Drexel Technical Evaluation Data Analysis

DESIRE MATOUBA

NOVEMBER 12TH 2021

Datasets

▶ 2 CSV files provided: Admissions and Financial Aid. I used Python to do the data analysis.

▶ Some columns in the Admissions file: Entry Term, Admission Status, Enrollment status, Standardized Test Score, High School GPA, Estimated Family Contribution, and Financial Aid.

▶ The Admissions file has data for 79,999 students who applied at Drexel University over 5 Fall semesters: from Fall 2017 to Fall 2021.

▶ The Financial Aid file contains data for 80,000 students.

▶ For each student there are two types of scholarships available: Aid Merit Based and Aid Need Based.

Data Wrangling

For the Admissions Table:

▶ The Admissions file has 80,000 rows. There is a duplicate Student ID in the Admissions table. I removed it so that there are 79,999 distinct students.

▶ Some columns have NA values. I replaced them with an empty space ''.

Data Wrangling

For the Financial Aid Table:

▶   There are 160,000 rows because there are 2 types of financial aid for each student.
▶   I used a pivot table to put each type of financial aid into its own column so that there are 80,000 rows.
▶   There is a Student ID in this file that is absent from the Admissions table. Tom instructed me to ignore this ID.

	StudentID	EntryTermUse	AidType	value
0	1	Fall 2018	AidMeritBased	16400
1	1	Fall 2018	AidNeedBased	0
2	2	Fall 2020	AidMeritBased	0
3	2	Fall 2020	AidNeedBased	0
4	3	Fall 2019	AidMeritBased	0

StudentID	AidMeritBased	AidNeedBased
1	16400	0
2	0	0
3	0	0
4	0	0
5	13800	30000

Data Wrangling

▶ Admissions and Financial Aid files have the Student ID column in common.

▶ I merge both files on the Student ID with an inner join.

▶ The merged file has all the original columns of the Admissions table plus the two financial aid type columns from the Financial Aid table.

Student Applications per Entry Term

▶ I could not match the color of the legends between the two graphs because they were created using 2 distinct codes.

▶ Bar plot shows a general increase of student applications over time except for the Fall 2020 semester.

▶ Pie chart shows that among the 79,999 students from the dataset, Fall 2021 has the highest percentage of student applications of almost 24%. Fall 2017 has the lowest percentage with 17.1%.
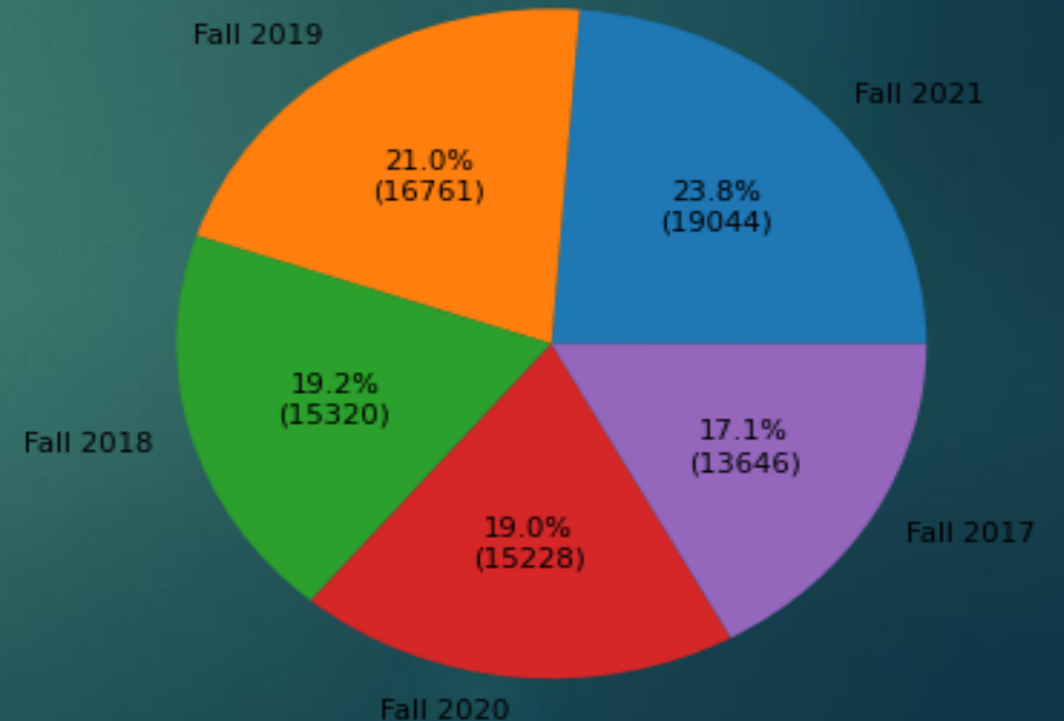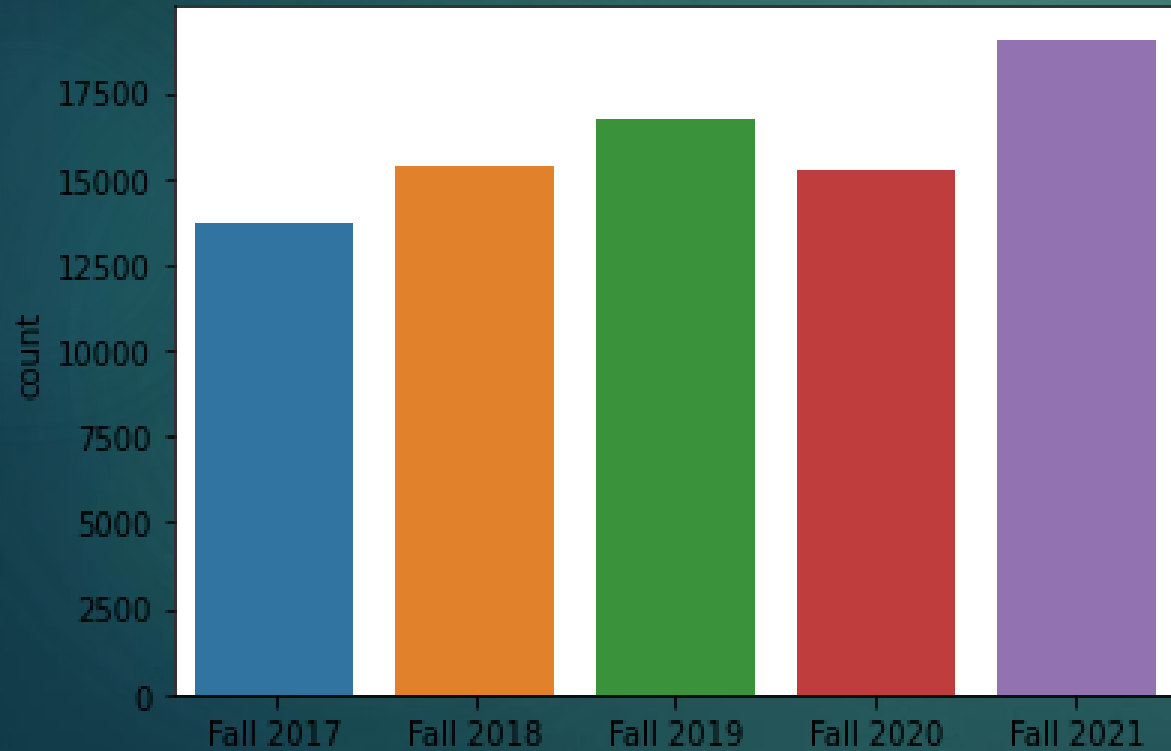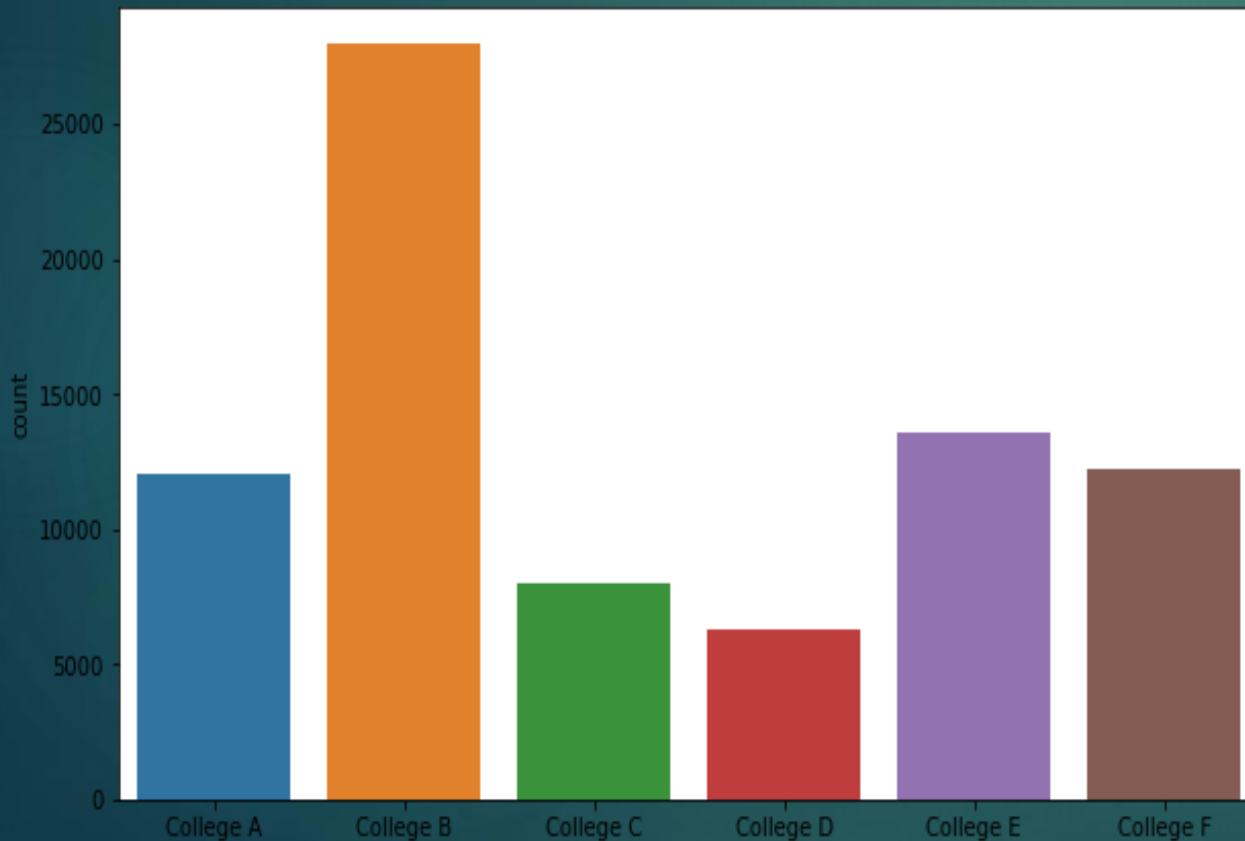
Student Applications per College

▶ I could not match the color of the legends between the two graphs because they were created using 2 distinct codes.

▶ Both graphs show that College B has by far the highest number of student applications. College B by itself represents over the third of all student applications.
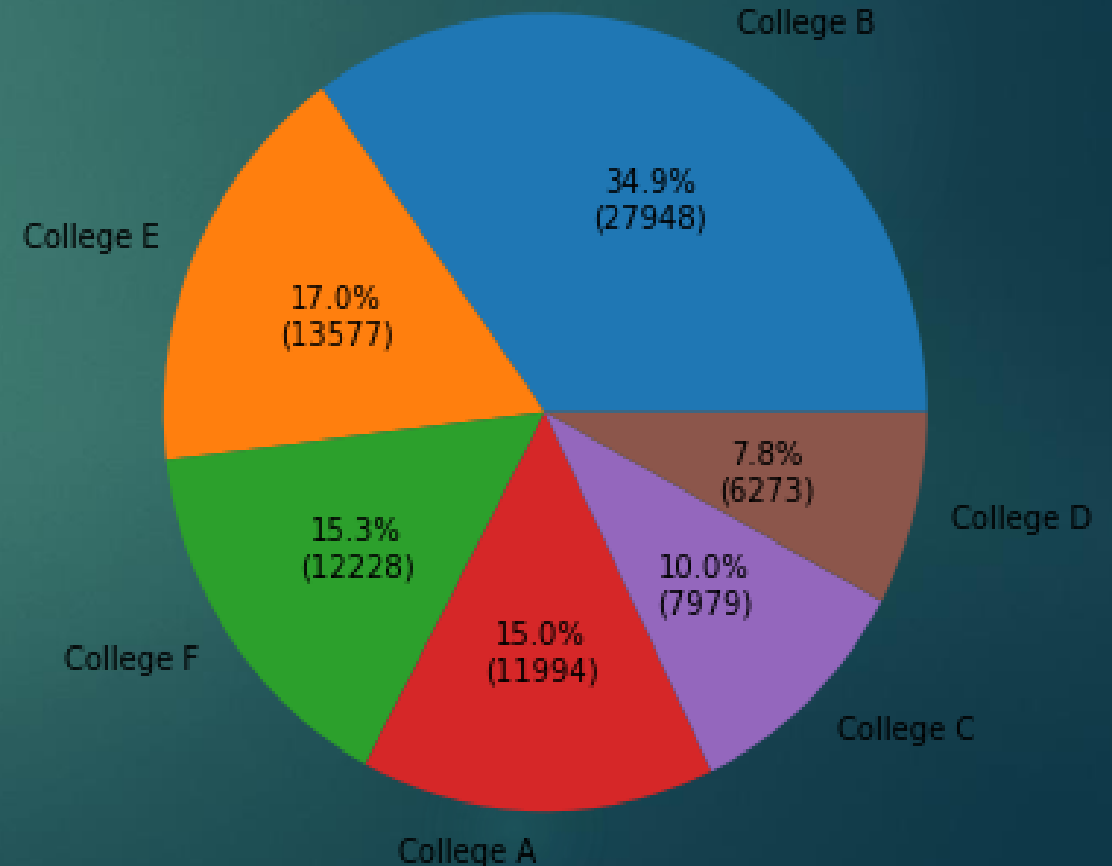
Gender Distribution

► This pie chart shows the Gender Distribution among the student applicants. About 55% are male and the remaining 45% female. The distribution is almost evenly spread.
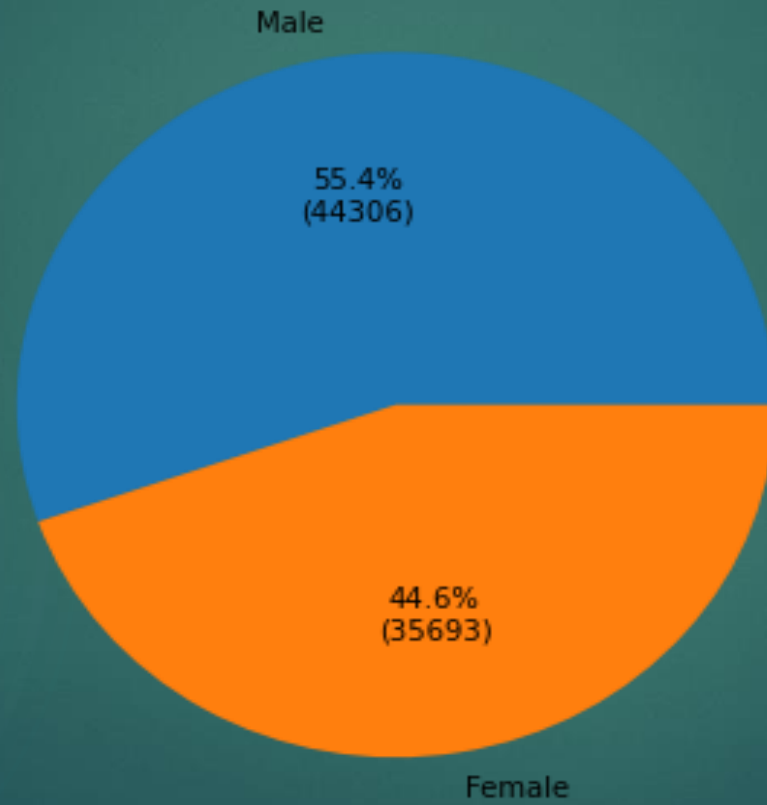
Rural-Urban Continuum Codes Distribution

▶ This pie chart shows the Rural-Urban Continuum Codes (RUCC) Distribution among the student applicants. Most of the students live in a metropolitan county where there is more than one million people.

Metro - 1 million +

84.8%
(67879)

3.0%
(2436)

Metro - Less than 250k

12.1%
(9684)

Metro - 250k to 1 million
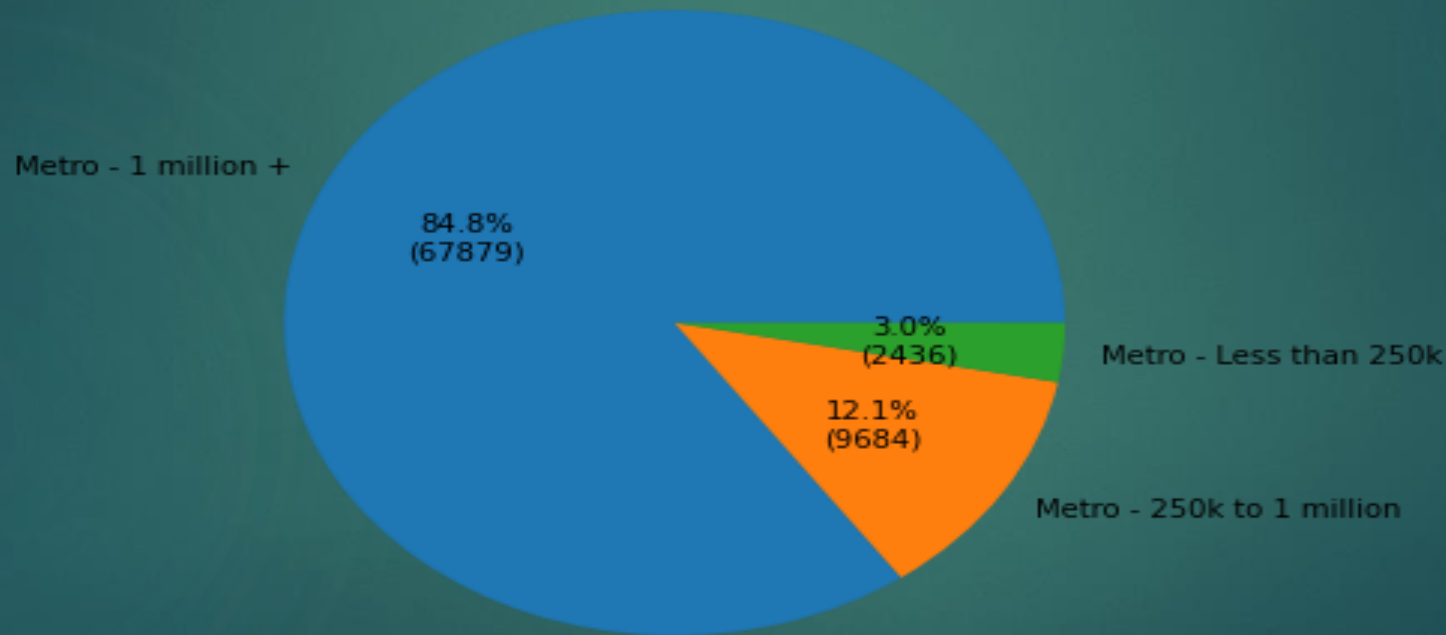
Aid Forms Application Status Distribution

▶ This pie chart shows the Aid Forms Application Status Distribution among the student applicants. Nearly half of the students who applied did not successfully file the aid forms. This is surprising.

Admission Status per College

▶ The first crosstab shows the admission status per college in raw data. 0 is for students who were not admitted and 1 for those who were.

▶ The second crosstab shows the row percentage for each college specifically. The overall Admission percentage is about 59%.

▶ The third crosstab shows the column percentage for each admission status specifically. It shows that College B has by far the highest percentages of both of admission and non-admission as more students applied to this college. The Student Applications per College bar plot confirms this.

Admission Status	0	1	All
College			
College A	4285	7709	11994
College B	13231	14717	27948
College C	2269	5710	7979
College D	2433	3840	6273
College E	5278	8299	13577
College F	5610	6618	12228
All	33106	46893	79999

Admission Status	0	1
College		
College A	35.73	64.27
College B	47.34	52.66
College C	28.44	71.56
College D	38.79	61.21
College E	38.87	61.13
College F	45.88	54.12
All	41.38	58.62

Admission Status	0	1	All
College			
College A	12.94	16.44	14.99
College B	39.97	31.38	34.94
College C	6.85	12.18	9.97
College D	7.35	8.19	7.84
College E	15.94	17.70	16.97
College F	16.95	14.11	15.29

Enrollment Status per College

► The first crosstab shows the enrollment status per college in raw data. 0 is for 2 groups of students: non-admitted students and admitted student who chose not to enroll. 1 refers to admitted students who enrolled.

► The second crosstab shows the row percentage for each college specifically. The overall enrollment percentage is about 12%.

► The third crosstab shows the column percentage for each enrollment status specifically. It shows that among the students who enrolled, more did so at College B and College E.

Enrollment Status	0	1	All
College			
College A	10581	1413	11994
College B	24661	3287	27948
College C	7139	840	7979
College D	5242	1031	6273
College E	11162	2415	13577
College F	11308	920	12228
All	70093	9906	79999

Enrollment Status	0	1
College		
College A	88.22	11.78
College B	88.24	11.76
College C	89.47	10.53
College D	83.56	16.44
College E	82.21	17.79
College F	92.48	7.52
All	87.62	12.38

Enrollment Status	0	1	All
College			
College A	15.10	14.26	14.99
College B	35.18	33.18	34.94
College C	10.19	8.48	9.97
College D	7.48	10.41	7.84
College E	15.92	24.38	16.97
College F	16.13	9.29	15.29

Admission Status per Gender

▶ The first crosstab shows the admission status per gender in raw data. 0 is for students who were not admitted and 1 for those who were.

▶ The second crosstab shows the row percentage for each gender specifically. There is no significant difference between the two.

▶ The third crosstab shows the column percentage for each admission status specifically. It shows that among those who were admitted, more students were men. It may be because more male students applied to Drexel University.

Admission Status	0	1	All
Gender			
Female	14445	21248	35693
Male	18661	25645	44306
All	33106	46893	79999

Admission Status	0	1
Gender		
Female	40.47	59.53
Male	42.12	57.88
All	41.38	58.62

Admission Status	0	1	All
Gender			
Female	43.63	45.31	44.62
Male	56.37	54.69	55.38

Enrollment Status per Gender

▶ The first crosstab shows the enrollment status per gender in raw data. 0 is for 2 groups of students: non-admitted students and admitted student who chose not to enroll. 1 refers to admitted students who enrolled.

▶ The second crosstab shows the row percentage for each gender specifically. The overall enrollment percentage is about 12%.

▶ The third crosstab shows the column percentage for each enrollment status specifically. It shows that among the students who enrolled, enrollment within male and female students is almost evenly spread.

Enrollment Status	0	1	All
Gender			
Female	31043	4650	35693
Male	39050	5256	44306
All	70093	9906	79999

Enrollment Status	0	1
Gender		
Female	86.97	13.03
Male	88.14	11.86
All	87.62	12.38

Enrollment Status	0	1	All
Gender			
Female	44.29	46.94	44.62
Male	55.71	53.06	55.38

Enrollment per Admission

▶ The first crosstab shows the enrollment status per admission status in raw data.

▶ The second crosstab shows enrollment per admission rate = enrollment/admission. Out of the 46,893 students who were admitted, only 21.12% or 9,906 students decided to enroll, which is less than the quarter of all admitted students.

▶ The third crosstab shows the column percentage for each enrollment status specifically. It confirms that that the 9,906 students who enrolled represent 100% of the admitted students community. This is the case because only admitted students are actually eligible for enrollment.

Enrollment Status	0	1	All
Admission Status			
0	33106	0	33106
1	36987	9906	46893
All	70093	9906	79999

Enrollment Status	0	1
Admission Status		
0	100.00	0.00
1	78.88	21.12
All	87.62	12.38

Enrollment Status	0	1	All
Admission Status			
0	47.23	0.0	41.38
1	52.77	100.0	58.62

High School GPA per Gender per College

▶ The overall average High School GPA across gender does not show a significant difference.

▶ The overall average High School GPA across college does not show a significant difference except for College C which is slightly lower than the 5 others.

▶ Within each college, the average High School GPA does not reveal any large variation between male and female students.

Gender College	Female	Male	All
College A	3.30	3.29	3.30
College B	3.31	3.31	3.31
College C	3.27	3.26	3.26
College D	3.35	3.33	3.34
College E	3.31	3.31	3.31
College F	3.29	3.30	3.30
All	3.31	3.30	3.30

High School GPA per Admission per College

▶ The difference in average High School GPA per college between admitted students and those who were not does not reveal a large variation in the data.

▶ The average High School GPA per college among admitted students is around 3.3 or higher. Among the non-admitted students, the GPA is always below 3.3.

Admission Status	0	1	All
College			
College A	3.24	3.32	3.30
College B	3.28	3.34	3.31
College C	3.19	3.29	3.26
College D	3.29	3.37	3.34
College E	3.26	3.34	3.31
College F	3.26	3.32	3.30
All	3.26	3.33	3.30

High School GPA per Admission per Entry Term

▶ The difference in average High School GPA per entry term between admitted students and those who were not does not reveal a large variation in the data.

▶ The average High School GPA per entry term among admitted students is around 3.33 or higher. Among the non-admitted students, the GPA is always below 3.3.

Entry Term Admission Status	Fall 2017	Fall 2018	Fall 2019	Fall 2020	Fall 2021	All
0	3.26	3.27	3.26	3.27	3.26	3.26
1	3.34	3.34	3.33	3.33	3.33	3.33
All	3.31	3.31	3.30	3.31	3.30	3.30

Estimated Family Contribution per College per Admission

▶ Looking at each college, the average estimated family contribution between admitted and non-admitted students is not drastically different.

▶ Admitted students from College D present an average estimated family contribution that is much lower than the average.

▶ Admitted students from Colleges A and C present an average estimated family contribution that is much higher than the average.

Admission Status College	0	1	All
College A	30445.0	29757.0	30003.0
College B	20602.0	21123.0	20876.0
College C	37127.0	34751.0	35427.0
College D	18514.0	18682.0	18617.0
College E	23847.0	25325.0	24751.0
College F	21461.0	21101.0	21266.0
All	23518.0	24742.0	24236.0

Estimated Family Contribution per Gender per Entry Term

▶ Across the 5 entry terms, there is no significant difference for the average estimated family contribution between male and female students.

▶ The average estimated average family contribution of female students is more or less constant over time, while that of male students gradually decreases over time.

Entry Term Gender	Fall 2017	Fall 2018	Fall 2019	Fall 2020	Fall 2021	All
Female	23933.0	24627.0	23952.0	25270.0	24503.0	24463.0
Male	26749.0	25628.0	23285.0	22511.0	22765.0	24053.0
All	25497.0	25178.0	23580.0	23752.0	23539.0	24236.0

Estimated Family Contribution per Admission per Enrollment

▶ Within the group of admitted students, the average estimated family contribution between students who enrolled and those who did not is not statistically significant.

▶ The NaN confirms that only admitted students are eligible for enrollment.

Enrollment Status	0	1	All
Admission Status			
0	23518.0	NaN	23518.0
1	24846.0	24354.0	24742.0
All	24219.0	24354.0	24236.0

Estimated Financial Need per College per Admission

▶ Looking at each college, the average estimated financial need between admitted and non-admitted students is not drastically different.

▶ The total cost of attendance (fees, room, board, personal expenses) is estimated at $70,000. This crosstab shows that the admitted students average need for financial support is in the $50,000-$55,000 range.

▶ These admitted students can consider taking loans along with the financial aid they receive. Loans data was not provided in the data I received to make an analysis on it.

Admission Status	0	1	All
College			
College A	51945.0	51947.0	51946.0
College B	54106.0	53783.0	53936.0
College C	50727.0	51288.0	51128.0
College D	54646.0	54703.0	54681.0
College E	52765.0	52652.0	52696.0
College F	53646.0	53940.0	53805.0
All	53343.0	53075.0	53186.0

Estimated Financial Need per Gender per Entry Term

▶ Across the 5 entry terms, there is no significant difference for the average financial need family between male and female students.

▶ The total cost of attendance (fees, room, board, personal expenses) is estimated at $70,000. This crosstab shows that the admitted students average need for financial support is in the $52,000-$54,000 range.

▶ These students can consider taking loans along with the financial aid they receive. Loans data was not provided in the data I received to make an analysis on it.

Entry Term	Fall 2017	Fall 2018	Fall 2019	Fall 2020	Fall 2021	All
Gender						
Female	53016.0	53094.0	53209.0	52967.0	52969.0	53050.0
Male	53214.0	53032.0	53304.0	53368.0	53496.0	53295.0
All	53126.0	53060.0	53262.0	53187.0	53261.0	53186.0

Estimated Financial Need per Admission per Enrollment

▶ Within the group of admitted students, the average estimated financial need between students who enrolled and those who did not is not statistically significant.

▶ The total cost of attendance (fees, room, board, personal expenses) is estimated at $70,000. This crosstab shows that the admitted students average need for financial support is in the $52,000-$54,000 range.

▶ These enrolled students can consider taking loans along with the financial aid they receive. Loans data was not provided in the data I received to make an analysis on it.

Enrollment Status	0	1	All
Admission Status			
0	53343.0	NaN	53343.0
1	53132.0	52860.0	53075.0
All	53232.0	52860.0	53186.0

Estimated Financial Aid per Gender per Entry Term

- The first crosstab is the average amount of Aid Merit Based financial aid. The second one is for the average amount of Aid Need Based financial aid. These are the 2 types of Financial Aid available for students.

- In each crosstab, across the 5 entry terms there is not a huge difference of average financial aid amount between male and female students.

- The average Aid Need Based amount has been historically higher than the average Aid Merit Based one.

Entry Term Gender	Fall 2017	Fall 2018	Fall 2019	Fall 2020	Fall 2021	All
Female	5532.0	5490.0	5555.0	5459.0	5387.0	5480.0
Male	5319.0	5283.0	5251.0	5308.0	5194.0	5266.0
All	5414.0	5376.0	5385.0	5376.0	5280.0	5362.0

| Entry Term Gender	Fall 2017	Fall 2018	Fall 2019	Fall 2020	Fall 2021	All
Female	8413.0	8569.0	8817.0	8618.0	8553.0	8600.0
Male	8491.0	8358.0	8438.0	8696.0	8359.0	8462.0
All	8457.0	8453.0	8605.0	8661.0	8445.0	8523.0

Estimated Financial Aid per Admission per Enrollment

▶ The first crosstab is the average amount of Aid Merit Based financial aid. The second one is for the average amount of Aid Need Based financial aid.

▶ Within the group of admitted students for each type of financial aid, there is no an alarming difference between students who enrolled and those who did not.

▶ The average Aid Need Based amount per student has been historically higher than the average Aid Merit Based one.

Enrollment Status Admission Status	0	1	All
0	0.00	NaN	0.00
1	9076.01	9411.41	9146.86
All	4789.27	9411.41	5361.62

Enrollment Status Admission Status	0	1	All
0	0.00	NaN	0.00
1	14463.87	14827.05	14540.59
All	7632.36	14827.05	8523.25

Heatmap

▶ This is a heatmap for numerical variables in the dataset.

▶ There are important positive correlation as well as negative correlations.
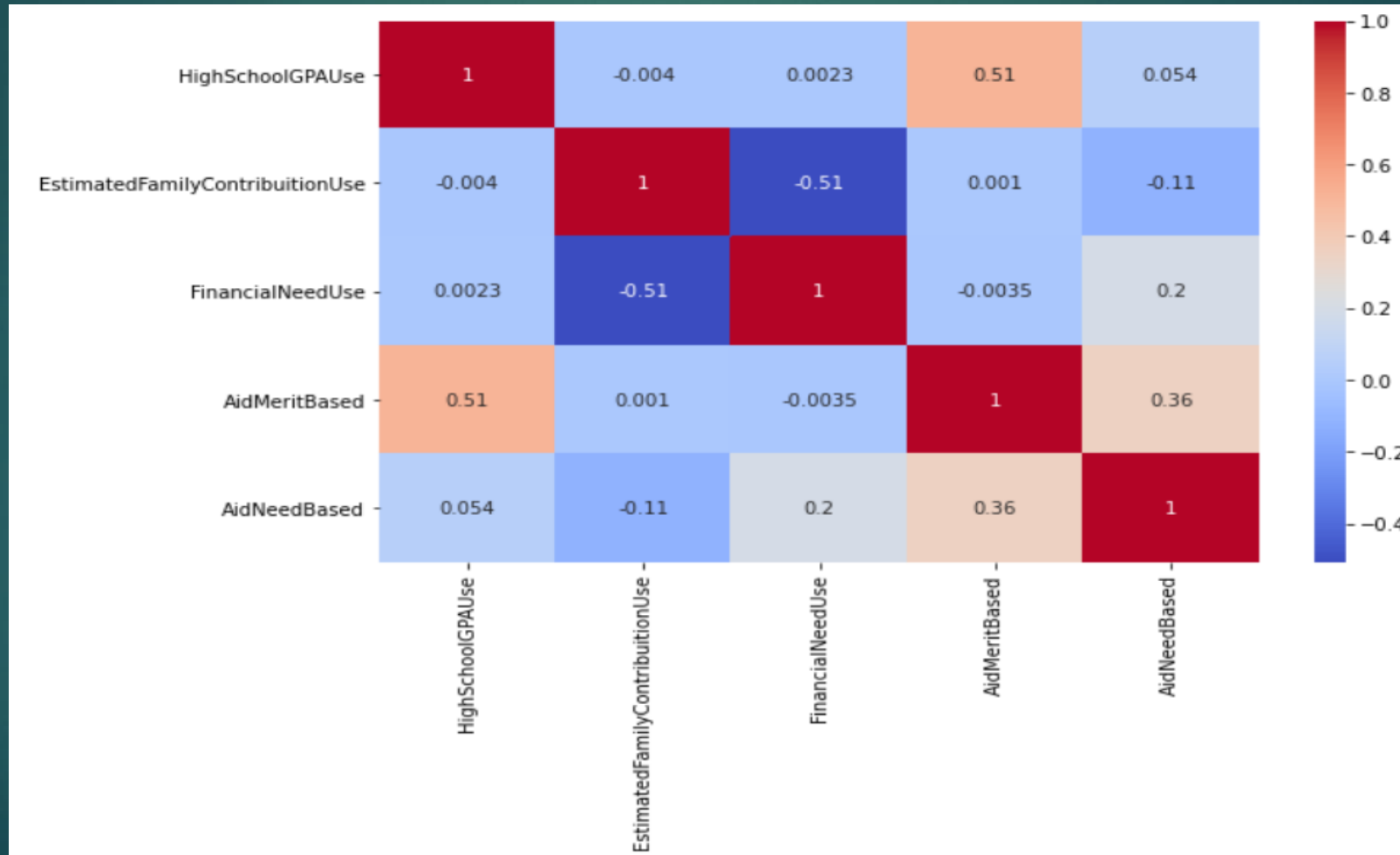
Conclusion

▶ This presentation presents an overview of the two provided datasets, the data wrangling process, and data visualization.

▶ The number of student applications has increased continuously for each entry term except for the Fall 2020 semester. The year 2020 has been deeply affected by the coronavirus pandemic. As many people lost their jobs and the financial situation was very hard, this may explain why a great number of students opted not to apply for college for the Fall. Although Drexel offered financial aid as in the other entry terms, students may have thought it was not enough to compensate for the difficult environment created by the pandemic.

▶ Students have applied at College B much more than any other College, which explains why College B has a higher number of enrolled students. It may be that College B is more known for its prestige and quality of majors that it attracts more students. Another reason is that students may tend to decline some majors more than others. This is a pure speculation.

▶ Within the group of admitted students, the ratio of enrollment/admission is about 21%. Some potential reasons why many students decline enrolling at Drexel University include a number of offers from other universities, the amount of financial aid available per student on average (about $14,000 - $16,000 for both types of financial aid available to students), and the average financial need in the $50,000 - $55,000.

Conclusion

- That being said, the loan data was not provided as to analyze how much loans students can have on average. But the dataset seems to imply that even if the loan data were provided, it would not be enough for many students to compensate for their average financial need.

- In terms of financial aid type and high school GPA, there is not a significant difference between male and female students.

- Most students who applied at Drexel University live in a Rural-Urban Continuum Codes (RUCC) where there is more than 1 million people.

- There is a strong positive correlation between High School GPA and Aid Merit Based type of financial aid.

- There is a strong negative correlation between estimated family contribution and financial need.

Areas of further analysis

▶ The Ethnicity data was not provided. This is a great area to study different admissions metrics for students and a way to see how diverse the student body is.

▶ Drexel University has a huge community of online students. The Admissions table does not specify the percentage of on-campus students and online students from the 79,999 students. It may be that different results may be seen from the online community of students.

▶ The 5 entry terms are all Fall semesters. There is no data for Spring semesters. It may be that most undergraduate students start during the Fall. That being said, analyzing Spring semesters data would also be interesting.