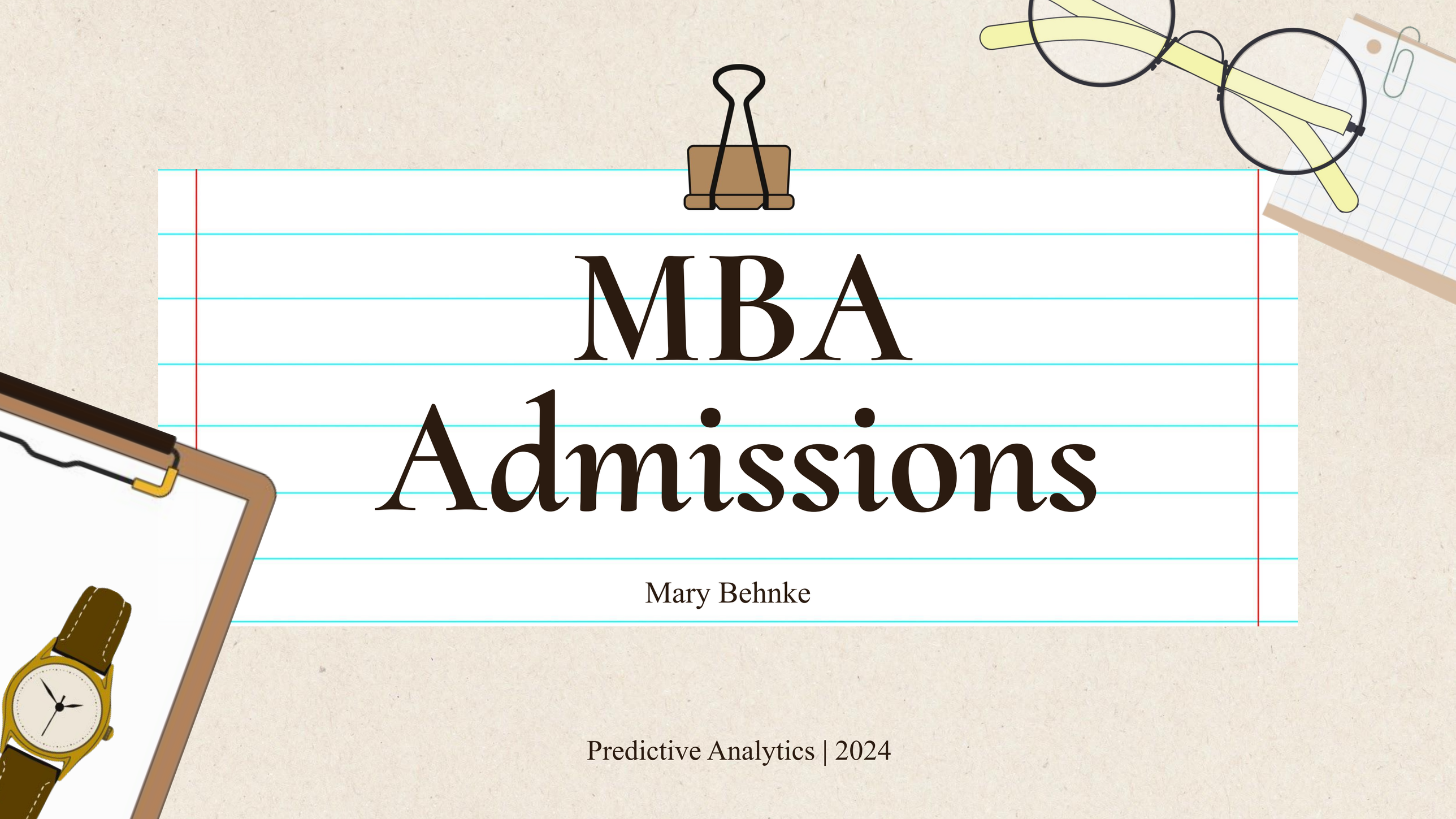
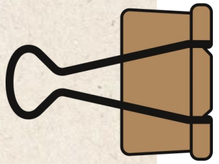


MBA Admissions

Mary Behnke

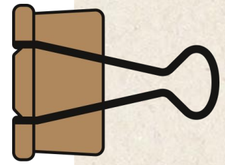
Predictive Analytics | 2024

AGENDA



- Dataset Information
- Question
- Preparation
- Exploratory Data Analysis

- Rejected vs Not Rejected
- Admitted vs Waitlisted
- Gender & International
- Takeaways



DATASET

MBA Dataset

source: Kaggle

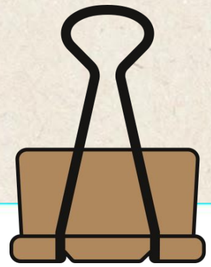
10

Variables

6134

Observations

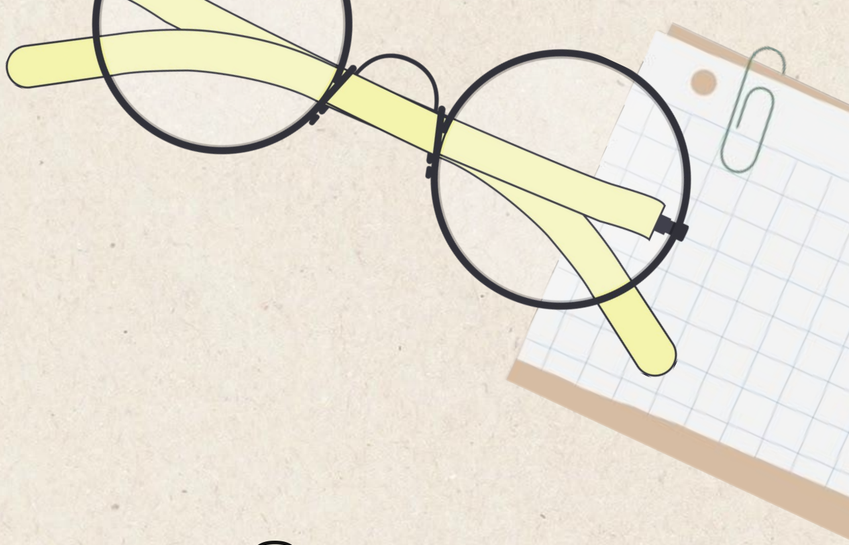
applicationID, gender, international, gpa, major, race, gmat, work_exp, work_industry, admission



Can a model predict whether an applicant will be rejected or admitted?

Is there any evidence for bias in admissions?

PREPARATION



Dealing with NA Values

- *race* column
  - imputed “Not Provided”
- *admission* column
  - imputed “Rejected”

Creating New Columns

- Four new columns (all factors):
  - *rejectID*
  - *admitID*
  - *genderID*
  - *intID*

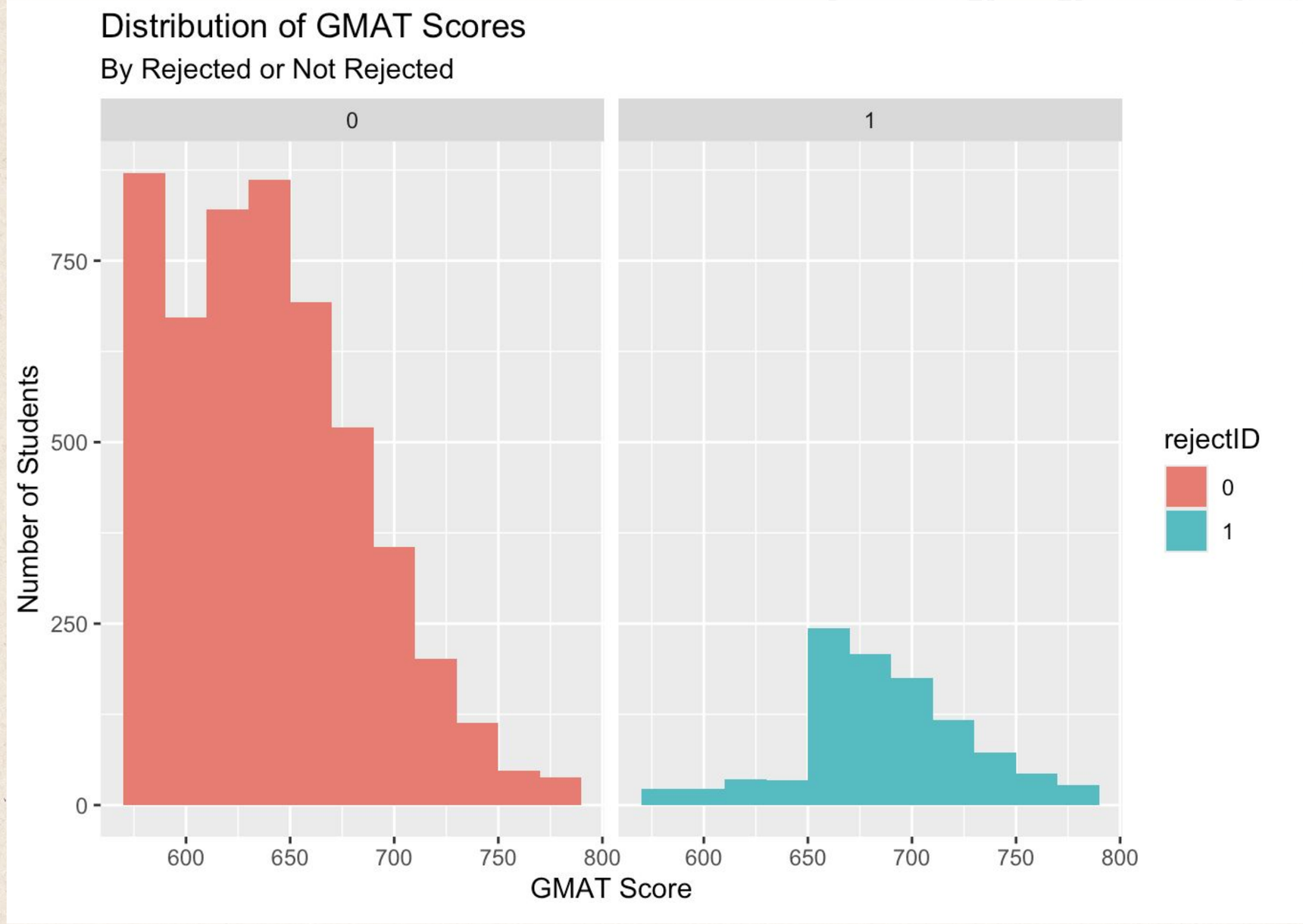
Characters to Factors

- *gender, international, major, work_industry, admission,* and *race*

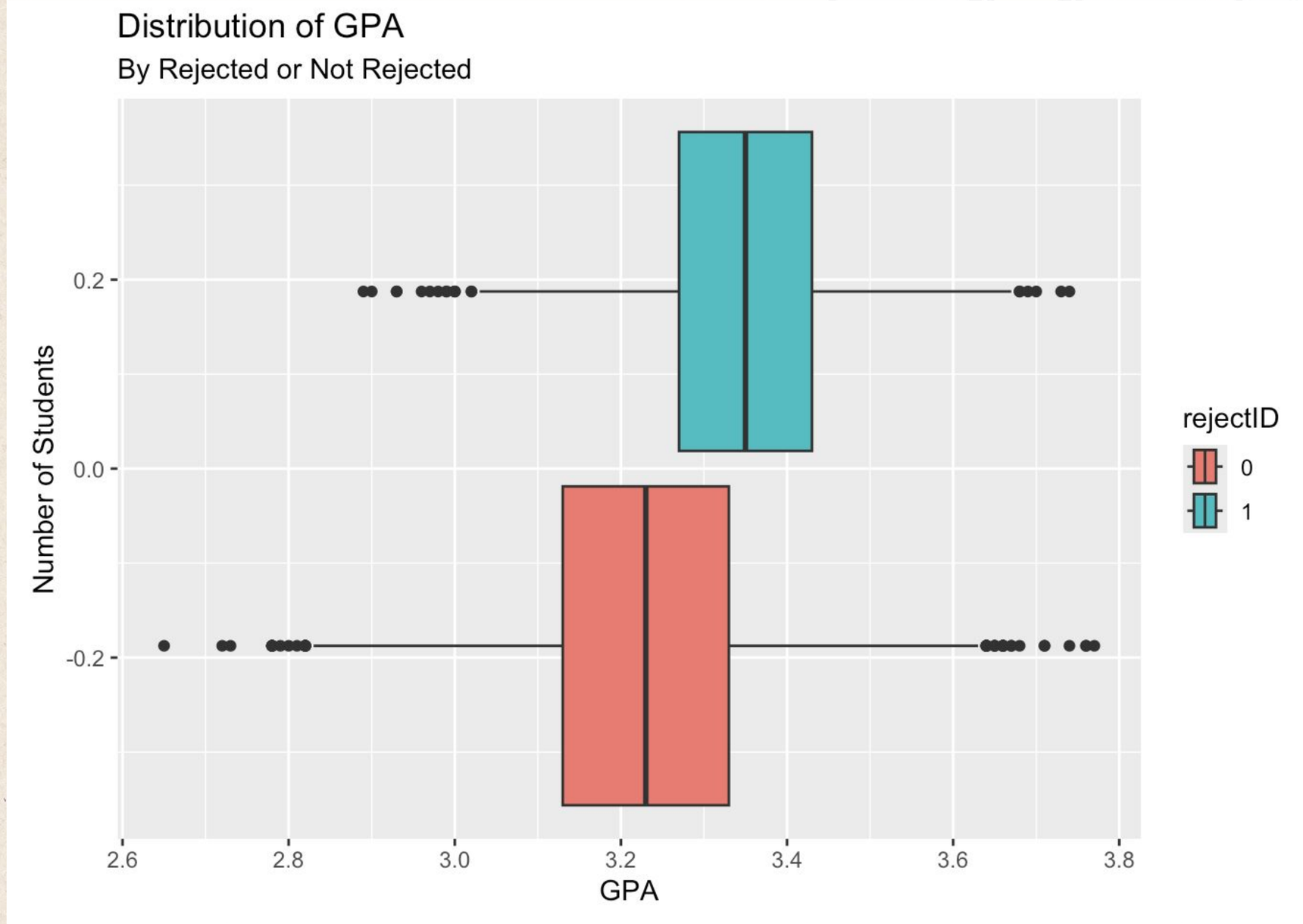
EDA

```
##  application_id    gender     international      gpa            major
##  Min.   :   1   Female:2251   False:4352    Min.   :2.650   Business  :1838
##  1st Qu.:1549   Male  :3943   True :1842    1st Qu.:3.150   Humanities:2481
##  Median :3098                               Median :3.250   STEM      :1875
##  Mean   :3098                               Mean   :3.251
##  3rd Qu.:4646                               3rd Qu.:3.350
##  Max.   :6194                               Max.   :3.770
##
##            race          gmat          work_exp
##  Asian        :1147   Min.   :570.0   Min.   :1.000
##  Black        : 916   1st Qu.:610.0   1st Qu.:4.000
##  Hispanic     : 596   Median :650.0   Median :5.000
##  Not Provided:1842   Mean   :651.1   Mean   :5.017
##  Other        : 237   3rd Qu.:680.0   3rd Qu.:6.000
##  White        :1456   Max.   :780.0   Max.   :9.000
##
##                          work_industry     admission
##  Consulting                     :1619   Admit   : 900
##  Private Equity/Venture Capital: 907   Rejected:5194
##  Technology                     : 716   Waitlist: 100
##  Nonprofit/Gov                  : 651
##  Investment Banking             : 580
##  Financial Services             : 451
##  (Other)                        :1270
```

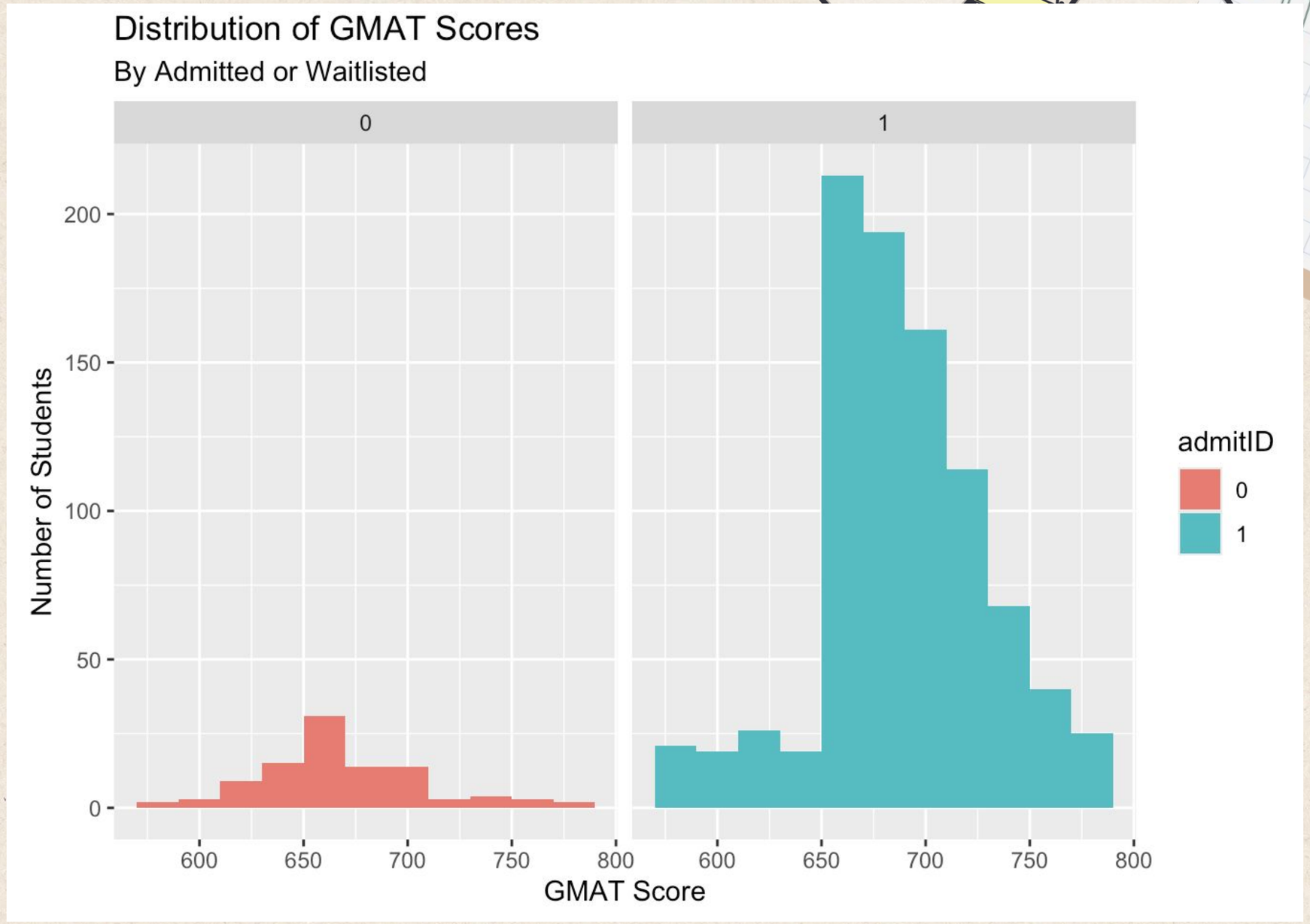
EDA


Distribution of GMAT Scores
By Rejected or Not Rejected
0
1
Number of Students
750
500
250
0
600
650
700
750
800
GMAT Score
rejectID
0
1

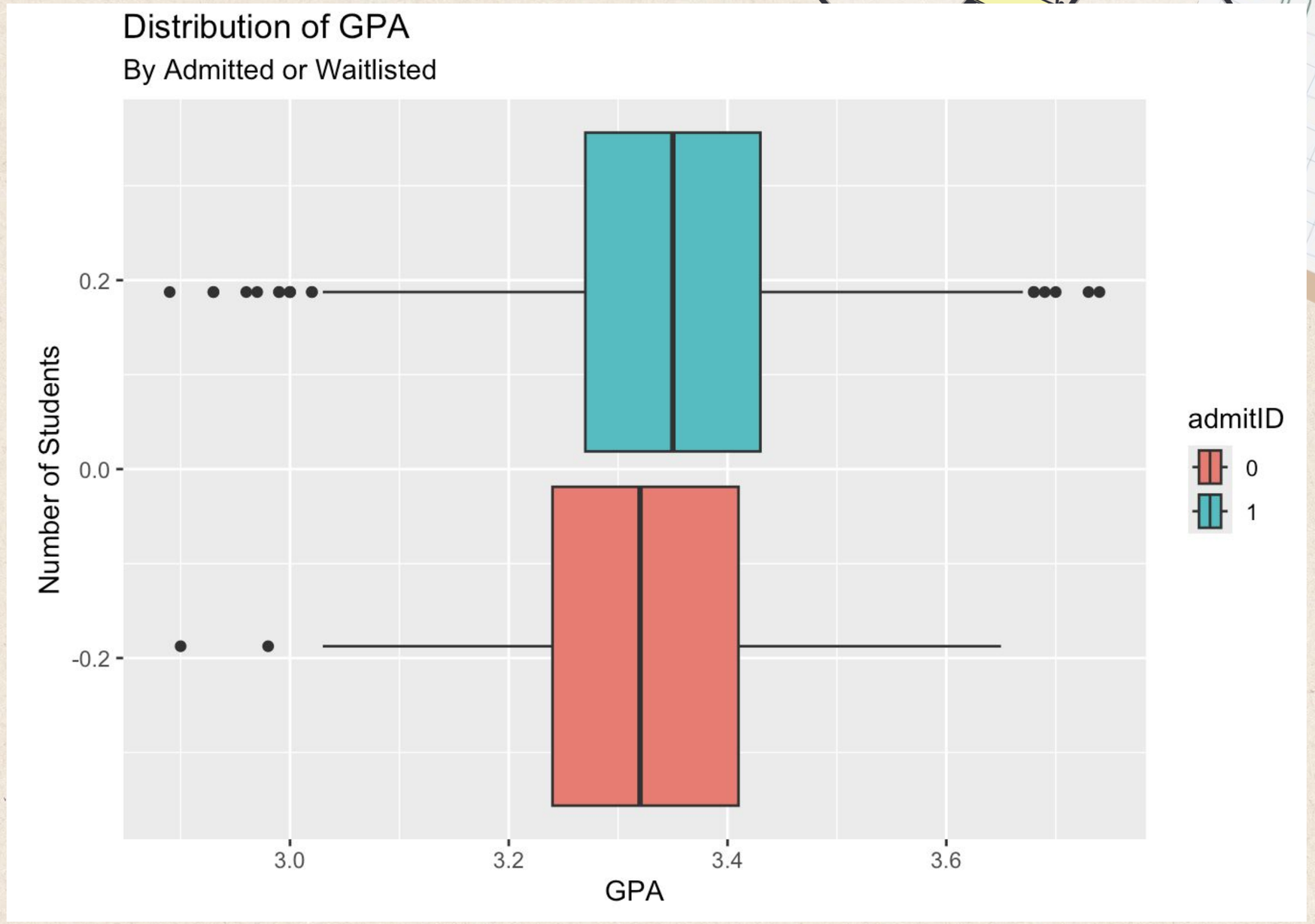
EDA


Distribution of GPA
By Rejected or Not Rejected
Number of Students
0.2
0.0
-0.2
2.6
2.8
3.0
3.2
3.4
3.6
3.8
GPA
rejectID
0
1

EDA


Distribution of GMAT Scores
By Admitted or Waitlisted
0
1
Number of Students
200
150
100
50
0
600
650
700
750
800
GMAT Score
admitID
0
1

EDA


Distribution of GPA
By Admitted or Waitlisted
Number of Students
0.2
0.0
-0.2
3.0
3.2
3.4
3.6
GPA
admitID
0
1

REJECTED

83.85%
Accuracy

Logistic Regression

Training set: 70%

Validation set: 30%

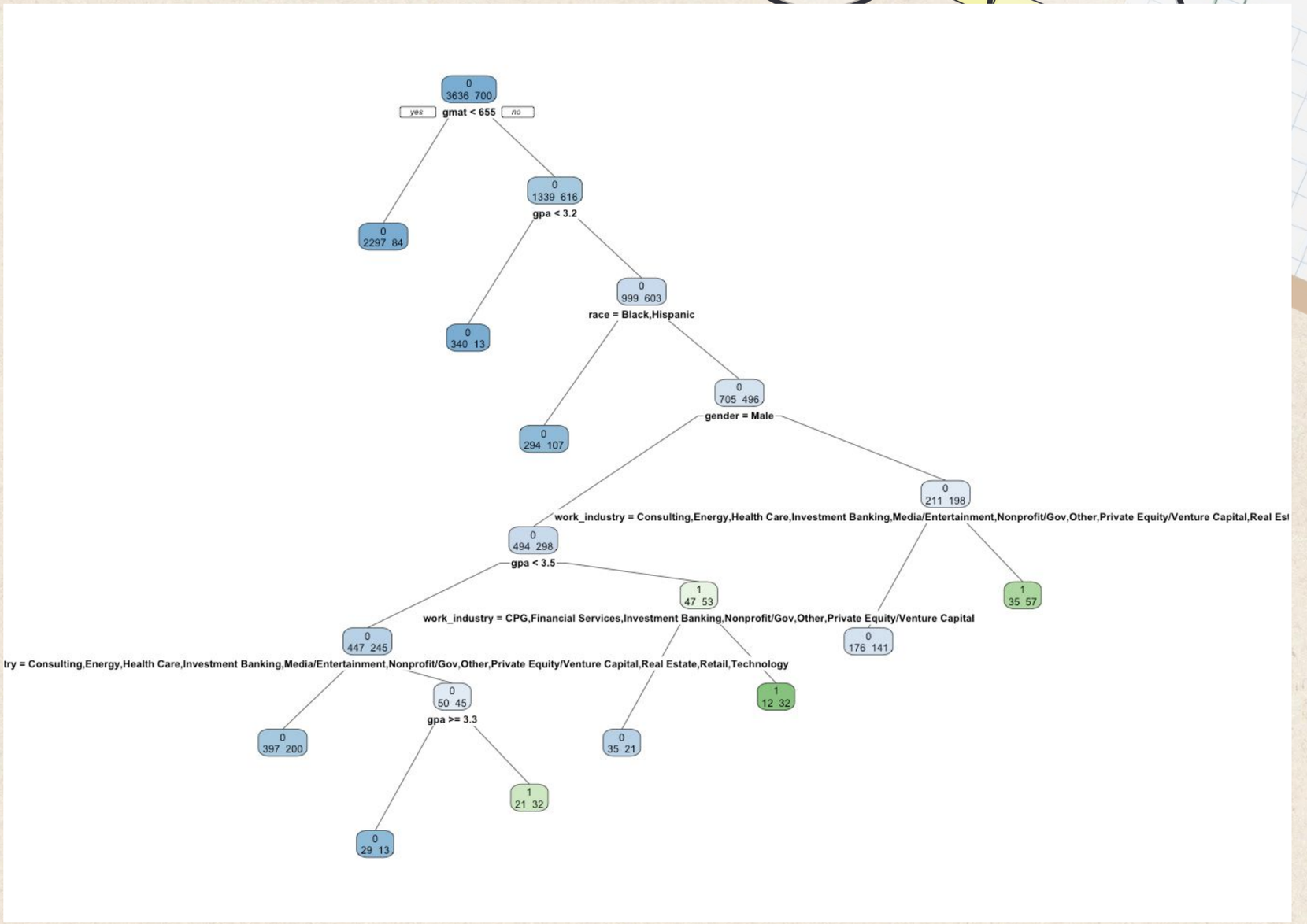
Predictors: *gender, international, gpa, major, race, gmat, work_exp, work_industry, gpa:race*

Significant predictors:

- (***) *genderMale, gmat*
- (*) *raceOther, work_industryInvestment Management, gpa:raceOther*

Decision Tree

84.18%
Accuracy



ADMITTED

88%
Accuracy

Logistic Regression

Training set: 70%

Validation set: 30%

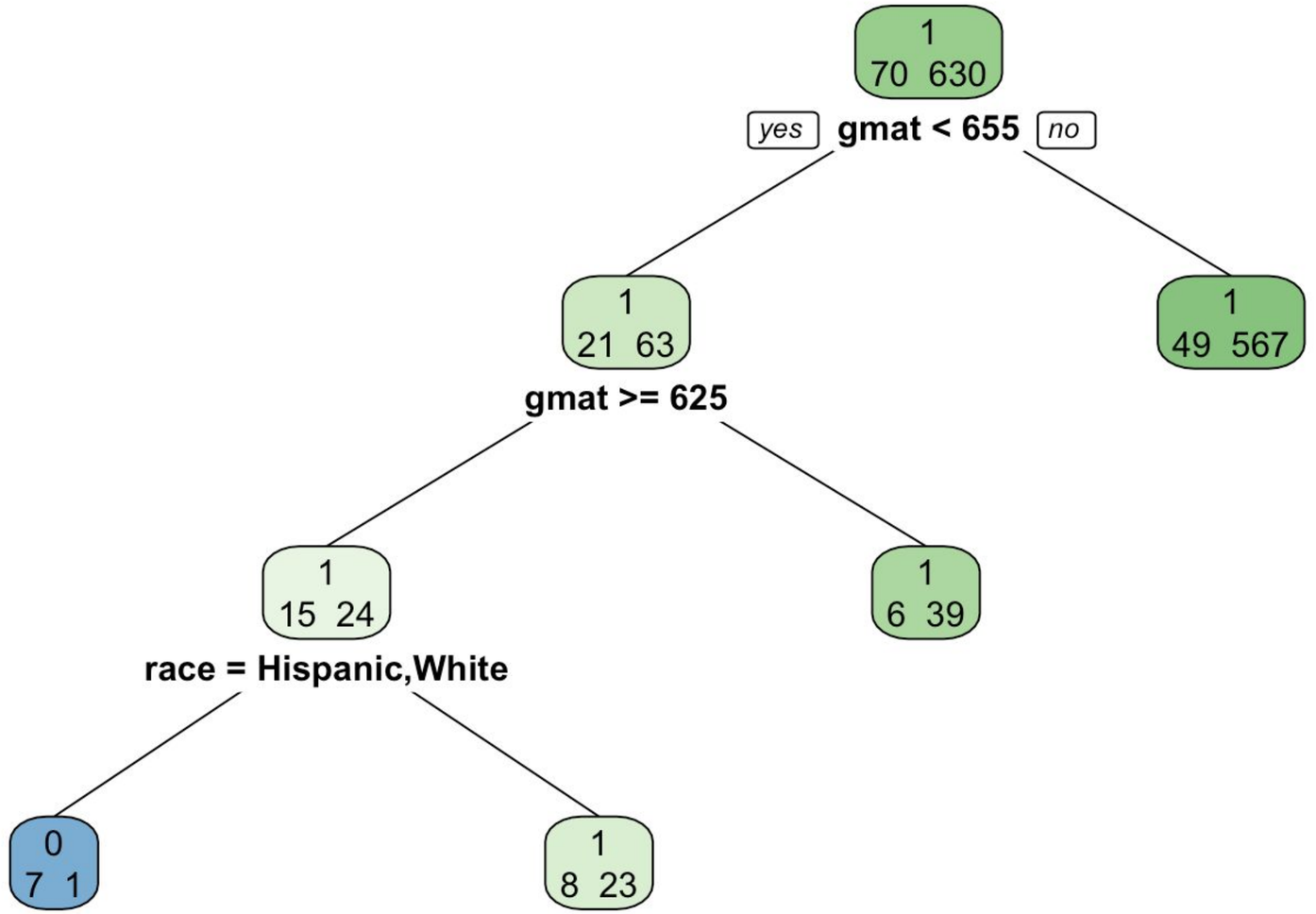
Predictors: *gender, international, gpa, major, race, gmat, work_exp, work_industry, gpa:race*

Significant predictors:

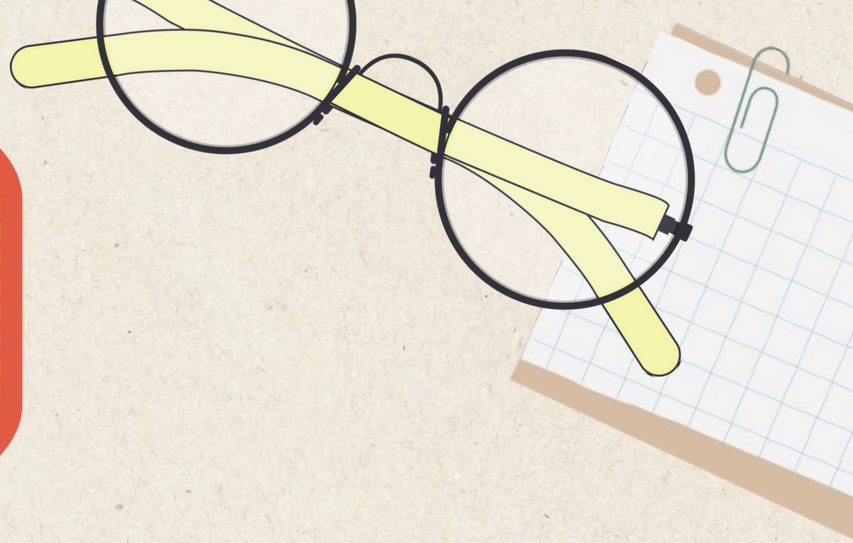
- (**) *gmat*
- (*) *genderMale, work_industryFinancial Services, work_industryTechnology*
- (.) *majorStem, work_industryInvestment Banking, work_industryInvestment Management, work_industryNonprofit/Gov*

Decision Tree

90.33%
Accuracy



UPLIFT MODELS



Gender

Rejected vs Not Rejected

- With all observations set to *male*, there was an average *-10%* in prediction for not rejected

Admitted vs Waitlisted

- With all observations set to *male*, there was an average *-5%* in prediction for admitted

International

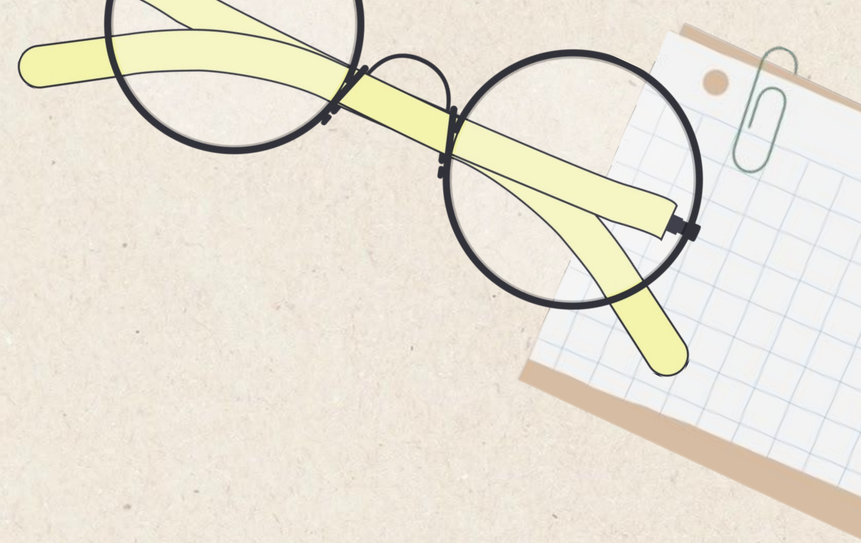
Rejected vs Not Rejected

- With all observations set to *international*, there was an average *-8%* in prediction for not rejected

Admitted vs Waitlisted

- With all observations set to *international*, there was an average *+31%* in prediction for admitted

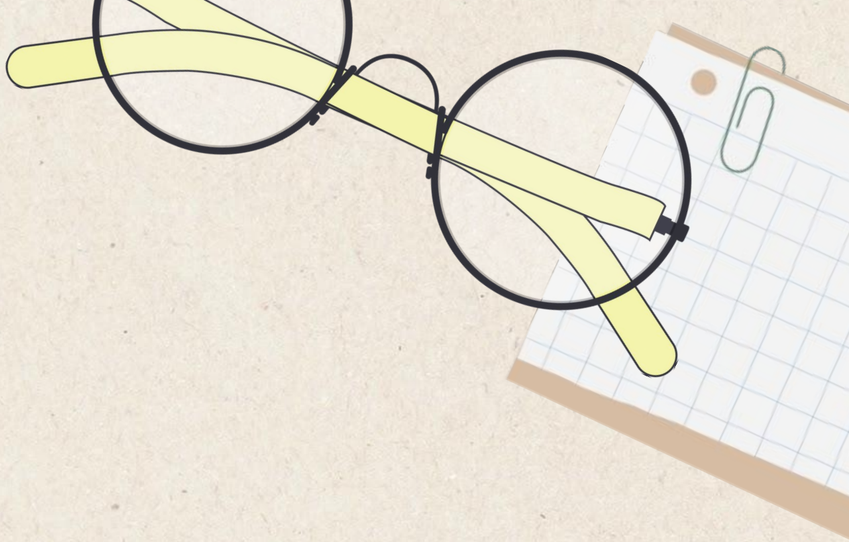
TAKEAWAYS



Relatively accurate models can be created

Possible bias for gender and international status

Could later look more into work industry and race factors



thank you!

Any Questions?