Capstone Project

AN ANALYSIS OF ONLINE ENROLLMENTS AT SMALL COLLEGES AND UNIVERSITIES IN THE UNITED STATES

Daniel Teodorescu

Mentored by: Branko Kovac

The Problem

- Anecdotal evidence shows that for small colleges and universities in the United States, adding online programs can boost enrollments and add new sources of revenues. Yet, this relationship has not been tested using large national datasets.

- Therefore, one goal of this project is to establish statistically whether the addition of online classes is associated with an increase in total enrollments at small colleges and universities, after controlling for relevant institutional characteristics.

The Problem

- Secondly, if the analysis shows that there is a significant relationship between total enrollment and the offering of online classes, it is hypothesized that the addition of new online programs or courses can also improve the enrollment at small Historically Black Colleges and Universities (HBCUs), making these institutions more attractive to non-traditional students.

- Therefore, a secondary goal of this project is to assess whether HBCUs are currently less likely than the rest of the higher education institutions in the U.S. to offer online classes or programs, after controlling for key institutional characteristics such as institution's size and location.

The Data Sets

- Fall 2017 Distance Enrollment Data contains 22,590 observations representing enrollment in distance education courses at colleges and universities in the U.S. There are multiple observations per institution, representing three levels: all students, undergraduate students only, and graduate students only.

- 2017 Institutional Characteristics Data contains variables that describe each higher education institution. There are 7,153 observations in the dataset, representing all higher education institutions in the U.S. and its territories.

Data – First Glance

- Fall 2017 Distance Enrollment – selected five variables:
  ◦ Total number of students enrolled
  ◦ Number of students enrolled exclusively in online courses
  ◦ Number of students enrolled in some online classes
  ◦ Number of students from out of state
  ◦ Number of international students

Data – First Glance

- Institutional Characteristics – selected nine variables:

HBCU	Indicates whether the institution is one of the Historically Black College or University (HBCU) institutions.
HOSPITAL	Indicates whether the institution has hospital.
MEDICAL	Indicates whether the institution offers a medical degree.
LOCALE	Locale identifies the geographic status of a school on an urban continuum ranging from "large city" to "rural."

GROFFER	Indicates whether the institution offers graduate programs
UGOFFER	Indicates whether the institution offers undergraduate programs
LANDGRNT	Indicates whether an institution is land grant
CONTROL	Indicates whether an institution is operated by publicly elected or appointed officials or by privately elected or appointed officials and derives its major source of funds from private sources.
HLOFFER	Highest degree offered

Data Wrangling

- Institutions with large enrollment in online programs (over 80%) are considered to be primarily online institutions. Therefore, they are excluded from analysis.

- The focus of analysis is small colleges and universities: enrollments between 500 and 5,000

- For profit institutions are excluded.

- The two datasets were merged by UNITID.

Data Wrangling

- Two new variables (dependent) were derived based on the variables in the first dataset:
  - perc100online - % students enrolled exclusively in online programs
  - perc_some_online - % students taking some of the classes online
- Where no values were reported for the two variables, it was assumed that online enrollment was 0.
- All categorical variables where reformatted from numeric to factors.

Exploratory Data Analysis (EDA)

Final data set (data_new2) contains 1,557 observations and 17 variables.

- EDA Approaches
  - Bar charts to compare median percentage of online enrollments by institution type
  - Scatterplots to examine relationships between continuous variables

EDA – Online Enrollments
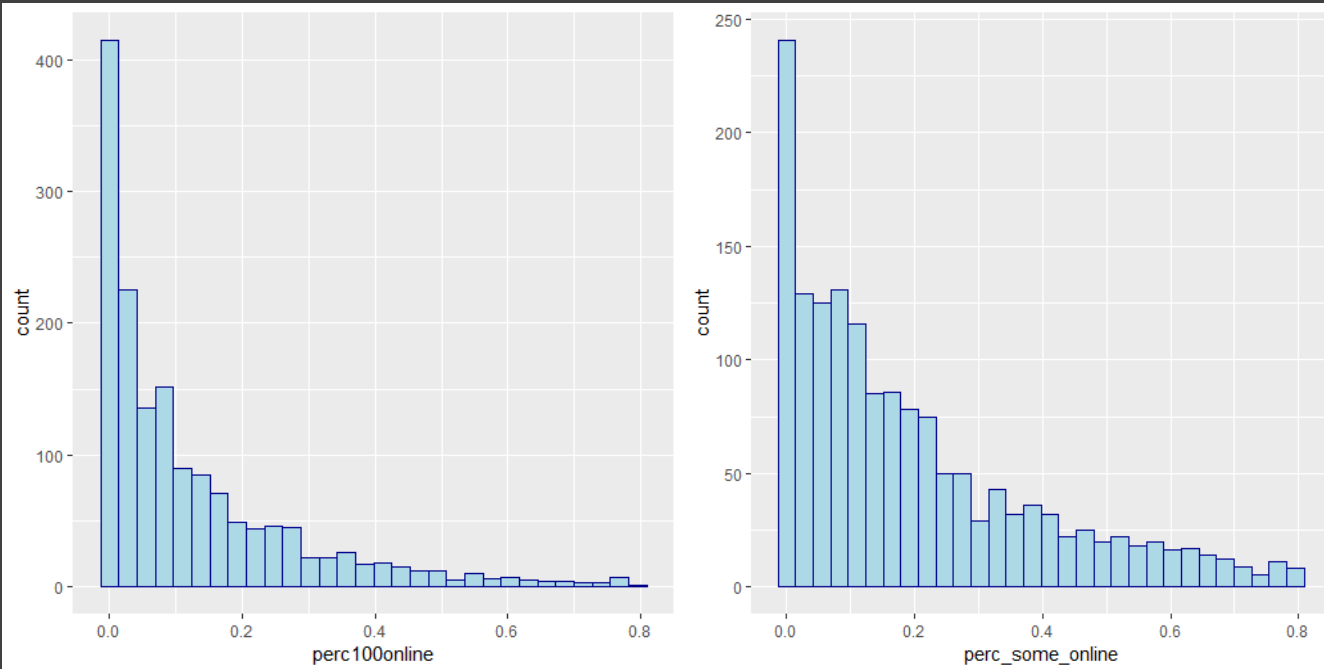
Variable	Min	Q1	Median	Mean	Q3	Max
perc100online (percent students enrolled exclusively in online programs)	0%	1%	7%	12%	18%	79%
perc_some_online (percent students taking some of their classes online)	0%	5%	13%	20%	29%	79%

- Overall, the median percentage of students enrolled exclusively in online programs across small institutions is 7%.

- Additionally, the median percentage of students taking a class online is 13%.

- Both variables are heavily skewed, with most institutions having little or no enrollment in online courses or online programs.
  - The lack of normality for these variables poses a limitation in the use of linear regression.

EDA– Online Enrollments by HBCU status
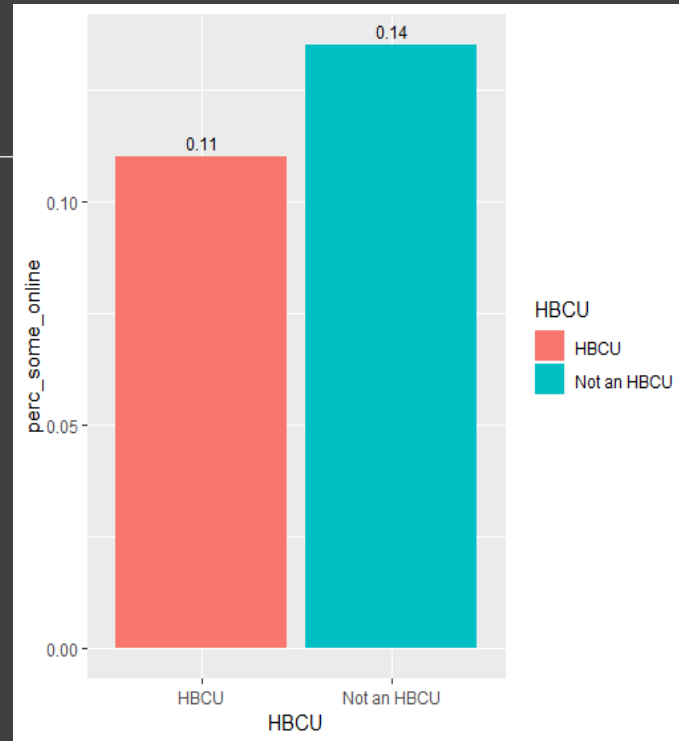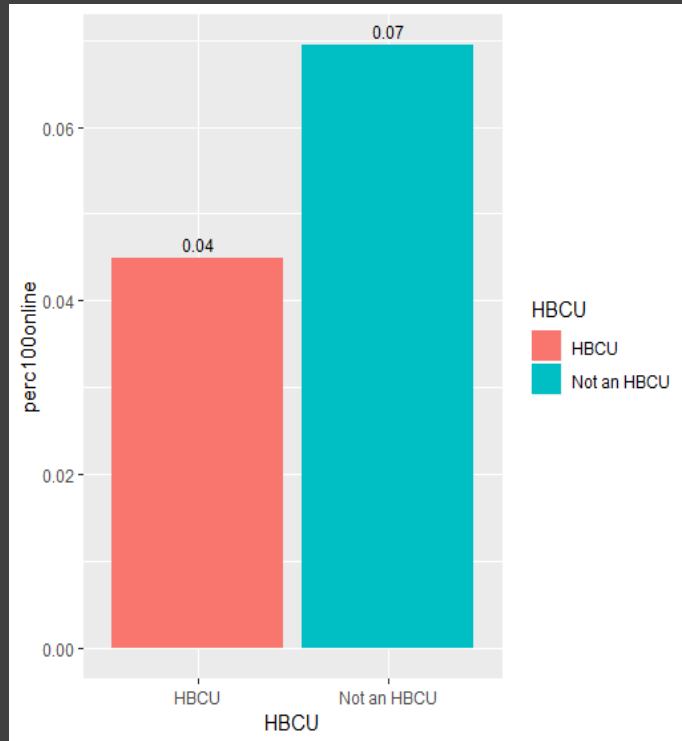
- On average, HBCSs enroll a lower percentage of students in online programs (4% vs. 7%).

- At the same time, the percentage of students taking some online classes is also lower at HBCUs than at non-HBCUs (11% vs. 14%).
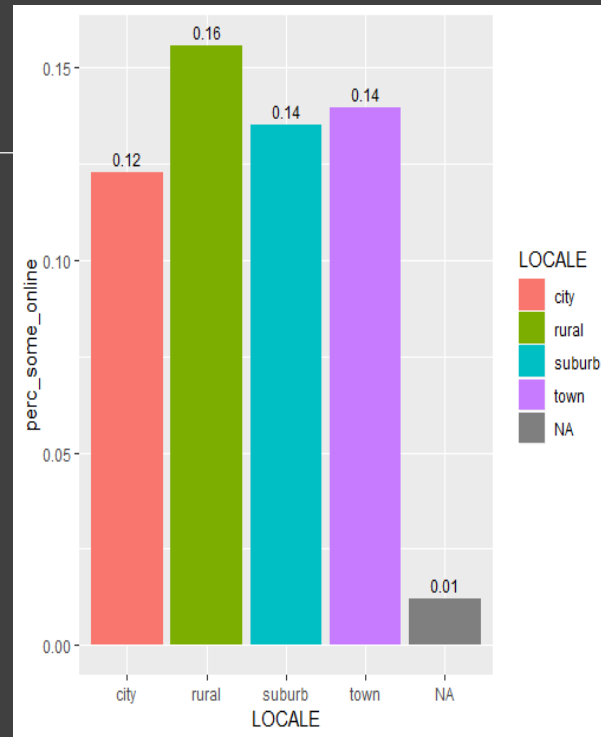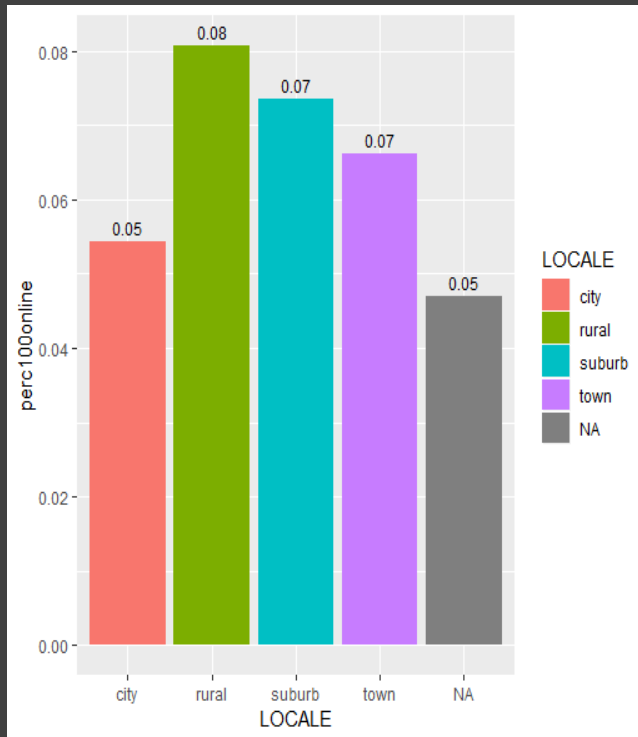
EDA– Online Enrollments by Land Grant status
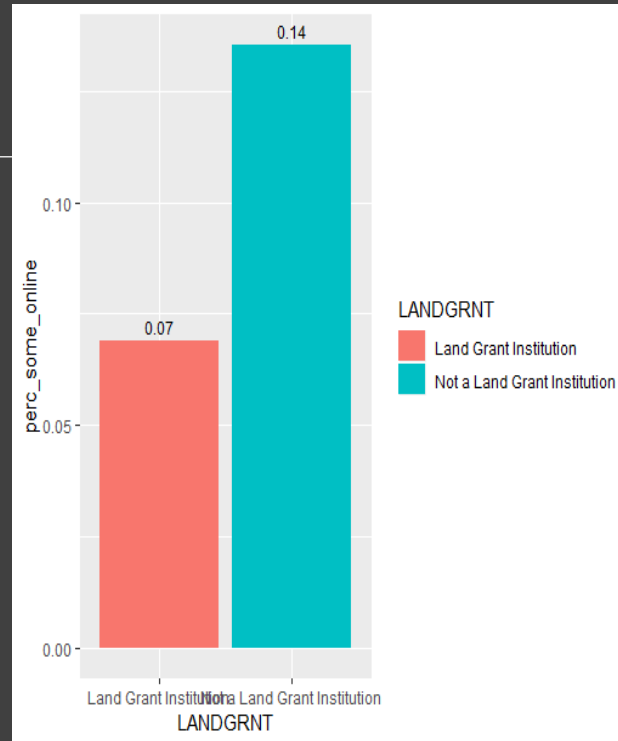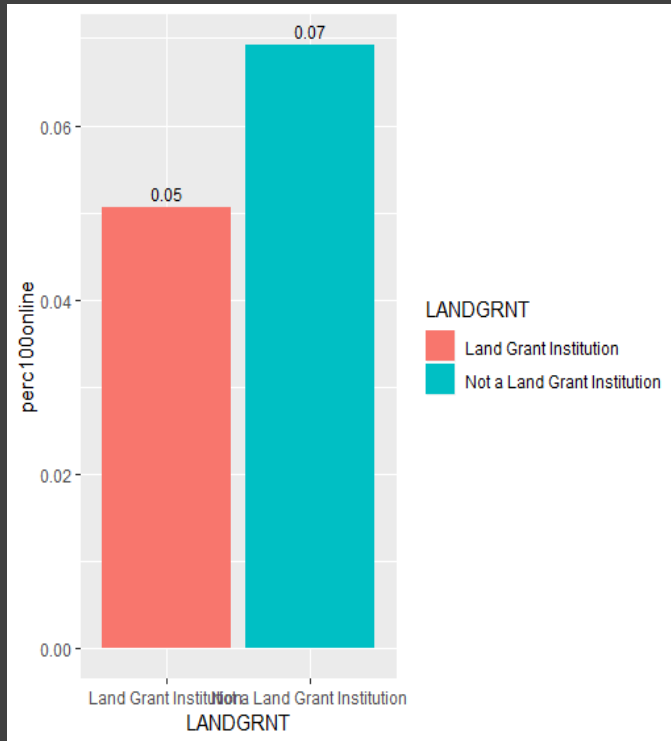
- Land grant institutions have a lower percentage of students enrolled in online degree programs than the rest of the institutions.

- They also have a lower percentage of students taking at least a class online.

EDA– Online Enrollments by CONTROL
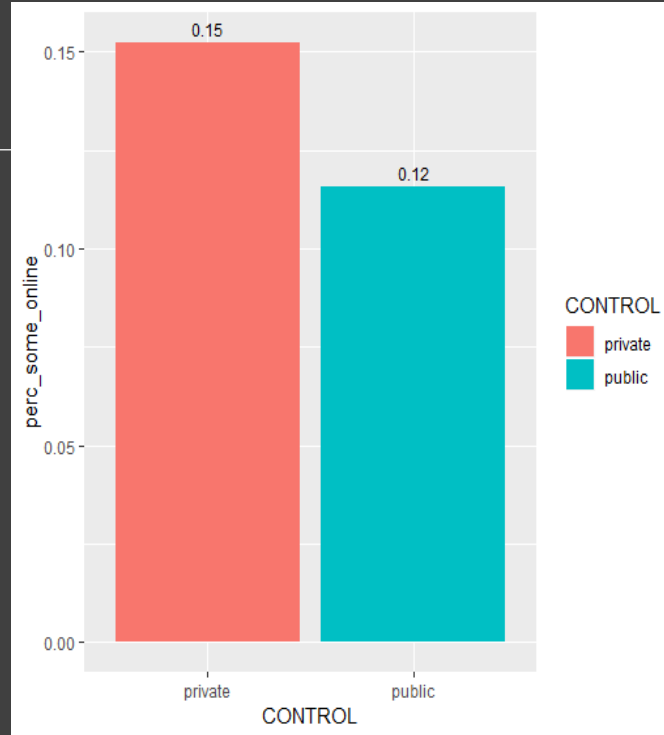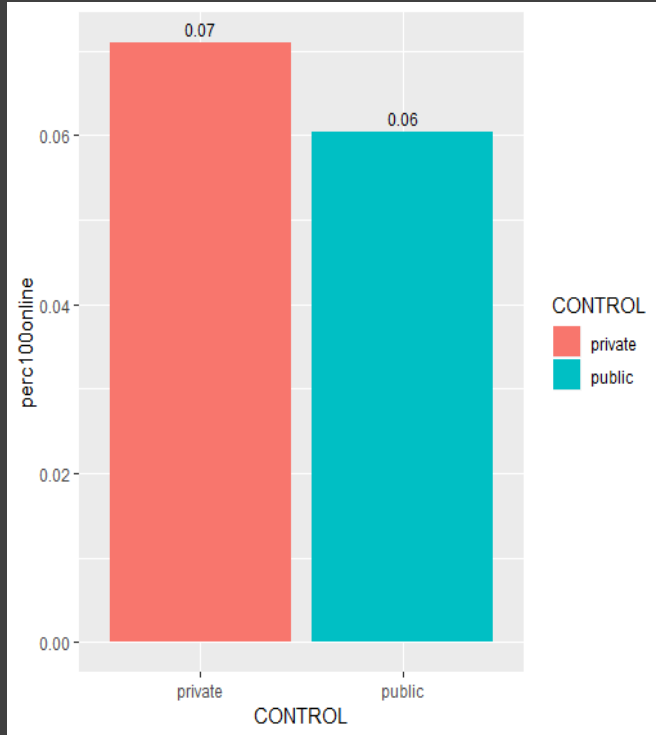
- Private institutions have a slightly higher median % of students enrolled in fully online programs than public institutions (7% vs. 6%).

- They also have a higher percentage of students taking at least one online course (15% vs. 12%).

EDA– Online Enrollments by LOCALE

- Higher education institutions that are located in rural areas tend to have a higher percentage of students enrolled in online degree programs than urban institutions.

- They also have a higher percentage of students taking at least a class online.

-

EDA– Online Enrollments by Highest Degree
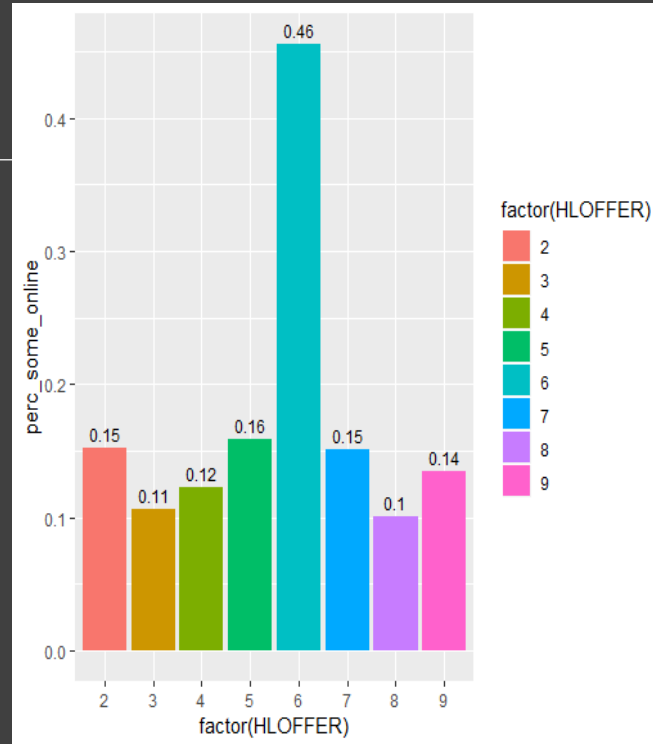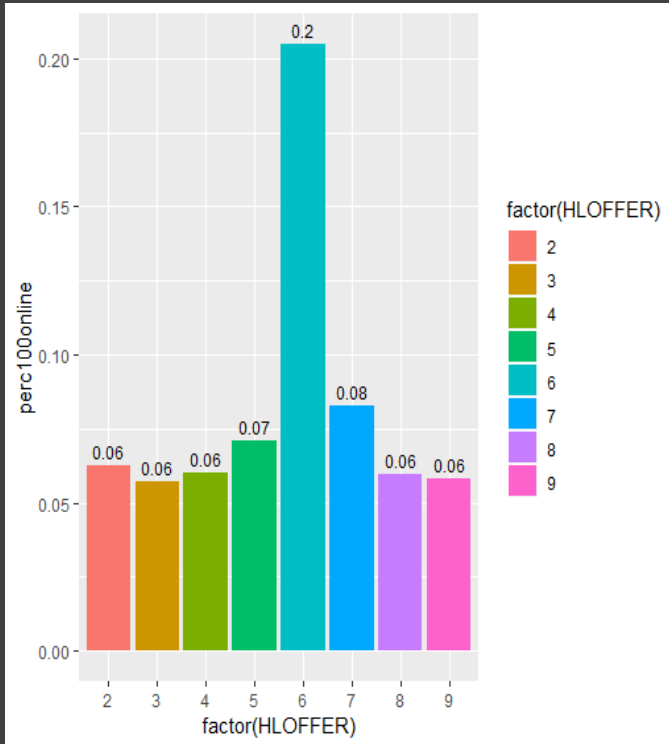
2 - Postsecondary award, certificate or diploma of at least one but less than two academic years
3 - Associate's degree
4 - Postsecondary award, certificate or diploma of at least two but less than four academic years
5 - Bachelor's degree
6 - Postbaccalaureate certificate
7 - Master's degree
8 - Post-master's certificate
9 - Doctor's degree

- Higher education institutions where the highest degree level offered is a Postbaccalaureate Certificate have a higher percentage of students enrolled in online degree programs than the rest of the institutions (20% vs. 6-8%).

- They also have a higher percentage of students taking at least a class online (46% vs 10-16%).

EDA– Online Enrollments by GROFFER and UGOFFER

- Institutions that offer graduate programs have about the same percentage of students enrolled exclusively in online programs/courses as institutions that have no graduate programs.

- Institutions that offer undergraduate programs have a higher percentage of students taking some of their classes online (19% vs. 13%).
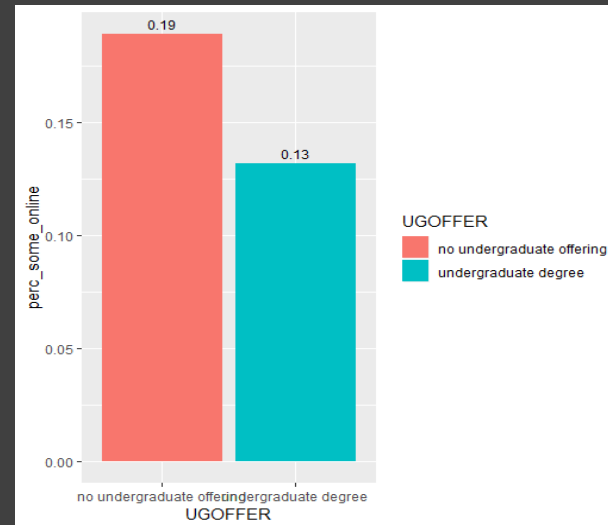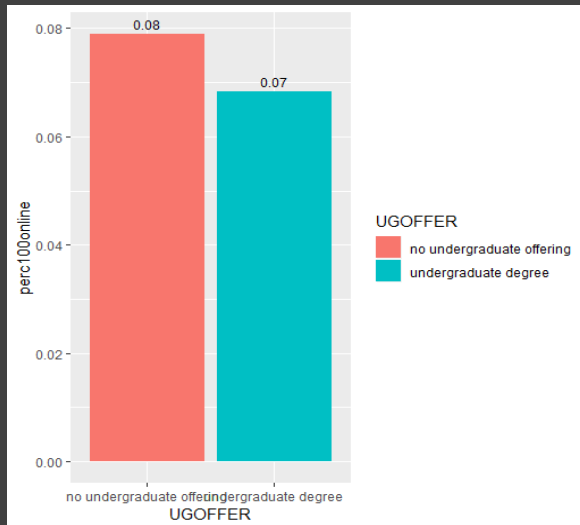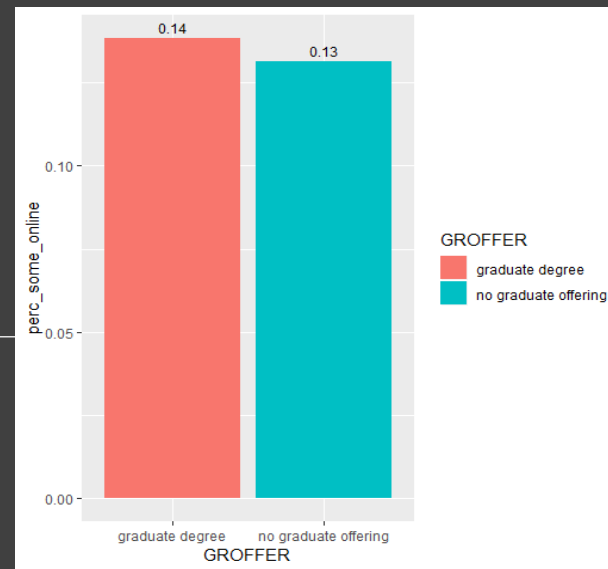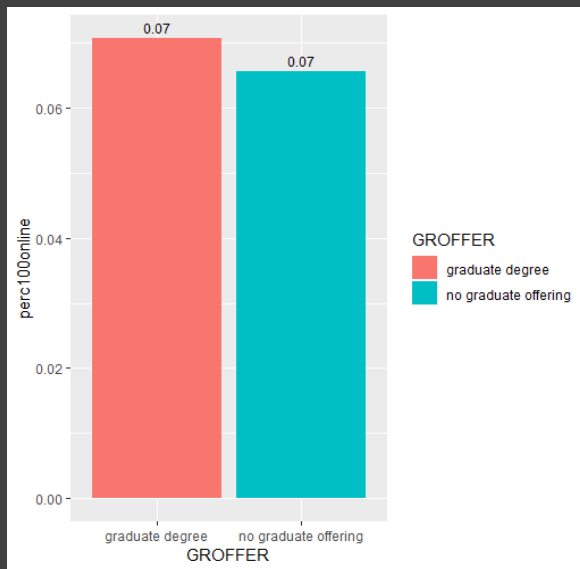
EDA– Online Enrollments by MEDICAL and HOSPITAL
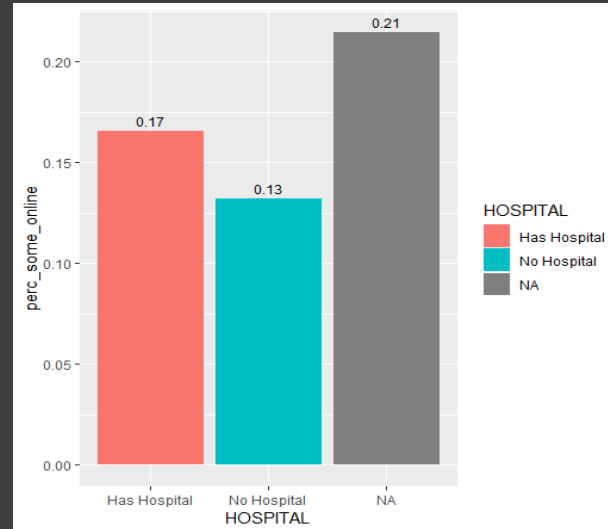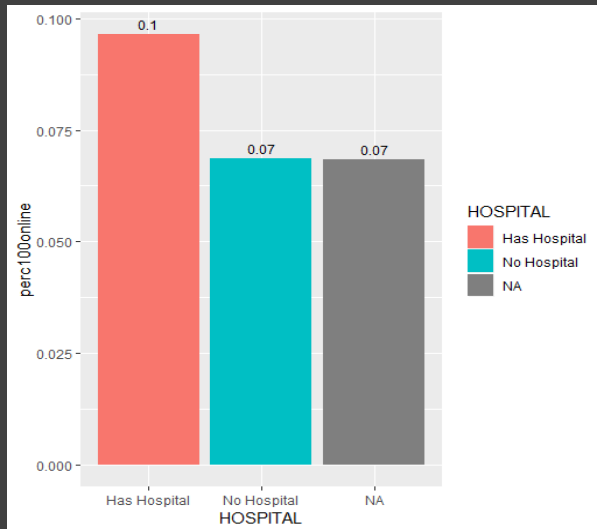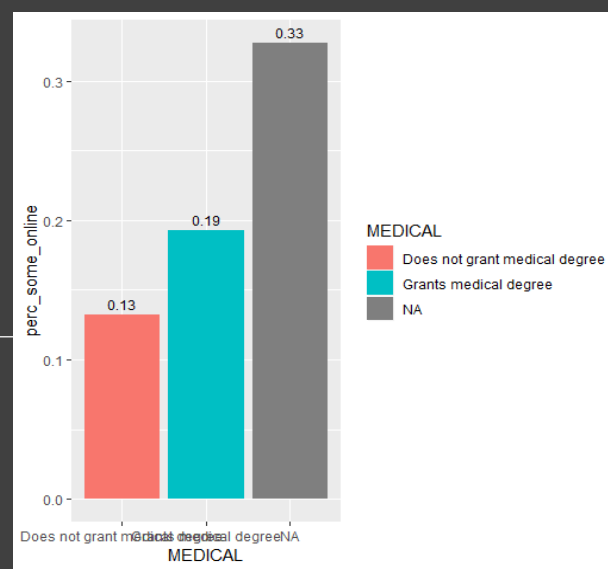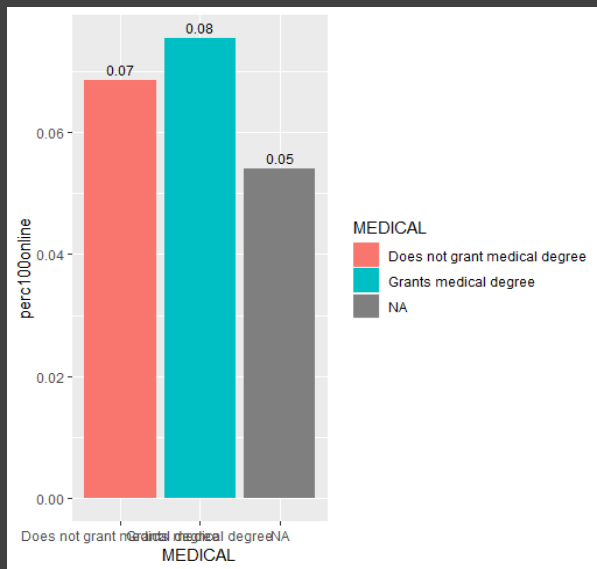
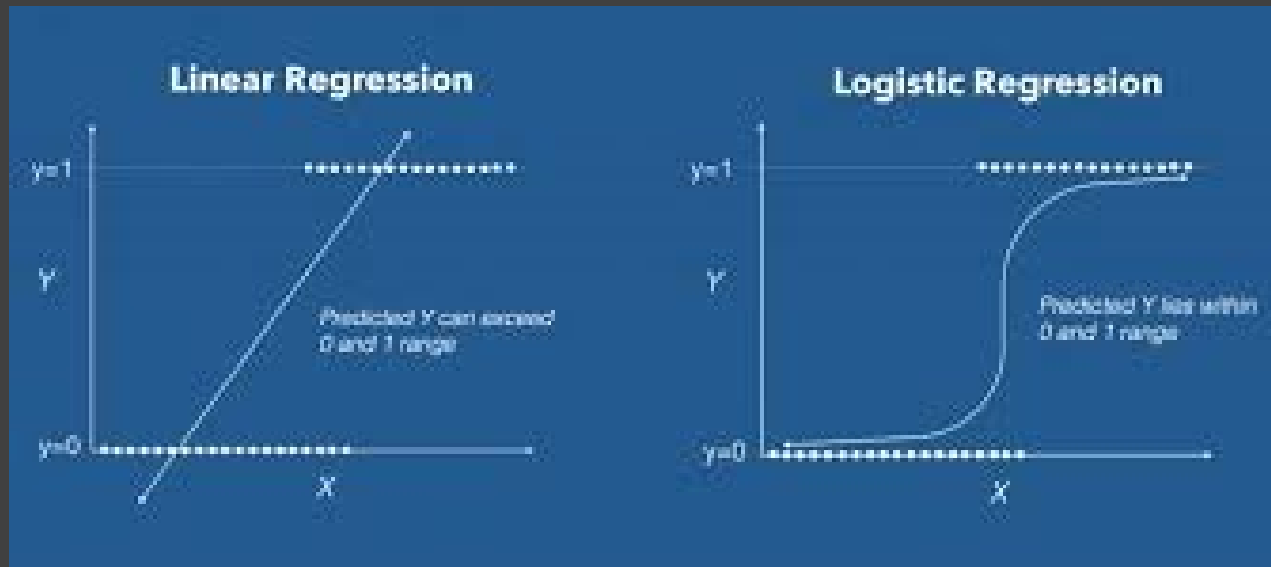- Institutions that offer medical programs or have a hospital have a higher % of students enrolled exclusively in online programs.
- They also have a higher % of students taking some of their classes online.

Machine Learning Approaches

Linear Regression — Predicted Y can exceed 0 and 1 range

Logistic Regression — Predicted Y lies within 0 and 1 range

Logistic Regression Model – predicts whether an institution offers online courses based on relevant characteristics.

Dependent Variable: some_online (institution has students enrolled in online classes)

Linear Regression Model – predicts total enrollment based on relevant institutional characteristics, including the offering of online classes.

Dependent Variable: EFDETOT (total enrollment)

Logistic Regression Results

- After controlling for key institutional characteristics such as size, location, and highest degree offered, non HBCUs are 20% more likely than HBCUs to have students enrolled in online classes.

- In addition, non land grant institutions are 6 times more likely than land grant institutions to have students enrolled in online classes.

- AUC = .56

Linear Regression Results

- After controlling for relevant institutional characteristics, offering online classes results in 322 additional students.

- Offering undergraduate programs adds on average 1185 students.

- Public institutions have on average 1328 more students than private institutions.

- Compared to institutions located in large urban areas, institutions located in towns enroll on average 407 fewer students.

- Institutions located in rural areas have 592 fewer students on average than urban institutions.

- R-Squared = .20;   min-max accuracy rate = 64%

Conclusions

- The logistic regression model revealed that, after controlling for size, location and other institutional characteristics, HBCUs are 20% less likely than other institutions to have students enrolled in online classes. A gap in the percentage of students taking online classes between the two groups of institutions was also found during the exploratory data analysis.

- The multiple linear regression model showed that, by offering online classes, higher education institutions can increase their enrollments on average by 322 students. This net gain is obtained after controlling for key factors such as institutional size, public vs. private, and location.

Recommendations

- Given the positive impact of offering online classes on total enrollment, small private institutions, especially those located in rural areas and small towns should plan to expand online programs and classes. They should also plan to offer post-bachelor's degrees since these programs will attract additional students.

- Given that the logistic regression results and the exploratory data analysis revealed a gap between HBCUs and non HBCUs in terms of online student enrollment, HBCUs should invest more in training faculty members on creating online courses and providing the necessary learning management systems infrastructure and technical support.

Further Work

- Explore whether students at colleges and universities on the West coast are more likely to take online classes than their peers in the rest of the country. The gmap function could be used to construct a map showing levels of perc_some_online by region.

- To increase R-Squared, add relevant variables such as total educational expenditures or student-faculty ratio, both available in other IPEDS data sets.