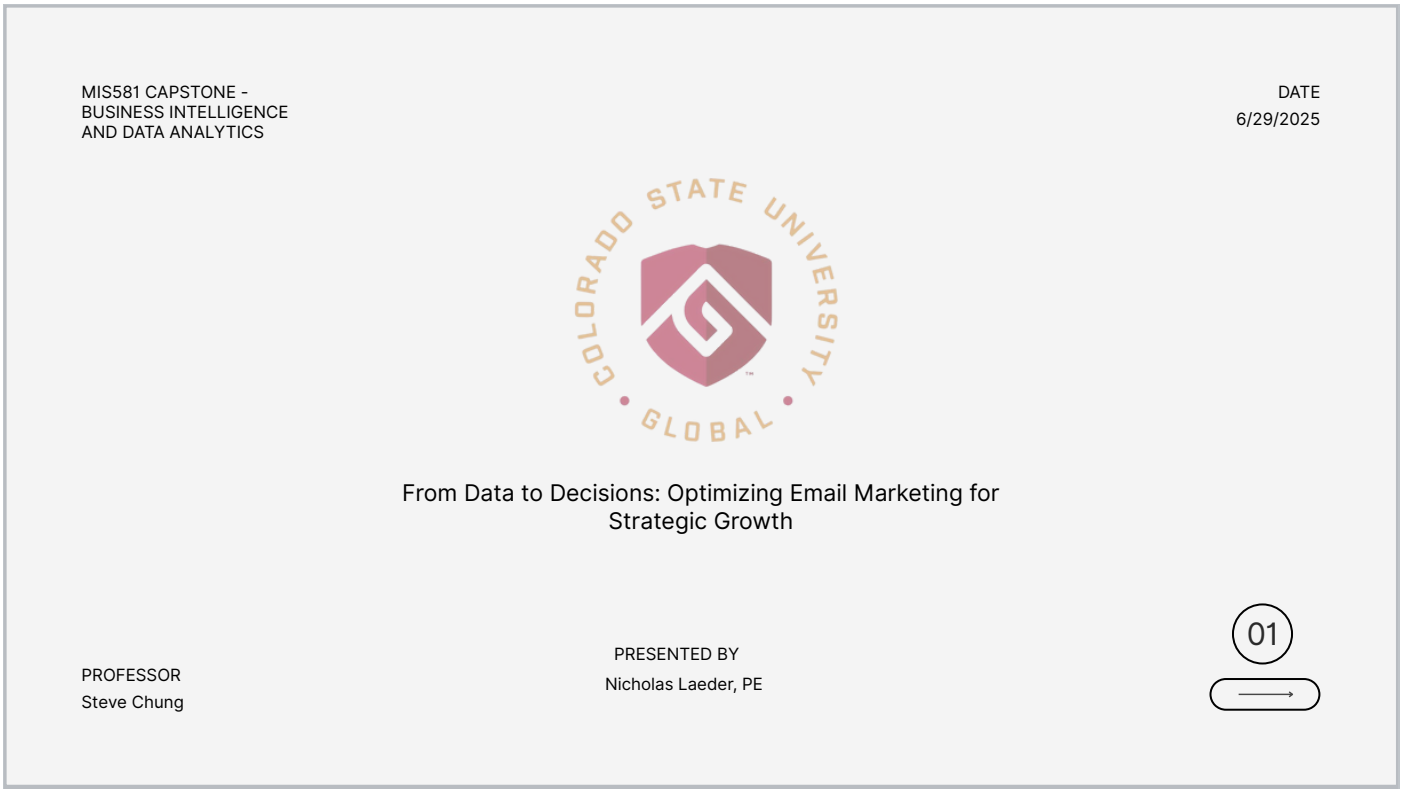
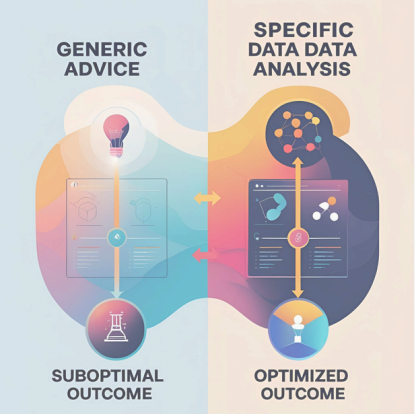


Good morning, my name is Nick Laeder. This presentation details my Capstone Project for MIS581, titled 'From Data to Decisions: Optimizing Email Marketing for Strategic Growth.' This research project applies data analytics techniques to a real-world business problem, aiming to derive actionable insights from historical marketing data for a partner organization.

I utilized over 18 months of email marketing data from a peer's small business. This business is a music and voice school for children in the metro Washington DC area.

- Problem: Small businesses often lack the resources for advanced analytics, leading to suboptimal email marketing strategies based on generic best practices rather than empirical evidence from their own customer base.
- Research Question: What specific, controllable attributes of an email campaign have a statistically significant impact on subscriber open rates?
- Hypotheses:
  - H_0: Email campaign attributes (send time, subject line characteristics) have no statistically significant effect on open rates.
  - H_A: At least one email campaign attribute has a statistically significant effect on open rates.



PROBLEM STATEMENT AND RESEARCH HYPOTHESES

02

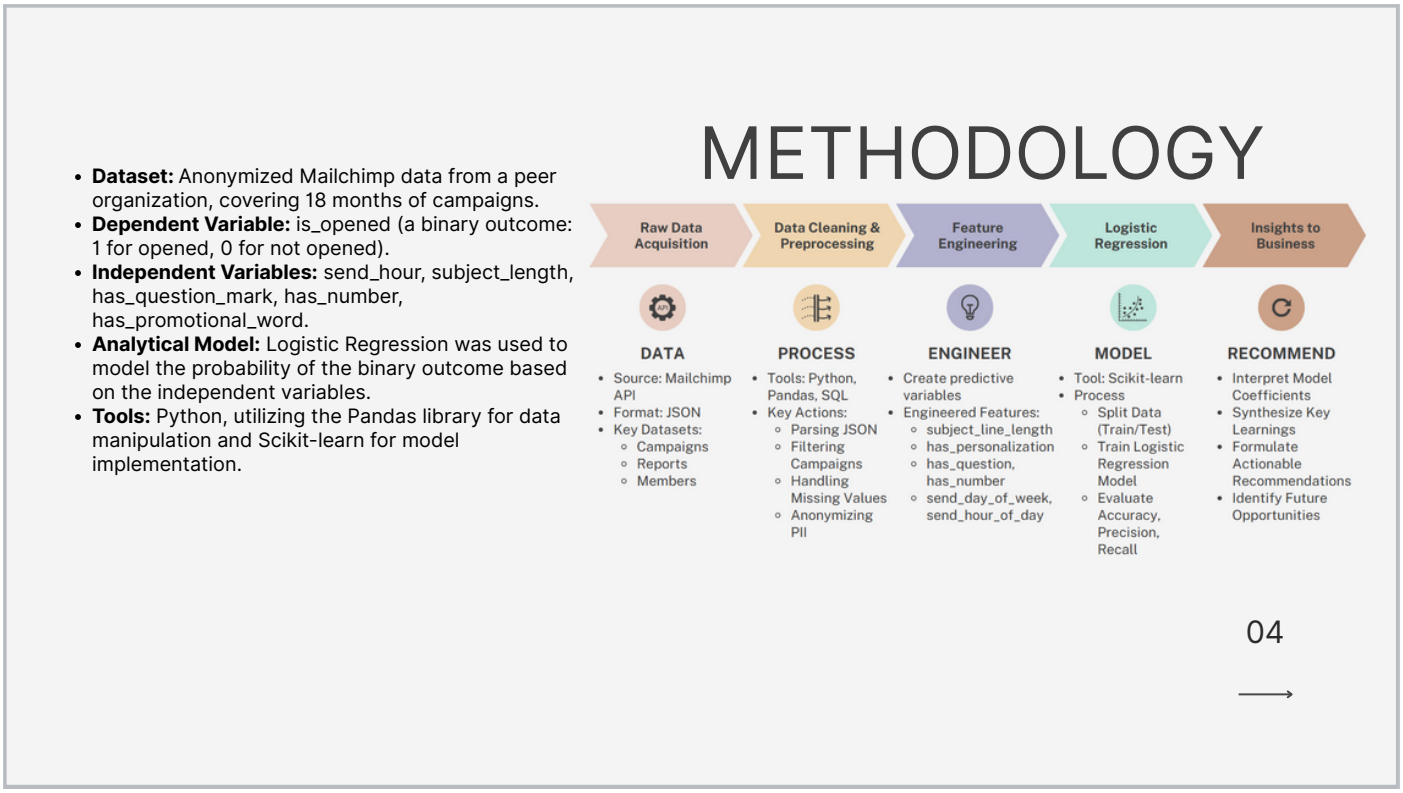
The problem this research addresses is a common one: small businesses are often data-rich but insight-poor. They collect vast amounts of marketing data but rely on generalized advice. This leads to my research question: What specific, controllable attributes of an email campaign have a statistically significant impact on subscriber open rates for this particular organization? To test this, I formulated a null hypothesis stating that no attributes have a significant effect, against an alternative hypothesis that at least one attribute does.

LITERATURE REVIEW:
SITUATING THE RESEARCH

- Existing research confirms the high ROI of email marketing (Kumar, 2018).
- Studies often focus on large-scale B2C enterprises, leaving a gap in small business-specific analytics (Chen & Lee, 2021).
- Data science literature emphasizes the power of regression modeling for identifying key business drivers (Provost & Fawcett, 2013).
- This project bridges the gap by applying formal data science techniques to a small business context.

03

To ground this project in existing knowledge, I conducted a literature review. The literature confirms that email marketing has a very high return on investment. However, most academic studies in this area focus on large corporations with dedicated analytics teams. There is a clear gap in research applying formal data science methods to the unique constraints and datasets of small businesses. This project aims to bridge that gap by applying proven regression techniques, as described by thought leaders like Provost and Fawcett, to a small business case study.



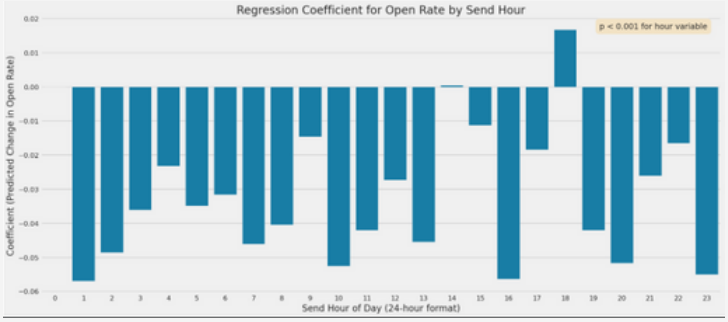
My methodology was structured to directly address the research question. I used an anonymized dataset from the partner organization. The dependent variable—the outcome we want to predict—was is_opened, a binary value. The independent variables were quantifiable attributes of each email that we can control. Given the binary nature of the outcome, a Logistic Regression model was the appropriate statistical tool to determine the impact of each variable. The entire analysis was performed in Python using standard data science libraries.

FINDING 1:

Send Time as a Primary Predictive Factor

MIS581 CAPSTONE
Nick Laeder

- The regression model identified send_hour as the most statistically significant variable ($p < 0.001$).
- The model's coefficients show a strong positive correlation between late afternoon/early evening send times (16:00-19:00) and the probability of an email being opened.
- Morning send times (09:00-11:00), often cited as a best practice, showed a non-significant or slightly negative correlation for this dataset.



05

The results of the regression analysis were conclusive. The model identified the send_hour variable as the most statistically significant predictor of whether an email would be opened, with a p-value well below the standard threshold of 0.05. Examining the model's coefficients, we see a strong positive correlation for sends between 4 PM and 7 PM. This finding provides empirical evidence that contradicts the generic advice the organization was previously following.

FINDING 2:
The Impact of Subject Line Characteristics

MIS581 CAPSTONE
Nick Laeder

Variable	Coefficient	P-Value
Subject Line Length	+0.001	$p < 0.05$
Has Number	-0.037	$p < 0.05$
Has Question Mark	-0.023	$p > 0.05$

- subject_length had a statistically significant positive coefficient, indicating longer, more descriptive subjects performed better.
- The presence of a number (has_number) in the subject line had a statistically significant negative coefficient.
- Variables for questions and promotional words were not statistically significant, suggesting they have a neutral impact.

06

⟶

The analysis of subject line content also yielded significant insights. The model showed a positive correlation between the length of the subject line and the open rate, suggesting that descriptive, informative titles are more effective than short, vague ones. Conversely, the presence of a number—a common clickbait tactic—was negatively correlated with opens. This suggests the audience has been conditioned to view such tactics as spam-like, reinforcing the need for a content strategy based on transparency.

HYPOTHESIS CONCLUSION

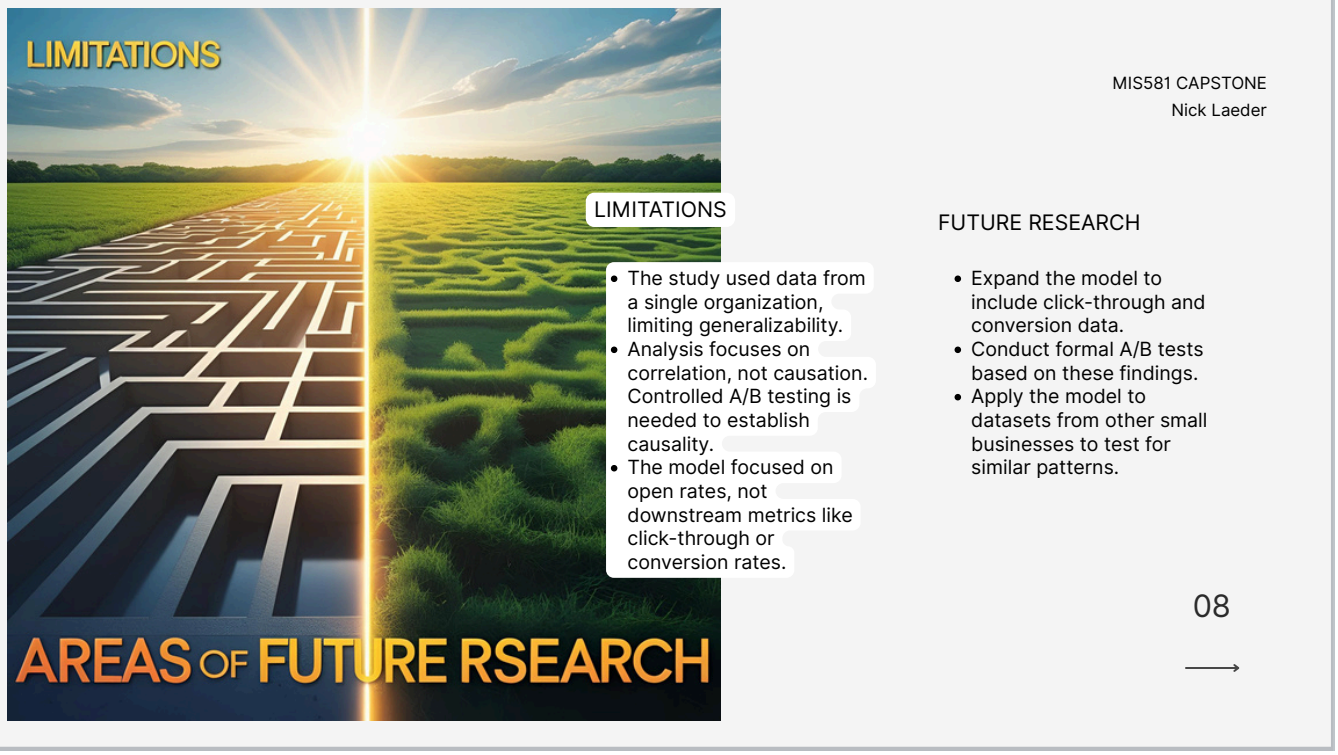
MIS581 CAPSTONE
Nick Laeder

Conclusion: The null hypothesis - H_0 - is rejected.

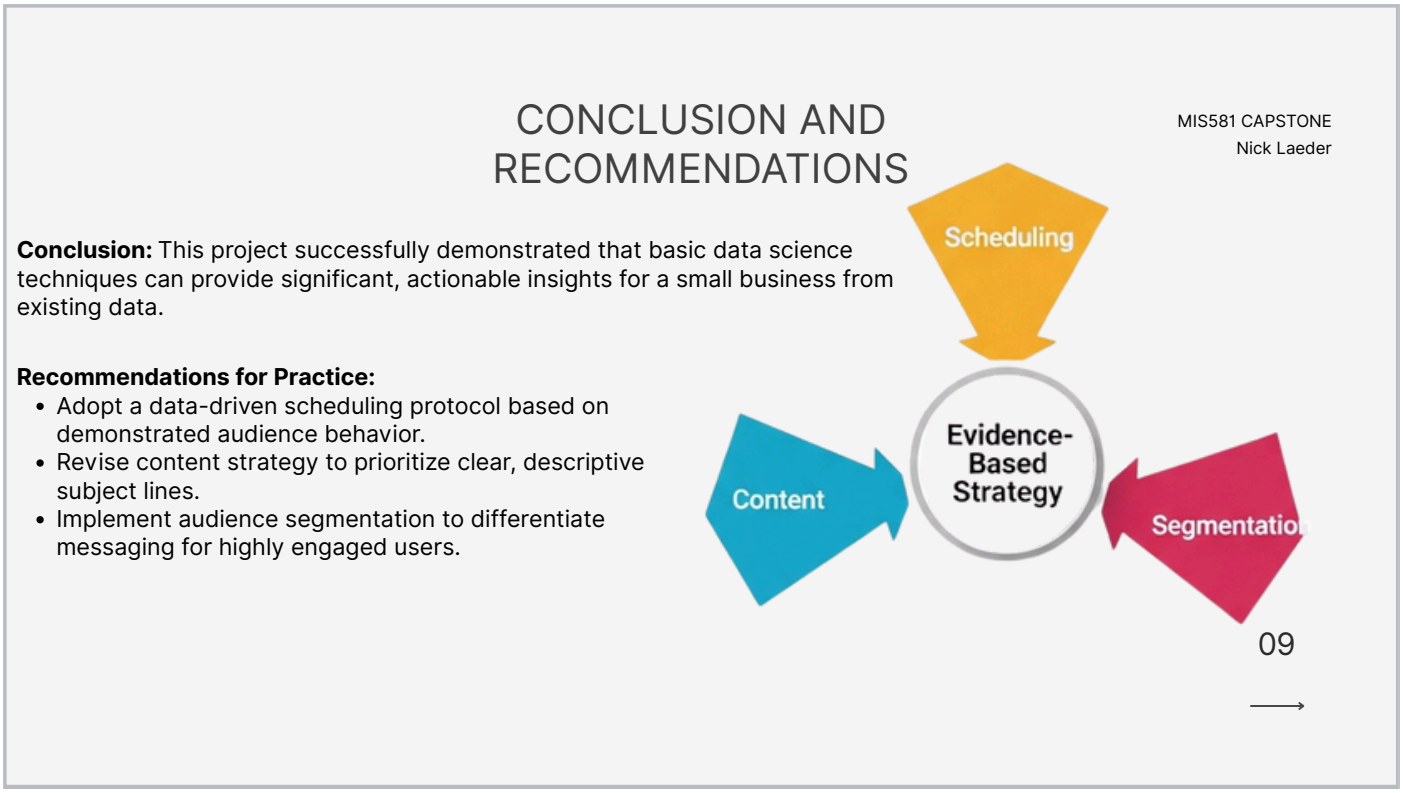
- The analysis provides strong statistical evidence that specific campaign attributes (send_hour, subject_length, has_number) have a significant effect on open rates.
- Discussion: The findings align with the principle that audience behavior is context-specific. Generic best practices are not a substitute for individualized data analysis. The negative perception of marketing gimmicks may reflect a broader trend in consumer skepticism.

07

Based on these findings, we can formally address our hypothesis. With multiple variables showing statistical significance, we reject the null hypothesis. The evidence strongly supports the alternative hypothesis: that controllable email attributes do have a significant impact on performance. In discussion, these results underscore a key theme from the literature: business analytics must be tailored. The data shows this specific audience is more sophisticated and less responsive to common marketing tactics, highlighting the importance of building trust through transparent communication.



It's important to acknowledge the limitations of this research. As a case study of a single organization, the results may not be generalizable to all businesses. Furthermore, this regression analysis identifies strong correlations, but cannot prove causation. Finally, the model was scoped to predict open rates, not deeper metrics like conversions. This points to several avenues for future research, including expanding the model to predict conversions and, most importantly, using these correlational findings to design formal A/B testing experiments to prove causality.



"In conclusion, this capstone project successfully achieved its objective. It demonstrated that even with limited resources, a small business can leverage data science to create a competitive advantage. The findings translate into three concrete recommendations for the partner organization: first, to adopt a new scheduling protocol; second, to revise their content strategy for subject lines; and third, to implement segmentation to better serve their most valuable audience members. These steps represent a move from intuition-based to evidence-based marketing."

REFERENCES

Dötlinger, L., Penz, M., Reiter, M., & Widauer, S. (n.d.). Lightweight DWH Data Analysis for SMEs. University of Innsbruck, Austria.

Dykes, B. (2024, February 27). The Future Of Data Storytelling Is Augmented, Not Automated. Forbes.

Gudfinnsson, K. (2019). Towards facilitating BI adoption in small and medium sized manufacturing companies (Doctoral dissertation, University of Skövde).

Jellypod. (n.d.). Grow Your Business with AI-Generated Podcasts.
Lennerholt, C., Van Laere, J., & Söderström, E. (2021). User-Related Challenges of Self-Service Business Intelligence. Information Systems Management, 38(4), 309–323.

Mally, P. K. (2023). Cloud Data Warehousing and AI Analytics: A Comprehensive Review of Literature. International Journal of Computer Trends and Technology, 71(10), 28–38.

Salqvist, P. (2023). Abstract: This thesis aimed to assess a given Data Warehouse against a well-suited Data Lakehouse in terms of read performance and scalability. (Master's thesis, KTH Royal Institute of Technology).

Yasser, M. F., & Alserafi, M. M. (2023). Cloud-Based Data Warehousing Solutions Capabilities and Challenges. 2023 International Conference on Computer Science and Information Technology (CSIT), 1–6.

09 ⟶

"In conclusion, this capstone project successfully achieved its objective. It demonstrated that even with limited resources, a small business can leverage data science to create a competitive advantage. The findings translate into three concrete recommendations for the partner organization: first, to adopt a new scheduling protocol; second, to revise their content strategy for subject lines; and third, to implement segmentation to better serve their most valuable audience members. These steps represent a move from intuition-based to evidence-based marketing."