/Latino get Admitted
- Majority of Black/African American applicants get denied, same as Alien applicants
- Asian applicant's majority get Admitted

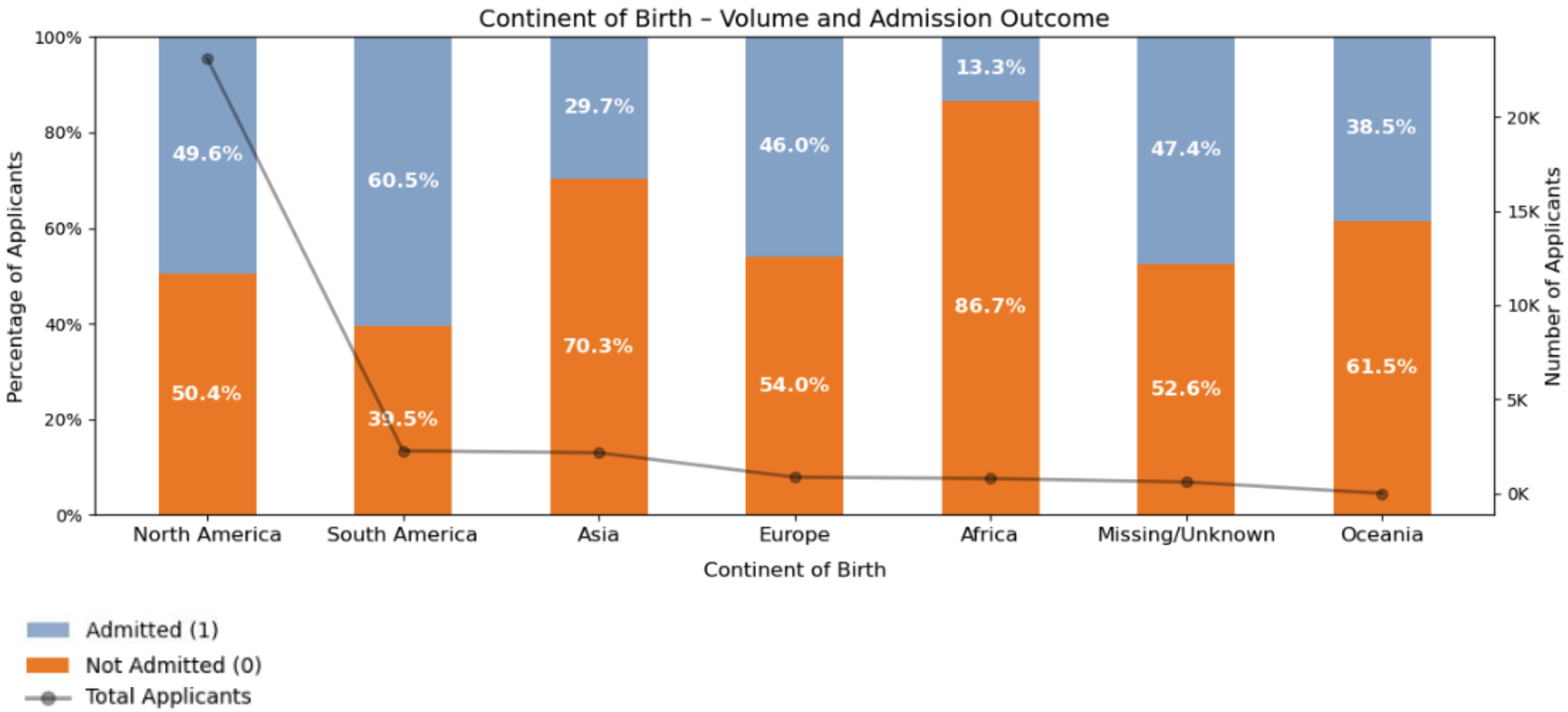
Age



Key Insights

- Traditional-age applicants dominate
- Lower admission rates for youngest applicants
- Peak admissions among young adults
- Older applicants show smaller volumes but decent admissions

Continent of Birth

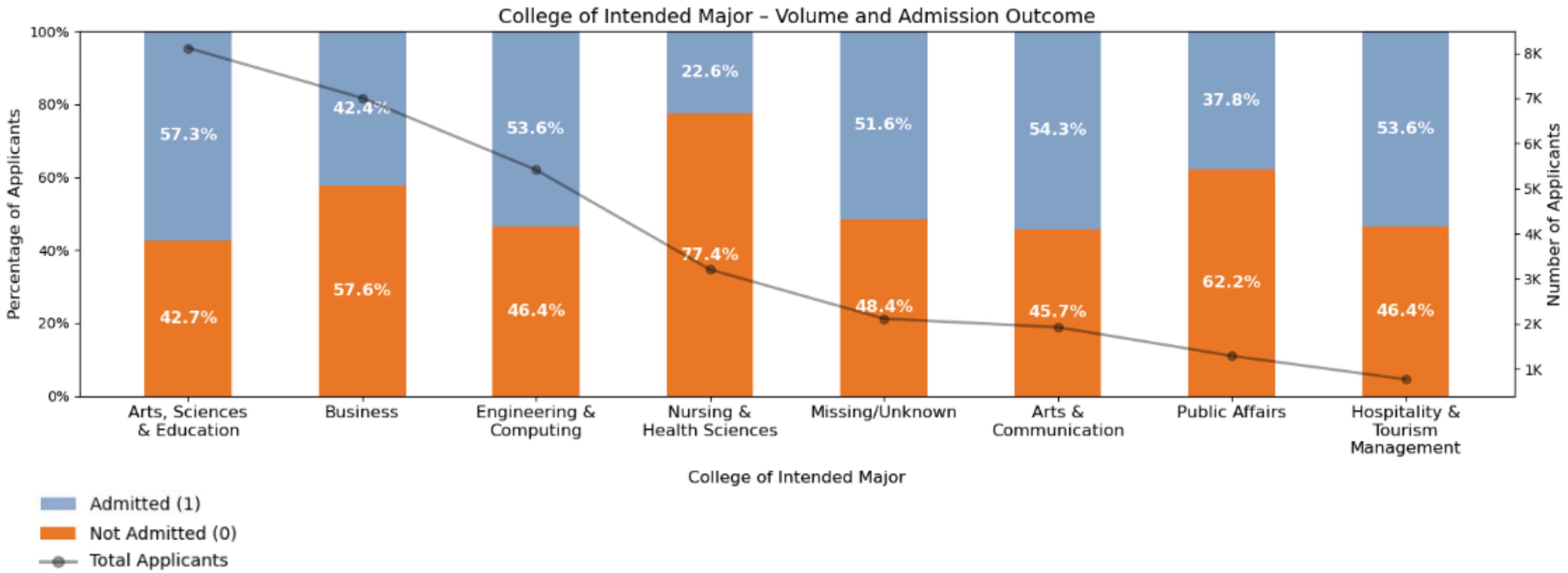


Key Insights

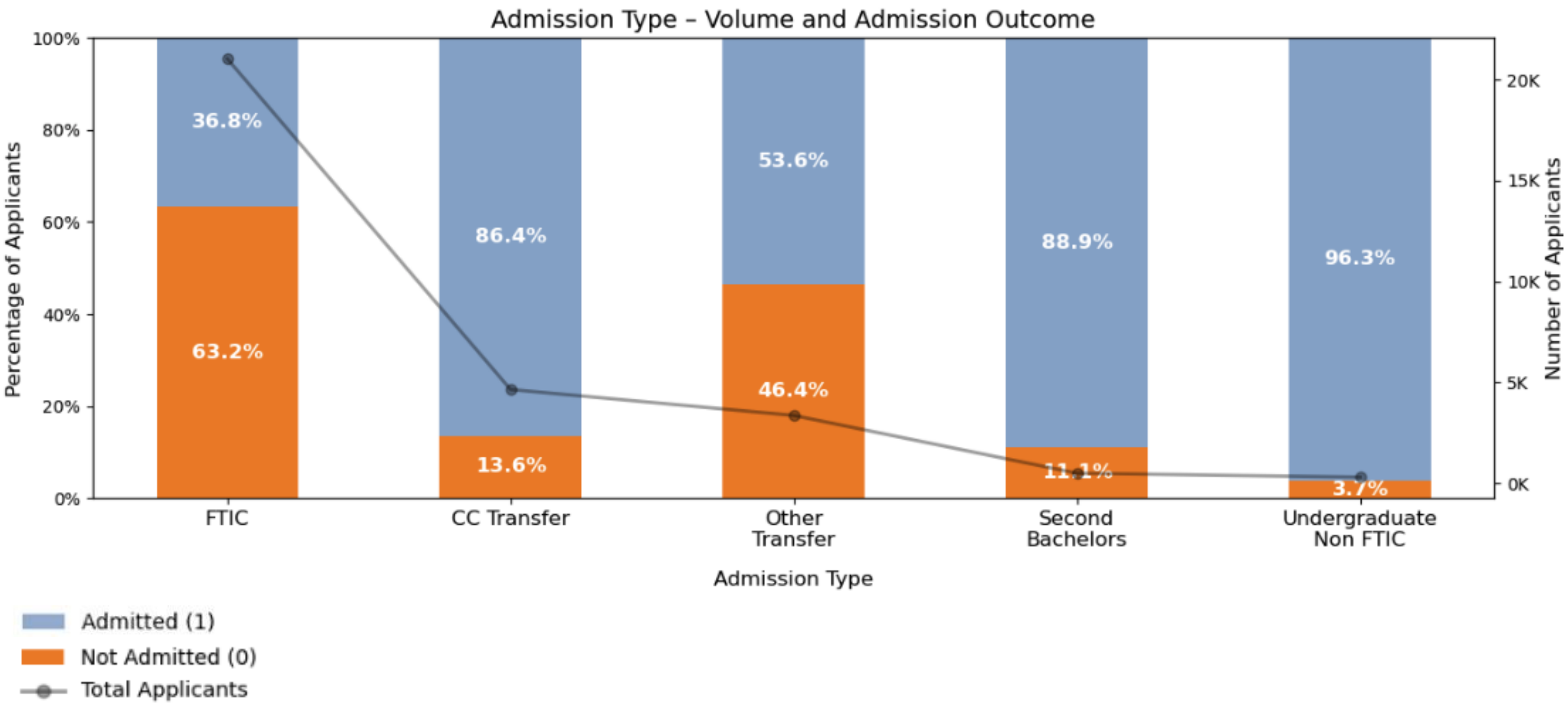
- Africa are more likely to not be admitted.
- Followed by Asia, *interesting because Asian ethnicity has the best admission rate*
- South American applicants have a higher acceptance rate.

College of Intended Major

- Nursing & Health Sciences shows the lowest admission rate, followed by Public Affairs



Application Type



Note: First Time in College (FTIC), Community College (CC) Transfer

Key Insights

- Most FTIC students are not admitted.
- Any other type of admission it is more favorable for the applicant to be admitted.

Summary of Visualization Findings

Feature	Largest	Smallest	Most Admitted	Least Admitted
Gender	Female	Male	Male	Female
FL Residency	In-State	Out-of-State	In-State	Out-of-State
Military Status	Non-Military	Military	Military	Non-Military
Ethnicity	Hispanic/Latino	Missing/Unknown	Asian	Black or African American
Age	18	30+	25 -30	<18
Country of Birth	North America	Oceania	South America	Africa
Major	Arts, Sciences & Education	Hospitality & Tourism	Arts, Sciences & Education	Nursing & Health Sciences
Application Type	FTIC	Undergrad Non FTIC	Undergrad Non FTIC	FTIC

Visuals to Quantification

Observed Disparities

Unequal selection rates

Group imbalance

Gaps in representation

Quantitative Metrics

Selection Rate Difference

Weighted SR Difference

Entropy Difference

Selection Rate (SR)

Definition

- The selection rate (or admission rate) for group *i* represents the proportion of individuals in that group who were admitted.
- Formula: $SR_i = \frac{Admitted_i}{Applicants_i}$
- Interpretation: used as the base measure to compare across groups *(significant difference between rates may indicate bias)*

SR Difference

- Measures disparity between groups. Larger differences show greater imbalance.
- Formula: $\Delta SR = \max(SR_i) - \min(SR_i)$
- Interpretation: admissions should be equally likely regardless of group membership (~0)

(Castelnovo et al., 2022; Mehrabi et al., 2022; Raftopoulos et al., 2024)

Weighted Selection Rate (WSR)

Definition

- Represents the proportion of admitted applicants in each group *i* relative to the entire applicant pool.

- Formula: $WSR_i = w_i \cdot SR_i = \frac{Admitted_i}{Total\ Applicants}$

where,

$$SR_i = \frac{Admitted_i}{Applicants_i} \qquad w_i = \frac{Applicants_i}{\Sigma_j Applicants_j}$$

WSR Difference

- Measures the gap between the groups that contribute the most and least to the admission total

- Formula: $\Delta WSR_i = \max(WSR_i) - \min(WSR_i)$

- Interpretation: groups show contribute equally (~0)

(Castelnovo et al., 2022; Mehrabi et al., 2022; Raftopoulos et al., 2024)

Entropy (H)

Definition

- Measures the diversity or balance of the applicant or admitted pool across groups.
  - Higher entropy = more diversity
  - Lower entropy = less diversity
- Formula: $H(p) = -\sum_i p_i \log_2(p_i)$

$$H(q) = -\sum_i q_i \log_2(q_i)$$

where $p_i = \frac{Applicants_i}{\Sigma_j Applicants_j}$ $\quad q_i = \frac{Admitted_i}{\Sigma_j Admitted_j}$

Entropy Difference

- Comparing the entropy of applicants against the entropy of admitted applicants measures if the admission process increased or decreased diversity.
- Formula: $\Delta H = H(q) - H(p)$
- Interpretation:
  - ~0: no meaningful change
  - Greater than 0: diversity increased
  - Less than 0: diversity decreased (potential bias)

(Castelnovo et al., 2022; Mehrabi et al., 2022; Raftopoulos et al., 2024)

Summary of Differences

- Neutral (no change)
- Diversity increased
- Diversity decreased

Feature	Selection Rate	Weighted Selection Rate	Entropy	Highlights
Gender	0.025	0.060	0.005	No meaningful difference, diversity stayed consistent after admissions
Florida Residency	0.428	0.276	-0.241	Large disparity, strong preference for in-state applicants
US Military Status	0.066	0.446	0.020	Minimal impact, slight increase in diversity, but group is too small (~3%) to affect outcomes
Ethnicity	0.338	0.229	-0.154	Moderate disparity, admissions reduced ethnic diversity
Age	0.463	0.208	0.138	Positive change as age diversity is slightly higher after admissions
Continent of Birth	0.472	0.384	-0.155	Moderate disparity as geographic diversity decreased

Methodology

Predictive modeling and baseline model performance

Experimental Set Up

Data Split (80/20)

- *Training **set** (80%) for model building and validation*
- ***Test set** (20%) held out for final evaluation*
- *Stratified split to preserve admission*

Data Pre-Processing

- *Discretize continuous features (k-means clustering)*
- *One-hot encoding*

Cross-Validation & Model Training

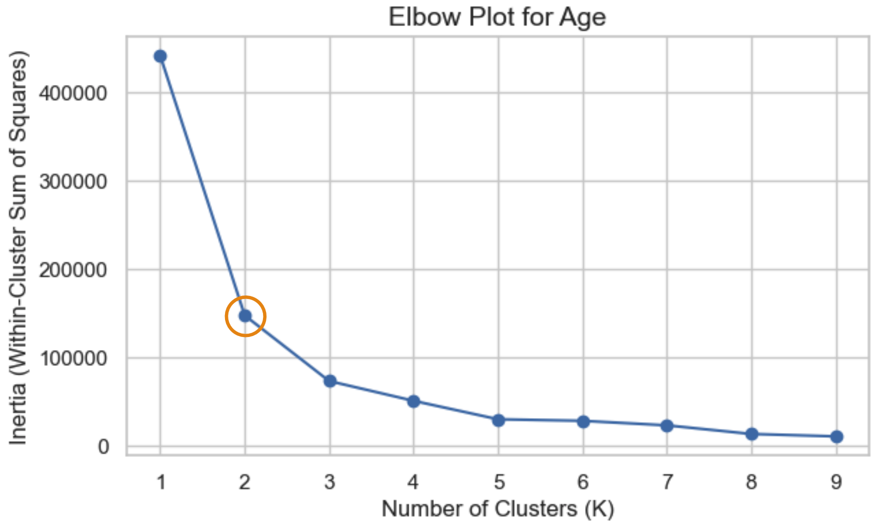
- *Hyper-parameter tuning using cross-validation (5-fold)*
- ***Re-train model with optimal hy per-parameters***

Final Evaluation Test Set

- *Evaluate on unseen test set*
- *Compare performance results*

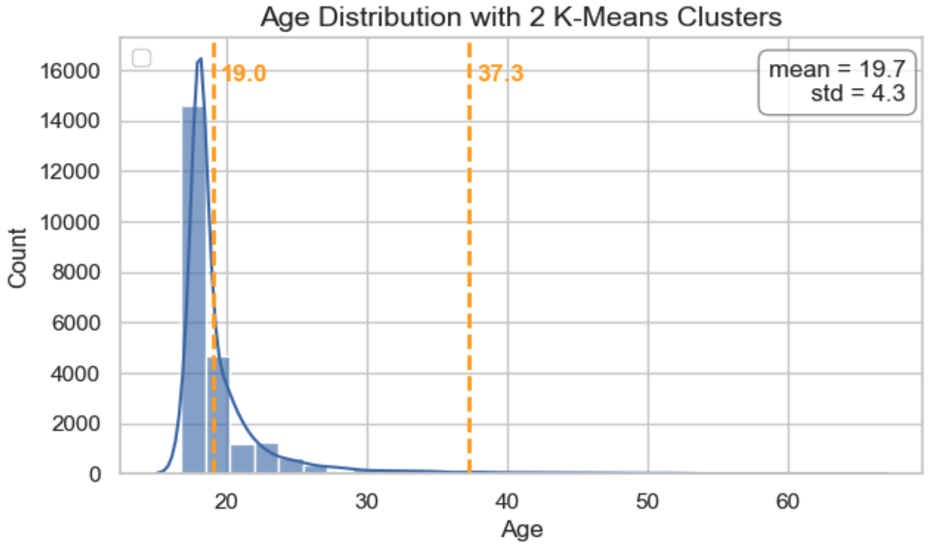
Age Feature Discretization

Optimal Num of Clusters



- The Elbow method shows a sharp drop at k = 2, indicating diminishing returns beyond two clusters

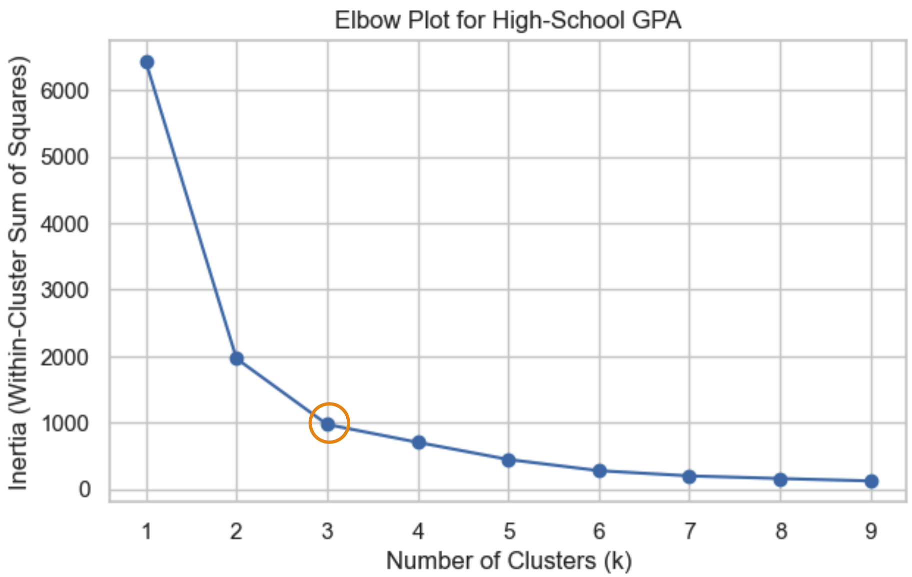
Distribution and Clusters



- 2 clusters capture most of the variation while maintaining balanced group sizes
- A 3rd cluster would form a very small, less meaningful subgroup

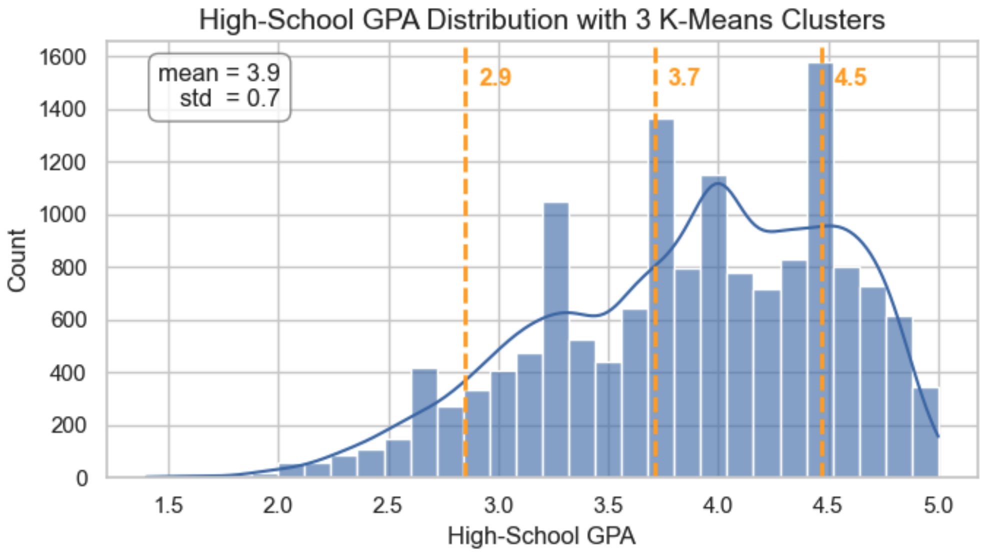
High-School GPA Discretization

Optimal Num of Clusters



- The Elbow method shows a noticeable drop in inertia at k = 3, after which gains become minimal

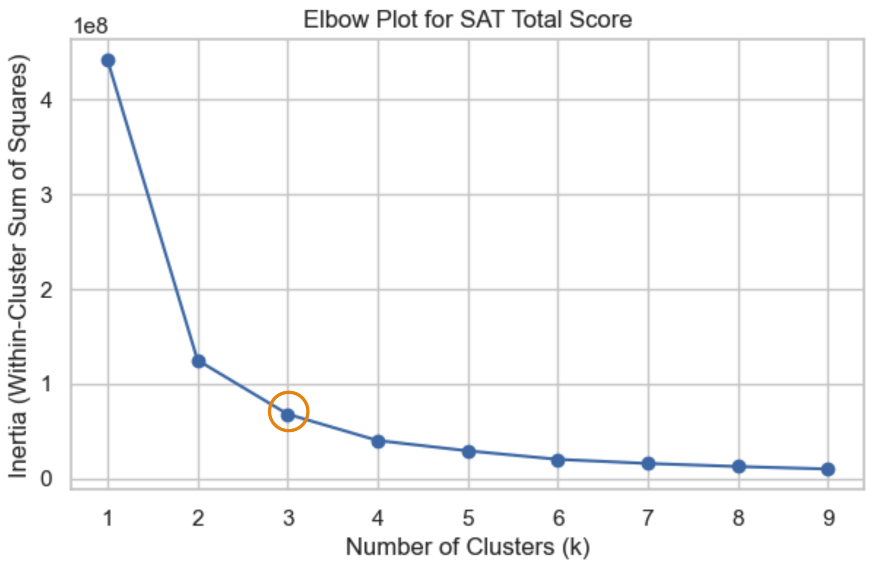
Distribution and Clusters



- 3 clusters capture most of the distribution
- Provides interpretable GPA bands aligned with real academic categories (low, mid, high), while preserving group balance

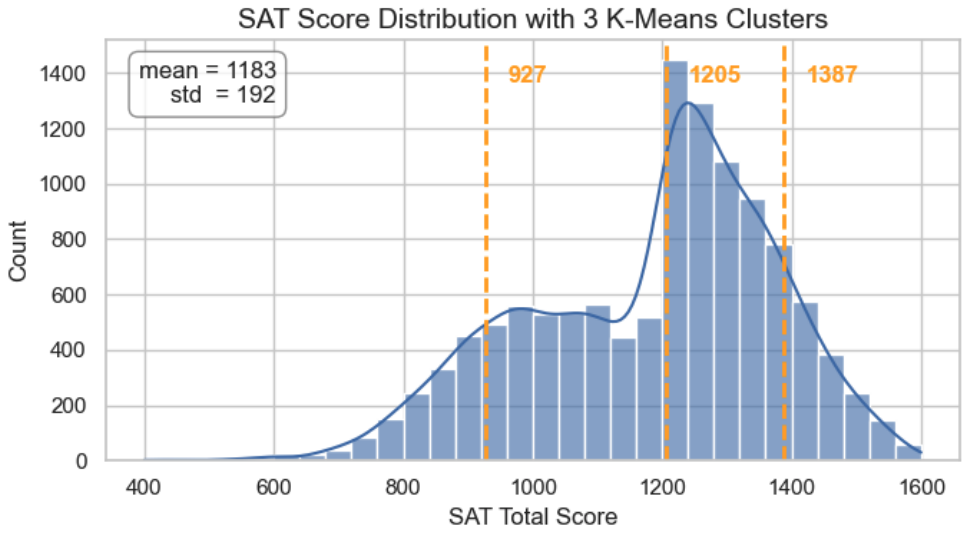
SAT Total Score Discretization

Optimal Num of Clusters



- The Elbow method shows a noticeable drop in inertia at k = 3, after which gains become minimal

Distribution and Clusters



- 3 clusters capture most of the distribution
- Provides interpretable SAT bands aligned with real academic categories (low, mid, high), while preserving group balance

Summary of Discretization

Feature	Number of Clusters	Range	Label	Note
Age	2	15 - 28 years old 29 - 68 years old	Younger Applicants Older Applicants	-
High-School GPA	3	1.4 - 3.2 3.3 - 4.0 4.1 - 5.0	Low GPA Mid GPA High GPA	Additional code of –1 used to identify missing values
SAT Total Score	3	400 - 1060 1070 - 1290 1300 - 1600	Low Score Mid Score High Score	Additional code of –1 used to identify missing values

Type of Models

Decision Trees	Random Forest	AdaBoost
Tree-structured model that repeatedly splits the data on the most informative features.	Ensemble of many decision trees trained on random subsets of data and features.	Boosting method that builds models sequentially, with each model focusing on correcting mistakes of previous ones.
	Pros	
• No need to normalize data • Handles missing values • Produces interpretable rules	• More robust than a single tree • Measures feature importance • Few preprocessing needs	• Often improves accuracy • Works well with weak learners (e.g., shallow trees)
	Cons	
• Can overfit without pruning • Biased toward features with many categories	• Harder to interpret • Larger computational footprint	• Sensitive to noise and outliers • Can overfit if boosting too many rounds

(Domingos, 2012; Hastie et al., 2009)

Type of Models

Logistic Regression	Gaussian Naïve Bayes	Bernoulli Naïve Bayes
Linear model that predicts the probability of the positive class.	Probabilistic model assuming features follow a normal distribution and contribute independently.	A Naïve Bayes variant for binary features, ideal for one-hot encoded datasets.
	Pros	
• Shows effect strength + direction • No distribution assumptions	• Works well with continuous features • Performs well on small data	• Works well with sparse data • Simple and computationally efficient
	Cons	
• Assumes linear relationships • Can underperform with complex, nonlinear patterns	• Assumes feat independence • Sensitive to distribution mismatches	• Assumes feat independence • Less effective when feature interactions matter

(Domingos, 2012; Hastie et al., 2009)

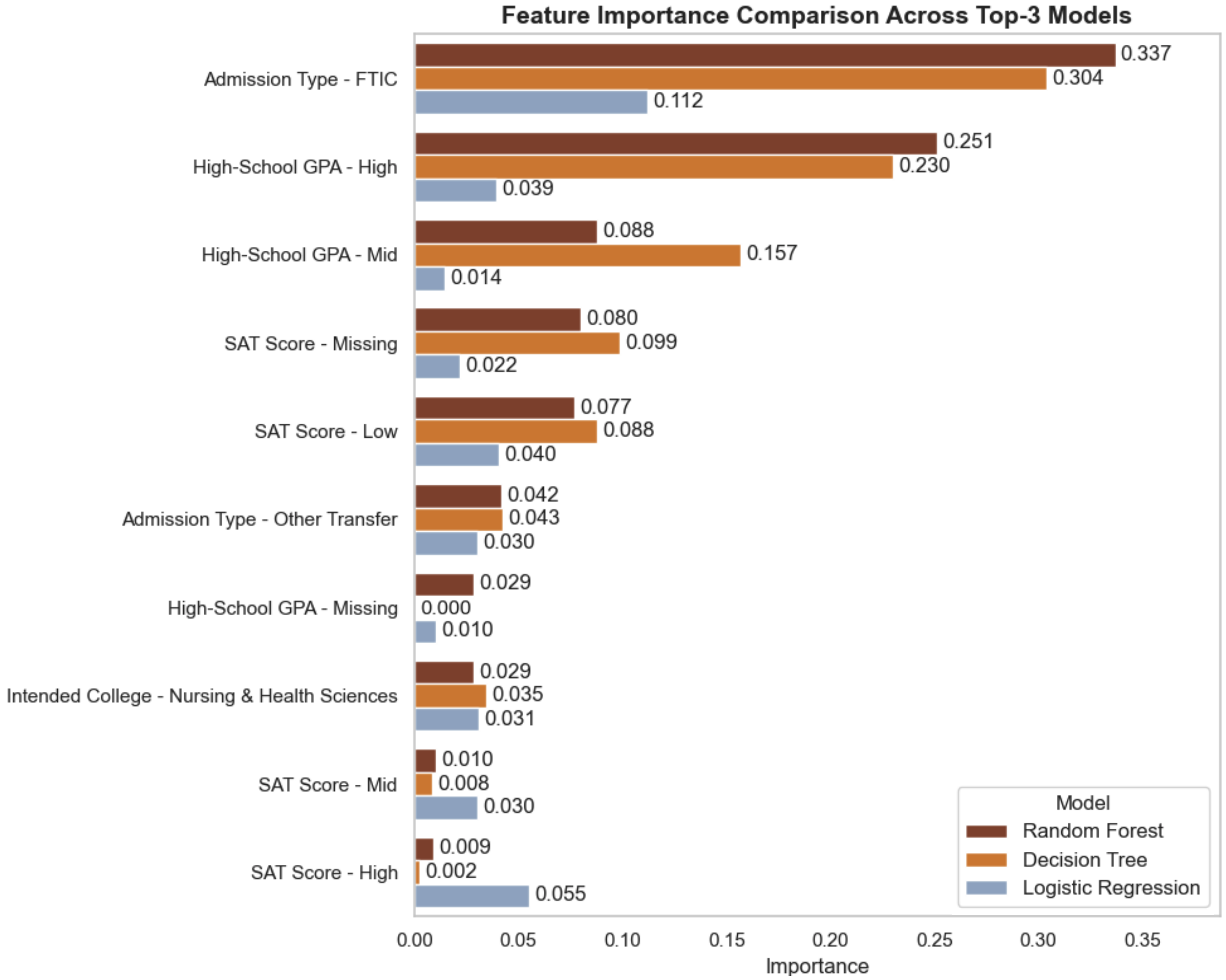
Model Performance Results

Highest performance

Model	Accuracy		Precision		Recall		F1-Score		ROC-AUC	
	Val	Test	Val	Test	Val	Test	Val	Test	Val	Test
Gaussian Naïve Bayes	0.866	0.869	0.859	0.860	0.863	0.869	0.861	0.864	0.921	0.925
Bernoulli Naïve Bayes	0.853	0.853	0.810	0.807	0.904	0.910	0.855	0.856	0.926	0.927
Logistic Regression	0.912	0.911	0.883	0.879	0.940	0.943	0.911	0.910	0.960	0.956
Decision Tree	0.916	0.914	0.891	0.884	0.941	0.943	0.915	0.913	0.971	0.968
Random Forest	0.917	0.915	0.890	0.886	0.942	0.944	0.915	0.914	0.972	0.969
AdaBoost	0.898	0.895	0.888	0.886	0.900	0.897	0.894	0.891	0.955	0.953

Feature Importance Comparison

- Academic factors and admission type are the strongest predictors of admission
- Demographic features are not among the top predictors (a positive indicator for fairness)



Note: *Logistic Regression coefficients were normalized to sum to 1 for comparison purposes.*

Knight Foundation School of Computing and Information Sciences

Fairness Evaluation

Identifying disparities across protected and unprotected groups

Protected & Unprotected Groups in Fairness Evaluation

Protected or Unprivileged Group

- Group that may face disadvantage or historical underrepresentation
- Used to check whether the model treats them unfairly
- Examples in our dataset:
  - *Females*
  - *Out-of-State*

Unprotected or Privileged Group

- Group typically receiving more favorable outcomes
- Serves as the "reference group" in fairness comparisons
- Examples in our dataset:
  - *Males*
  - *In-State*

Fairness Metrics

Statistical Parity or Demographic Parity

- Probability of receiving a positive prediction should be the same across groups.
- **Formula:**

$$P(\hat{Y} = 1 \mid A = \text{unpriv}) - P(\hat{Y} = 1 \mid A = \text{priv})$$

- **Interpretation:**
  - Fair: ~0 difference
  - Potential disparity if negative

Disparate Impact

- Measures the **ratio of positive prediction** rates between unprivileged and privileged groups.
- Formula: $\dfrac{P(\hat{Y} = 1 \mid A = \text{unpriv})}{P(\hat{Y} = 1 \mid A = \text{priv})}$
- **Interpretation:**
  - Fair or acceptable: ≥ 0.8
  - Potential disparity if < 0.8

Fairness Metrics

Equal Opportunity

- True positive rate (**TPR**) must be equal across groups.
- **Formula:** $TPR_{unpriv} - TPR_{priv}$

  where, $TPR = \dfrac{TP}{TP + FN}$
- **Interpretation:**
  - ~0 = fair
  - Potential disparity if negative

Equalized Odds

- True positive rate (**TPR**) and false positive rate (**FPR**) must match across groups.
- Formula:

  $|TPR_{priv} - TPR_{unpriv}| + |FPR_{priv} - FPR_{unpriv}|$

  where, $FPR = \dfrac{FP}{TN + FP}$
- **Interpretation:**
  - ~0 = fair

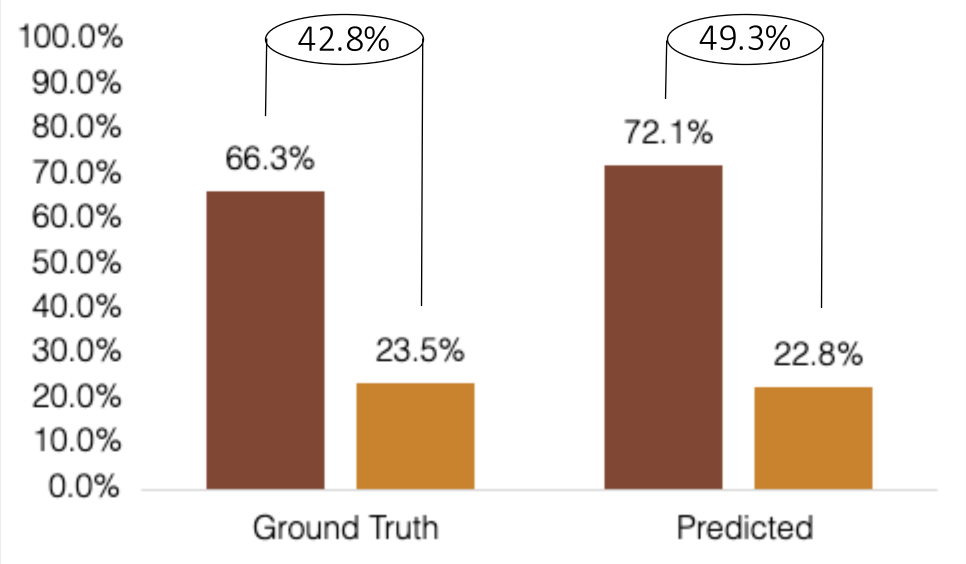
Model Fairness Performance Results

Category	Protected Group	Statistical Parity (↓)	Disparate Impact (↑)	Equal Opportunity (↓)	Equalized Odds (↓)
Gender	Female	0.0452	0.9156	0.0127	0.0443
Florida Residency	Out-of-State	0.4931	0.3158	0.1115	0.2877
Military Status	Non-Military	0.0538	0.9043	0.0832	0.1917
Age	Younger (15 – 28 years)	0.2358	0.6797	0.0446	0.2108
Ethnicity	Black/African American or Nonresident	0.2546	0.5761	0.0576	0.1369
Continent of Birth	Africa, Asia or Oceania	0.2871	0.4659	0.0453	0.1500

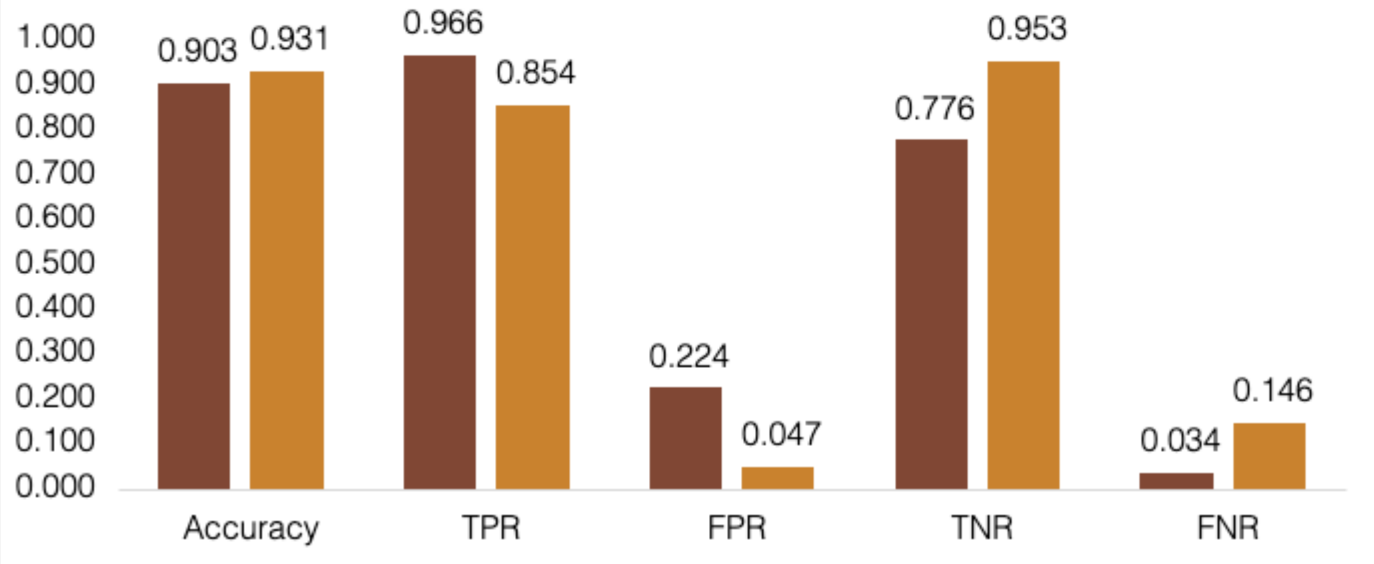
Best Performing   Worst Performing

Deep-Dive into Florida Residency

Selection Rates by Group



Classification Rates by Group



Privileged Group (In-State Applicants)

Unprivileged Group (Out-of-State Applicants)

- The model increases the selection gap, widening the disparity
- Exhibits predictive bias favoring privileged group (falsely admits in-state versus out-of-state applicants)

Bias Mitigation Techniques

Pre-processing strategies and their impact on predictive performance and fairness

Pre-Processing Bias Mitigation

Pre-Processing Techniques

- Modify the training data **before** model training
- Methods:
  - Random Oversampling (duplicate samples)
  - SMOTE Oversampling (synthetic samples)
- **Goal:** increase representation of under-represented/protected groups

Steps

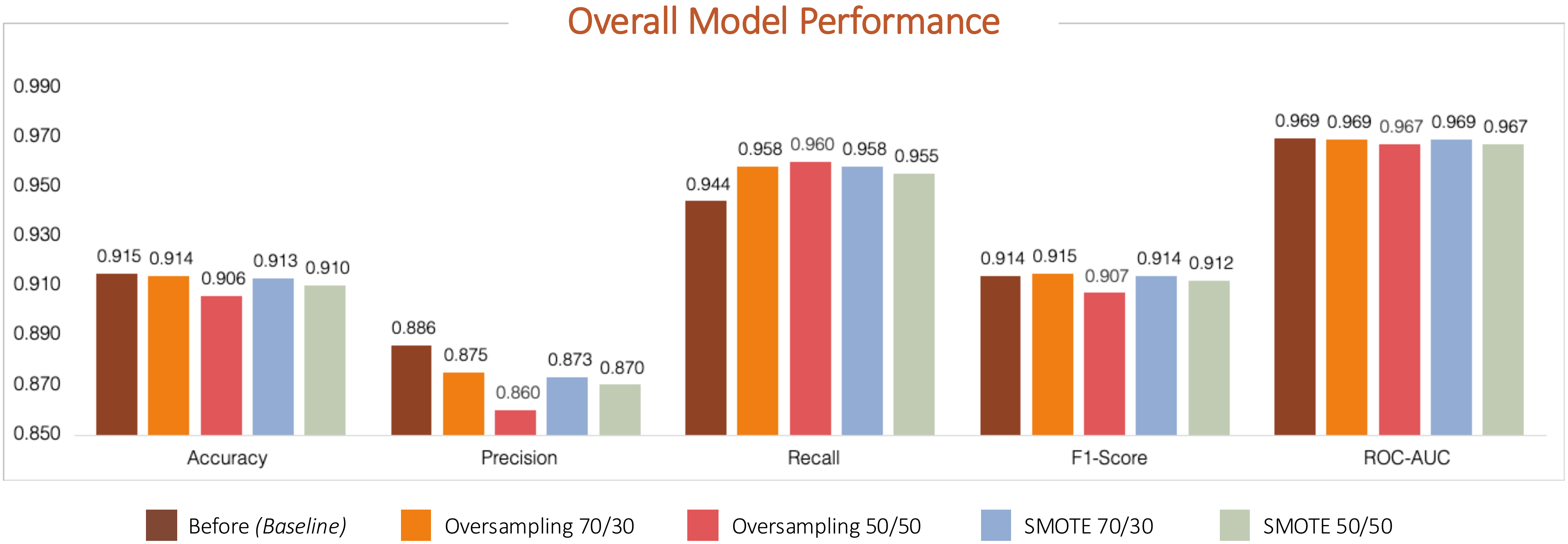
1. Rebalance unprivileged group in the training set only

FL Residency Group	Before	After	
In-State (1)	78%	70%	50%
Out-of-State (0)	22%	30%	50%

2. Re-train the Random Forest model
3. Test model on unseen data (untouched test set)
4. Evaluate model performance and fairness impact on Florida Residency groups

(Lumenova, 2024)

Pre-Processing Bias Mitigation Results



Pre-Processing Bias Mitigation Results

Metrics for FL Residency		Before *(Baseline)* 78/22		Random Oversampling 70/30		Random Oversampling 50/50		SMOTE 70/30		SMOTE 50/50	
		Priv	Unpriv	Priv	Unpriv	Priv	Unpriv	Priv	Unpriv	Priv	Unpriv
Accuracy	↑	0.903	0.931	0.00%	-0.17%	-0.10%	-2.06%	-0.10%	-0.29%	-0.10%	-1.09%
TPR	↑	0.966	0.854	-0.23%	9.65%	-0.82%	13.55%	0.05%	8.42%	-1.00%	11.29%
FPR	↓	0.224	0.047	-1.98%	54.26%	-5.95%	122.34%	1.59%	51.06%	-7.54%	86.17%
TNR	↑	0.776	0.953	0.57%	-2.70%	1.71%	-6.09%	-0.46%	-2.54%	2.17%	-4.29%
FNR	↓	0.034	0.146	6.41%	-56.63%	23.08%	-79.52%	-1.28%	-49.40%	28.21%	-66.27%
Statistical Parity	↓	0.493		-8.38%		-16.34%		-6.77%		-13.24%	
Disparate Impact	↑	0.316		17.34%		32.94%		15.08%		25.50%	
Equal Opportunity	↓	0.111		-75.93%		-89.07%		-64.13%		-95.19%	
Equalized Odds	↓	0.288		-39.91%		-59.30%		-32.03%		-56.95%	

● Improvement vs Baseline ● Least Worst vs Baseline

Summary of Bias Mitigation Results

- Key Findings
  - Of the two techniques, Oversampling performed the best vs SMOTE
  - Oversampling 50/50 split performs the best for FL Residence
  - However, 70/30 split still give us fairness gains and gives us the best overall model performance

- Limitations
  - SMOTE synthetic samples may not perfectly reflect real applicants
  - Oversampling often leads to overfitting

Conclusion

Key Findings

- Predictive performance was consistently strong across models
- Fairness analysis revealed consistent disparities across certain groups
- Indicates data-driven structural patterns, not model-specific issues
- Demographic features were not top predictors

Impact of Bias Mitigation

- Pre-processing techniques significantly improved fairness
- Predictive performance remained strong (overall and for specific group)
- Fairness can be improved without sacrificing accuracy

Future Work

Data & Model Improvement

- Expand experiments to additional demographic features
- Incorporate additional semesters to increase data samples
- Explore additional fairness-aware algorithms and bias mitigation techniques

Expand Evaluation

- Conduct intersectional fairness analysis
- Split analysis by type of application and/or intended college
- Evaluate trends overtime or by semester cohort

References

[1] Barocas, S., Hardt, M., & Narayanan, A. (2023). Fairness and machine learning: Limitations and opportunities. *The MIT Press.* https://fairmlbook.org/

[2] Binkley, C. (2025, August 15). Critics say Trump's push for fairness in college admissions is leaving out Legacy Preferences. *AP News.* https://apnews.com/article/trump-legacy-admissions-status-affirmative-action-ad4ab9b2cc2f16730685076c07aa2146

[3] Bird, K. A., Castleman, B. L., & Song, Y. (2024). Are algorithms biased in education? Exploring racial bias in predicting community college student success. *Journal of Policy Analysis and Management*, *44*(2), 379–402. https://doi.org/10.1002/pam.22569

[4] Castelnovo, A., Crupi, R., Greco, G. et al. (2022). *A clarification of the nuances in the fairness metrics landscape.* Scientific Reports 12, 4209. https://doi.org/10.1038/s41598-022-07939-1

[5] Chouldechova, A. (2016). *Fair prediction with disparate impact: A study of bias in recidivism prediction instruments.* ArXiv. https://doi.org/10.48550/arXiv.1610.07524

[6] Domingos, P. (2012). A few useful things to know about machine learning. *Communications of the ACM, 55*(10), 78–87. https://doi.org/10.1145/2347736.2347755

[7] Dressel, J., & Farid, H. (2018). The accuracy, fairness, and limits of predicting recidivism. *Science Advances, 4*(1). https://doi.org/10.1126/sciadv.aao5580

[8] Feldman, M., Friedler, S. A., Moeller, J., Scheidegger, C., & Venkatasubramanian, S. (2015). *Certifying and removing disparate impact.* ArXiv. https://arxiv.org/pdf/1412.3756

References

[9] Gándara, D., Anahideh, H., Ison, M. P., & Picchiarini, L. (2024). Inside the black box: Detecting and mitigating algorithmic bias across racialized groups in college student-success prediction. *AERA Open, 10*. https://doi.org/10.1177/23328584241258741

[10] Hardt, M., Price, E., & Srebro, N. (2016). *Equality of Opportunity in Supervised Learning*. ArXiv. https://arxiv.org/pdf/1610.02413

[11] Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer. https://hastie.su.domains/ElemStatLearn/

[12] Irving, J. M. (2024, August 2). Part 1: Creating Dataset Objects in AI Fairness 360 and Exploring Fairness Metrics. *Medium*. https://medium.com/@james.irving.phd/blog-post-series-ai-fairness-360-mitigating-bias-in-machine-learning-models-c1ec744c91c4

[13] Irving, J. M. (2024, August 2). Part 2: Implementing Bias Mitigation Techniques with AI Fairness 360. *Medium*. https://medium.com/@james.irving.phd/blog-post-series-ai-fairness-360-mitigating-bias-in-machine-learning-models-2268e01584bd

[14] Kleinberg, J., Ludwig, J., Mullainathan, S., & Rambachan, A. (2018). Algorithmic Fairness. *AEA Papers and Proceedings, 108*, 22–27. https://www.jstor.org/stable/26452698

[15] Lumenova. (2024, July 23). *Fairness and bias in Machine Learning: Mitigation Strategies.* Lumenova AI. https://www.lumenova.ai/blog/fairness-bias-machine-learning/

[16] Lawyers' Committee for Civil Rights Under Law. (2023, August). Summary of the Supreme Court's Decision in SFFA v. Harvard and SFFA v. UNC. https://www.lawyerscommittee.org/wp-content/uploads/2023/08/LC_Harvard-UNC-Cases_D.pdf

References

[17] Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2019). *A survey on bias and fairness in machine learning*. ArXiv. https://doi.org/10.48550/arXiv.1908.09635

[18] Raftopoulos, G., Davrazos, G., & Kotsiantis, S. (2024). Fair and transparent student admission prediction using machine learning models. *Algorithms, 17*(12), 572. https://doi.org/10.3390/a17120572

[19] *U.S. Secretary of Education Linda McMahon directs National Center for Education Statistics to collect universities' data on race discrimination in admissions.* U.S. Department of Education. (2025, August 7). https://www.ed.gov/about/news/press-release/us-secretary-of-education-linda-mcmahon-directs-national-center-education-statistics-collect-universities-data-race-discrimination-admissions

[20] Van Busum, K., & Fang, S. (2025). Interactive mitigation of biases in machine learning models for Undergraduate Student Admissions. *AI, 6*(7), 152. https://doi.org/10.3390/ai6070152

[21] Varshney, K. (2018, September 19). Introducing AI Fairness 360. *IBM*. https://research.ibm.com/blog/ai-fairness-360

Appendix

- List of original features
- Data quality issues and challenges
- Feature importance

Original Features by Category

Demographic Information

- Gender
- Age
- Ethnicity
- Country of Birth
- US Military Status
- Florida Residency

High-School & Academic Background

- High-School State
- High-School Type
- Highest Level of Education
- High-School GPA
- Undergraduate GPA
- Graduate GPA

College Application Details

- Admitted
- Admission Type
- Intended Major
- Honors College
- Enrolled
- Denied
- Incomplete
- Matriculated

Standardized Test Scores

- ACT Composite Score
- ACT Sub-scores (English, Math, Reading, Science, Writing)
- ACT-to-SAT Conversions (English/Reading/Writing and Total)
- SAT Old (Composite, Math, Verbal, Writing Sub-Scores)
- SAT New (Cumulative, Math, Reading & Writing Sub-scores, Retake Indicator)

Data Preparation Workflow

Original Dataset

- **Total features:** 42
- **Features mix:**
  - Demographic information: 6
  - High-school and academic background: 6
  - Standardized test scores: 22
  - College application details: 8

Data Cleaning & Preprocessing

- **Standardized test scores:** combined into unified features
- **Missing values:** added binary flags & consistent handling
- **Categorical features:** grouped categories for clarity
- **Indicators:** converted key features into binary variables
- **EDA:** explored distributions & identified patterns

Feature Set for Bias Investigation & Modeling

- Clean dataset
- **Total features:** 14
- **Features mix:**
  - Demographic information: 6
  - High-school and academic background: 3
  - Standardized test scores: 2
  - College application details: 3

High Cardinality Features

Intended Major to FIU College	Country of Birth to Continent	High-School State
• 134 original possible values • Decreased to 9	• 196 original possible values • Decreased to 7	• 131 original possible values • Decreased to 4

	Total Count	Total %
Arts, Sciences & Education	8111	27.2
Business	7004	23.5
Engineering & Computing	5412	18.1
Nursing & Health Sciences	3203	10.7
Missing/Unknown	2114	7.1
Arts & Communication	1925	6.5
Public Affairs	1290	4.3
Hospitality & Tourism Management	770	2.6

	Total Count	Total %
North America	23098	77.4
South America	2267	7.6
Asia	2168	7.3
Europe	871	2.9
Africa	798	2.7
Missing/Unknown	614	2.1
Oceania	13	0.0

	Total Count	Total %
Missing/Unknown	15086	50.6
Florida	12938	43.4
Other U.S.	1570	5.3
International/Other	235	0.8

Missing Value Analysis

13 features with 90% or more missing values

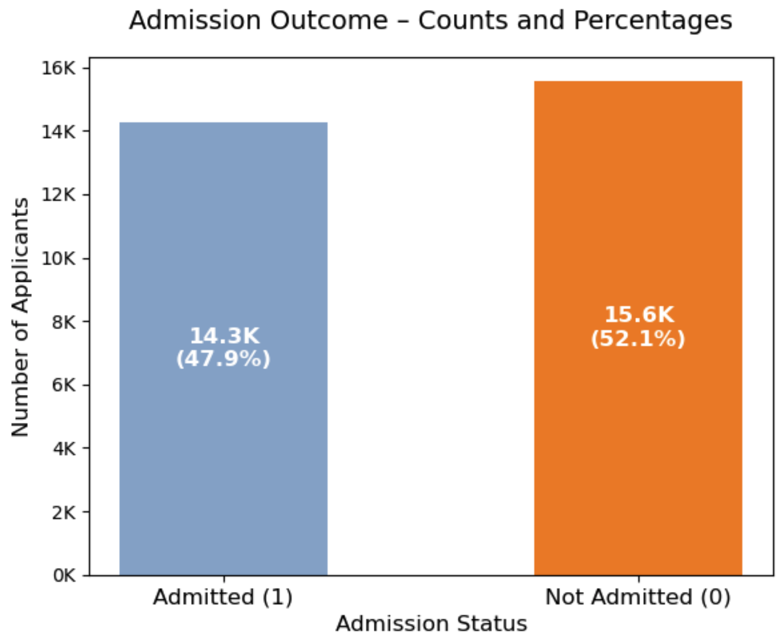
- Undergraduate and Graduate Cumulative GPA (*100%*)
- SAT (Old Format) Scores (*99% -100%*)
    - Composite/Cumulative
    - Sub-scores: Writing, Math, Verbal
- ACT Scores (*92%- 99%*)
    - Composite
    - Sub-scores: English, Math, Reading, Science, Writing

13 features with missing values between 50% to 90%

- ACT to SAT Conversion Scores (*86%*)
- SAT (New Format) Scores (*59%-64%*)
    - Composite/Cumulative
    - Sub-Scores: Reading, Math, Reading & Writing, Writing & Language
    - SAT Retake/Report Indicator
- High School Type and State (*51% -53%*)

Applicants Neither Admitted nor Denied

- Exclude denied feature, treat the applicants as not admitted
- Avoid losing 12% of our data, still helpful for analysis and modeling



		Denied 0	Denied 1	
Admitted	0	3.5K	12.1K	15.6K
	1	14.3K	0.0K	14.3K
		17.7K	12.1K	29.8K

Non-Resident Alien Ethnicity

- Non-resident Alien is not a true "ethnicity" but more of a migratory/residency status
- Challenging to accurately understand this cohort as country of birth might not be a good proxy

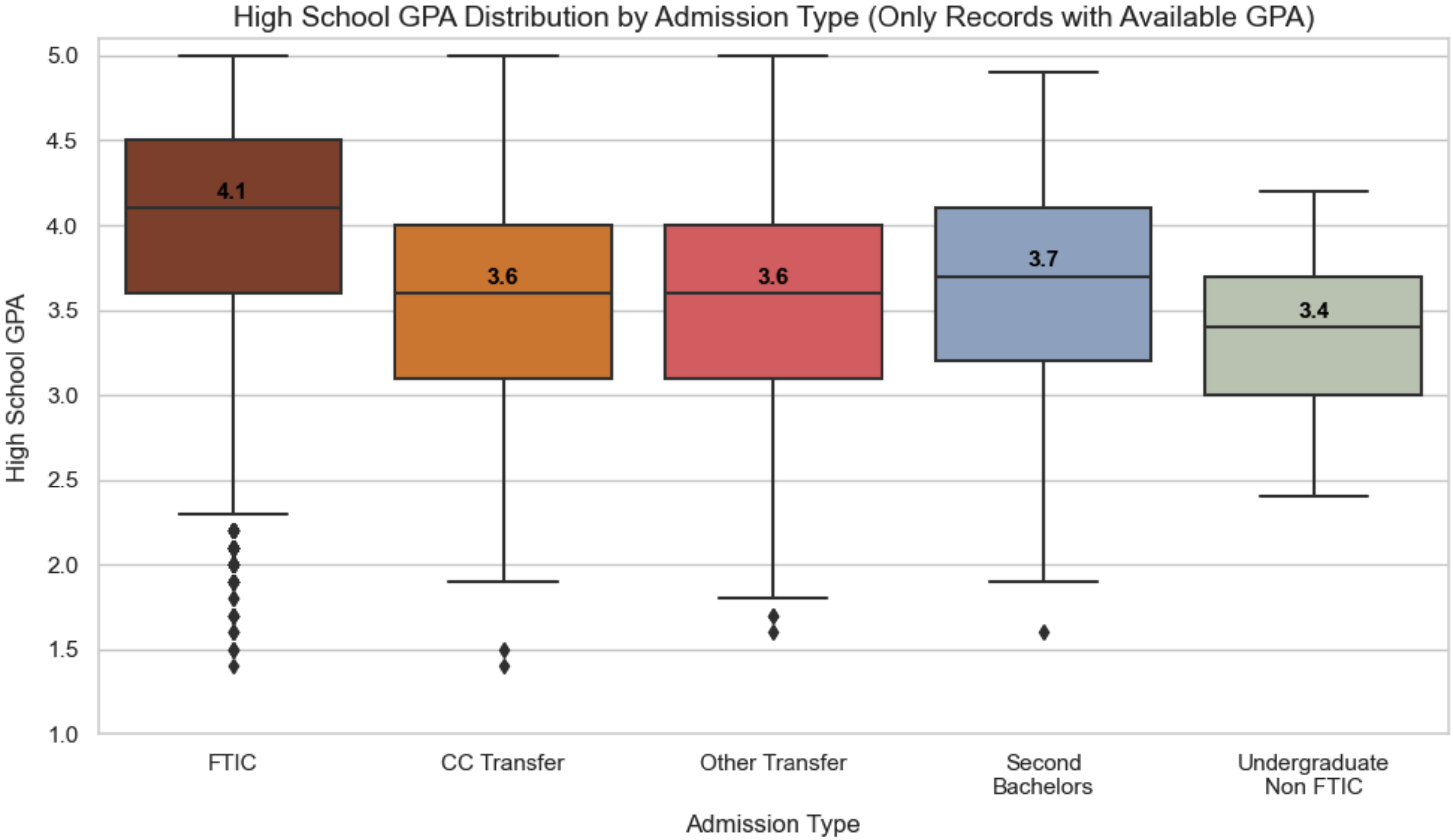
Ethnicity	Continent							Total
	North America	South America	Asia	Europe	Africa	Missing/Unknown	Oceania	
Hispanic/Latino	47%	31%	1%	10%	0%	24%	15%	**39%**
Nonresident Alien	6%	68%	85%	69%	86%	63%	62%	**22%**
White	19%	0%	3%	17%	4%	5%	0%	**16%**
Black or African American	18%	0%	0%	1%	9%	2%	8%	**15%**
Asian	3%	0%	9%	1%	0%	4%	8%	**3%**
Two or Mor Races	4%	0%	1%	2%	0%	1%	0%	**3%**
Missing/Unknown	3%	0%	1%	0%	1%	0%	0%	**2%**
Pacific Islander	0%	0%	0%	0%	0%	0%	8%	**0%**
American Indian or Alaska Native	0%	0%	0%	0%	0%	0%	0%	**0%**
Total	**77%**	**8%**	**7%**	**3%**	**3%**	**2%**	**0%**	**100%**
	23.1K	**2.3K**	**2.2K**	**0.9K**	**0.8K**	**0.6K**	**0.0K**	**29.8K**

Admission Type vs High-School GPA

- Almost all FTIC admitted students had the High-School GPA
- Followed by Undergrad Non FTIC (however this group is very small)

Admission Type	High-School GPA Missing				Admitted	Not Admitted
	No	Yes	Total	% Missing	% Missing	% Missing
FTIC	13,055	7,977	21,032	37.9%	0.5%	59.7%
CC Transfer	3,249	1,396	4,645	30.1%	29.5%	33.8%
Other Transfer	1,723	1,637	3,360	48.7%	39.5%	59.3%
Second Bachelors	274	222	496	44.8%	39.5%	83.6%
Undergraduate Non FTIC	262	34	296	11.5%	10.9%	27.3%
Total	**18,563**	**11,266**	**29,829**	**37.8%**	**15.0%**	**58.7%**

Admission Type vs High-School GPA



Key Insights

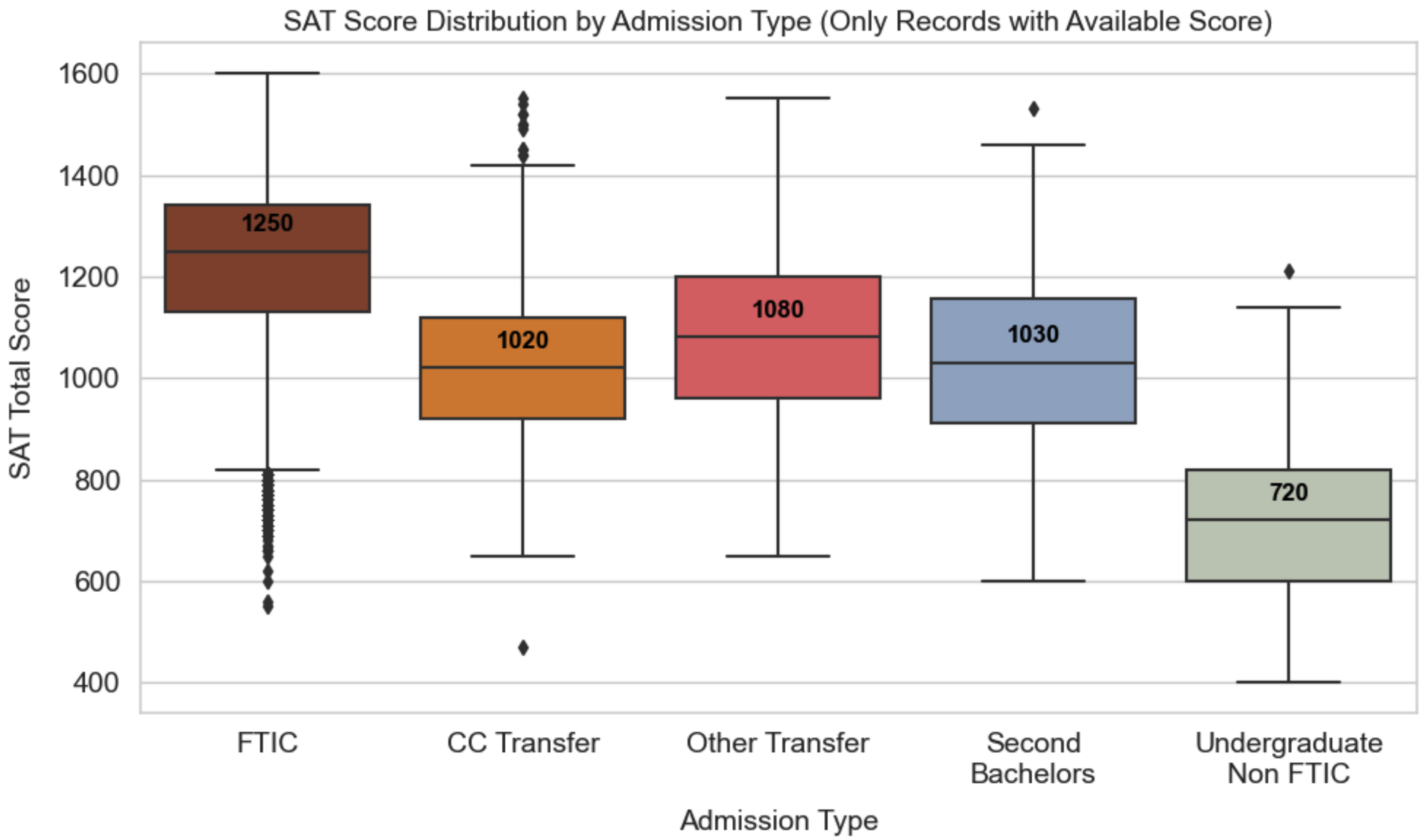
- FTIC applicants dominate and show stronger academic preparation
- Transfer students are more diverse in GPA distribution
- Smaller groups (Second Bachelors, Non-FTIC) present lower averages

Admission Type vs ACT/SAT Score

- All FTIC admitted students had the SAT or ACT to SAT Conversion Score

Admission Type	ACT/SAT Missing				Admitted	Not Admitted
	No	Yes	Total	% Missing	% Missing	% Missing
FTIC	11,781	9,251	21,032	44.0%	0.0%	69.6%
CC Transfer	1,748	2,897	4,645	62.4%	62.4%	62.3%
Other Transfer	1,203	2,157	3,360	64.2%	52.7%	77.4%
Second Bachelors	99	397	496	80.0%	77.6%	100.0%
Undergraduate Non FTIC	138	158	296	53.4%	52.3%	81.8%
Total	**14,969**	**14,860**	**29,829**	**49.8%**	**27.6%**	**70.2%**

Admission Type vs ACT/SAT Score



Key Insights

- FTIC have the highest scores
- Transfers have moderate scores
- Second Bachelors show similar medians, but with a wider spread due to smaller sample size
- Undergraduate Non-FTIC have the lowest SAT scores

Feature Importance

Decision Tree & Random Forest

- Based on how much each feature reduces impurity (Gini or entropy) when it's used to split the data.
- The model tracks how much impurity reduction each feature contributes across the whole tree or across all trees in the forest.
- Those reductions are normalized so that importances sum to 1.

Logistic Regression

- Coefficients represent the strength and direction of association between each feature and the log-odds of the positive class (admitted = 1).
  - Positive coefficients: more likely to get admitted
  - Negative coefficients: less likely to get admitted
- Magnitude reflects the relative strength (in log-odds space).

(Domingos, 2012; Hastie et al., 2009)

Top 10 Features by Model

- Summary of top 10 features for the best performing 3 models
  - Academic factors and admission type are the strongest predictors of admission
  - Demographic features are not among the top predictors (a positive indicator for fairness)

Decision Tree

Feature	Importance
Admission Type - FTIC	0.304
High-School GPA - High	0.230
High-School GPA - Mid	0.157
SAT Score - Missing	0.099
SAT Score - Low	0.088
Admission Type - Other Transfer	0.043
Intended College - Nursing & Health Sciences	0.035
Highest Education - Bachelors or Higher	0.011
SAT Score - Mid	0.008
Highest Education - Associate Degree	0.007

Random Forest

Feature	Importance
Admission Type - FTIC	0.337
High-School GPA - High	0.251
High-School GPA - Mid	0.088
SAT Score - Missing	0.080
SAT Score - Low	0.077
Admission Type - Other Transfer	0.042
High-School GPA - Missing	0.029
Intended College - Nursing & Health Sciences	0.029
SAT Score - Mid	0.012
SAT Score - High	0.009

Logistic Regression

Feature	Coefficient
Admission Type - Second Bachelors	6.593
Admission Type - FTIC	-5.762
Highest Education - Bachelors or Higher	-4.565
SAT Score - High	2.841
Highest Education - No Degree	2.667
Admission Type - Undergrad Non-FTIC	2.257
Highest Education - Associate Degree	2.141
SAT Score - Low	-2.078
High-School GPA - High	2.026
Intended College - Nursing & Health Sciences	-1.588

Sample Metrics Breakdown

Gender	Admitted		Total	Weight	SR	ΔSR	WSR	ΔWSR	Entropy (H)				
									Applicants		Admitted		ΔH
	No	Yes							p	H(p)	q	H(q)	
Female	9.1K	8.0K	17.2K	0.575	0.468	0.025	0.269	0.060	0.575	0.984	0.563	0.989	0.005
Male	6.4K	6.2K	12.7K	0.425	0.493		0.209		0.425		0.437		

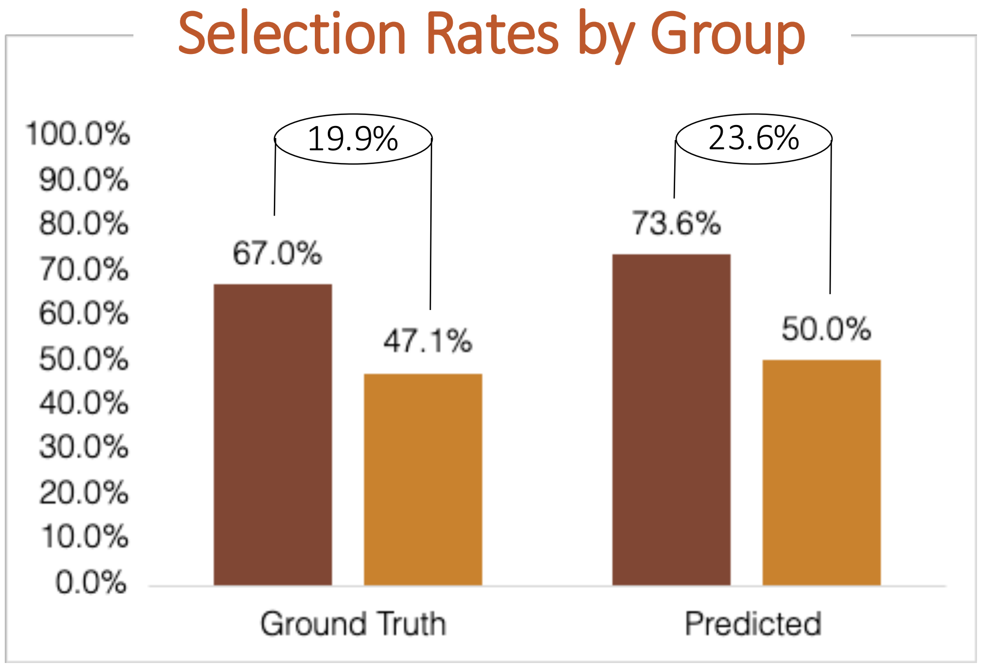
Small difference, meaning balanced admission rates

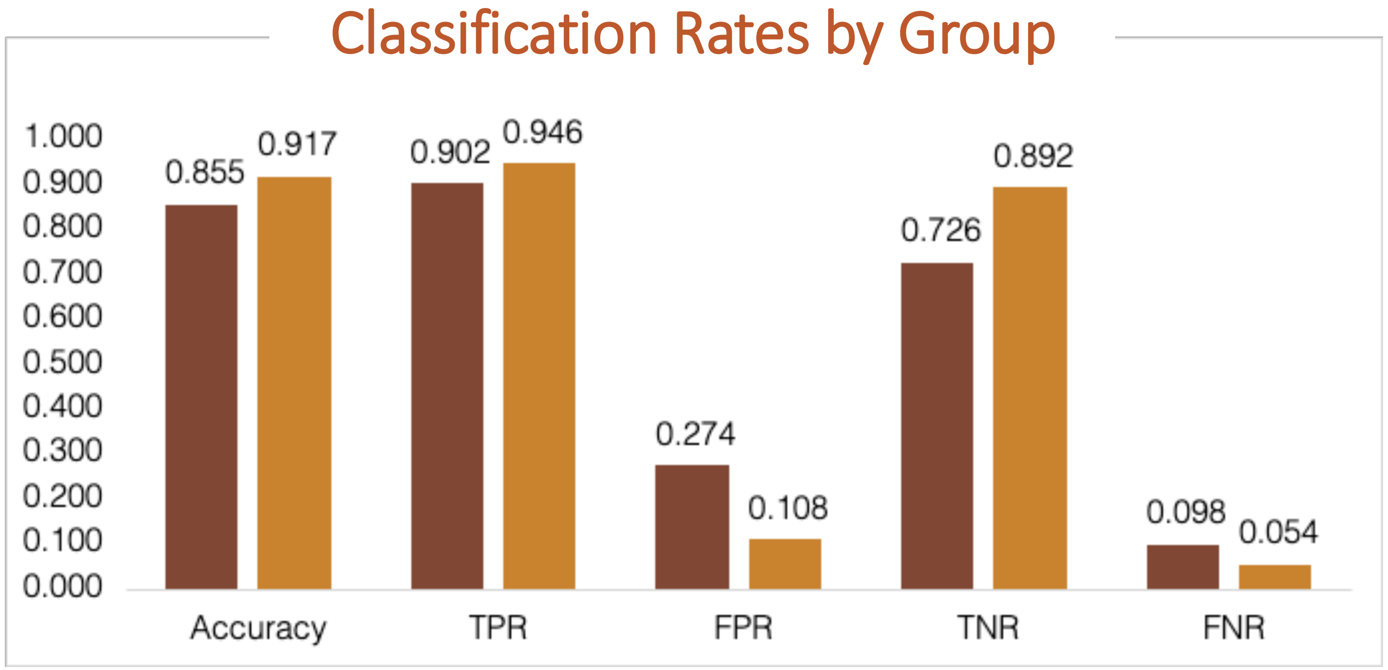
Low disparity, similar contribution to total admissions

No change in diversity

Note: for binary features (n = 2), entropy values near 1 indicate high diversity and equal representation across groups vs values near 0 reflect concentration in a single group.

Deep-Dive into Age





Privileged Group (Older Applicants)

Unprivileged Group (Younger Applicants)

- The model increases the selection gap, but does not show accuracy bias in favor of the privileged group
- However, it still gives benefit to the older applicants (higher FPR)

Pre-Processing Bias Mitigations Results

Metrics for FL Residency	Before *(Baseline)* 78/22		Random Oversampling 70/30		Random Oversampling 50/50		SMOTE 70/30		SMOTE 50/50	
	Priv	Unpriv	Priv	Unpriv	Priv	Unpriv	Priv	Unpriv	Priv	Unpriv
Accuracy ↑	0.903	0.931	0.903	0.929	0.902	0.911	0.902	0.928	0.902	0.920
TPR ↑	0.966	0.854	0.964	0.937	0.958	0.970	0.966	0.926	0.956	0.951
FPR ↓	0.224	0.047	0.219	0.073	0.210	0.105	0.227	0.072	0.207	0.088
TNR ↑	0.776	0.953	0.781	0.927	0.790	0.895	0.773	0.928	0.793	0.912
FNR ↓	0.034	0.146	0.036	0.063	0.042	0.030	0.034	0.074	0.044	0.049
Statistical Parity ↓	0.493		0.452		0.413		0.460		0.428	
Disparate Impact ↑	0.316		0.371		0.420		0.363		0.396	
Equal Opportunity ↓	0.111		0.027		0.012		0.040		0.005	
Equalized Odds ↓	0.288		0.173		0.117		0.196		0.124	

● Improvement vs Baseline ● Least Worst vs Baseline