BU Athletics: Team-B

What have I worked on?

We have been working on the questions that BU Athletics has requested from us and plan on having them all answered by checkpoint A

What will I be working on next?

We are going to finish answering the questions that BU Athletics has asked from us by the first checkpoint. These questions are not too complicated and many of them only involve more basic statistics (mean, average, range, min, max). After answering all these questions we will start to ask more complex questions that may involve some of the data science models we have worked on (e.g. clustering)

Have I run into any issues? Do I need help?

No issues for the team. We have been working on individual questions and contributing to our branch on the github repo with the questions we have been answering and getting plots for.

Team Lead: Have I talked to the client recently? When are we meeting with them next?

I talked to Matthew, our TPM, on Thursday. This will be the most common time we will meet and discuss how the project is moving forward. He was satisfied with where we were at in the project. Next week we plan on discussing more 'branch-off' projects after the first checkpoint.

Some Work/Questions Answered for BU Athletics and Plots to go Along start on next page

We plotted the distribution of FIRSTMAJOR (figure 1) and the distribution of LASTMAJOR (figure 2) over the number of students. We found that initially Business Admin & Mgt has the most students and later more students transferred to this major.

The number of students who changed their major is 232 and the percentage of students who changed their major over all students is 23.4%.

For the next step, we would like to find out why those students want to change their majors, does it relate to GPA or other features.

(Note: The cleared graph is displayed in the github repo)

Figure 1

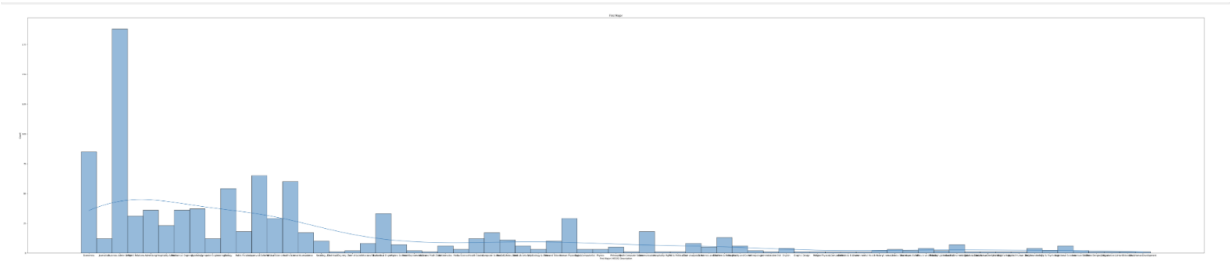
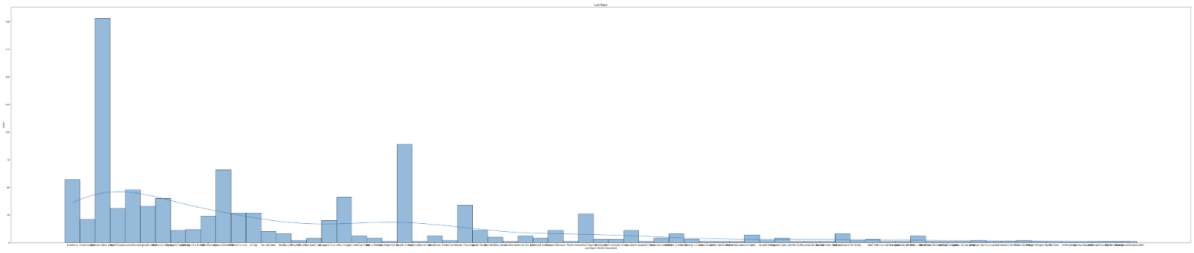
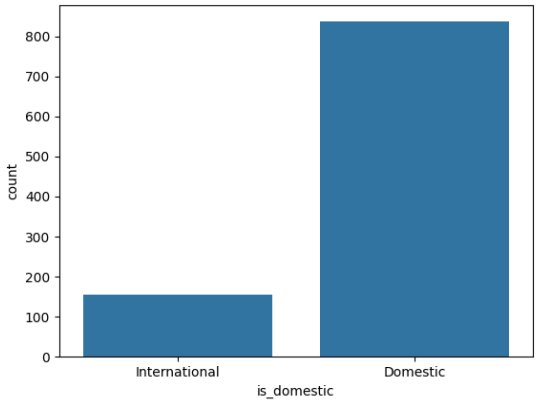
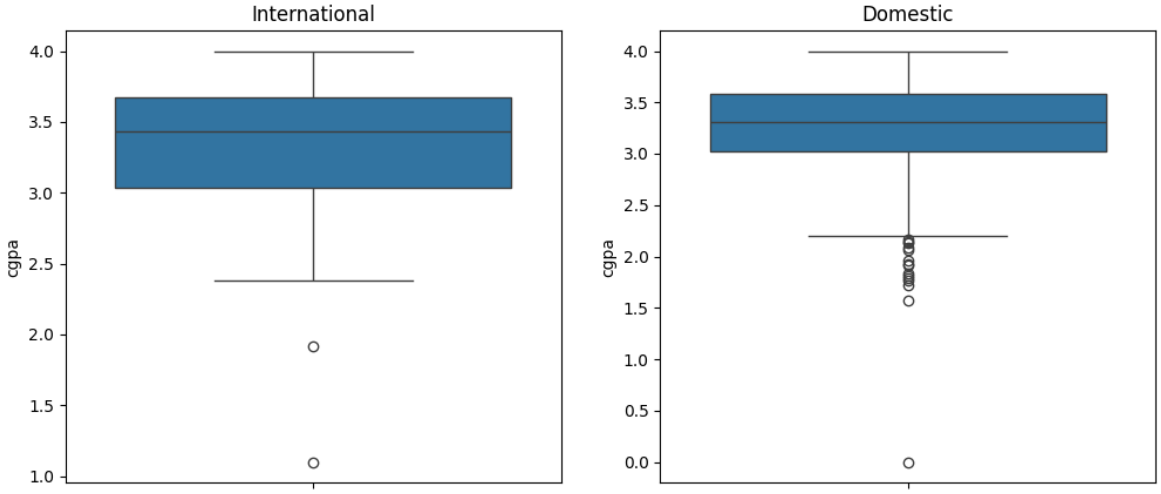


Figure 2



DOMESTIC vs INTERNATIONAL

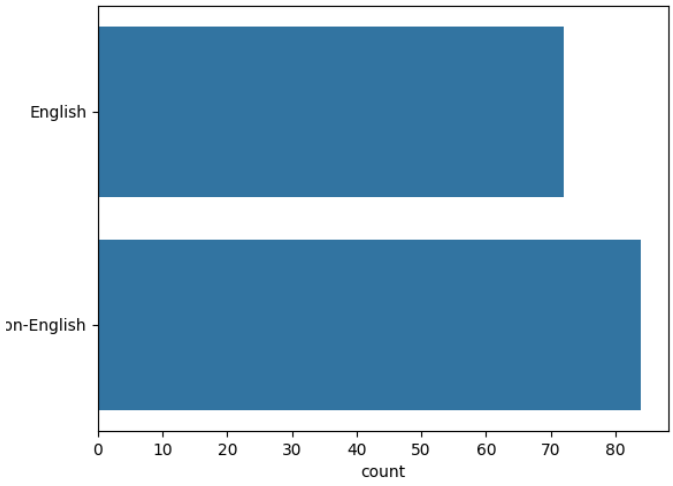


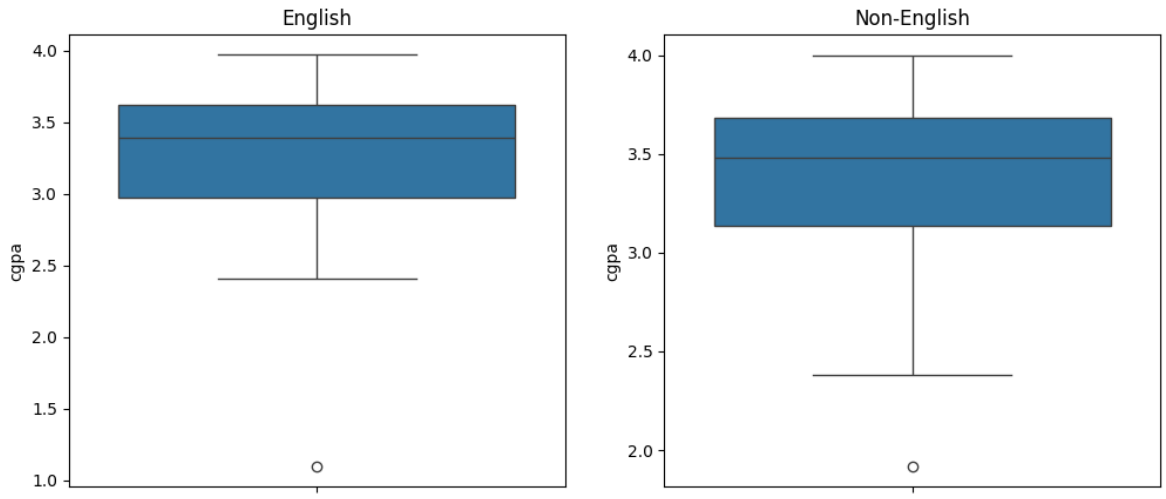


- Median is almost the same for both, suggesting an average performance of international and domestic students are almost similar.
- But a lot of outliers in Domestic. Some athletes even got less than 2 CGPA. So grades for domestic students are a little less consistent.
- Also, values between median and upper quartile range are better for International students, which may suggest that they have a good number of higher grades.

NOTE : Since the data is imbalanced, we cannot come to a conclusion on the above insights.

ENGLISH vs NON ENGLISH (International Students)

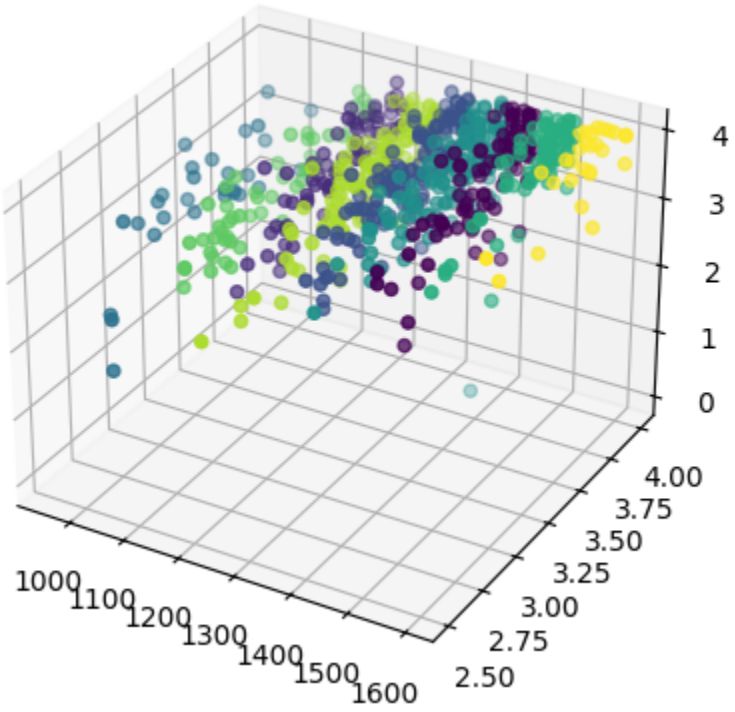




If you ignore the outliers, the distribution is pretty much the same, suggesting to us that English as the primary language or not, is not making any necessary for athletes grades, in fact the lower quartile range is slightly better for Non-English students.

NOTE : The above distribution is taken from the records of international students only.

How HS Stats Compare to College Stats



- This is a 3D scatter plot of students, x = SAT, y = HS GPA, z = C GPA
- The number of clusters as seen above was chosen based on the best silhouette score for cluster numbers in range [5, 10]
- Some stats for each of the clusters are provided below on the next page and in the GitHub repo
- SAT seems to be a better indicator on college performance than HS GPA as shown
- However, after the highest performance in HS clusters, (roughly the two clusters with the highest HS SAT), college GPA is extremely similar for the rest of the clusters

```
Cluster 1 stats:
SAT Average: 1296 Median: 1300.0
HS GPA Average: 3.708 Median: 3.8
C GPA Average: 3.334 Median: 3.37


Cluster 2 stats:
SAT Average: 1551 Median: 1540.0
HS GPA Average: 3.879 Median: 4.0
C GPA Average: 3.625 Median: 3.7


Cluster 3 stats:
SAT Average: 1107 Median: 1110.0
HS GPA Average: 3.388 Median: 3.4
C GPA Average: 3.07 Median: 3.085


Cluster 4 stats:
SAT Average: 1414 Median: 1410.0
HS GPA Average: 3.771 Median: 3.9
C GPA Average: 3.386 Median: 3.45


Cluster 5 stats:
SAT Average: 1236 Median: 1240.0
HS GPA Average: 3.623 Median: 3.7
C GPA Average: 3.262 Median: 3.29


Cluster 6 stats:
SAT Average: 1355 Median: 1360.0
HS GPA Average: 3.741 Median: 3.8
C GPA Average: 3.311 Median: 3.41


Cluster 7 stats:
SAT Average: 1009 Median: 1005.0
HS GPA Average: 3.313 Median: 3.3
C GPA Average: 3.105 Median: 3.105


Cluster 8 stats:
SAT Average: 1477 Median: 1470.0
HS GPA Average: 3.863 Median: 4.0
C GPA Average: 3.547 Median: 3.58


Cluster 9 stats:
SAT Average: 1177 Median: 1180.0
HS GPA Average: 3.634 Median: 3.7
C GPA Average: 3.247 Median: 3.275
```