STEM Pathways: From Data Insights to Higher Ed Protocols

Jasmine Huang, and Professor Irving Epstein

Department of the Provost, Brandeis University, Waltham, MA

Introduction

Background: Universities can use data to improve student retention, success, and institutional efficiency, yet many institutions struggle to collect and use data effectively.

Key Issues:

- Fragmented systems and siloed teams
- Inconsistent data standards
- Limited analytical expertise

These barriers make it hard to trace student pathways, especially in introductory STEM courses, which often determine whether students continue in STEM fields.

Hypothesis: By standardizing data protocols and automating data collection and analysis, universities can improve their data analysis to uncover academic bottlenecks and make faster, data-driven decisions that support student success.

Methods

Data Collection: Partnered with Brandeis University's Office of the Provost to access STEM course data on enrollment, performance, and retention, focusing on student pathways from entry to graduation or dropout.

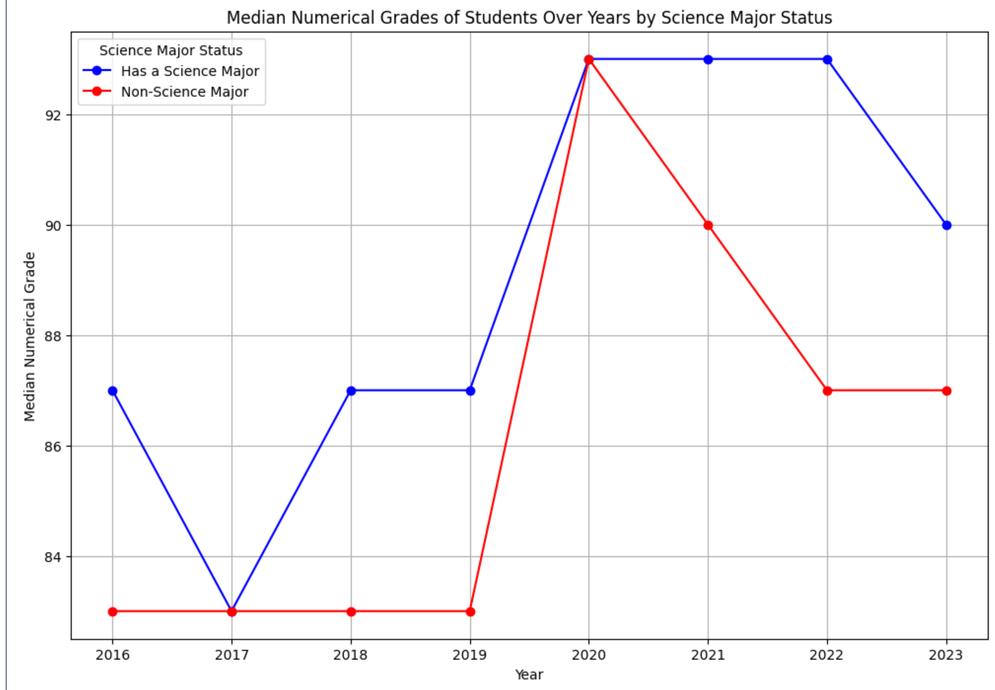
Data Preparation: Used Python (Pandas, NumPy) for cleaning, merging, and transforming datasets. Standardized column names and formats to resolve inconsistencies.

Analysis: Calculated retention rates, grade distributions, and course transition patterns. Examined links between early course performance and STEM persistence.

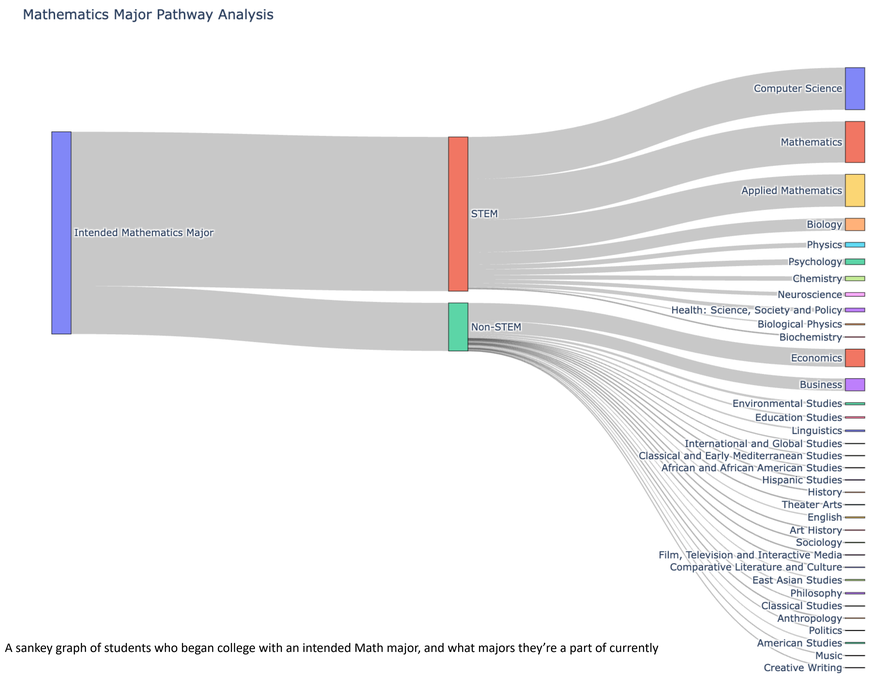
Visualization: Built Sankey diagrams and interactive plots (Matplotlib, Plotly) to map student flow and dropout points.

Framework Development: Outlined data protocols and automated workflows to improve reproducibility and efficiency in institutional analysis.

Results



A graph of the median numerical grade for students in the dataset over the last 8 years based off of their current status as a STEM major or non-STEM major



A sankey graph of students who began college with an intended Math major, and what majors they're a part of currently

Conclusions

Conclusions & Broader Insights

- The main barriers to data-driven decision-making are organizational, not technical — siloed departments, inconsistent practices, and limited collaboration.
- A shortage of technical and analytical infrastructure limits the university's ability to turn data into insight.
- Visual tools (e.g., Sankey diagrams) make patterns and bottlenecks visible, helping translate complex data into actionable insights. Being able to recognize patterns and warning signs with student pathway data will help establish better programs and preventative measures to increase STEM retention earlier.

Future Directions:

- Automate data flows: Workday should automatically anonymize and distribute data instead of siloing it within departments. This can be governed centrally by the data trustee
- System integration: Make all major systems—Moodle, myHousing, advising, and others—interoperable, or centralize key student information through Workday.
- Automation focus: Seek to automate as much of the data collection and reporting process as possible, minimizing manual effort.
- Structural improvement: Split the academic IT and administrative IT teams to better support both educational and institutional data needs.

References

1. [1] Komljenovic, J., Sellar, S., and Birch, K. 2024. Turning universities into data-driven organisations: Seven dimensions of change. *Higher Education*. https://doi.org/10.1007/s10734-024-01277-z.
2. [2] Morse, C. 2014. Visualization of student cohort data with Sankey diagrams via web-centric technologies. M.S. Thesis, University of New Mexico. Available at: https://digitalrepository.unm.edu/ece_etds/180.
3. Daniel, Ben. (2014). Big Data and analytics in higher education: Opportunities and challenges. British Journal of Educational Technology. 46. 10.1111/bjet.12230.

Acknowledgments

This research was conducted as a part of the Brandeis DSII Lab Internship. I would like to thank Professor Irving Epstein, Mark Hewitt, and Haley Rosenfield for their support through supplying data and being my advisor. We would also like to thank the Brandeis SciComm Lab for their help preparing this poster.