DS 4002 – Instructor: Loreto Alonzi
Due Date & Submission Format up to professor's discretion.

Individual Assignment

Preparatory Assignments: None

Why am I doing this? Students, schools, and others involved in the college admissions process know that it is an ever-changing landscape. Students choose their high school activities, and how they divide their time on college applications into various parts of the application. Colleges often college students' GPA, extracurriculars, and essays, but one unique variable is standardized tests. Testing has changed due to the COVID-19 pandemic, and it is important for students to be able to identify whether they need to take these tests anymore.

What am I going to do? You will be analyzing trends in admissions characteristics of universities from 2014-2023, looking specifically at the COVID-19 pandemic. From the provided data, the student will produce a final analysis report based on a validated logit regression model capable of predicting the use of admissions criteria with an AUC > 0.75 and identifying up to three admissions criteria that exhibit statistically significant ($p<0.05$) shifts in influence following the COVID-19 pandemic, if they exist. From this, the students should be able to properly identify which admissions criteria have shifted because of the Covid-19 pandemic at universities in the United States.

How will I know I have Succeeded? You will have succeeded when you meet the criteria laid out in the below rubric:

Spec Category	Task	Criteria for "Meets Spec"
I. Project Goal	The student will replicate the key preprocessing, modeling, and evaluation steps outlined in the case study's technical documentation to determine if admissions criteria importance shifted post-2020.	• The student has successfully completed the full analysis pipeline, including data merging, cleaning, feature engineering, and logit regression analysis.
II. Data & Preprocessing	The student must merge the yearly IPEDS ADM and HD files (2014-2023). They must clean the data, handle missing values, and engineer the PostCOVID binary feature (1 for 2020 and later, 0 otherwise).	• Data is successfully merged and cleaned according to the provided script. • The PostCOVID feature and binary criterion indicators are correctly engineered. • The unit of analysis is correctly maintained as the institution-year.
III. Modeling & Analysis	The student will run a series of logit regression models for various admissions criteria (e.g., SAT, GPA, Essay) as the dependent variable. They must include institutional characteristics (CONTROL,	• Logit regression models are correctly estimated with clustered standard errors at the institutional level.

	REGION) and the PostCOVID variable as key independent variables.	• Coefficients for the PostCOVID variable are correctly interpreted in terms of odds or probability shifts. • Statistical significance is evaluated using a 5% level (alpha = 0.05).
IV. Evaluation & Presentation	The student must evaluate the models using AUC and report the pseudo-R^2. They must create at least one time-series plot and a coefficient plot to visualize the trends and the post-COVID effects.	• The final model has a reported AUC. • At least one admissions criterion with a statistically significant shift ($p < 0.05$) post-COVID is identified, if one exists. • High-quality visualizations clearly illustrate the shift in admissions criteria importance over the 2014-2023 period.

Acknowledgements: Special thanks to Jess Taggart from UVA CTE for coaching on making this rubric. This structure is pulled from Streifer & Palmer (2020). I also want to thank Gianluca Guadagni, who I learned this format of rubric from.