BU Athletics - Deliverable 3

TEAM B

Mounika Chowdary Akuraju, Robert Hannon, Bargav Jagatha,

Jeff Tedi, Shangzhou Yin

Background and Motivation

- Core Project

  - Analyze the correlation between admissions data and academic performance of BU Student Athletes

  - Analyze how BU Student Athletes Perform throughout their time at BU

- Extension Project

  - Make Correlations between trends in Student Athlete Performance and larger discussions about society and social dynamics as a whole

Extension Goals

- Further Analysis of GPA (the only definitive metric we have of academic performance)

- Using "Athletics Annual Surveys Data" to find deeper reasoning for GPA trends

  - Deeper dive into certain survey data questions and finding potential sociological impacts on student academic performance

- Classification of how student athletes will perform - try to classify if a student will be high performing (latin honors threshold) or will potentially be underperforming (GPA <2.5)

New Dataset for Extension

BU Athletics Data 22-23

	G	H	I	J
1	IPEDS Race Description	Athletic Code 1	Team Gender	Athletic Code Description 1
2	Non-Resident Alien	HC	M	MEN'S ICE HOCKEY
3	African American or Black	SC	M	MEN'S SOCCER
4	White	MC	M	MEN'S ROWING
5	Non-Resident Alien	SC	M	MEN'S SOCCER
6	Non-Resident Alien	MC	M	MEN'S ROWING
7	White	HC	M	MEN'S ICE HOCKEY
8	White	HC	M	MEN'S ICE HOCKEY
9	Non-Resident Alien	MC	M	MEN'S ROWING

Athletics Annual Surveys Data

	A	B	C	D	E	F	G
1		Department Average	MB	WB	CC	CW	FH
2	Effectiveness of head coach's leadership and direction for the team	3.10	2.88	1.77	4.00	3.63	3.00
3	Effectiveness of head coach's communication skills	2.90	3.18	2.15	3.71	3.11	3.24
4	Effectiveness of assistant coach's communication skills	3.37	3.59	2.69	3.63	3.53	2.90
5	Effectiveness of student-athlete leadership within your team	3.26	2.29	2.85	3.75	3.68	3.38

Merged on team name

- Categorical data also turned numerical
- Filled in Nan values for SAT and HSGPA by the average of the entrants team

Issues

- Lack of data and variables
  - Dataset given only has 993 entries
  - There is no concrete way to define student athlete academic success besides GPA
  - Issues with trying to collect/retrieve data
    - Had to go to BU Athletics Office to receive information
    - Requested more data on average GPA of BU Students, average GPA per major, and latin honors thresholds by college but were not able to receive these values
  - The "Athletics Annual Surveys Data" only provides averages per team and not individuals scores
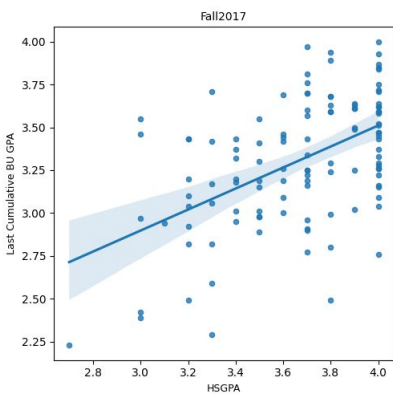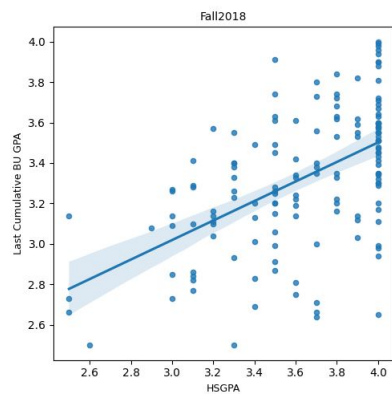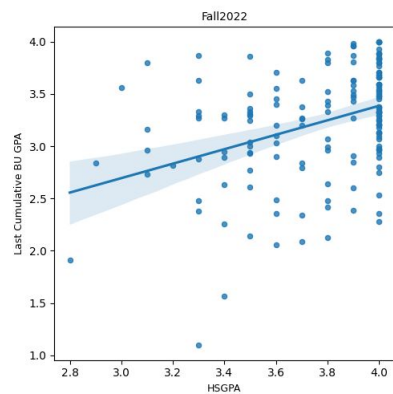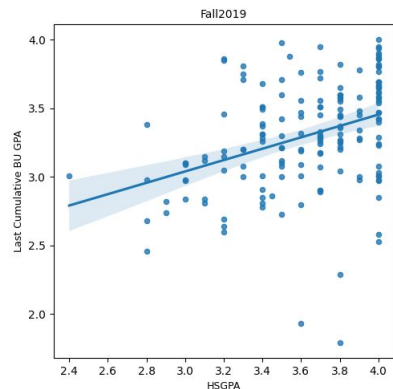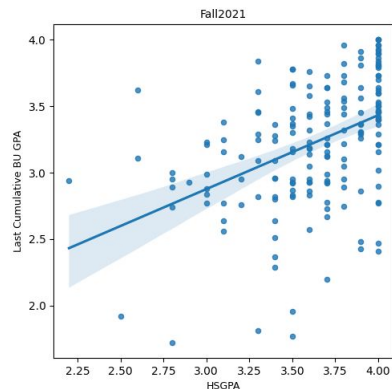 - Furthermore, the questionnaire yields only one significant indicator, as shown next
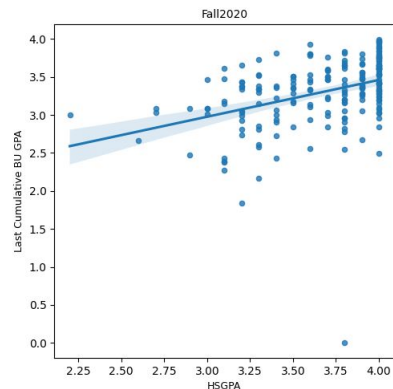
HSGPA

vs
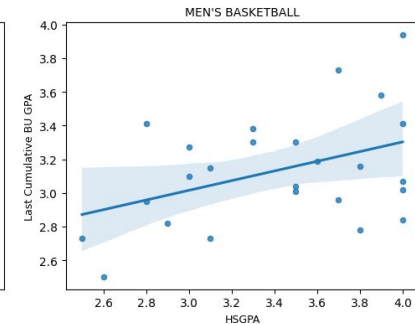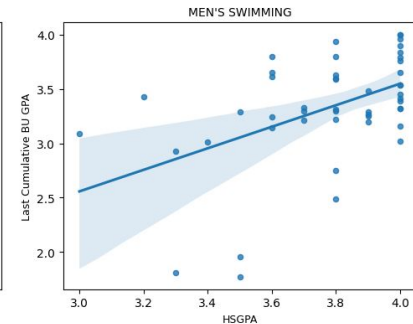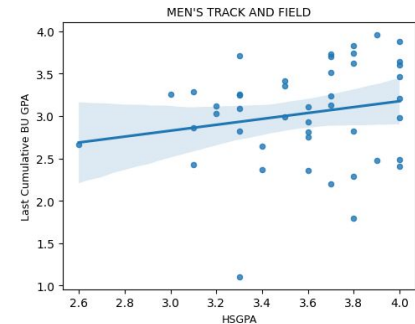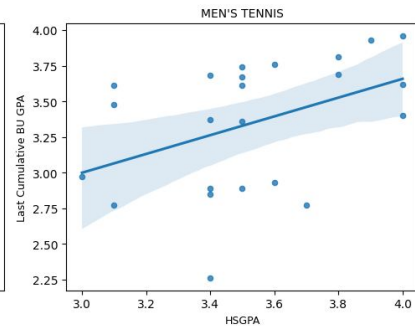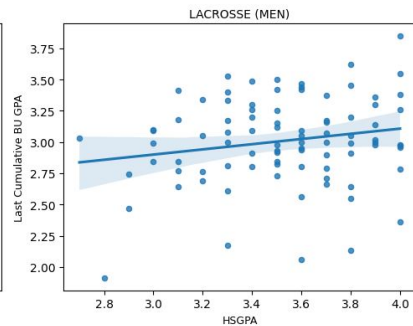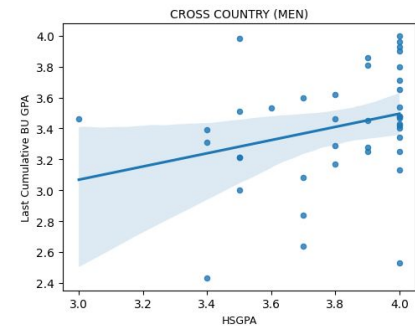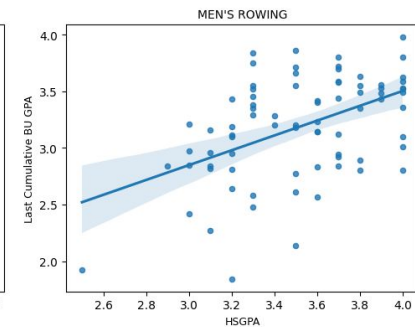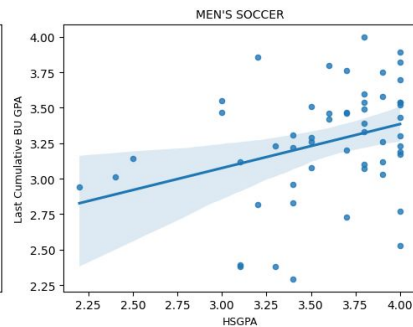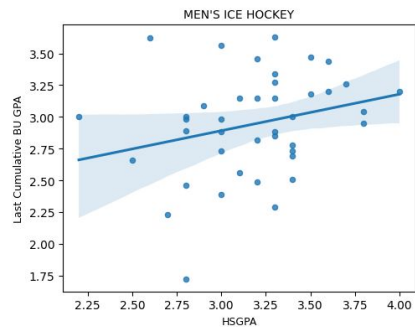
College GPA

across Teams

HSGPA

vs

College GPA

across Terms

HSGPA

vs
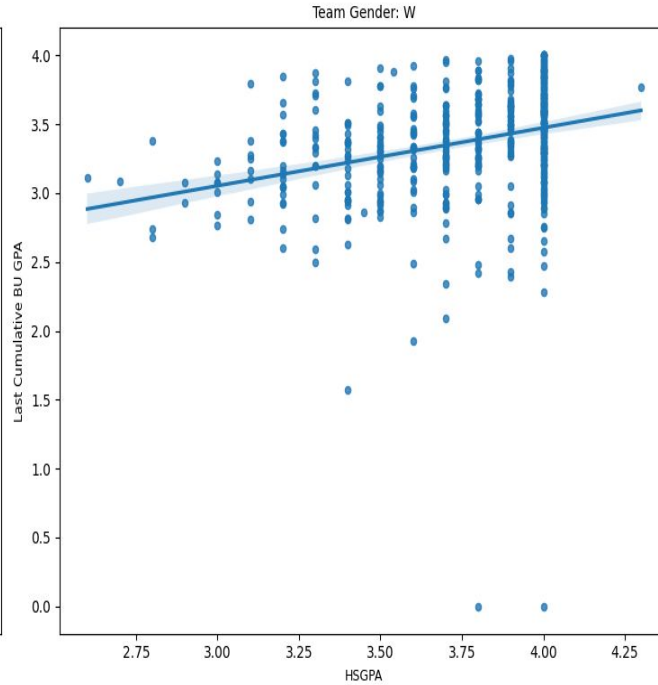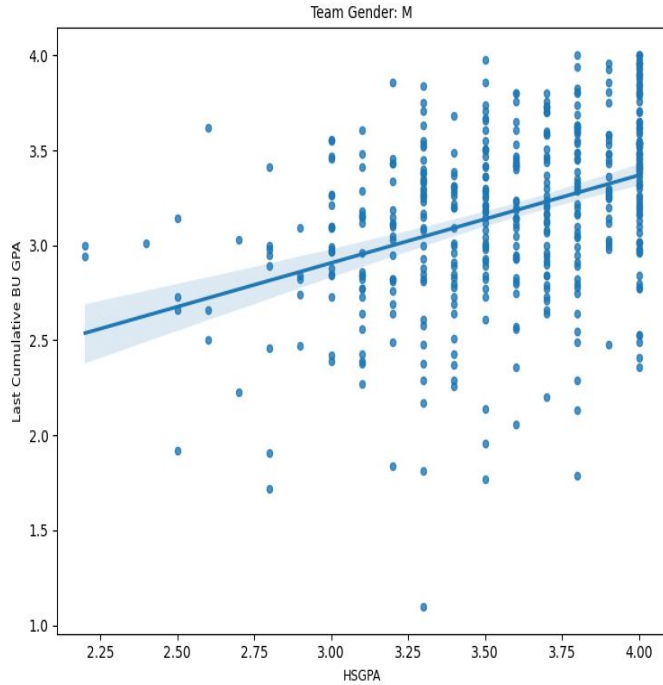
College GPA

across
Team Gender

The GPA range looks better for Woman Gender. Also the best fit line tells too that their consistency with high school GPA is little better than Men.

Last Cumulative BU GPA by Country
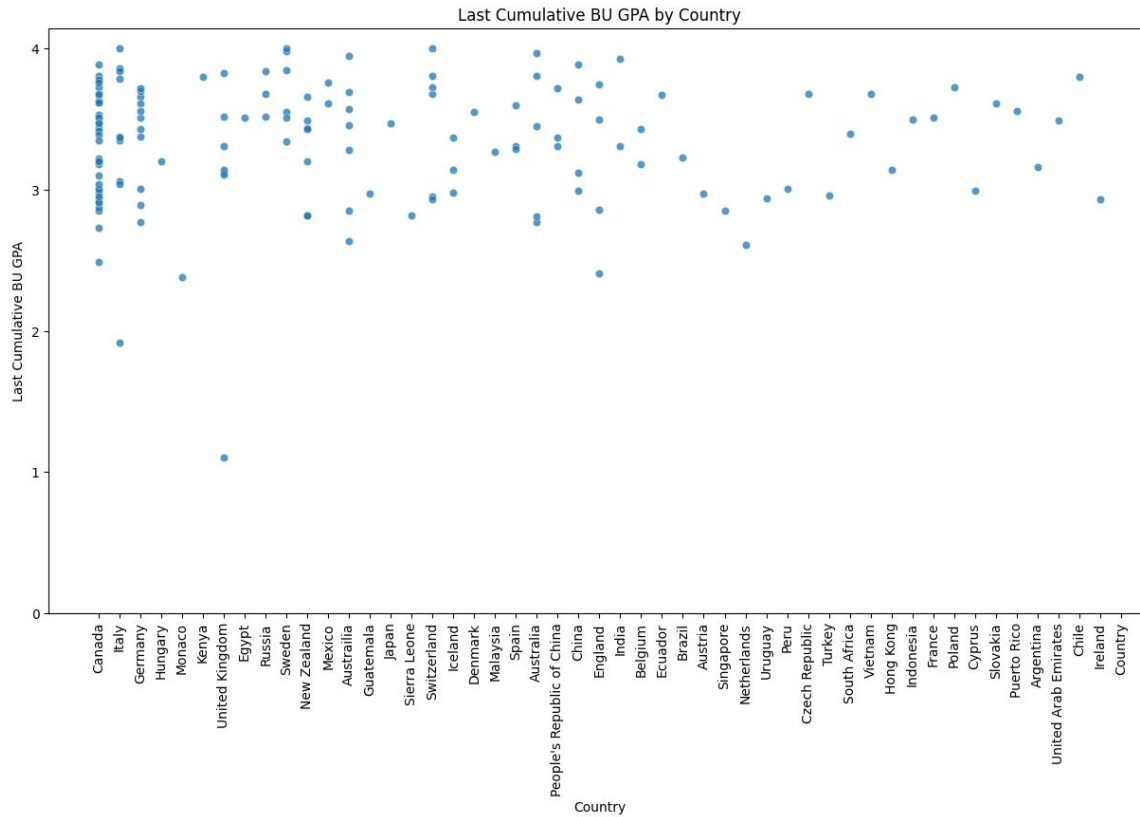
Canada has lot of students and their GPA is diverse too, but one or two students that came from Singapore, Uruguay, Netherlands etc. GPAs are not that good. But again nothing can be concluded since the data from those countries is minimal.
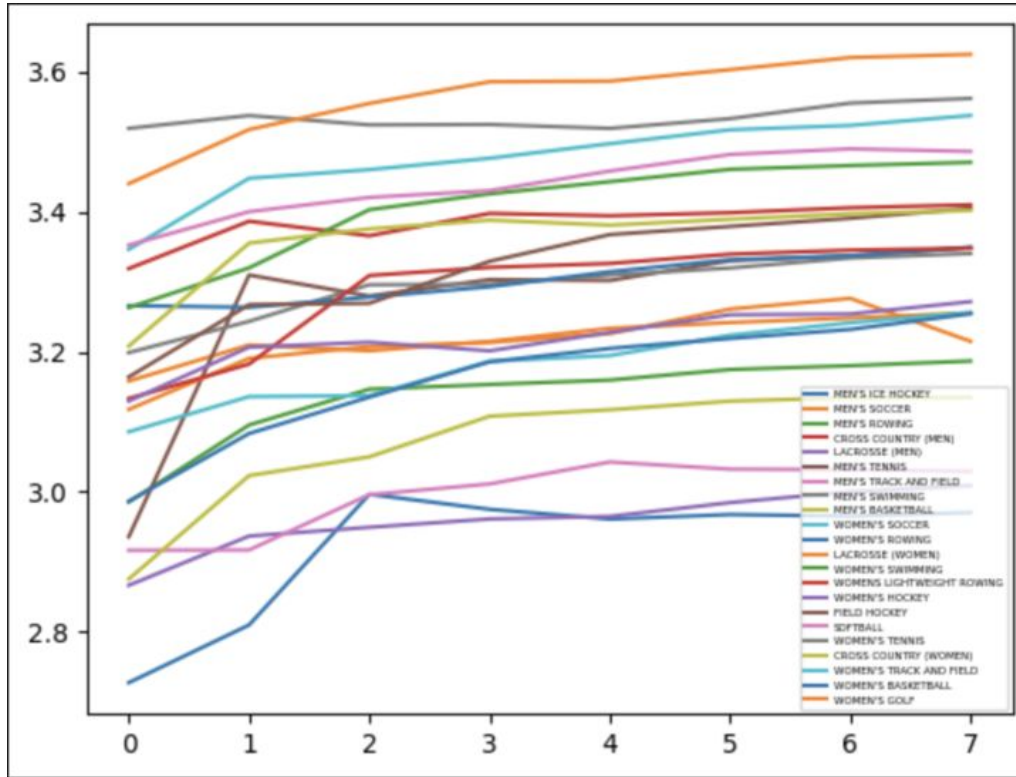
How GPA changes for student athletes by semester - broken down by team
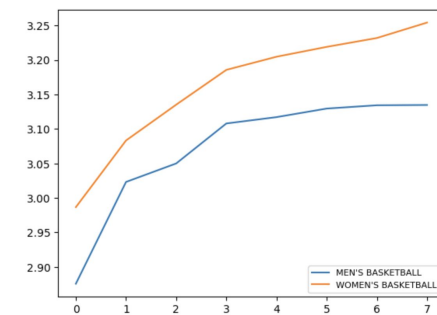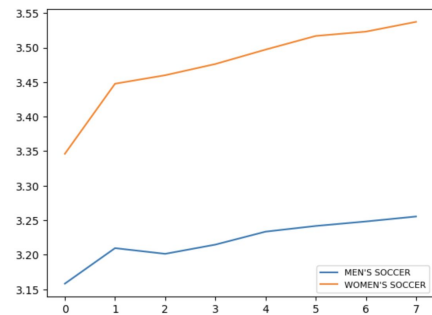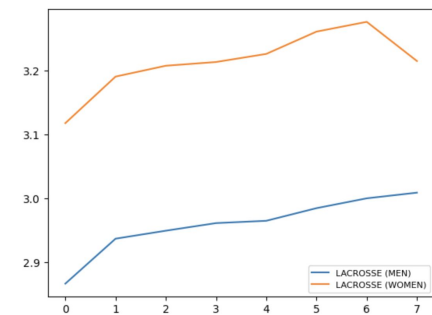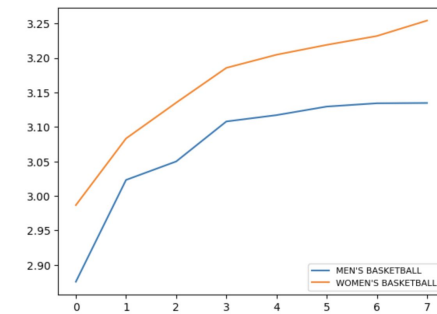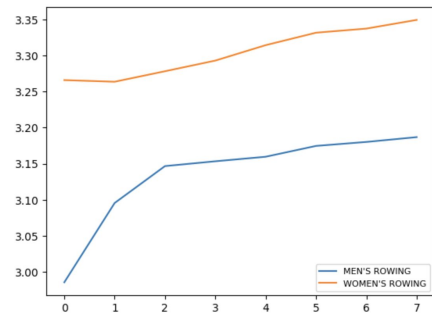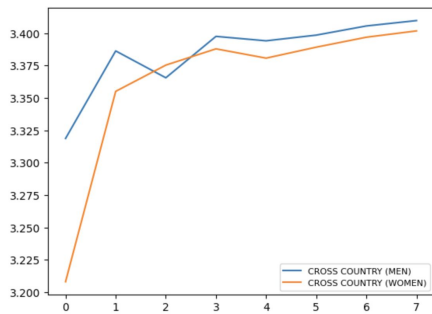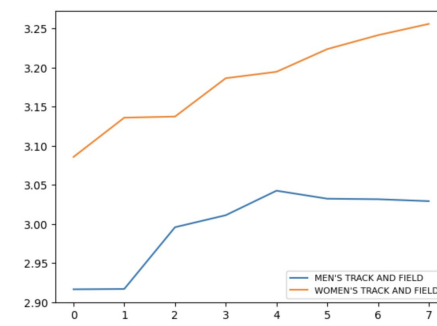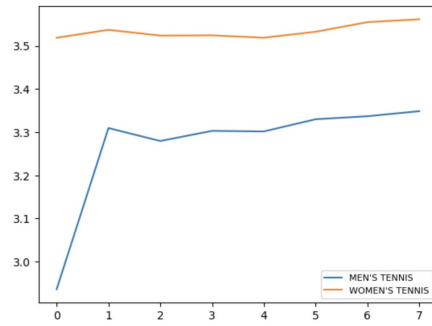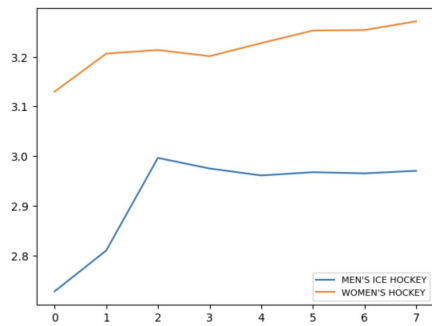
How GPA changes for student athletes by semester - broken down by team and gender

GPA by Major for Men and Women
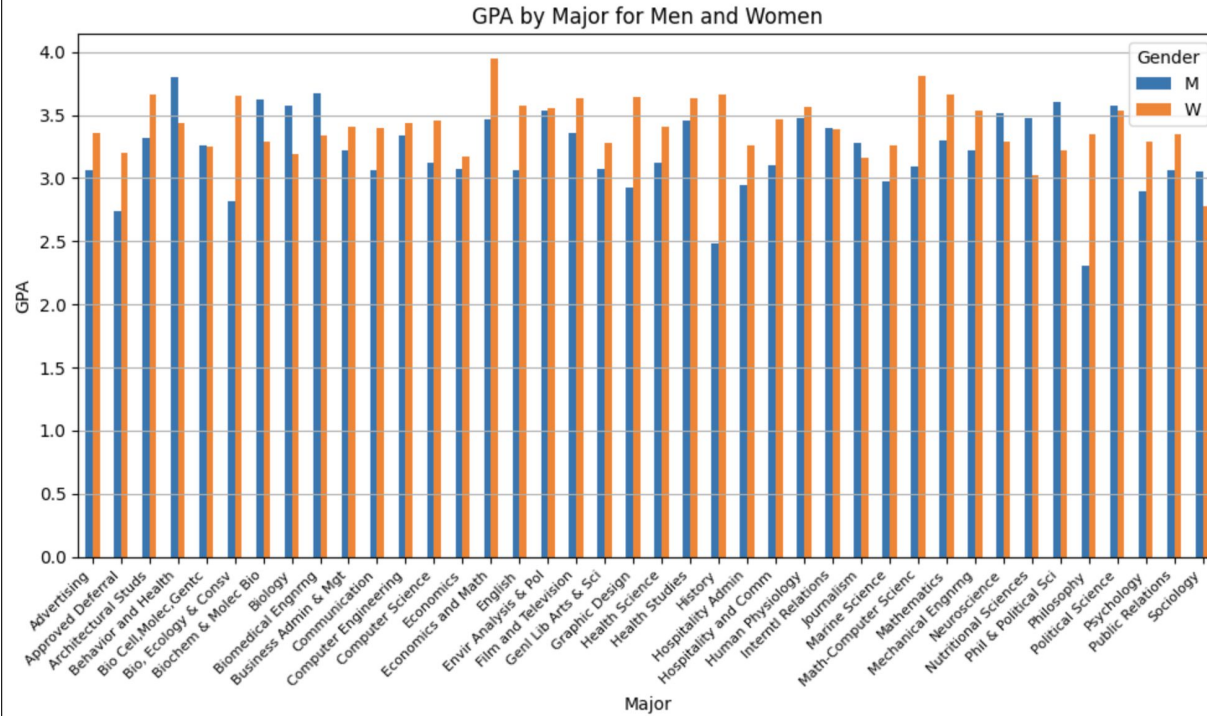
Females outperform Males for 28/40 majors both genders take

Count by Major for Men and Women
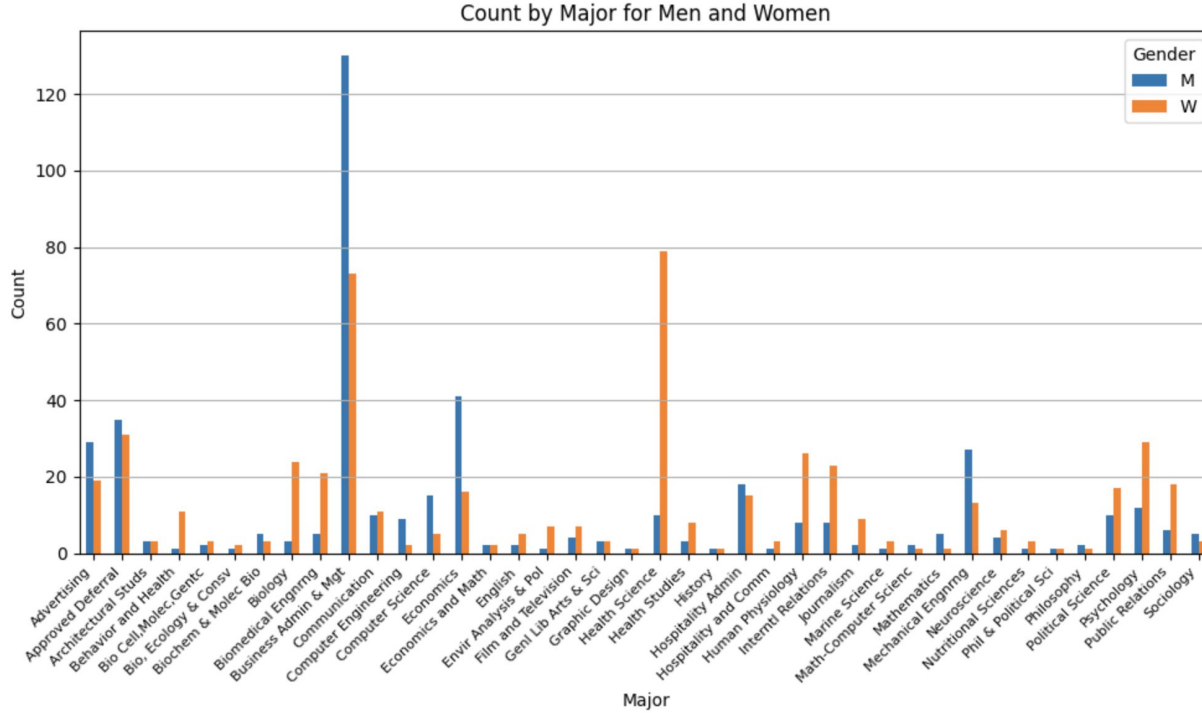
Males Most Popular Majors: Business, Econ, Advertising, Mech. Eng., Hospitality

Females Most Popular Majors: Health Science, Business, Psychology, Human Phys. Bio

Majors that Males take much more than Females: Math, Comp. Eng., Comp. Sci., Econ., Mech. Eng.

Majors that Females take much more than Males: Behavior and Health, Bio, Health Science, Envir. Analysis and Politics, Journalism

Correlation of Each Polling Feature with Averge Cumulative BU GPA of that team
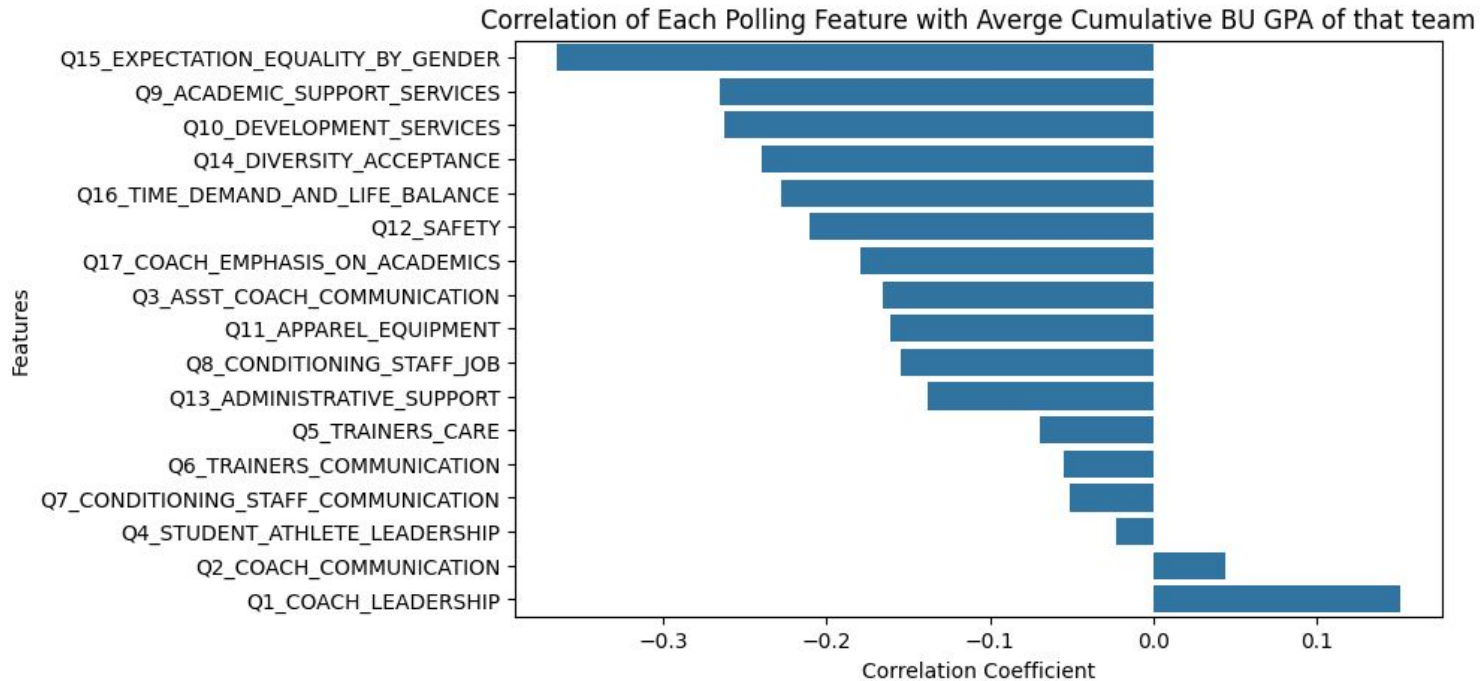
Features such as Gender Equality, Academic Support, and Development Services are observed to be inversely correlated with the average BU GPA of teams.

```
                        OLS Regression Results
==========================================================================
Dep. Variable:    Last Cumulative BU GPA   R-squared:              0.215
Model:                             OLS     Adj. R-squared:         0.208
Method:                  Least Squares     F-statistic:            34.33
Date:                 Tue, 05 Dec 2023     Prob (F-statistic):   9.  e-37
Time:                        18:54:08      Log-Likelihood:       -295.
No. Observations:                 761      AIC:                    605.2
Df Residuals:                     754      BIC:                    637.7
Df Model:                           6
Covariance Type:              nonrobust
==========================================================================
                                  coef   std err      t    P>|t|    [0.025    0.975]
--------------------------------------------------------------------------
const                           1.4312    0.333    4.297   0.000    0.777     2.085
Q15_EXPECTATION_EQUALITY_BY_GENDER -0.1049  0.035  -2.966  0.003   -0.174    -0.035
Q9_ACADEMIC_SUPPORT_SERVICES   -0.0237    0.054   -0.441   0.659   -0.129     0.082
Q10_DEVELOPMENT_SERVICES        0.0044    0.090    0.049   0.961   -0.173     0.182
Q1_COACH_LEADERSHIP             0.0173    0.032    0.541   0.588   -0.045     0.080
HSGPA                           0.4075    0.046    8.882   0.000    0.317     0.498
SATCFORAI                       0.0006    0.000    4.931   0.000    0.000     0.001
==========================================================================
Omnibus:               375.576   Durbin-Watson:             1.9
Prob(Omnibus):           0.000   Jarque-Bera (JB):      5285.690
Skew:                   -1.863   Prob(JB):                  0.00
Kurtosis:               15.362   Cond. No.               3.42e+04
==========================================================================
```

From the above p values, except "Q15_EXPECTATION_EQUALITY_BY_GENDER", no other attribute from polling data is a significant predictor, in determining college GPA, which tells us Gender Equality is important.
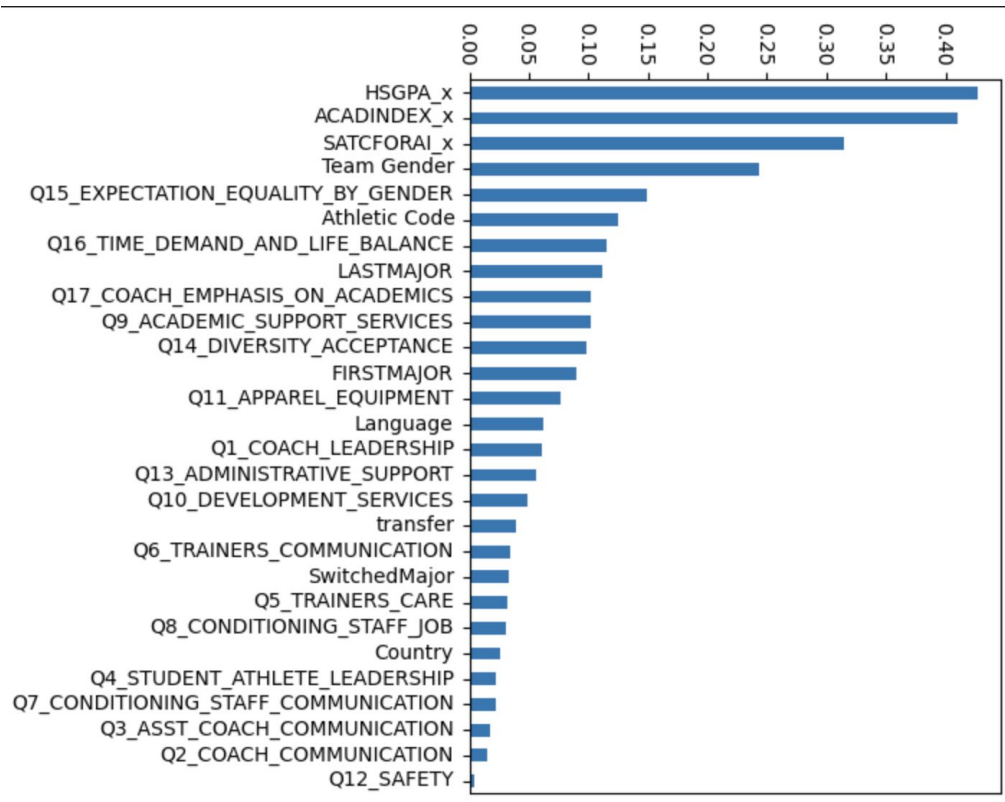
Top Correlated Features (all features)
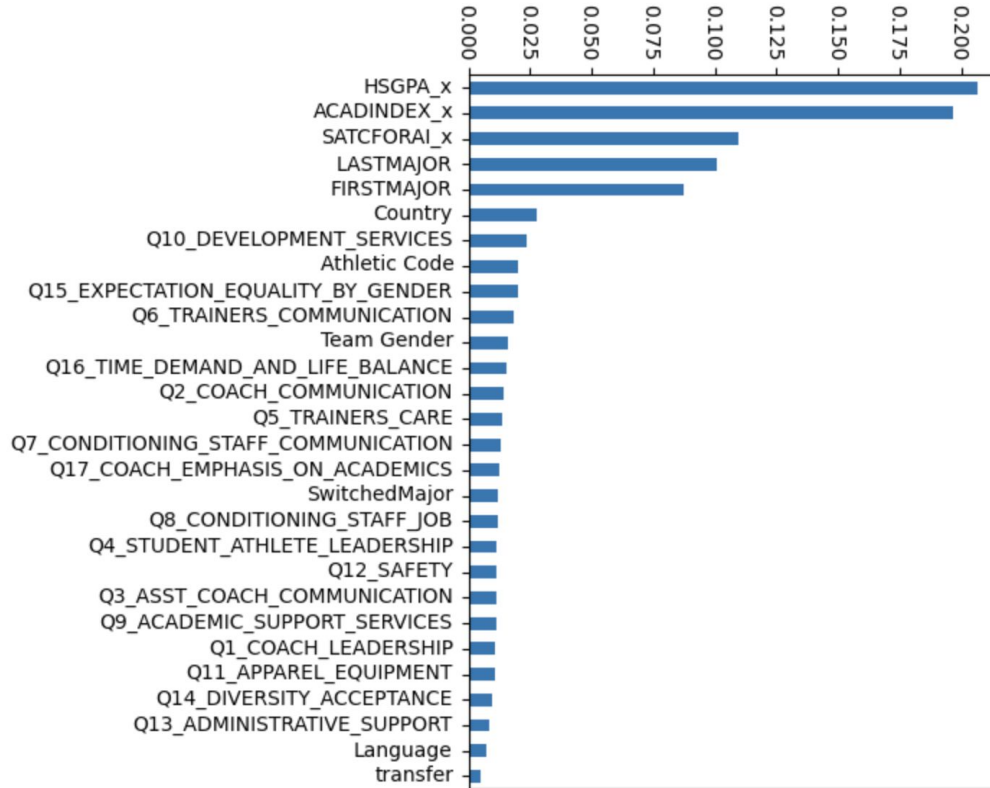
Feature Importance Using Random Forest

Objective: Predict the GPA performance of each student athlete

	Last Cumulative BU GPA
0	3.00
1	3.46
2	3.53
3	3.86
4	3.66
5	2.89
6	2.39
7	3.20
8	2.38
9	3.46
10	3.62
11	3.80
12	3.12
13	3.09
14	2.91
15	3.03
16	3.51

class 1: [0 - 2.5]
class 2: (2.5 - 4.0]

Top 15 Feature Importance
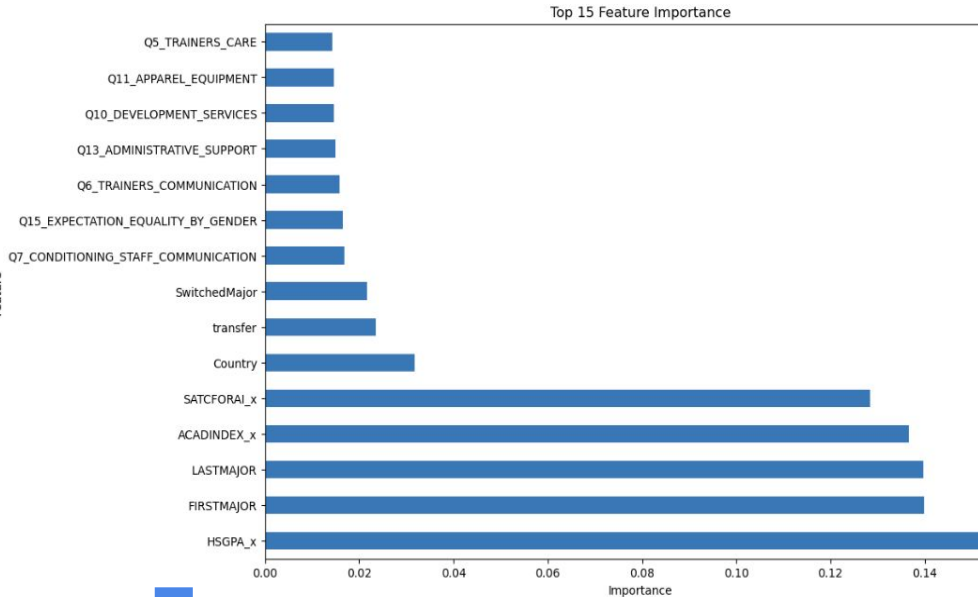
Logistic Regression

```
Accuracy: 0.9593908629441624
Classification Report:
              precision    recall  f1-score   support

           1       0.00      0.00      0.00         7
           2       0.96      0.99      0.98       190

    accuracy                           0.96       197
   macro avg       0.48      0.50      0.49       197
weighted avg       0.93      0.96      0.94       197
```

Objective: Predict the performance of each student athlete

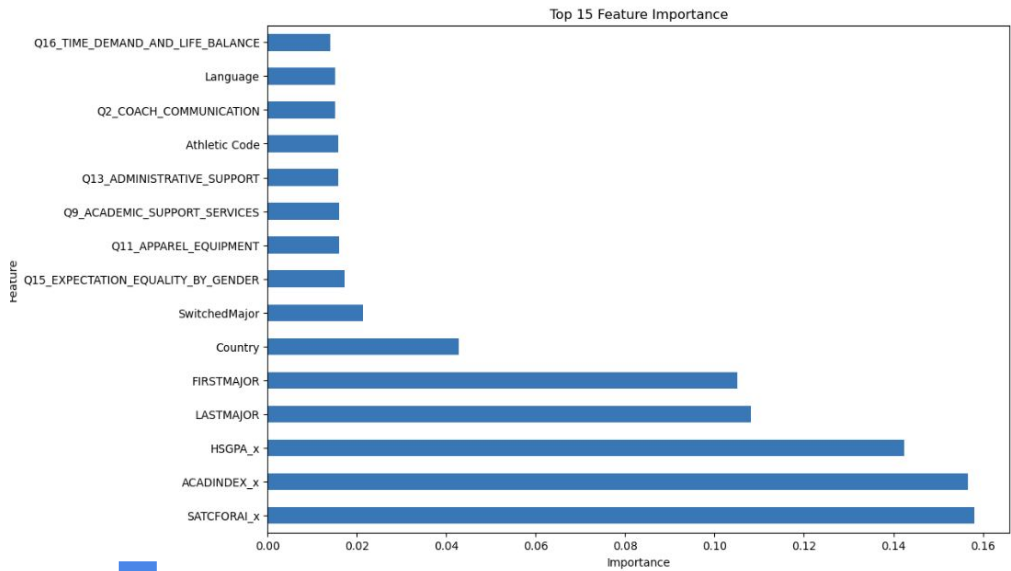|	Last Cumulative BU GPA
0	3.00
1	3.46
2	3.53
3	3.86
4	3.66
5	2.89
6	2.39
7	3.20
8	2.38
9	3.46
10	3.62
11	3.80
12	3.12
13	3.09
14	2.91
15	3.03
16	3.51

class 1: [0 - 3.5]
class 2: (3.5 - 4.0]

Top 15 Feature Importance

Logistic Regression

```
Accuracy: 0.7208121827411168
Classification Report:
              precision    recall  f1-score   support

           1       0.75      0.88      0.81       130
           2       0.64      0.42      0.50        67

    accuracy                           0.72       197
   macro avg       0.69      0.65      0.66       197
weighted avg       0.71      0.72      0.70       197
```

Conclusions Thus Far

- Female Student Athletes far outperform their Male counterparts

- Negative Correlation Coefficient but high feature Importance of Gender Equality indicates desire of female student athletes to outperform male counterparts (potentially for equal recognition) is a driving force in academic achievement

- Male student athletes take more math heavy majors as opposed to female counterparts while females take many more health (mental and physical) than male counterparts

- It is easier to predict if someone will be high performing or low performing rather than trying to predict their actual GPA (since the given feature space was small)