

**Exploring MBA Admission Patterns: An Analysis of Wharton's Class of 2025**

Saifuzzaman Tanim

Department of Computer Science, Brooklyn College

CISC 3225: Data Tools and Algorithms

Instructor: Matthew McNeill

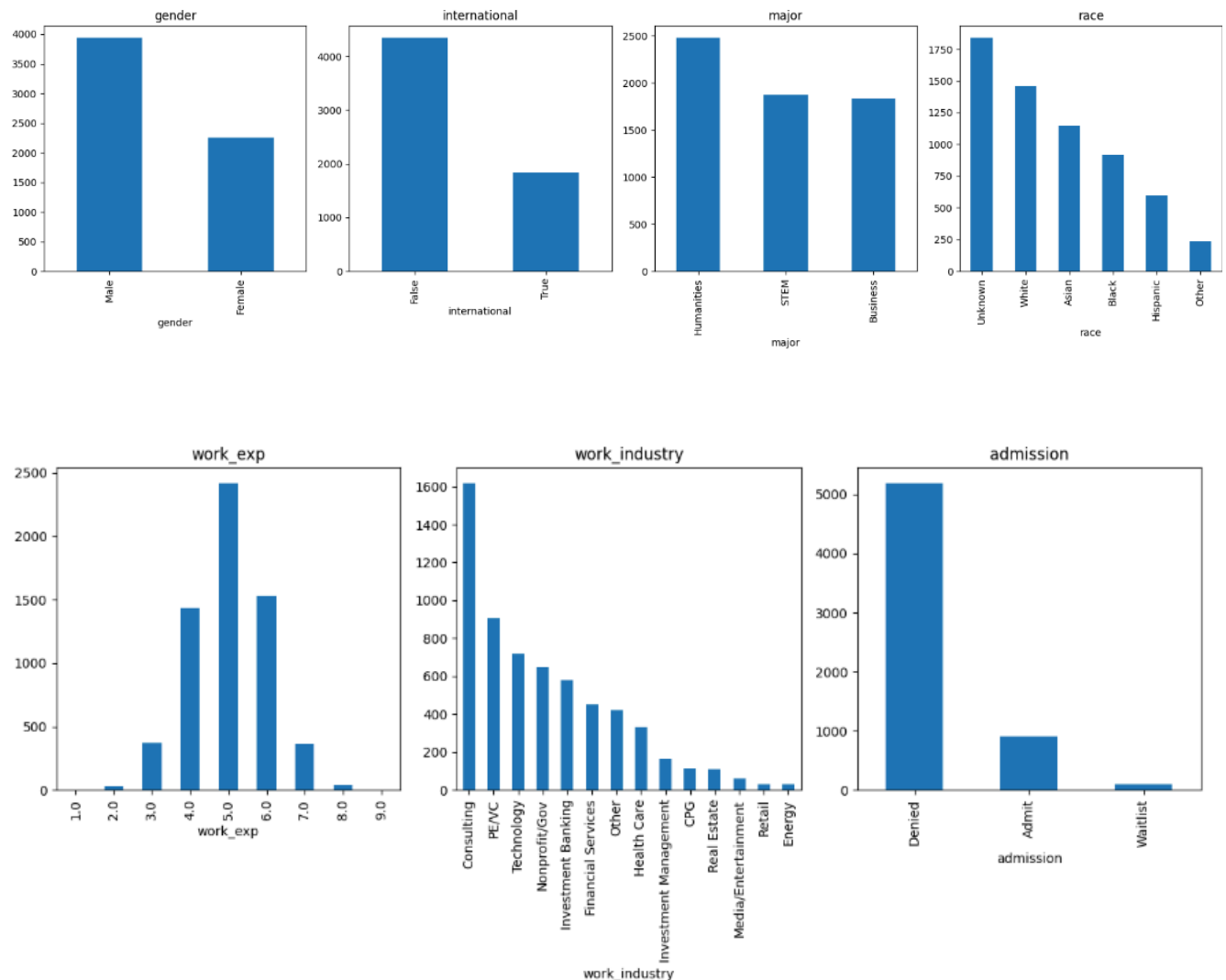
May 21, 2025

**Introduction:**

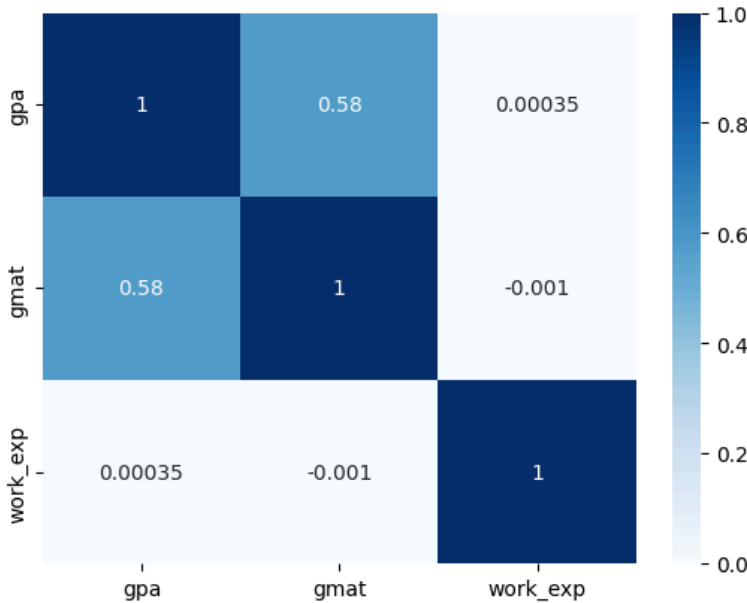
The dataset used for this project contains information about the Wharton Class of 2025 MBA admissions. The purpose of this dataset might be to keep data on previous years' MBA applicants. They might perform analysis to find patterns in admitted students and how well they do in university. The dataset contains 10 columns, such as gender, international status, GPA, major, race, GMAT scores, work experience, work industry, and admission status for more than 6 thousand students. For my analysis, I tried to find patterns between the columns and how they affect admission status. Additionally, I looked into whether certain types of students prefer specific majors or industries. Lastly, I attempted to predict if a student with a certain quality would be admitted or not.

## Exploratory Analysis:

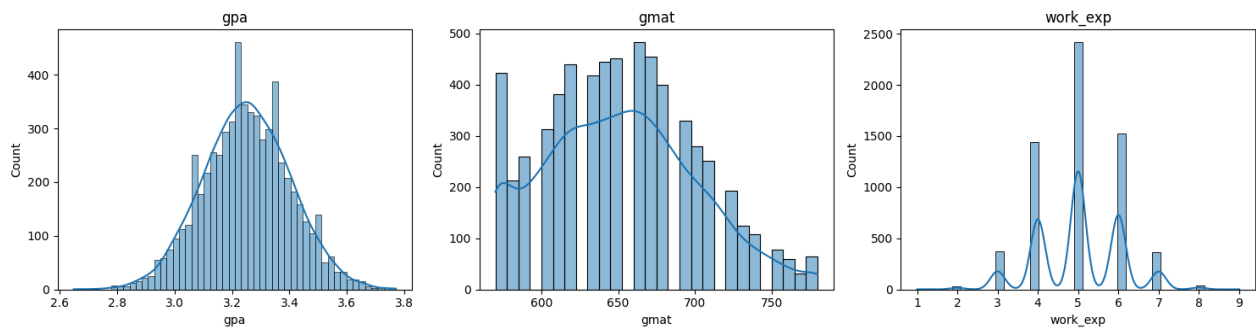
The dataset contains 6,194 MBA applications with various information such as GPA, GMAT scores, work experience, gender, major, race, work industry, and admission status. There are missing values in the "race" and "admission" columns. Missing race data might mean that some applicants did not report their race, also missing admission data likely represents denied applicants. These missing values were handled by filling "admission" with "Denied" and "race" with "Unknown."



The bar plots above show information about the categorical columns. It shows that there were more male applicants than female applicants. Also, more domestic students than international students. There was a slight difference in their choice of major. Finally, the majority of applicants got denied, which indicates that their admission policy is pretty strict. However, more analysis and statistical testing are needed to determine what caused these outcomes or if they were just random.



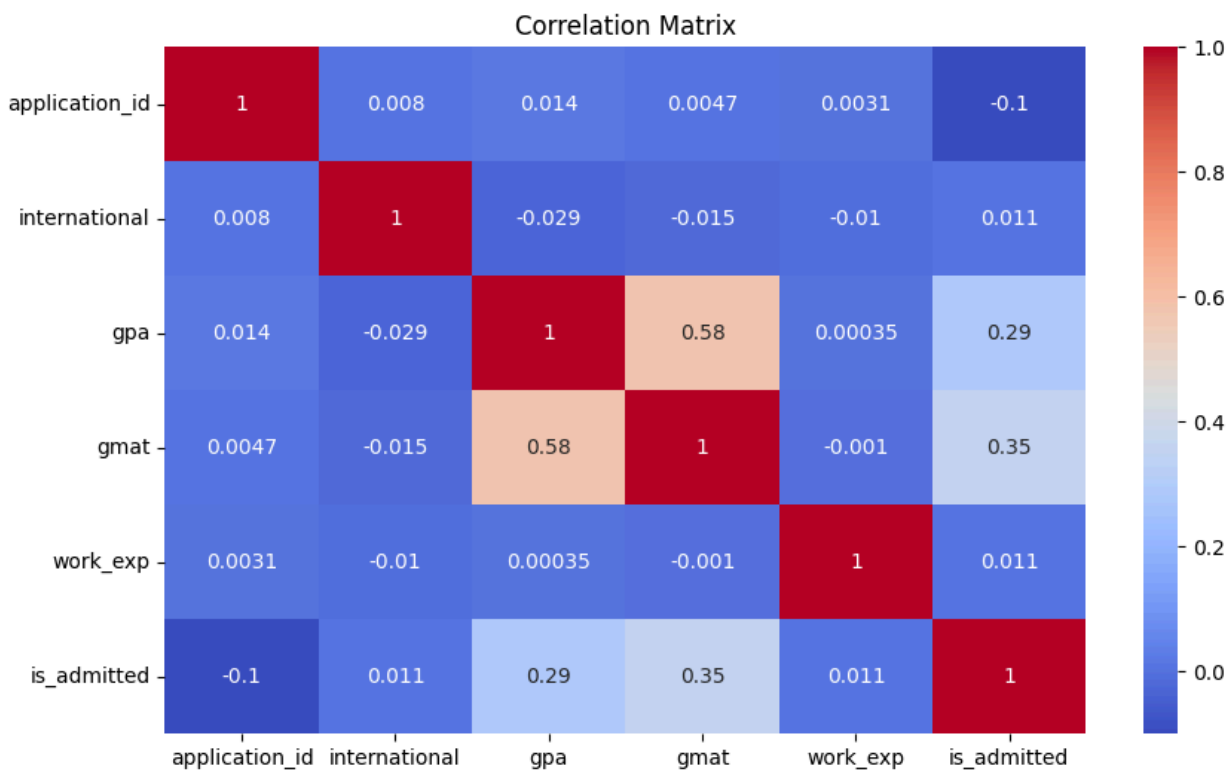
We can use a correlation matrix to see if any of the continuous variables affect each other. The correlation matrix shows a moderate positive relationship (0.58) between GPA and GMAT scores, suggesting that applicants with higher GPAs also tend to have higher GMAT scores. This makes sense since students who do well in undergraduate programs tend to do well on the GMAT, as the test includes questions based on topics learned in undergrad.



The histograms provide insights into the distribution of numeric columns. GPA and work experience are roughly normally distributed, and that makes sense as students generally have this type of distribution. Most of them have a GPA around the 3.2 mark. Some tend to do badly, while some are exceptional. The same goes for work experience. Most of them have certain experience, in this case, 5 years. Some have very little, who are probably fresh graduates, and some have way more. They might have been working for a while and now want to get a master's. However, GMAT scores are slightly skewed to the right. This might mean that it is easy to get some score in the exam, but very difficult to get a high score, just like the SAT. However, more analysis and statistical testing are needed to determine what caused these outcomes or if they were just random.

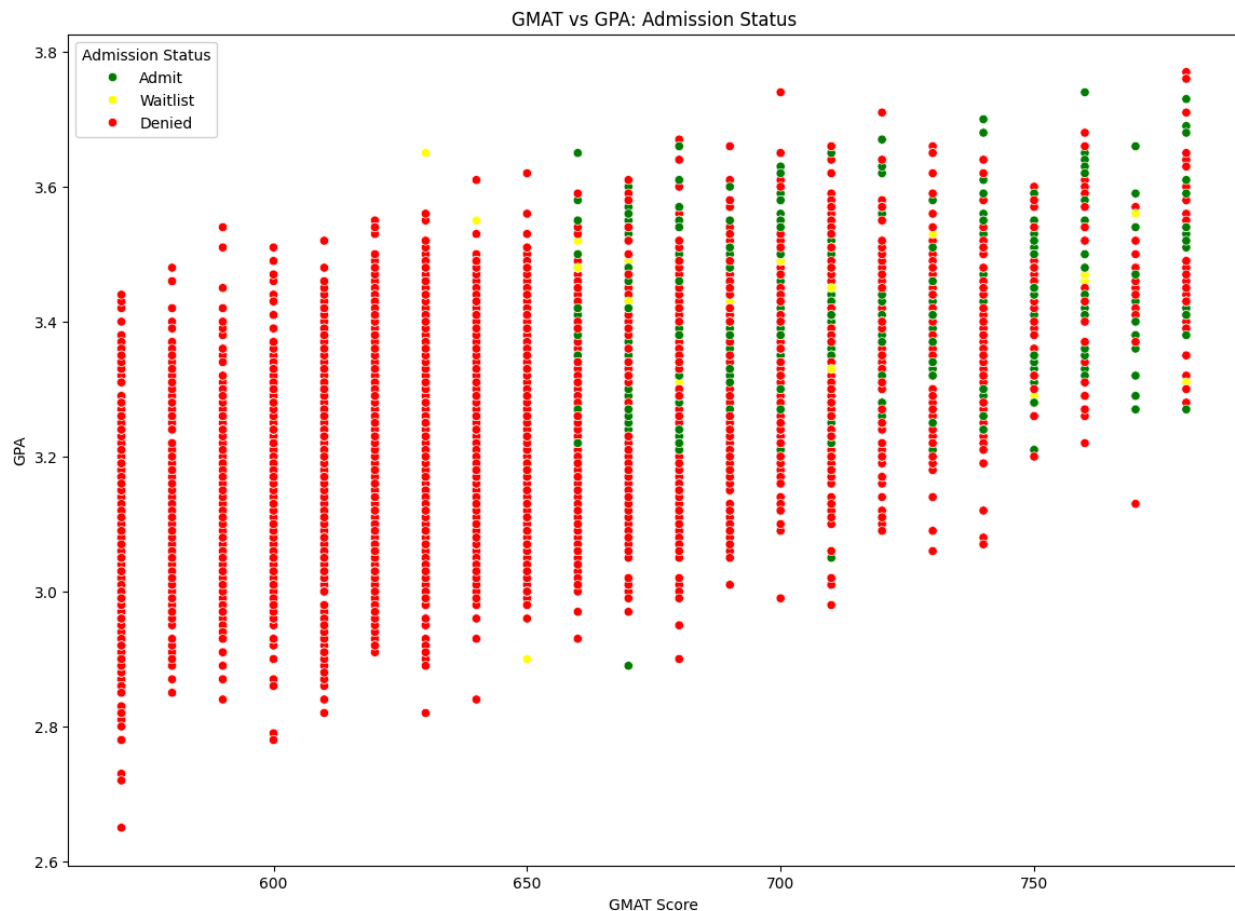
### Correlation Matrix Analysis (GMAT, GPA, Admission)

A correlation matrix was created to determine if there are any relationships between all continuous variables and being admitted. A new column, `is_admitted`, was added to the dataframe with a binary digit so we can compute the correlation. The correlation between GPA and GMAT was the highest, and we explained it above. The correlation between GMAT and admission is 0.35, and GPA and admission is 0.29, which is moderately positive. This makes sense, in general, if a student has a higher GPA or GMAT score, their chance of being accepted increases. But that is not all, there might be other factors that affect admission status even more. More in-depth research and statistical tests are needed to confirm these claims.



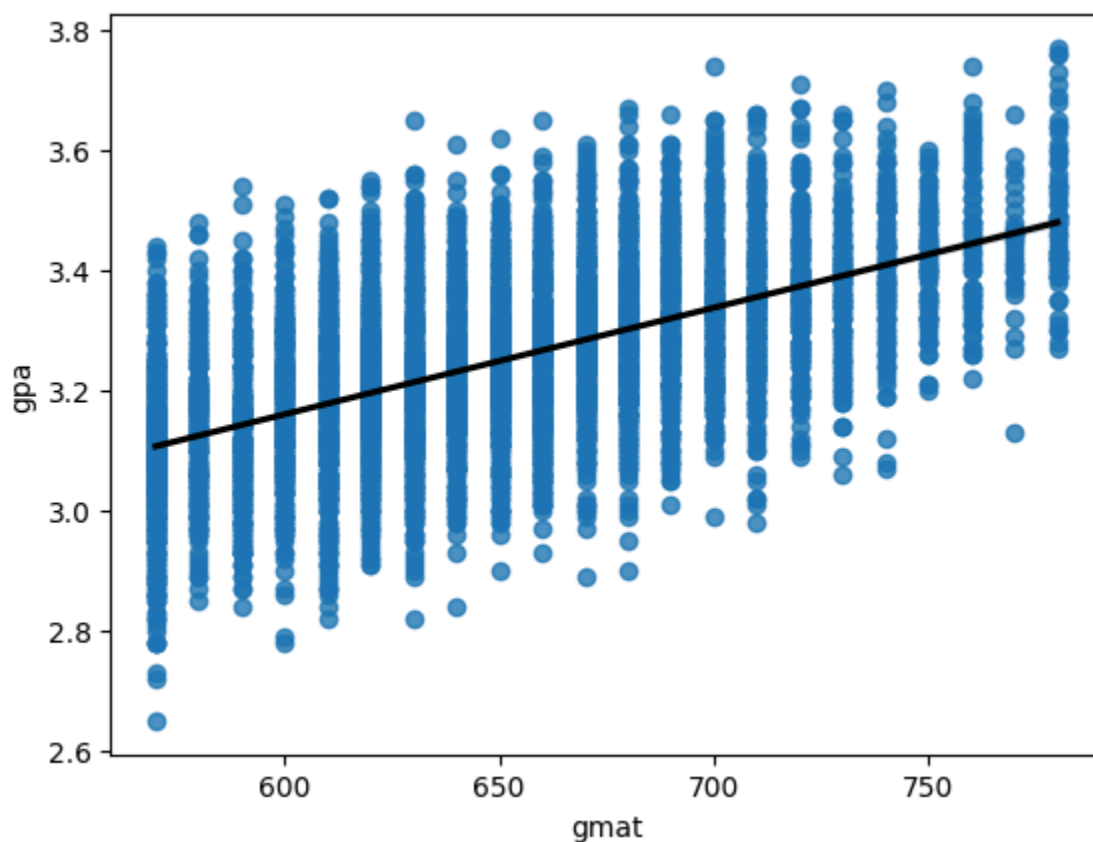
## Visualizing Admission Trends: GMAT vs. GPA (Scatter Plot)

To have a better understanding of how GPA and GMAT scores affect being admitted, a scatter plot was created. We can clearly see that all the students below a 660 GMAT score were denied, even if they had a great GPA. Although there were four who were waitlisted. This might indicate that there is a threshold for GMAT scores. Also, for GPA, there might be a threshold around 3.2. The students who were selected were mostly in the top right corner, which supports the assumption that there is some relationship between GPA, GMAT score, and being admitted. Although there were a lot of students who were denied in the top right section as well. This might mean that other factors were taken into consideration. That makes sense since they have such a strict admission policy. More research is needed to confirm what other factors played a role in being admitted, but as of now, we can say that GPA and GMAT scores did have some effect, and there was a threshold.



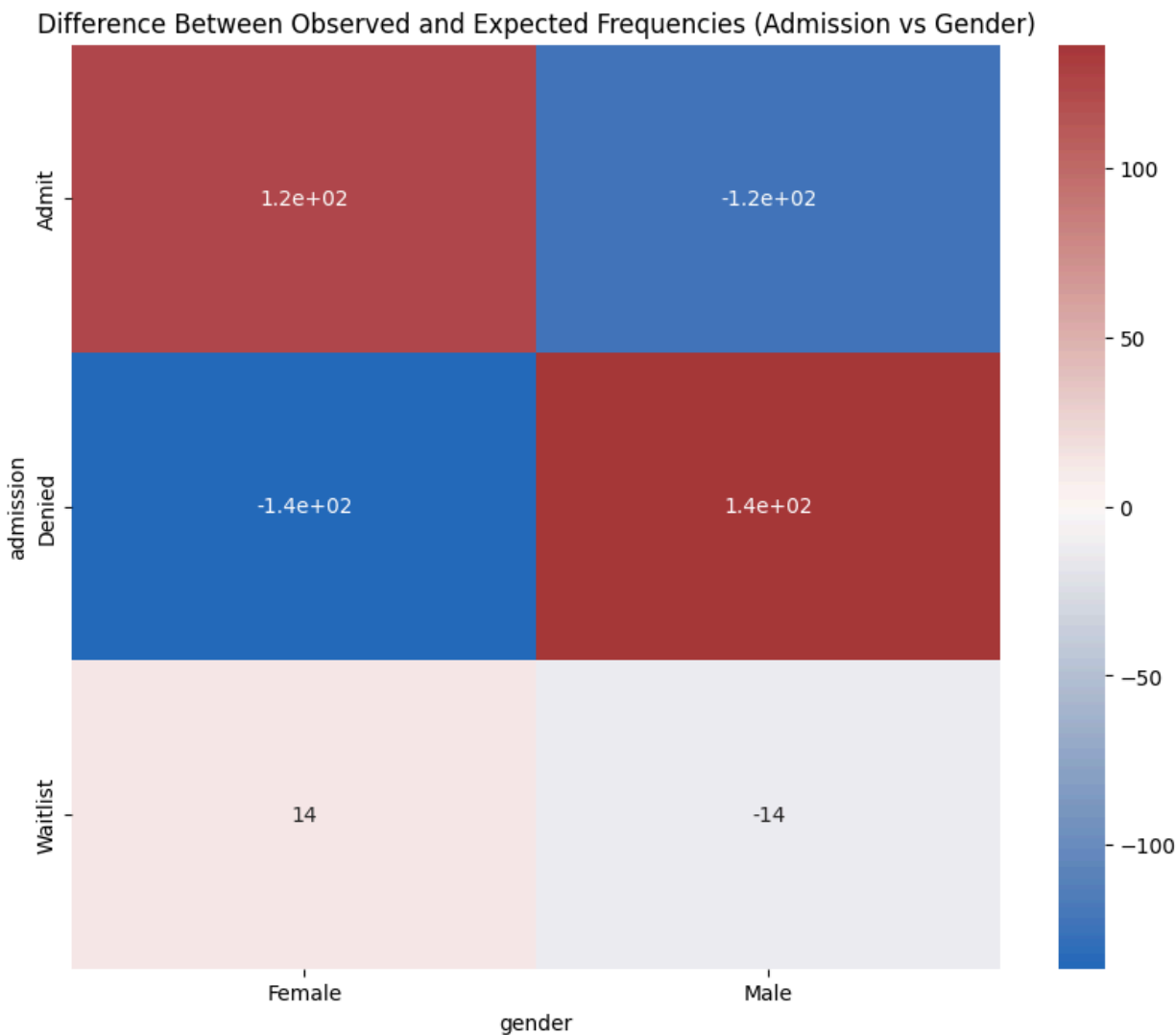
## GMAT vs. GPA (Pearson Correlation and Linear Regression)

A linear regression was done to help us predict unseen data. We can see a slope that is slightly positive. Although we cannot tell the exact number from the graph, we can use the `linregress` function to get the exact number, which was 0.0018. It might not seem large, but GPA is at a much smaller scale compared to GMAT scores, so even a small change has significant meaning. In this case, it might not make sense to find data that is outside the range of the data. This is because the data we have is most likely the range of possible values. However, we can do other useful tests from the `linregress` function. We can do hypothesis testing and see if there is a linear relationship between GPA and GMAT scores. The null hypothesis was that there is no linear relationship between GMAT and GPA scores. The p-value was 0.0, which means that the probability of getting the observed correlation or a stronger one between GMAT and GPA just by random chance is nonexistent. So, we reject the null hypothesis, and there is statistically significant evidence that GPA and GMAT scores have a linear relationship.



Gender and Admission Analysis (Chi-Square Test)

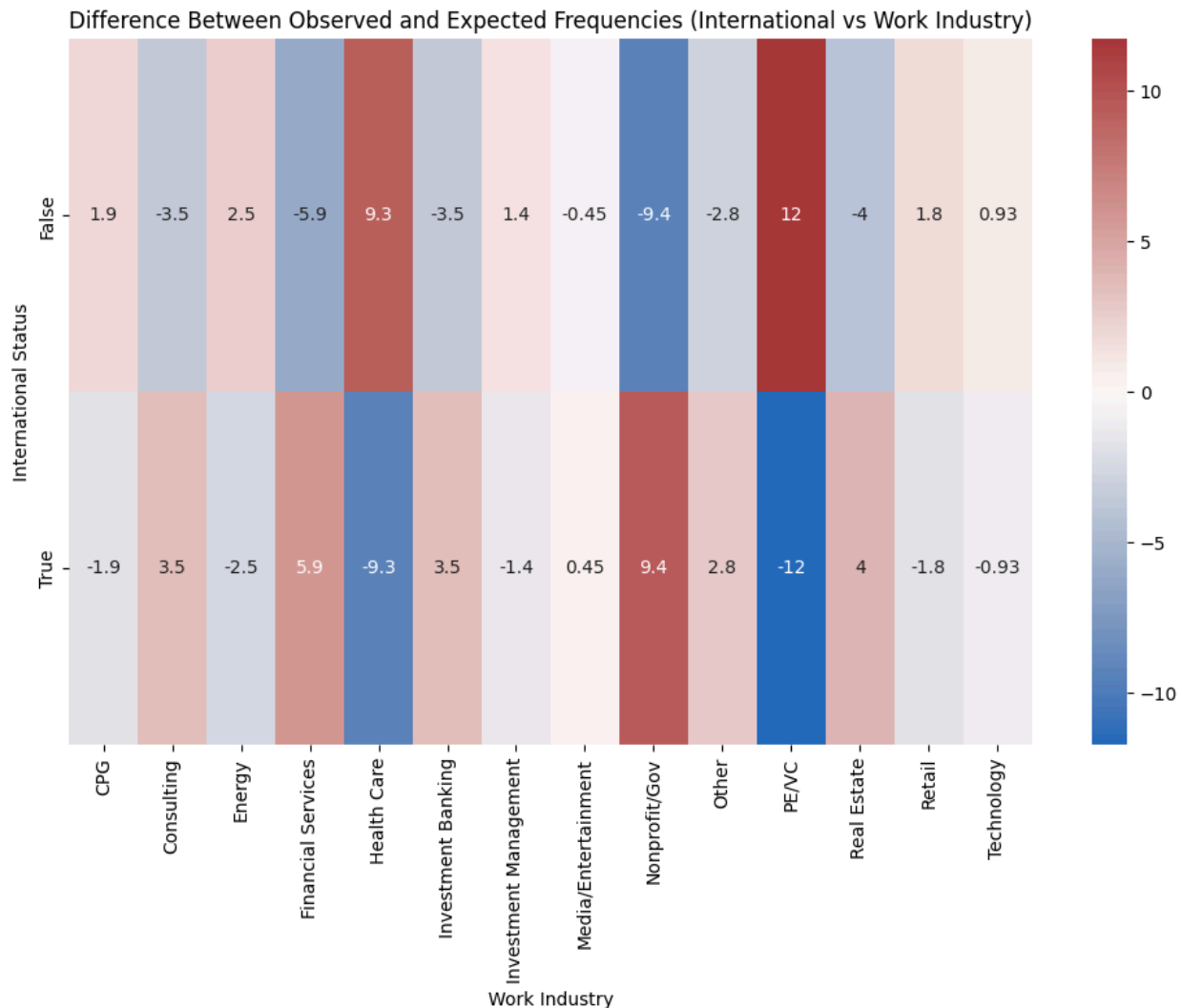
A chi-square test was performed to examine the relationship between gender and admission status. The heatmap shows the differences between observed and expected frequencies for each admission category. Male applicants experienced higher rejection rates than expected, with 140 more denials and 120 fewer admissions than anticipated. Female applicants had 120 more admissions and 140 fewer denials than expected. The waitlist category showed minimal differences, indicating that gender did not significantly impact waitlist placements. The chi-square test resulted in a p-value of 1.31e-21, indicating a statistically significant relationship between gender and admission status. So, the null hypothesis that gender and admission status are independent is rejected, suggesting that gender may influence admission outcomes.





## International and Industry Trends (Chi-Square Test)

Another chi-square test was conducted to examine the relationship between international status and their work industry. The heatmap shows the differences between observed and expected frequencies for each work industry category. From the heatmap, we can see that 12 more Domestic students work in PE/VC than expected, 9.4 worked less in Nonprofit/Gov than expected. And the opposite goes for international students. These numbers might be statistically significant that domestic or international students prefer to work in certain industries. So we do hypothesis testing again, where our null hypothesis is that there is no association between international status and work industry. We get a p-value of 0.9608, which means that we fail to reject the null hypothesis. And that suggests that the observed differences in work industry distribution by international status are likely due to chance.



## **k-Nearest-Neighbors Classification**

A k-Nearest Neighbors (KNN) model was built to predict admission status using GMAT, GPA, and work experience as input features. The model has an accuracy of 81%, meaning that it is able to successfully predict a student's admission status based on their GMAT, GPA, and work experience 81% of the time. However, 19% of predictions were incorrect. This might be because we did not have enough data columns, like other factors that go into consideration while making the decision. In addition, the scatter plot we did above for GPA, GMAT scores, and their admission status showed that a lot of students were denied even if they did well in both of these. This might have been the reason behind those 19% incorrect predictions. Since KNN uses the Euclidean distance formula, and we used  $k = 7$ , if more students near the test student we are using were denied, our KNN model would predict them as denied too, but in reality, they were accepted. So in general, it is pretty difficult to make predictions from just these 3 columns, but we got an accuracy of 81%, which is pretty good.

## **Conclusion**

From this project, I learned that while GMAT and GPA play important roles in MBA admissions, but they are not the only factors that determine the outcome. GMAT and GPA had a stronger connection to admission status than other features. However, the connection was still moderate, suggesting that other factors also impact the decision-making process. The KNN model had an accuracy of 81%, indicating that it was pretty effective in predicting admission based on GMAT, GPA, and work experience. However, its accuracy was limited because it only used these three features. Gender analysis showed that female applicants were admitted more often than expected, while male applicants faced more denials, hinting at a possible gender influence in admissions. More research is needed to explore the relationship between gender and admission rate. If gender is a big factor making a difference in admission status, it is a concern that requires deeper investigation. On the other hand, international status, work industry, or work experience did not show any strong connections to admission status. Overall, the findings suggest that academic scores are important but not the only indicators of admission outcomes. Future studies could include more features, like more exam grades, extracurricular activities, and recommendation letters, to provide a fuller picture of what influences MBA admissions.