Decoding College Admission and Tuition Rates

Group 3: Bianca Linares (Leader), Connor Overbay, Maansi Taori

DS 4002

October 30, 2024

50% increase in college applications to southern schools

Less than ***30%*** rise of application to New England schools

Identifying trends in admission and tuition rates can inform future decision making

Project Details

01 **Hypothesis**

Admission and tuition rates for public and private universities in the Southern region are expected to differ significantly from those in the Northern region, based on comparative analyses using **Paired T-Tests** .

02 **Research Question**

How have university admission and tuition rates changed over the past 20 years, and how can these patterns and trends help predict future changes?

03 **Modeling Approach**

To analyze and forecast university admissions data over the past two decades, both the **Autoregressive Integrated Moving Average (ARIMA)** and **Seasonal Exponential Smoothing (ETS)** models were applied. ARIMA captures short-term fluctuations and non-seasonal patterns, while Seasonal ETS identifies long-term trends and recurring cycles. Together, these models provide a comprehensive approach for understanding past trends and predicting future shifts in admissions and tuitions.

Data Collection

Data Acquisition

Us Department of Education, College Scorecard

No ethical or licensing concerns

Data Cleaning

Removed NA values

Filtered to Northern and Southern schools

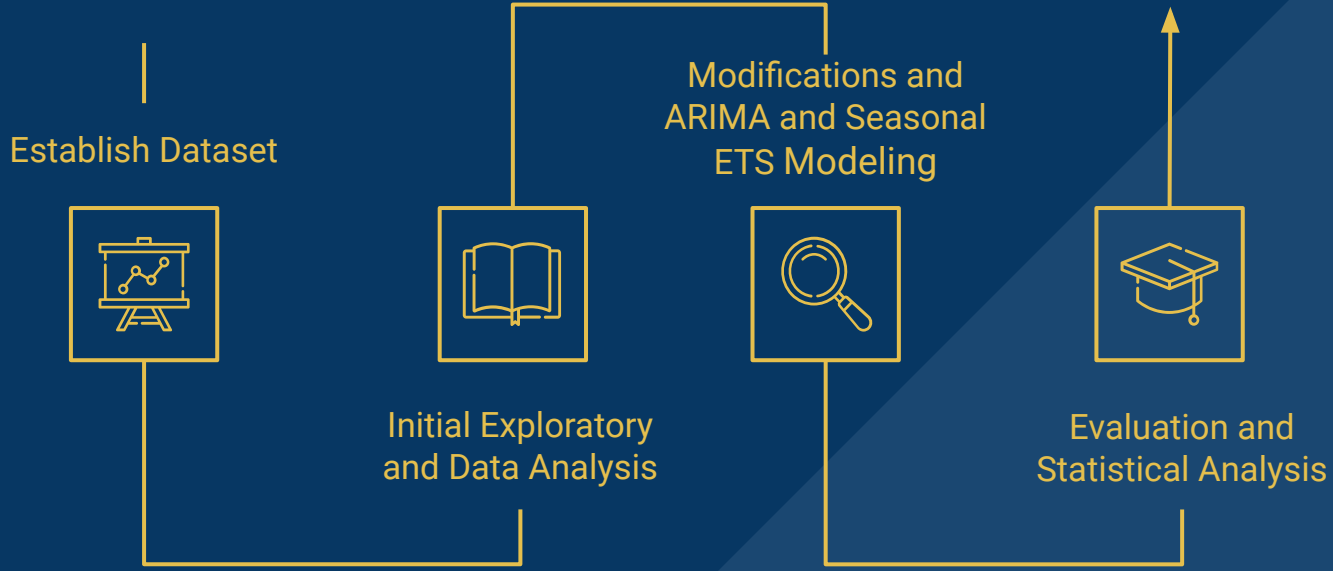
Data Selection

2001-2023 years, self-selected columns

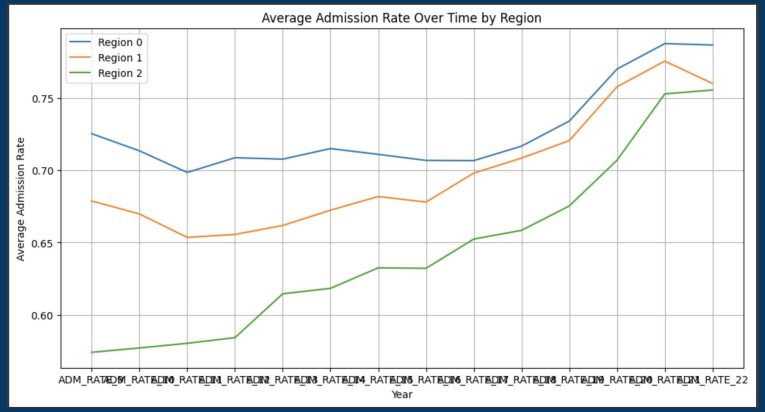
Data Dictionary

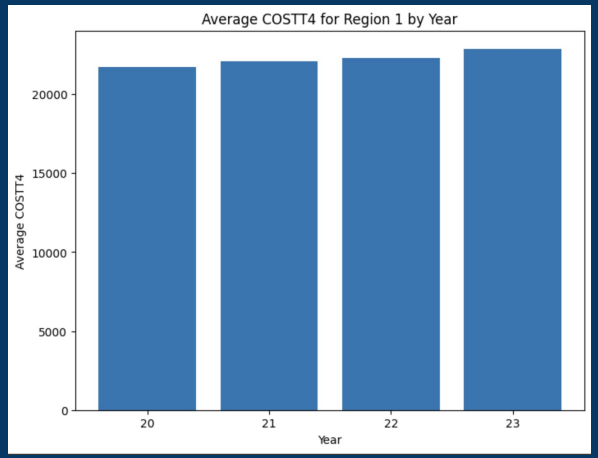
1. University Name
2. University State
3. Cost of Tuition
4. Admission Rate
5. Academic School Year

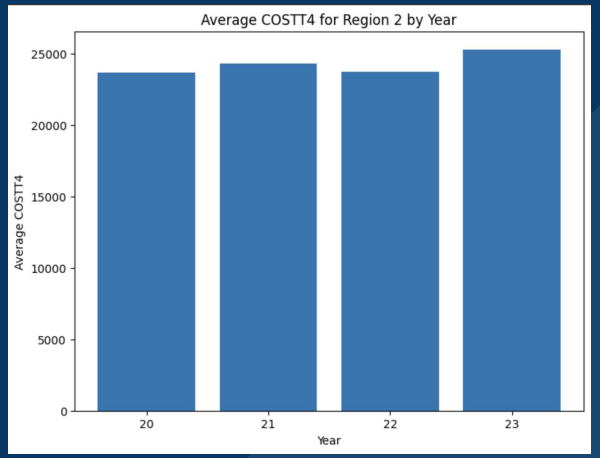
Pipeline


Establish Dataset
Initial Exploratory and Data Analysis
Modifications and ARIMA and Seasonal ETS Modeling
Evaluation and Statistical Analysis

Diagnostic and Validation Work







- After cleaning data, ran an analysis with average admission rates and tuition rates to understand general trends
  - What region of schools has higher admission rates?
  - Which schools tend to have higher costs of attendance?
- Groups were then used to analyze trends in tuition and admission rates

Challenges

Dataset was not perfect for ARIMA model

The ADF test showed the admissions data was non-stationary, making ARIMA unsuitable, so Seasonal ETS model was used for better handling of trends and seasonal patterns in the forecasts of admissions data

Scope of project remained unclear

The scope of the project was limited by data availability, restricting the external factors we could analyze. We focused on regional trends within the past 20 years, where complete data was available

Statistical Analysis Metrics

Changed ANOVA test with ADF testing to assess stationarity and Paired T-Tests to compare admissions and tuition rates between regions

Biases and Uncertainty

Dataset Source

Comes from a site dedicated to all institution data and research, making it more objective and less prone to bias

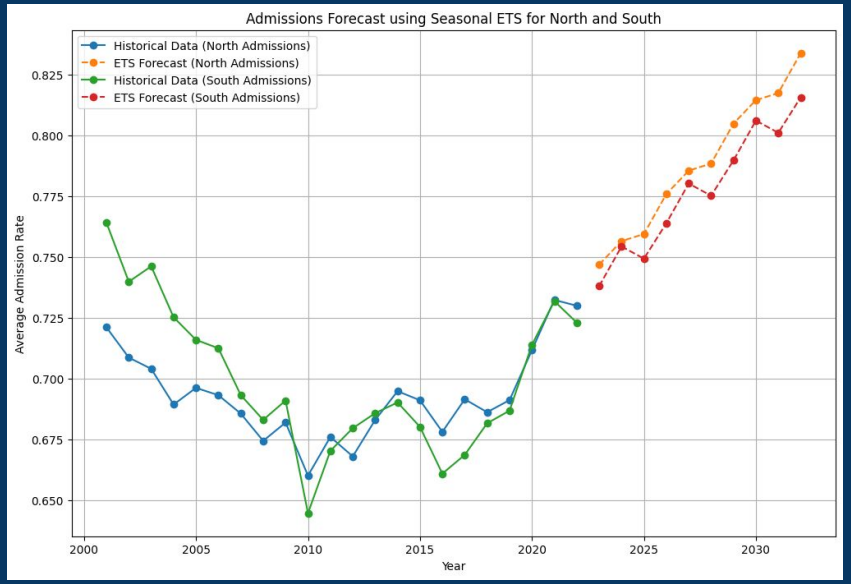
Institution Type

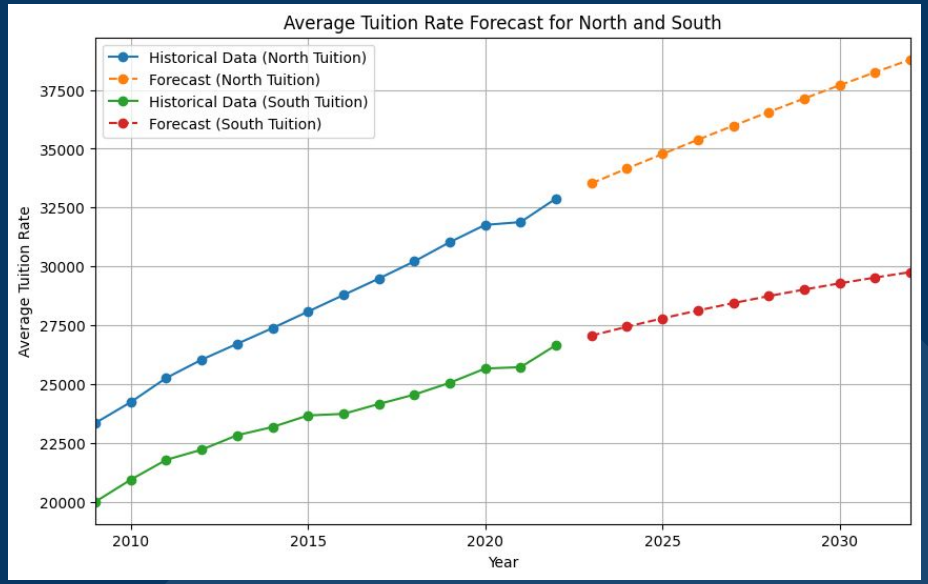
Looking at private and public institutions which have different processes for tuition and admission rates

Year

External factors such as recessions and COVID-19 could have affected rates

Results





Admissions Rates

t-statistic = -1.59, **p-value** = 0.13

No statistically significant difference in admissions rates between North and South regions

Tuition Rates

t-statistic = 15.89, **p-value** = 6.77e-10

Statistically significant difference in tuition rates between North and South regions, with one region having significantly higher tuition rates

Conclusion

Admissions Hypothesis: There would be a significant difference in admission rates between North and South regions.
→ Result: The paired t-test showed **no statistically significant difference** in admissions rates between the two regions (p-value = 0.13).

Tuition Hypothesis: There would be a significant difference in tuition rates between North and South regions.
→ Result: The paired t-test confirmed a **statistically significant difference** in tuition rates, with the South having higher tuition rates (p-value = 6.77e-10).

Modeling:
→ The ADF test showed that the admissions data was non-stationary, making it unsuitable for the ARIMA model.
→ The Seasonal ETS model was applied, and it successfully captured the trends and seasonal fluctuations in admissions data.

Future Direction

New lines of exploration:

- Observing rates by institution type
- Change in rates before and after COVID-19

Possible Improvements:

- Extending the types of regions considered
- Analyzing different external factors

New Questions:

- What causes the discrepancies between tuition rates?
- Are there measurable shifts in admission rates that occurred during and after certain events, compared to prior periods?

References

[1]Wall Street Journal, “Sorry Harvard, everyone wants to go to college in the south now,” *Wall Street Journal*, Sept. 27, 2024. https://www.wsj.com/us-news/education/sorry-harvard-everyone-wants-to-go-to-college-in-the-south-now-235d7934

[2]”US Department of Education College Scorecard” *collegescorecard.ed.gov.* https://collegescorecard.ed.gov/

Thank you! Questions?