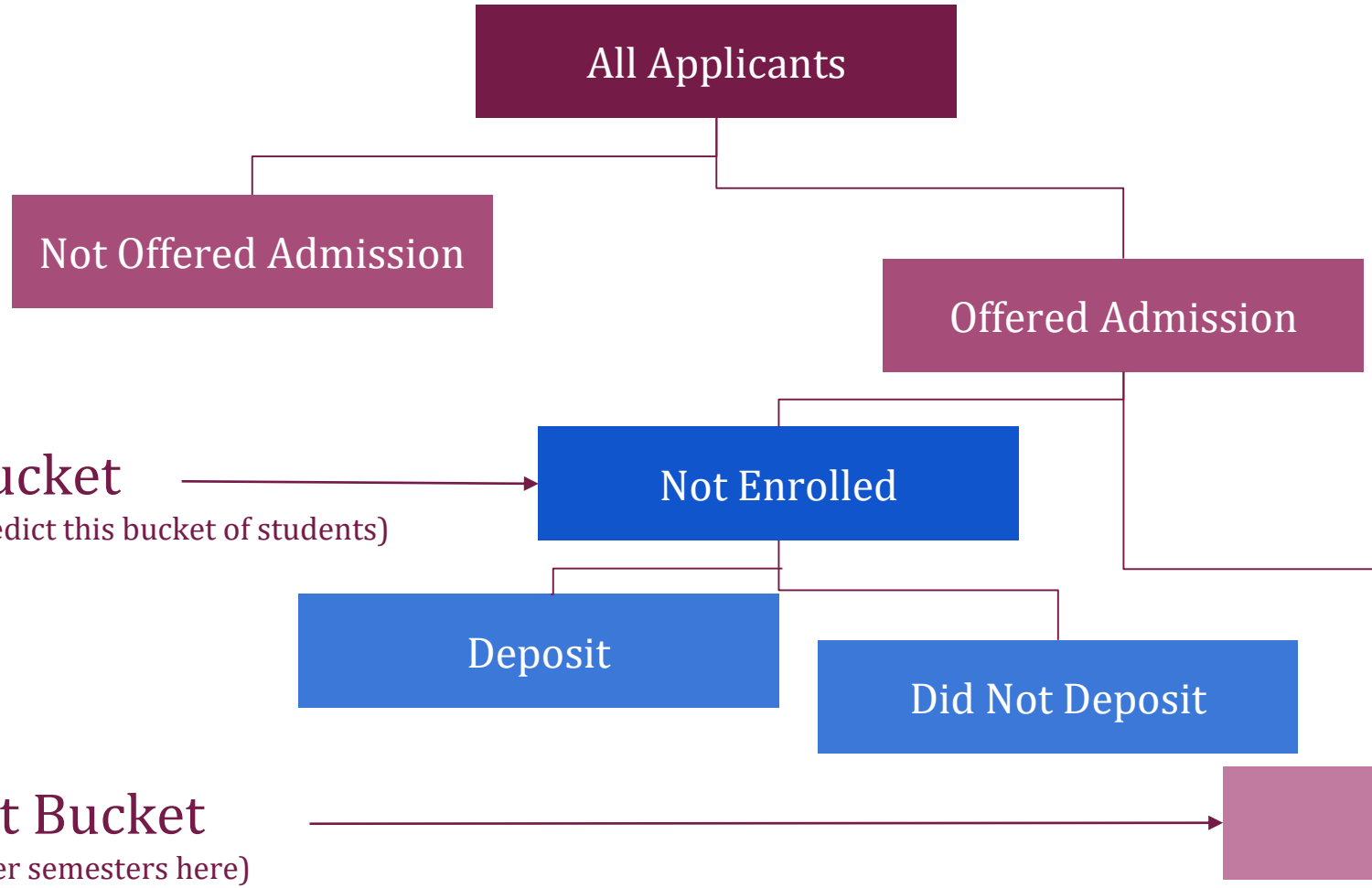
Applicant Data

TEACHERS COLLEGE

7 Relational Databases | Over 45 Features/Variables | Over 200k records

Shreya Goel

Applicant Data Buckets



School Loses this Bucket
(To research on factors that can predict this bucket of students)

School Final Student Bucket
(Not considering retention in further semesters here)

Current Research Focus

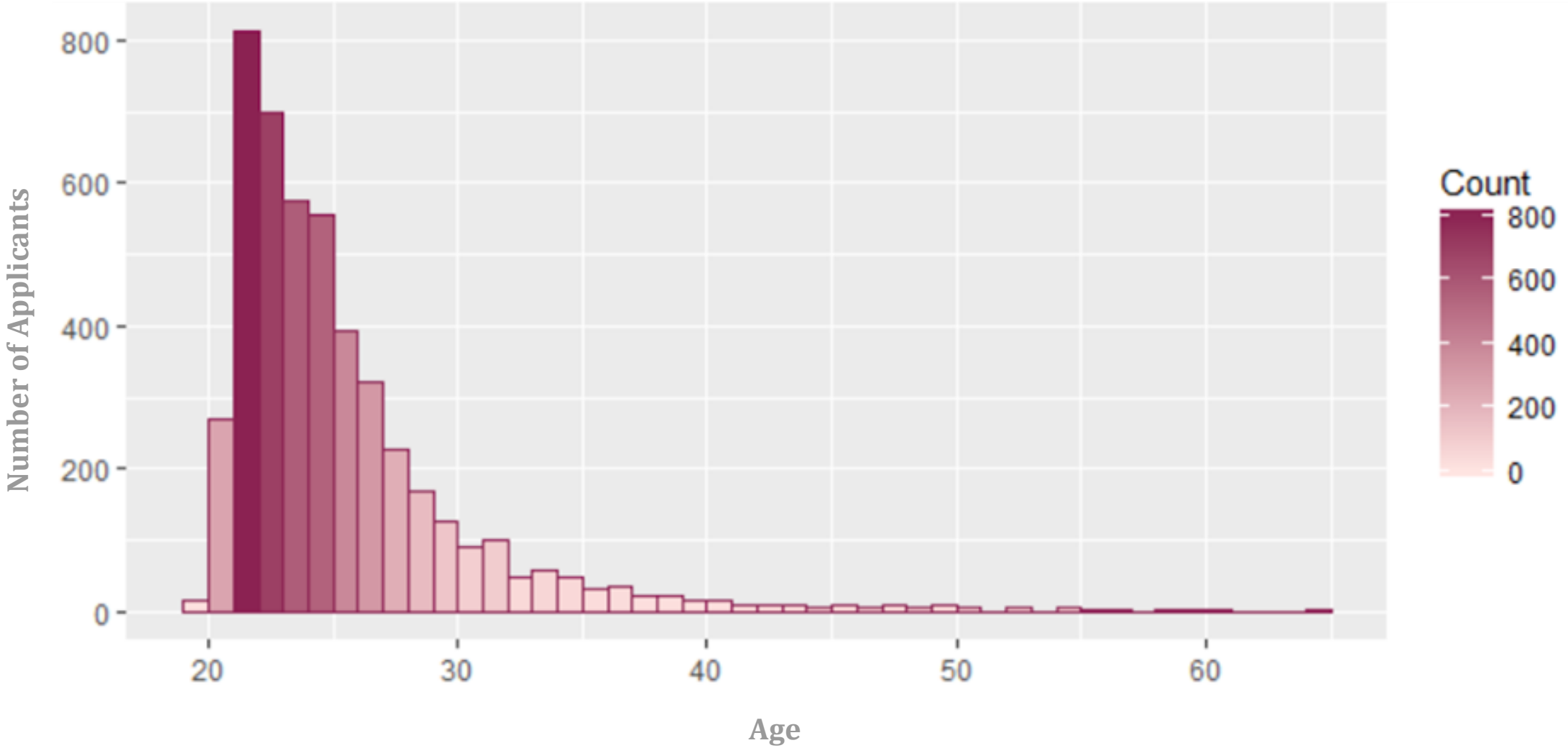
Build a classification scheme to identify the attributes that distinguish between students who deposit but ultimately don't enroll and those who never deposited in first place.

Initial Analysis

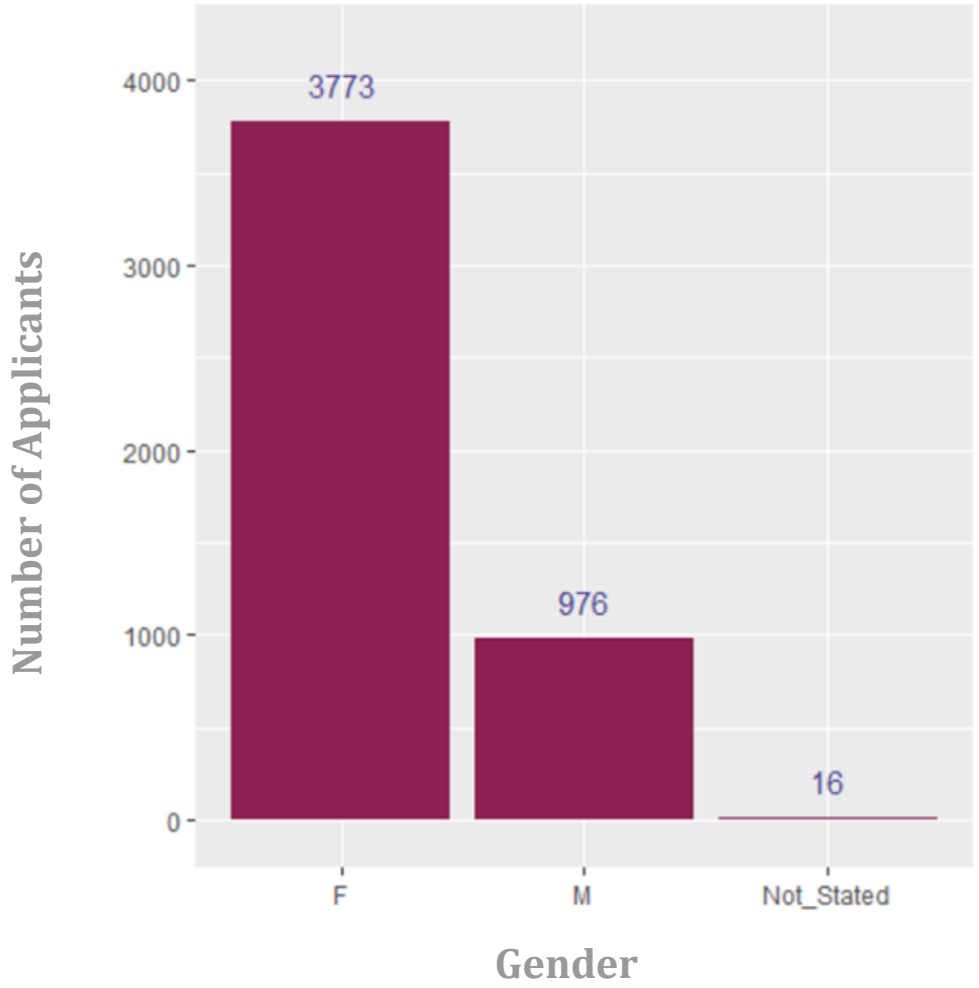
(Facts)

4765 Applications | 4470 Unique Applicants

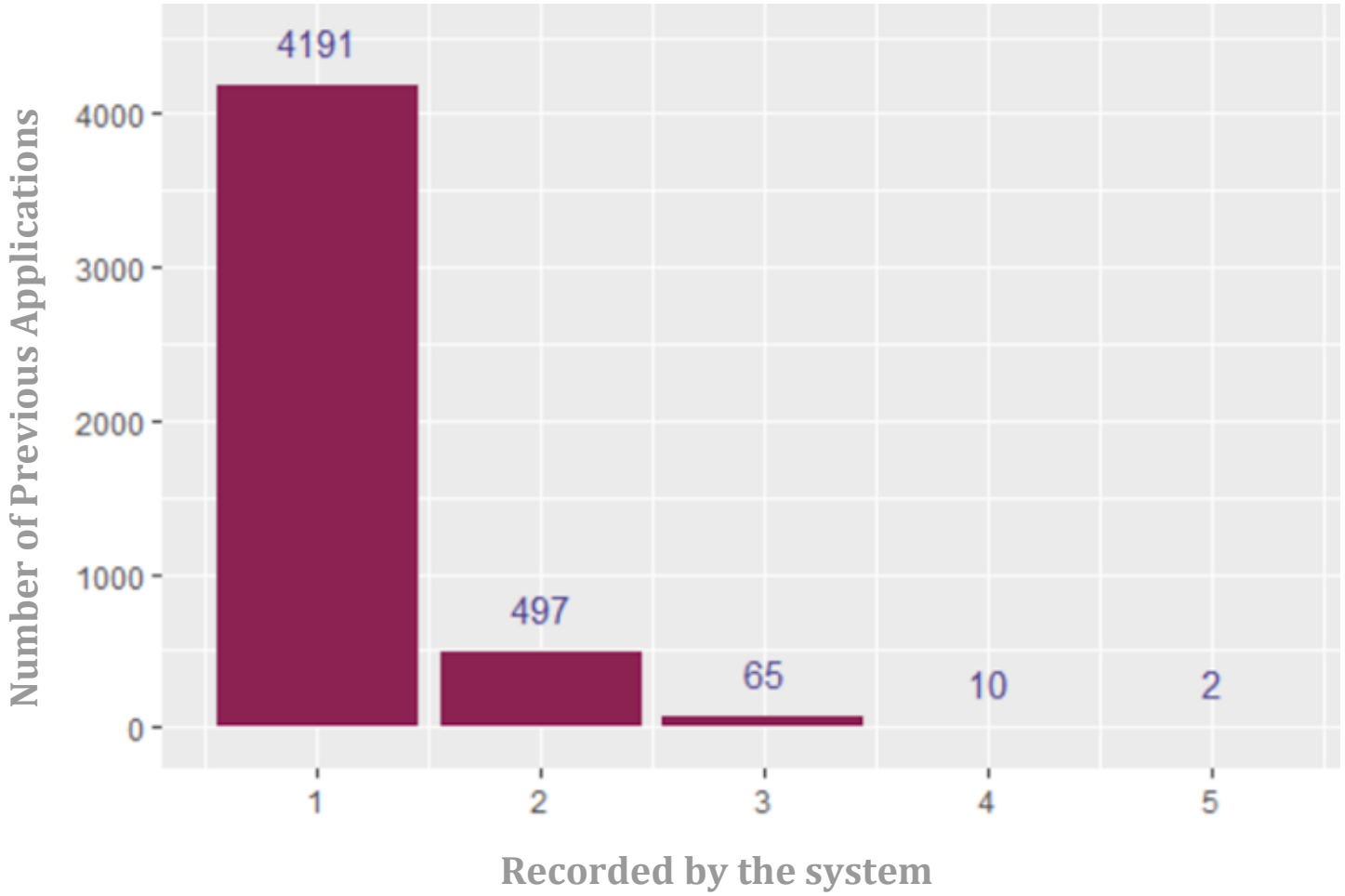
Age Group of Applicants

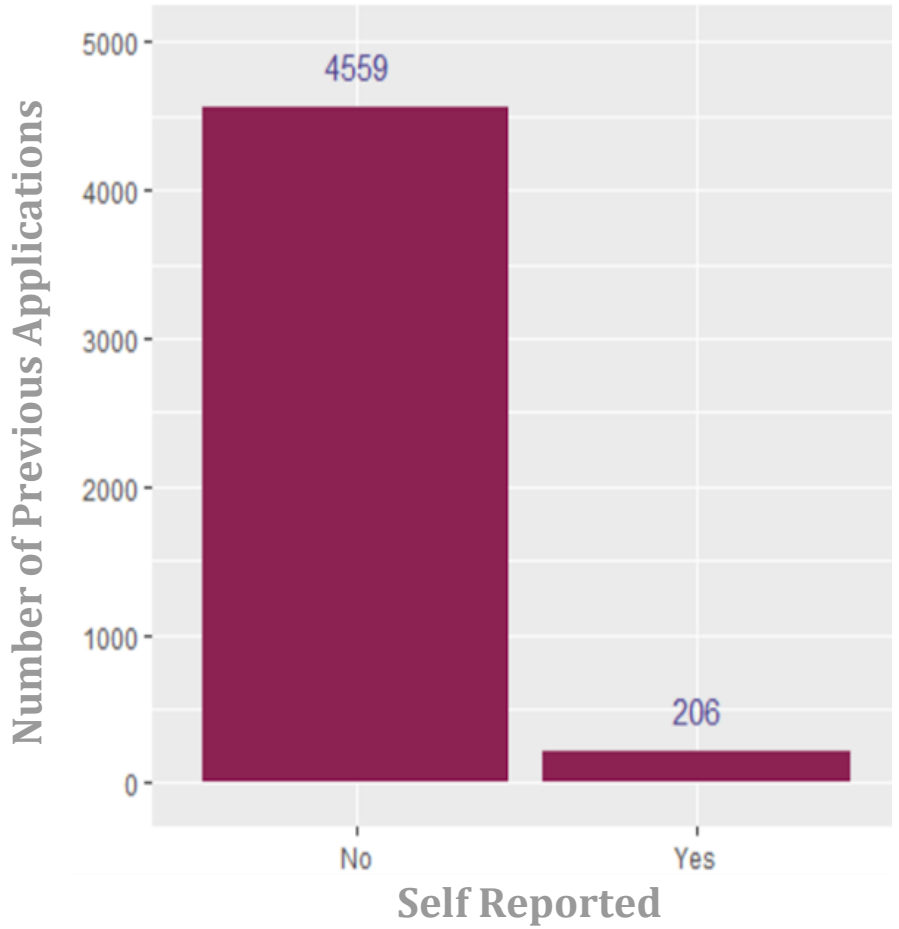


Gender Dominance Amongst Applicants

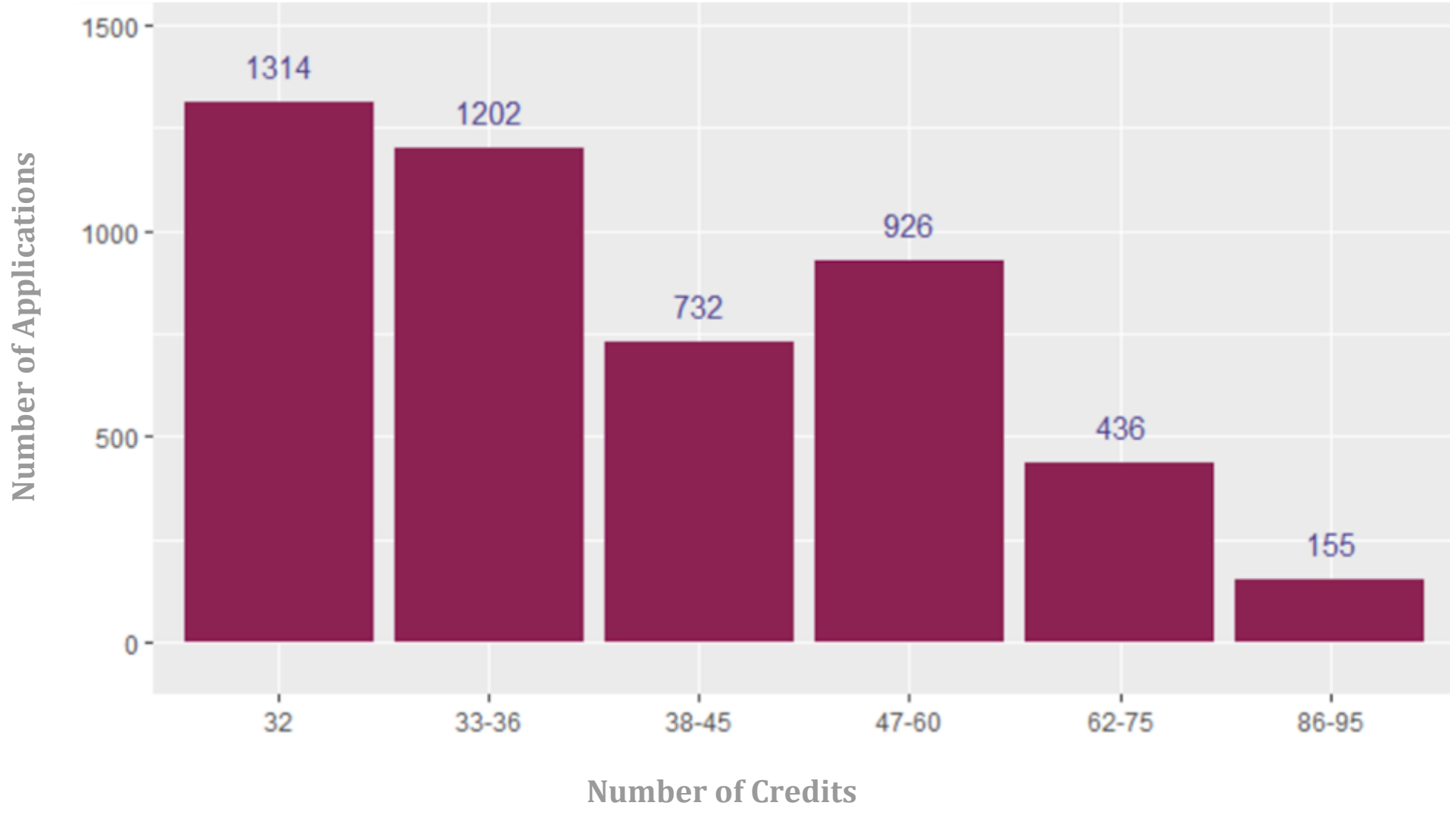


Number of Previous Application

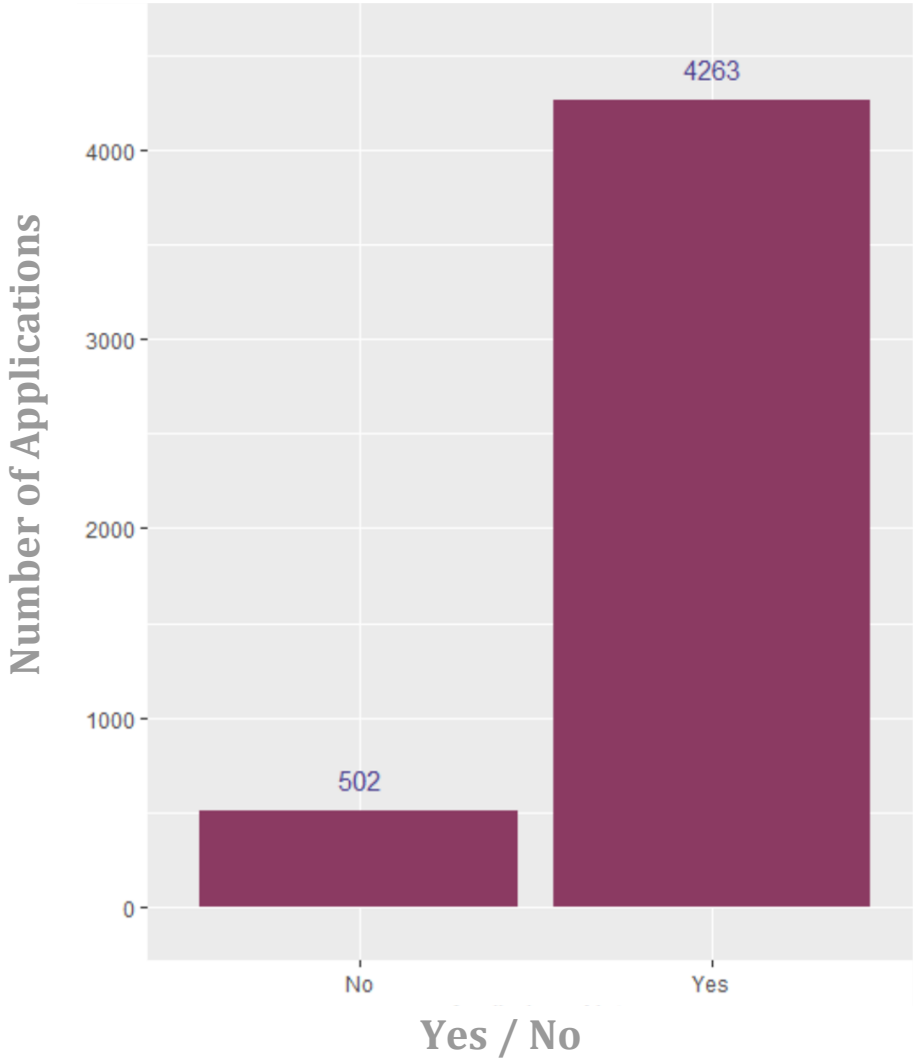




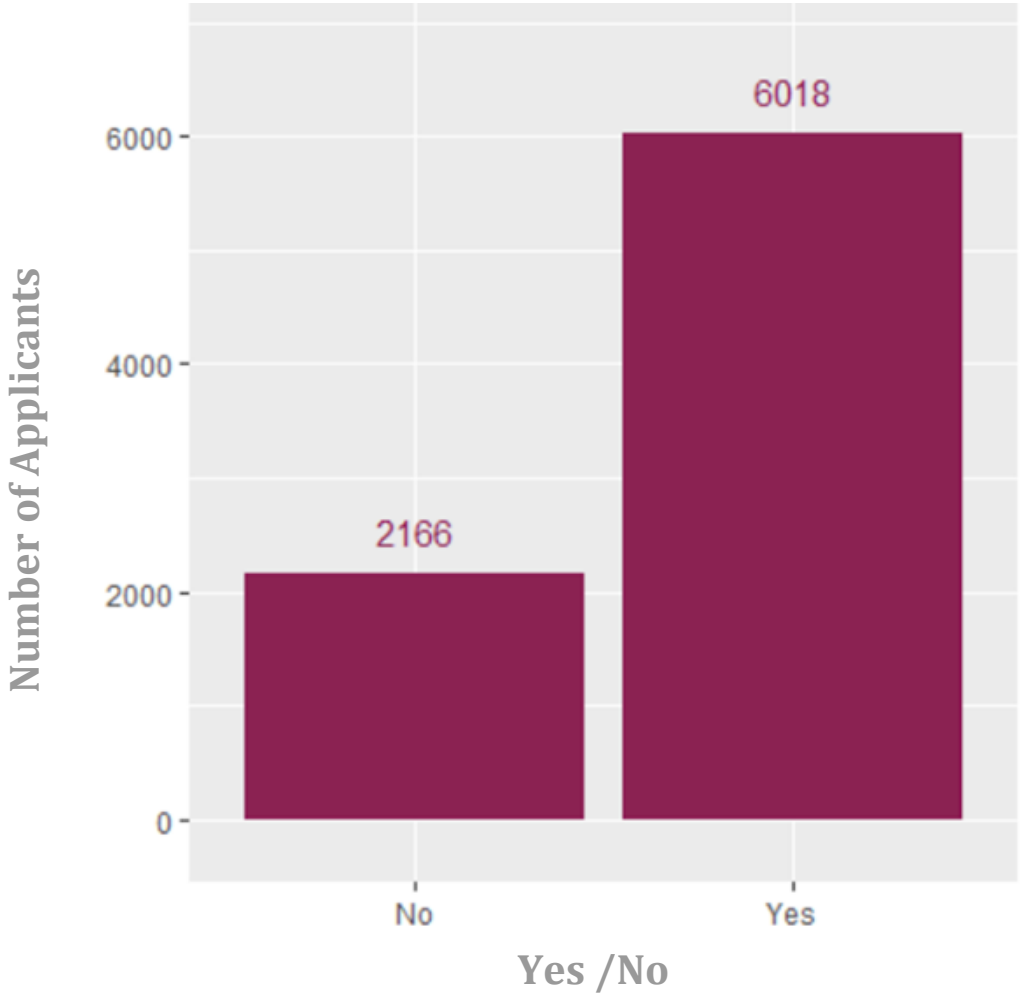
Degree Credits Applied for, by Applicants



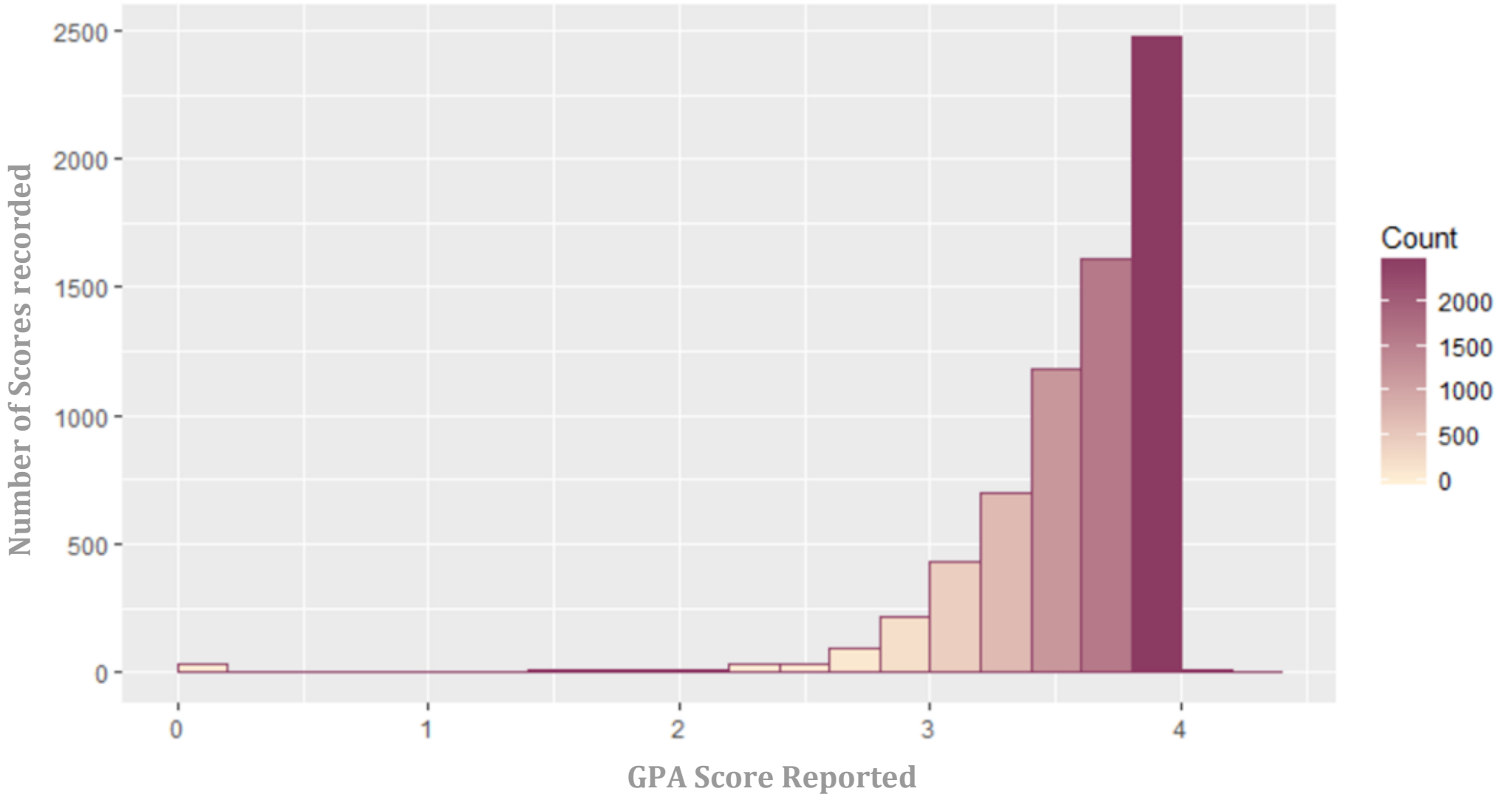
Number of Scholarship Applications



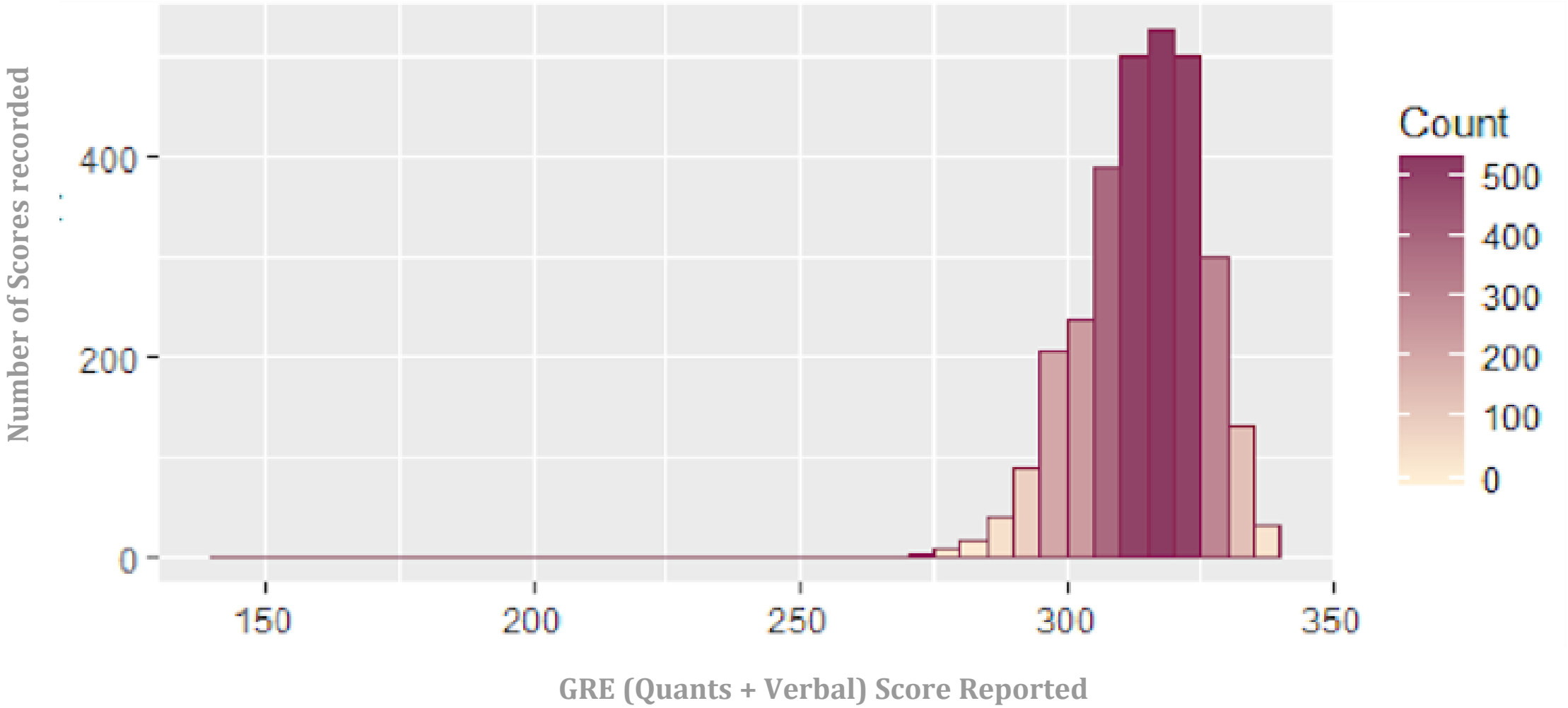
Previous Degree or Not (Dropped) as reported by the Applicants



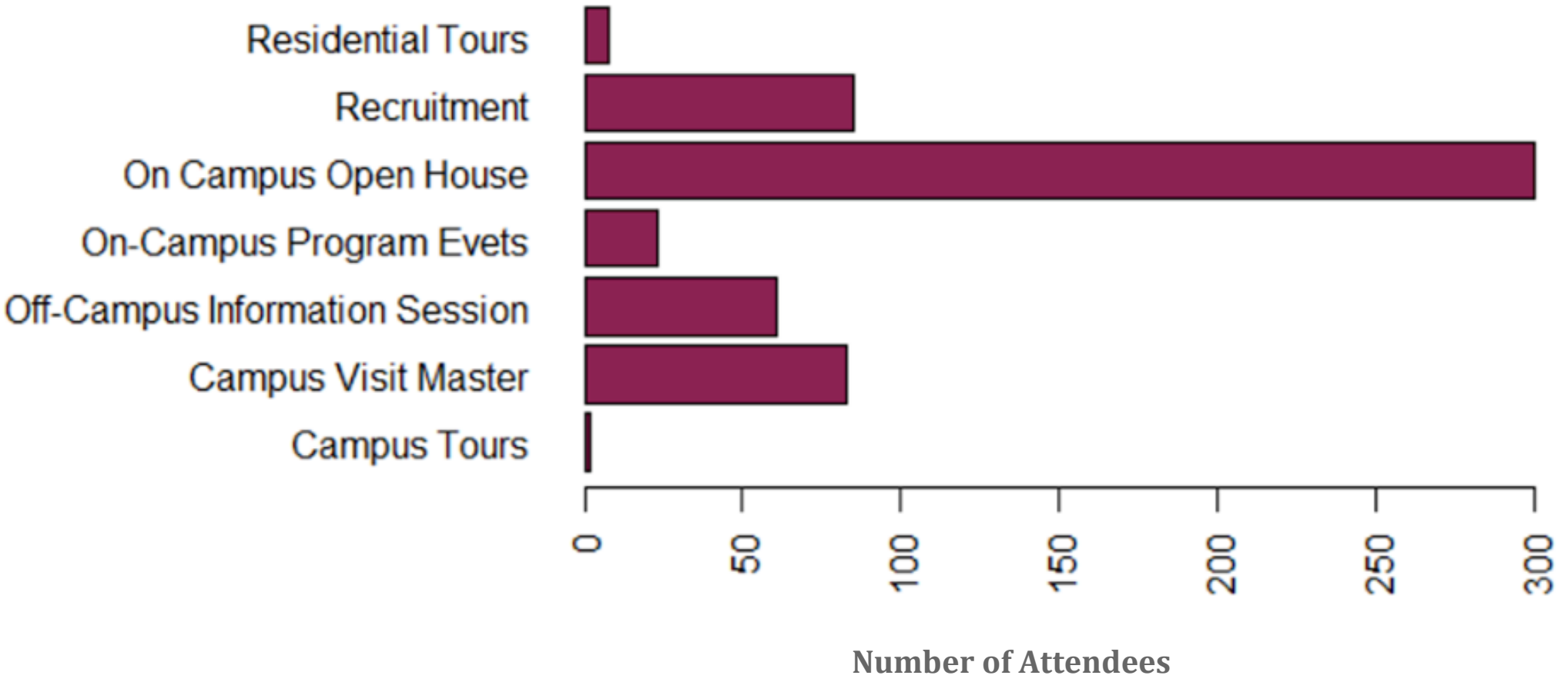
Previous School GPA Scores as reported by the Applicants



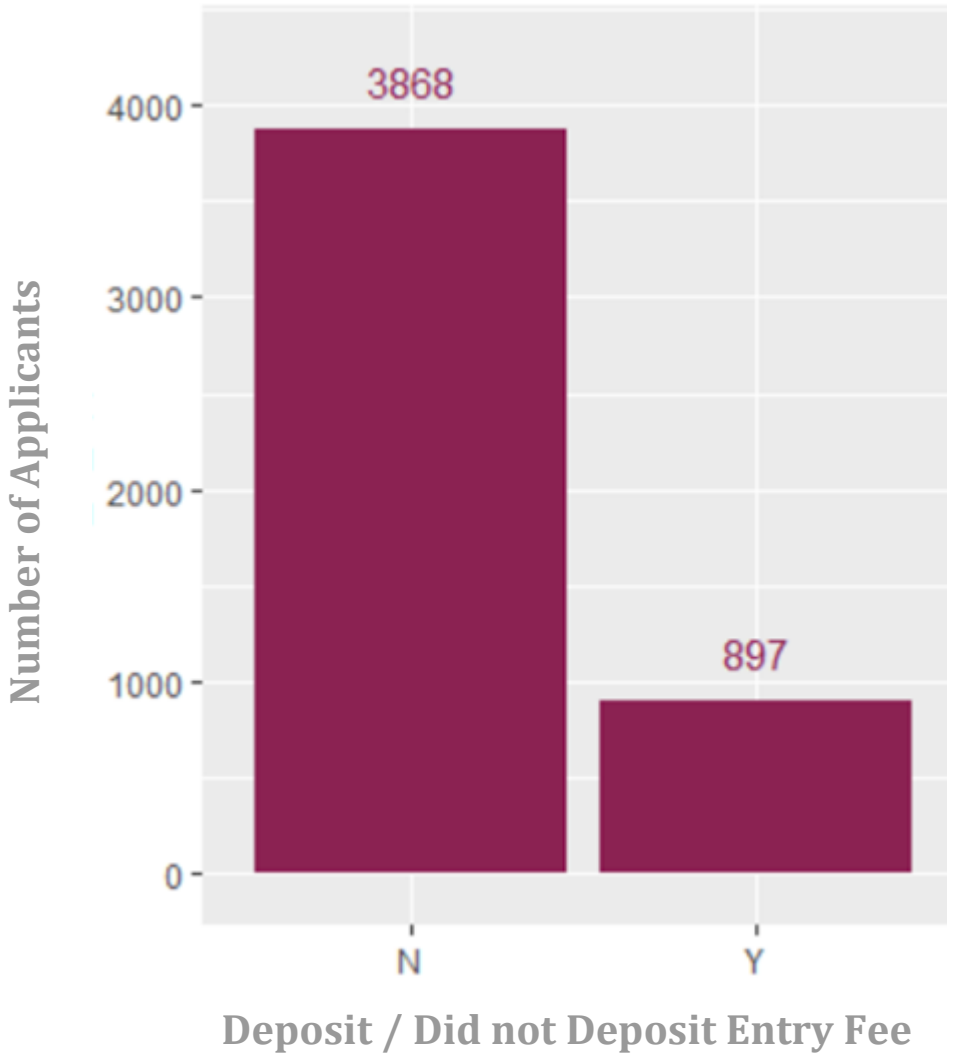
GRE Total Scores as reported by the Applicants



Engagement with Events



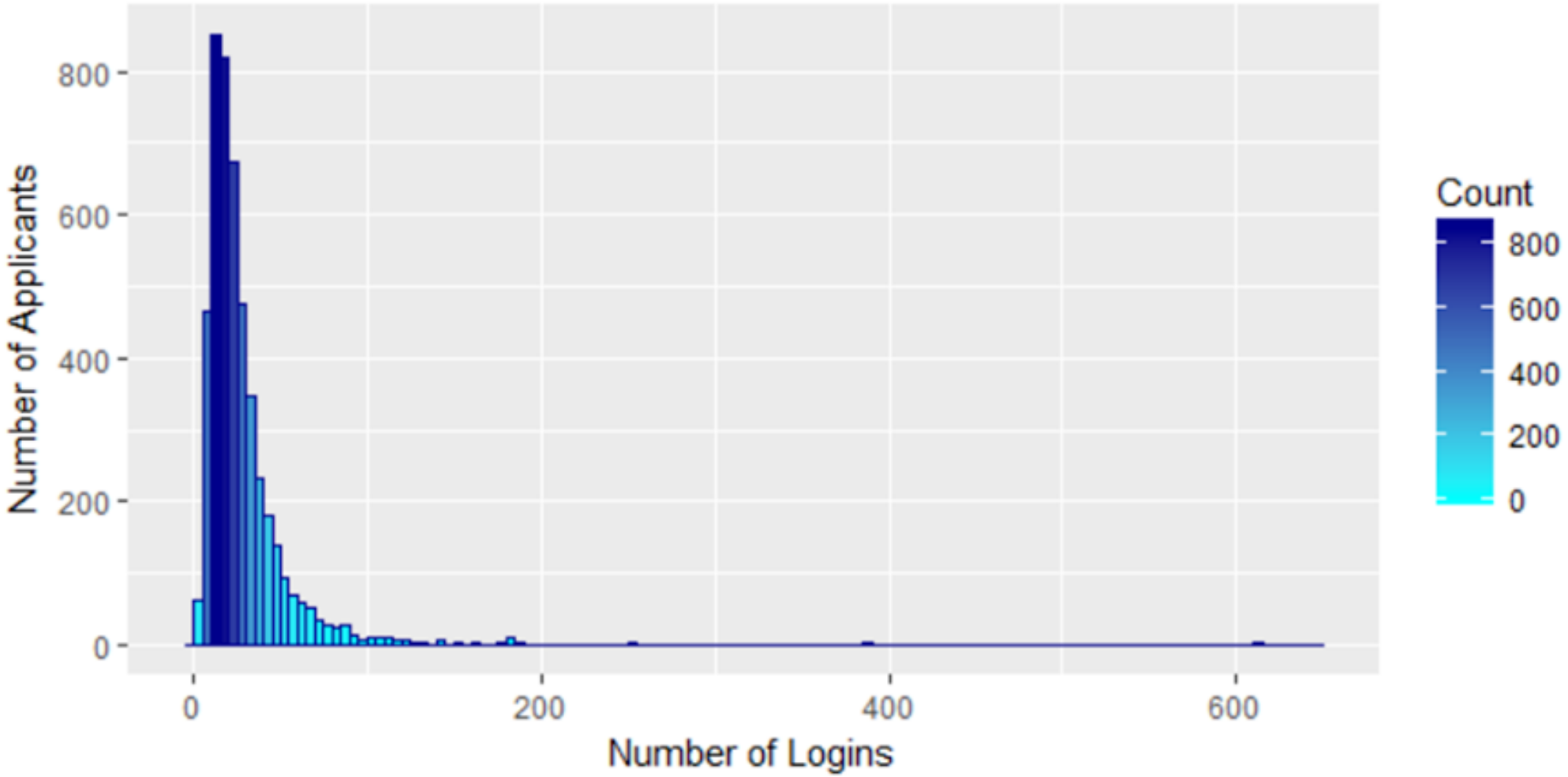
Deposit/Did not Deposit the Entry Fee



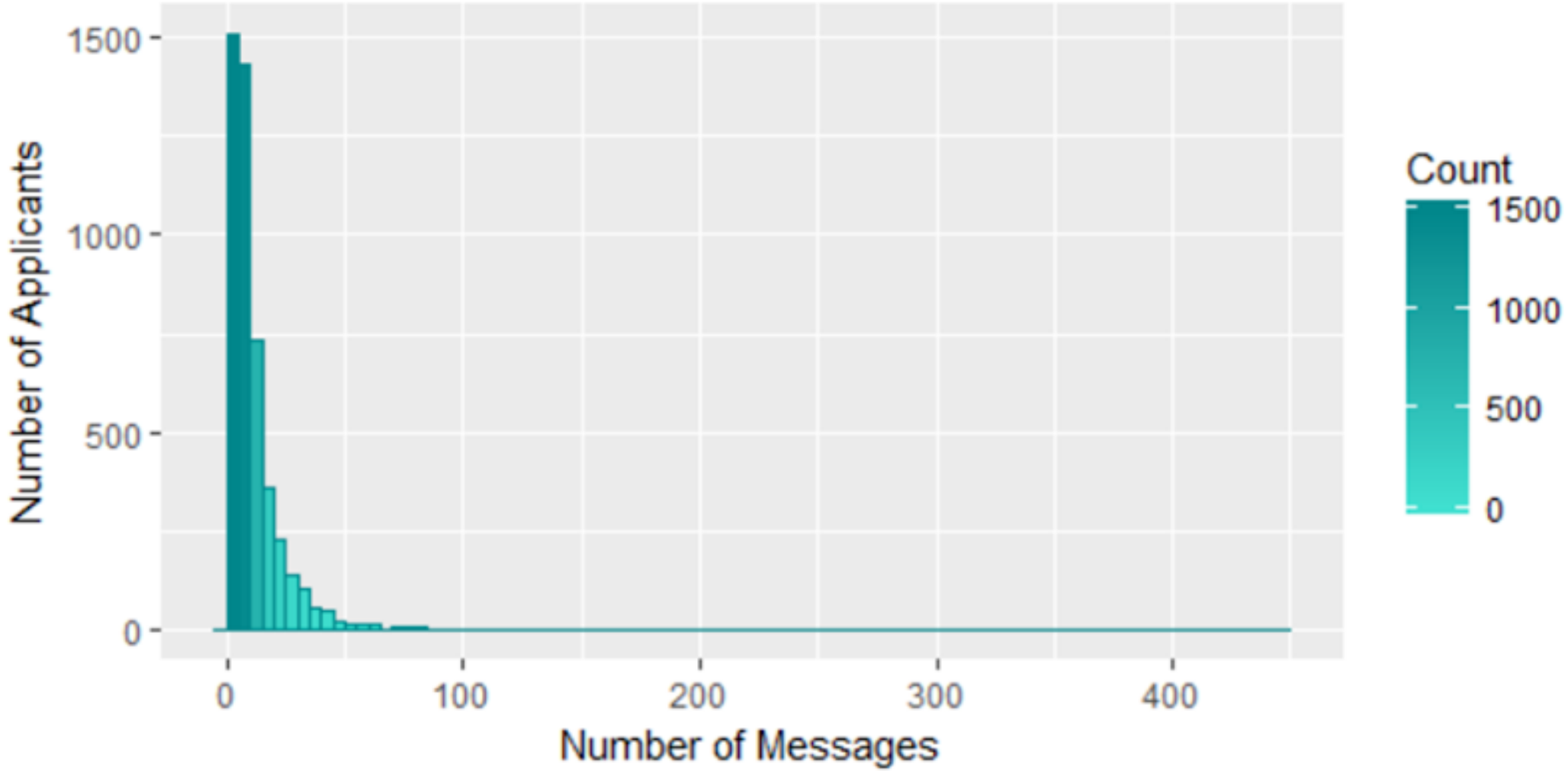
Login and Messages

Deposited & Not Deposited

Number of Logins per Applicant



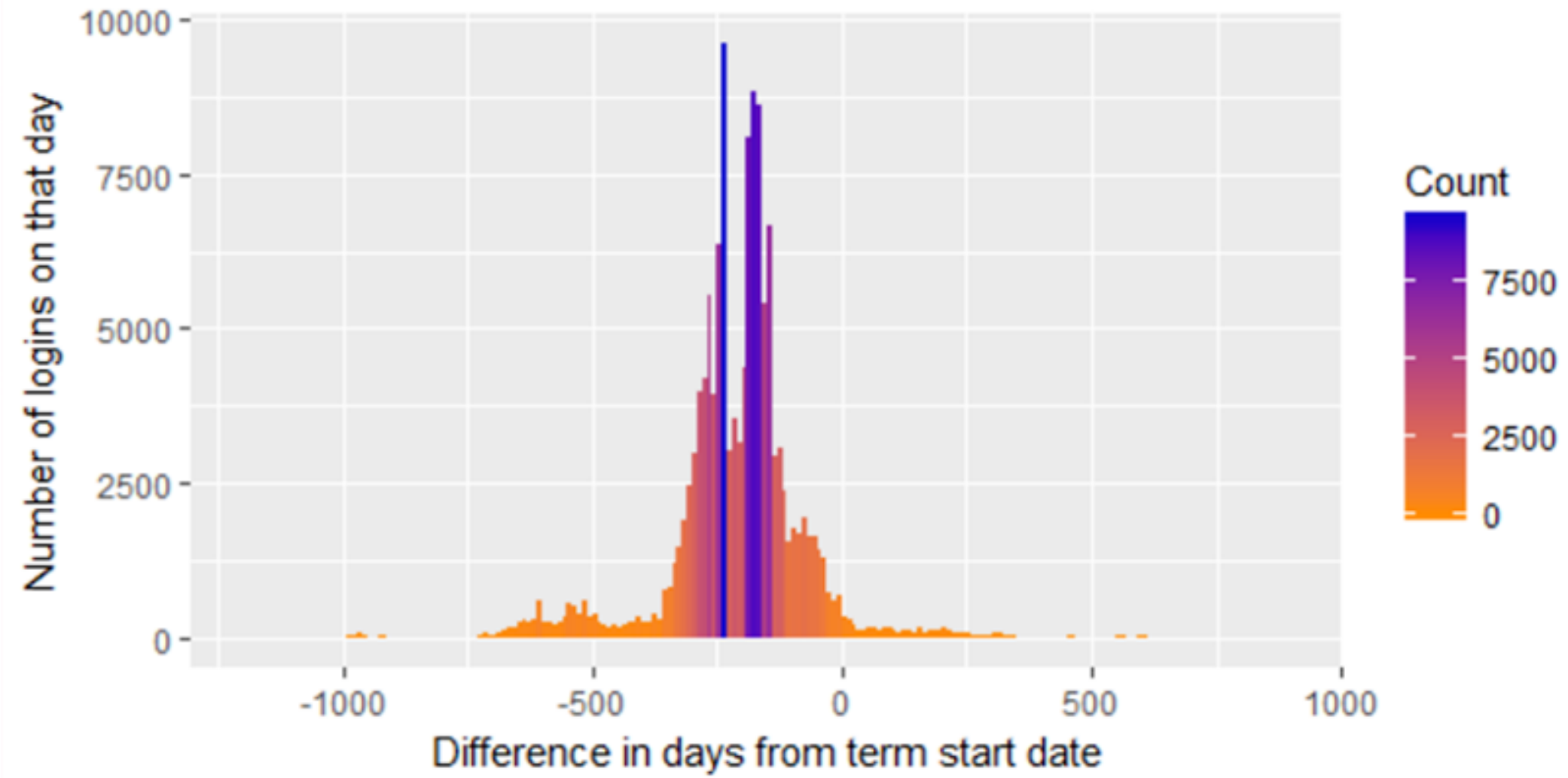
Number of Messages Sent



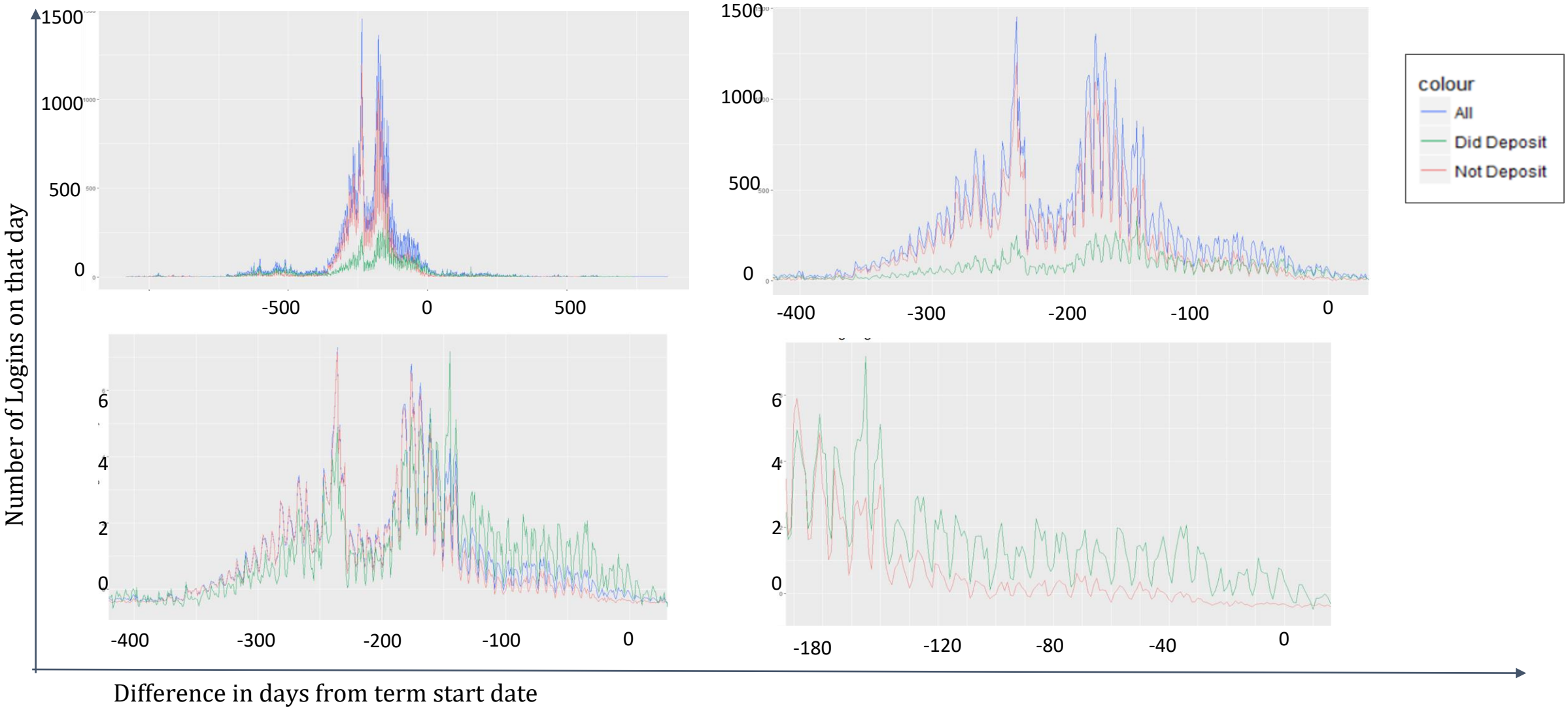
Logins and Messages Data (per Applicant)

Term start date was chosen to be the date from which other dates were subtracted to know about the applicant behavior, certain number of days away from that date.
Time was currently separated and ignored from the analysis.

Difference in days from Term Start Date



Clustering of Logins over Time



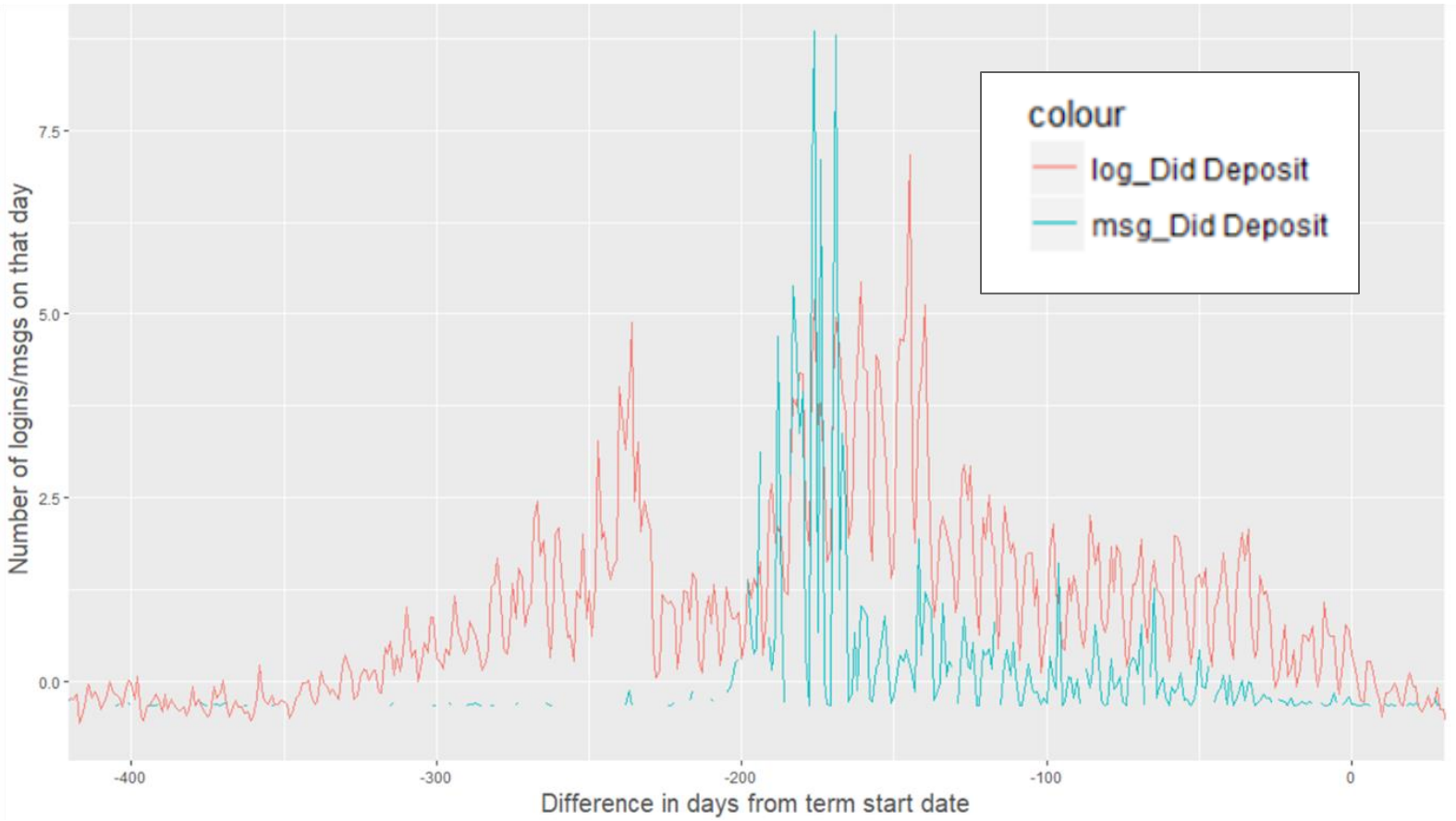
Clustering Logins & Messages over Time (Did Deposit)

Number of messages and logins on a particular day from term start date.

Two major bumps in login activity (red).

- Application submission/result
- Deposit fee to confirm they are still interested

Bump in message sent by TC-Reminders to deposit fee (blue).



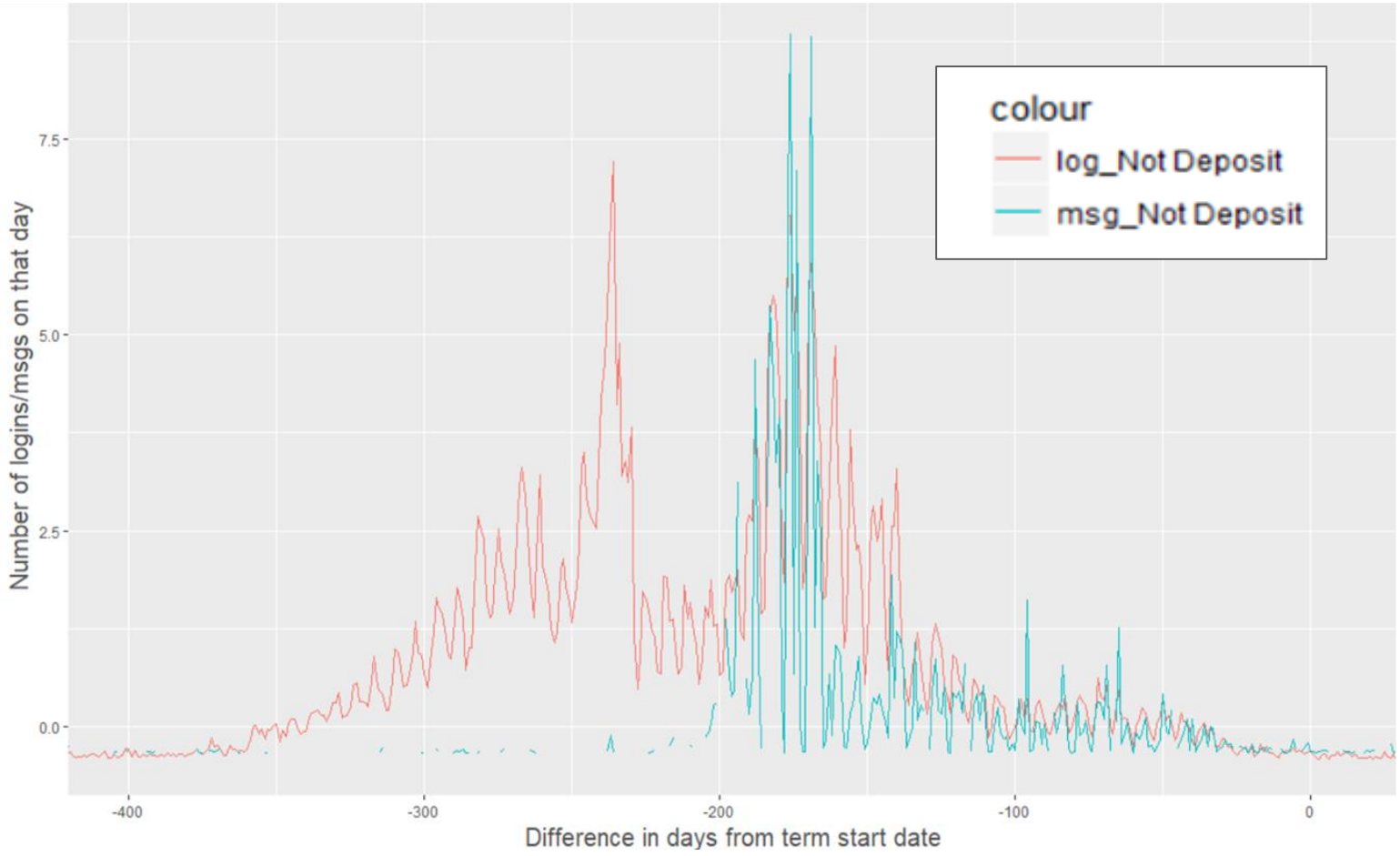
Clustering Logins & Messages over Time (NO Deposit)

Number of messages and logins on a particular day from term start date.

Two major bumps in login activity (red).

- Application submission/result
- Deposit fee to confirm they are still interested (very less activity as compared to previous one)

Bump in message sent by TC-Reminders to deposit fee (blue).



Current Research Focus

Build a classification scheme identifying the attributes that distinguish between students who deposit but ultimately don't enroll and those who never deposited in first place.

Recommendations for Future Research

If given access to the complete data pool of applicants, these are some of the ideas I would like to explore:

- Were there any particular semesters or years where applicants chose not show up increasingly than the other years, considering the political or financial situation worldwide.
- Hierarchical Cluster Analysis
- Differentiate behaviour of applicants in different applicant buckets
- Predict the applicants who will not enroll if offered admission
- Predict the applicants who will not show up after they deposit fee

Thank You

Shreya Goel

M.S. Learning Analytics | *Columbia University*
Marketing Analytics Intern | *Columbia University*
Mob.: (929) 248-2764
email: shreya.goel@columbia.edu