

Enhancing DAEN Program Effectiveness: Analysis of Alumni Interview Feedback at George Mason University

Lingyun Dai, James Baldo

December 4, 2024

Abstract

This study examines the effectiveness of the DAEN program at George Mason University's College of Computing and Engineering through comprehensive alumni feedback analysis. The research employs structured interviews with fourteen program alumni who graduated between 2019 and 2023, collecting data on employment trajectories, technical skill utilization, and program effectiveness. Interview data was processed using AWS services and natural language processing techniques to generate standardized analysis datasets. Through Tableau visualizations, the study identifies key patterns in employment outcomes, technology usage, and program satisfaction, with alumni rating the program an average of 3.7 out of 5. The findings reveal opportunities for curriculum enhancement, particularly in cloud computing and hands-on project work, while highlighting the program's strengths in database systems and machine learning education.

Keywords: DAEN, Alumni, Employment, Program, Data, Analysis, Skill, Interview, Industry

Contents

1	Introduction	4
1.1	Purpose	4
1.2	Readership	4
1.3	Document Structure	4
2	Problem Statement	5
2.1	Problem	5
2.2	Scope	5
3	Data	6
3.1	Data Product - Questions	6
3.2	Collection Process	7
3.3	Data Process	8
3.4	Data Quality	8
4	Analysis	11
5	Visualization	12
6	Findings	20
7	Next Steps	21
8	Lessons Learned	22
A	Background	23
B	References	23
C	Research Biases and Data Quality Considerations	23

List of Abbreviations

DAEN	Data Analytics Engineering
AWS	Amazon Web Services
M.S.	Master of Science
NLP	Natural Language Processing
MS	Microsoft
ML	Machine Learning

1 Introduction

1.1 Purpose

The primary objective of this research is to evaluate the DAEN program’s effectiveness through comprehensive alumni interviews. Through structured feedback obtained from program alumni, this study conducts a detailed analysis of employment outcomes and program performance metrics. These insights are instrumental in enabling the DAEN department to implement strategic enhancements to the program’s curriculum development and technology integration.

1.2 Readership

The report is intended for the DAEN program academic and administrative personnel including leaderships, stakeholders, and development team. The findings presented in this report will provide valuable insights for improving program effectiveness, curriculum design, and student outcomes.

1.3 Document Structure

This report is organized into seven main sections. The Introduction establishes the research context and objectives. The Problem Statement formulates the research questions and scope. The Data section outlines the interview-based data collection methodology, data processing methodology, data quality interpretation, and interview question design. The Analysis section details the data processing pipeline and techniques employed. The Visualization section presents the derived data visualizations and their interpretations. The Findings section synthesizes key insights obtained through visualizations. Finally, the Next Steps and Lessons Learned section proposes future research directions and methodological improvements.

2 Problem Statement

2.1 Problem

The DAEN program faces significant challenges in evaluating program effectiveness as this represents the first systematic effort to gather alumni feedback. The absence of a centralized alumni database and limited alumni engagement mechanisms create substantial obstacles in reaching program graduates. The research revealed several critical issues:

- **Limited Cloud Technology Exposure:** While most alumni work extensively with cloud platforms in their current roles, many received insufficient exposure to these technologies during their coursework.
- **Theoretical vs. Practical Balance:** Alumni consistently highlighted a need for more hands-on projects, indicating a gap between theoretical knowledge and practical application.
- **Emerging Technology Integration:** The curriculum requires updates to incorporate emerging industry trends such as AI Ethics, DataOps, and Machine Learning Operations.
- **Alumni Engagement:** Establishing and maintaining contact with program graduates poses a significant challenge, impacting the program's ability to gather comprehensive feedback.

2.2 Scope

This 15-week research project encompassed the development of interview methodology, conduct of alumni interviews, and comprehensive data analysis. The project faced significant initial challenges in establishing contact with alumni, as no centralized database existed. Over 50 alumni were contacted through LinkedIn, resulting in 14 successful interviews. The scope included:

- Development of structured interview questions
- Implementation of data collection protocols
- Creation of data processing pipeline
- Analysis and visualization of collected data
- Compilation of program improvement recommendations

A valuable outcome of this research is the creation of an initial alumni contact database, which will be shared with the GMU alumni office to facilitate future engagement efforts.

3 Data

3.1 Data Product - Questions

The research design centered on a carefully structured set of interview questions designed to capture comprehensive insights about alumni experiences and outcomes. The questions were developed to gather both quantitative and qualitative data across several key domains:

- Academic Profile Assessment (Questions 1-2)
- Career Progression Tracking (Questions 3-6)
- Technical Skill Utilization (Questions 4, 7)
- Program Value Assessment (Questions 8-10)
- Post-Graduation Development (Question 11)

The interview questions were structured as follows:

- **Graduation Timeline:** "What year and semester did you graduate from the DAEN program?"
Purpose: Establishes temporal context for program experience. Enables trend analysis across different graduation cohorts.
- **Degree Classification:** "Did you receive an M.S. or Certificate?"
Purpose: Identifies academic credential level. Allows for segmentation of outcomes by program type.
- **Initial Employment:** "What was the title of your first job, the name of the company, and the general responsibilities?"
Purpose: Maps initial career placement outcomes. Enables analysis of program-to-career transition. Provides insights into entry-level positions for DAEN graduates. Helps identify common career entry points.
- **Initial Technical Requirements:** "What technologies and tools did you use for this job title?"
Purpose: Documents technology requirements for entry-level positions. Helps align curriculum with industry entry requirements. Identifies potential gaps in technical preparation. Provides baseline for technology evolution tracking.
- **Career Progression:** "What is your current job title, the name of the company and the general responsibilities?"
Purpose: Tracks career advancement patterns. Enables comparison between entry and current positions. Identifies common career progression paths. Helps understand long-term career outcomes.
- **Job Mobility:** "How many jobs have you had since you graduated?"
Purpose: Measures career mobility patterns. Indicates job market dynamics for DAEN graduates. Helps understand typical career progression timelines.

- **Technology Stack:** "List the most used technologies/tools in your career. (E.g. Programming language, framework, cloud, ML)"
Purpose: Identifies critical technical skills for career success. Maps evolution of technology requirements over time. Helps prioritize curriculum updates. Provides insights into industry technology trends.
- **Program Value Assessment:** "What knowledge and skills that you acquired in the DAEN program have been the most valuable to your career? Can you specify the concepts/methodologies/technologies that were most valuable?"
Purpose: Evaluates curriculum effectiveness. Identifies most impactful program elements. Helps validate current course offerings. Guides future curriculum development.
- **Gap Analysis:** "If DAEN program provided these specific courses, topics, or training, I would have been more prepared in my career..."
Purpose: Identifies curriculum gaps. Captures emerging industry needs. Guides program enhancement efforts. Provides direct feedback for improvement.
- **Program Effectiveness:** "How well did the DAEN courses prepare you for your career? (Scale: 1 – Not well at all, 5 – Very good)"
Purpose: Quantifies overall program effectiveness. Provides comparative metric across graduation years. Enables objective program assessment. Facilitates trend analysis over time.
- **Continuous Learning:** "Have you completed any courses/certifications since you graduated from the DAEN program?"
Purpose: Identifies post-graduation learning needs. Reveals industry certification preferences. Helps understand skill gap mitigation strategies. Guides potential certification partnership opportunities.

3.2 Collection Process

The data collection process followed a systematic approach designed to ensure consistency and quality:

- **Alumni Identification**
 - LinkedIn network search using DAEN program keywords
 - Referral requests from identified alumni
 - Collaboration with faculty for alumni contacts
- **Interview Platforms**
 - MS Teams (primary platform)
 - Zoom (secondary platform)
 - LinkedIn messaging (alternative option)
- **Recording Methods**

- Audio recordings (Teams and Zoom)
- Automated transcription (Teams)
- Direct text responses (LinkedIn)
- **Data Security**
 - Secure storage of recordings
 - Confidential handling of personal information
 - Anonymous processing of responses

3.3 Data Process

The data processing pipeline consisted of several structured stages:

- **Storage and Security**
 - Upload to AWS S3 bucket
 - Encryption of sensitive data
 - Access control implementation
- **Transcription**
 - AWS Transcribe processing
 - Manual verification of transcripts
 - Format standardization
- **Data Extraction**
 - NLP toolkit implementation
 - Text tokenization and pattern matching
 - Keyword extraction and categorization
- **Structure and Format**
 - CSV file creation with 15 primary columns
 - Data transformation for visualization
 - Quality control checks

3.4 Data Quality

Interview-based research inherently presents various potential biases and data quality challenges that must be carefully considered and addressed. Drawing from established research methodology literature (Bergelson et al., 2022; Nikiforova, 2020) and implementing practical measures in our research process, we established the following quality control framework:

- **Interview Structure Implementation**

- Utilized a standardized set of 12 structured questions asked in the same order for all participants
- Questions were designed to be neutral and open-ended to avoid leading the respondents
- Interviewers maintained consistent protocols by not prompting or suggesting answers
- Each interview followed the same format regardless of platform (MS Teams, Zoom, or written surveys)
- Responses were accepted as given without attempting to influence or modify them

- **Data Collection Quality Controls**

- Recorded all interviews (audio or text) with participant permission
- Used AWS Transcribe consistently across all audio recordings to ensure uniform transcription quality
- Applied the same NLP toolkit for all text extraction to maintain processing consistency
- Implemented systematic data extraction procedures using local natural language processing tools
- Maintained standardized formats for data storage and processing

- **Data Processing Standards**

- Created a uniform CSV structure with 15 standardized columns for all interview data
- Applied consistent text preprocessing tasks across all responses
- Used automated Python scripts in AWS SageMaker for standardizing string formats
- Implemented systematic handling of multiple responses through column expansion
- Maintained data anonymity through consistent identifier formatting (CEC-DAEN-[Graduation Year]-[MS/Cert]-[Sequence Number])

- **Quality Verification Procedures**

- Conducted human review of ambiguous responses
- Filtered out unclear or inconsistent terminology
- Applied standardized categorization for job titles based on industry classifications
- Mapped similar technologies to standardized categories through comprehensive research
- Verified data transformations through multiple review cycles

To ensure consistency and minimize bias in our data collection and processing:

- Interviewers proceeded systematically through questions without deviation
- No additional prompting or clarifying questions were used that might influence responses
- Raw responses were preserved in their original form before any processing
- All data processing steps were documented and applied uniformly
- Standardization procedures were based on established industry classifications rather than subjective interpretations

This systematic approach to data quality management was implemented to ensure that our findings would accurately represent the alumni feedback while minimizing potential biases or inconsistencies in data collection and processing.

4 Analysis

The analysis followed a structured approach to process and interpret the collected interview data:

Step 1: Data Organization

- Compilation of raw interview transcripts
- Creation of standardized data templates
- Initial categorization of responses

Step 2: Data Processing

- Text analysis using NLP techniques
- Standardization of job titles and technologies
- Category development for skills and tools

Step 3: Visualization Development

- Selection of appropriate visualization types
- Implementation in Tableau
- Iterative refinement based on patterns

Step 4: Pattern Analysis

- Identification of trends across graduation years
- Analysis of technology usage patterns
- Assessment of program satisfaction metrics

5 Visualization

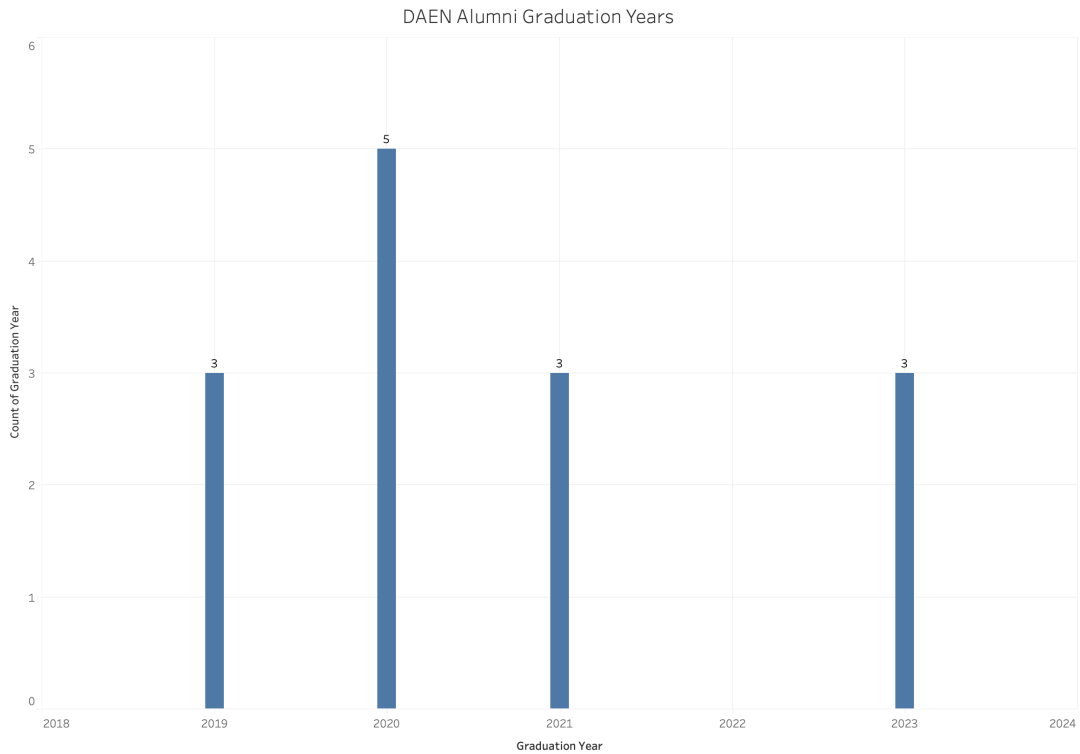


Figure 1: DAEN Alumni Graduation Year Distribution

The graduation year distribution reveals a concentration of respondents between 2019-2023, with the peak in 2020 (5 alumni). Notable challenges were encountered in reaching alumni who graduated before 2018.

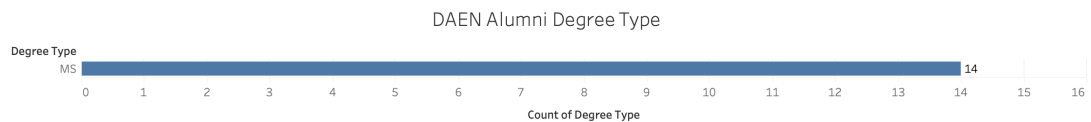


Figure 2: DAEN Alumni Degree Type Distribution

All respondents held MS degrees, indicating our sample represents the full-degree program experience. The absence of certificate holders suggests a need for targeted outreach to this alumni group.

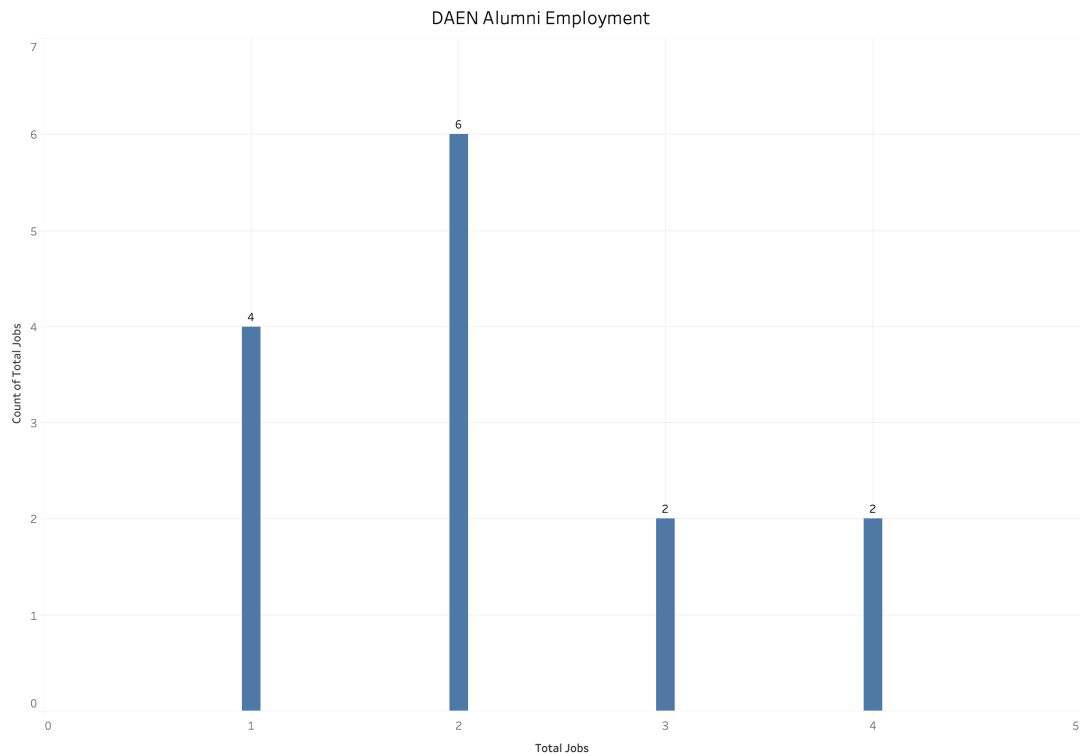


Figure 3: Total Jobs Held by DAEN Alumni

The analysis reveals that approximately 43% of alumni have held two jobs since graduation, suggesting active career progression.

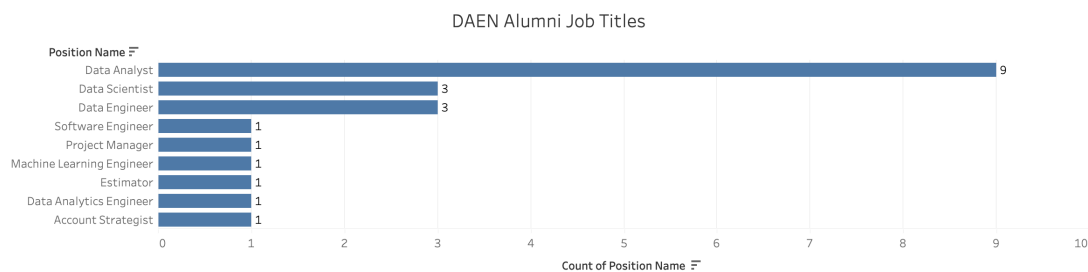


Figure 4: Job Titles of DAEN Alumni

The alumni job titles were consolidated, with roles like Business Analyst, Senior Data Analyst, and Lead Data Analyst grouped under "Data Analyst", which accounted for 43% of jobs. This focused on core functions rather than seniority. Senior Data Scientist roles were similarly consolidated under the broader Data Scientist title.

Genius Plaza Capital One
 East West Bank Medica
 Potomac Economics Fannie Mae Codice
 CloudNova Technology FINRA Choice Hotels International
 LLC R.B. Hinkle Construction Inc. Ashling Partners
 Association Analytics Activision Blizzard Media Verizon
 C&M Associates Comscore National Hockey League
 Databuoy Corp Data Economy
 Market Star

Figure 5: Companies Employing DAEN Alumni

The diversity in employing companies demonstrates broad industry recognition of the program, with no single industry dominating employment outcomes.

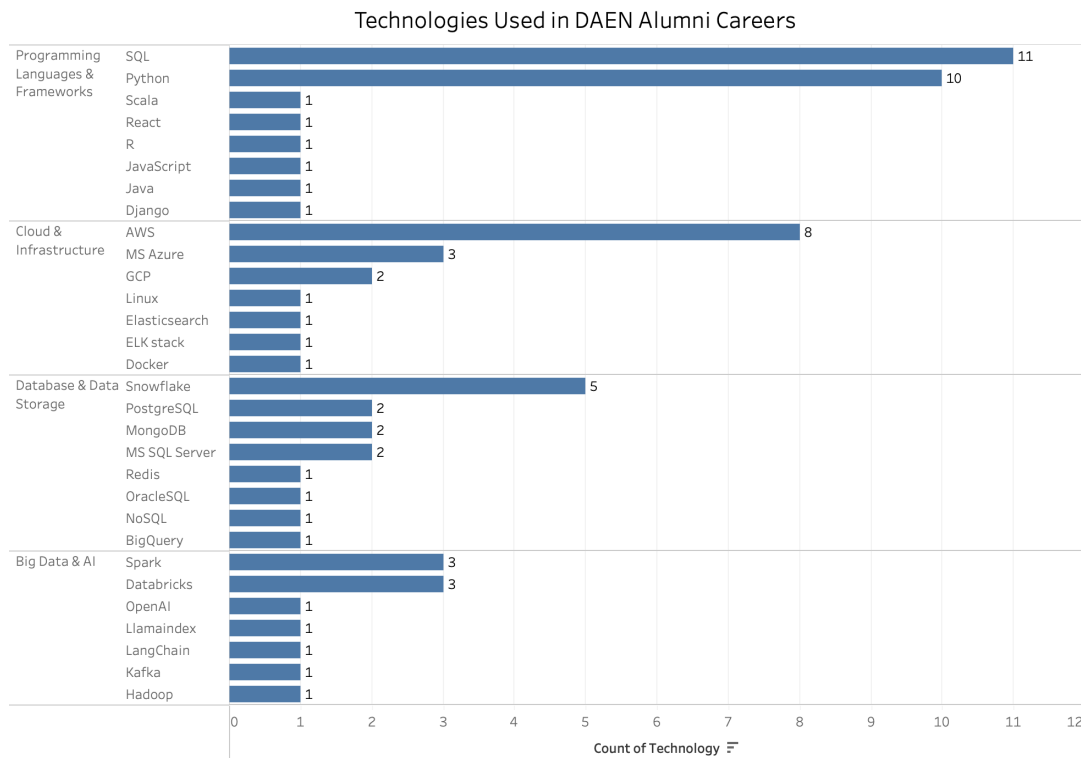


Figure 6: Technologies Used in DAEN Alumni Careers

Technology usage shows clear patterns:

- Programming Languages & Frameworks: Python and SQL dominate (78% usage)

- Cloud & Infrastructure: AWS leads with 47% adoption
- Database & Data Storage: Snowflake is increasingly popular (33% usage)
- Big Data & AI: Spark and Databricks (55% usage)

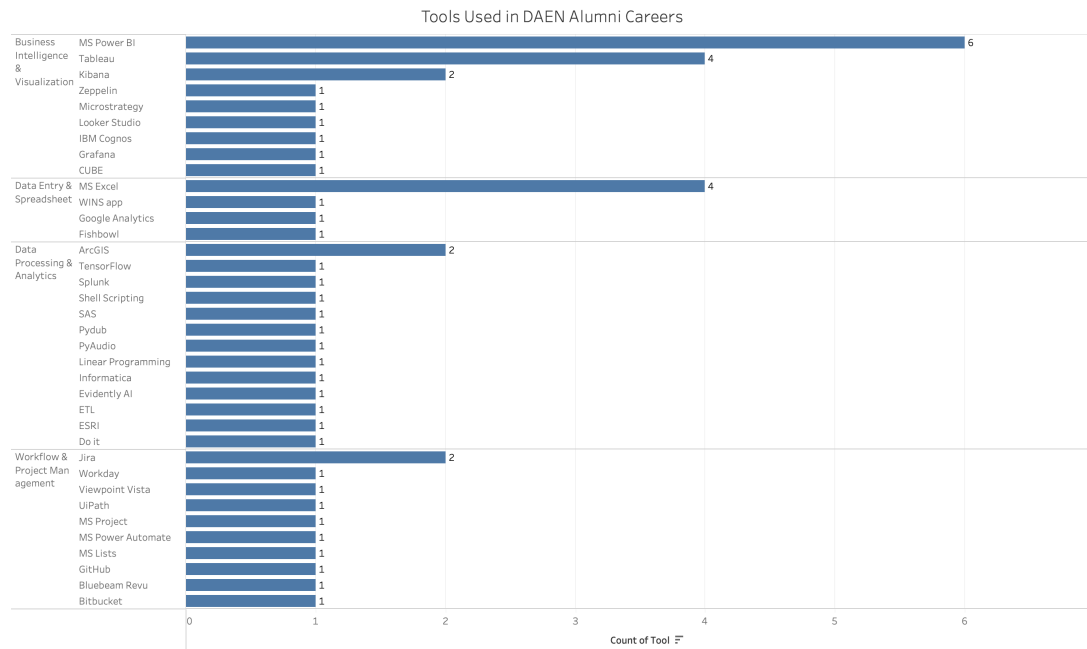


Figure 7: Tools Used in DAEN Alumni Careers

Visualization and analytics tools show strong preferences:

- Business Intelligence & Visualization: Power BI and Tableau lead adoption
- Data Entry & Spreadsheet: Heavy use of MS Excel
- Data Processing & Analytics: Popular in ArcGIS
- Workflow & Project Management: Popular in Jira

Most Valued DAEN Courses by Alumni



Figure 8: Valuable Program Courses

Course evaluations highlight key program strengths:

- DAEN 690, Data Analytics Project (Capstone course)
- Database courses
- Required DAEN core courses showed strong practical value (AIT 580, Data Mining, STAT 515, Operations Research, AND DAEN 690)

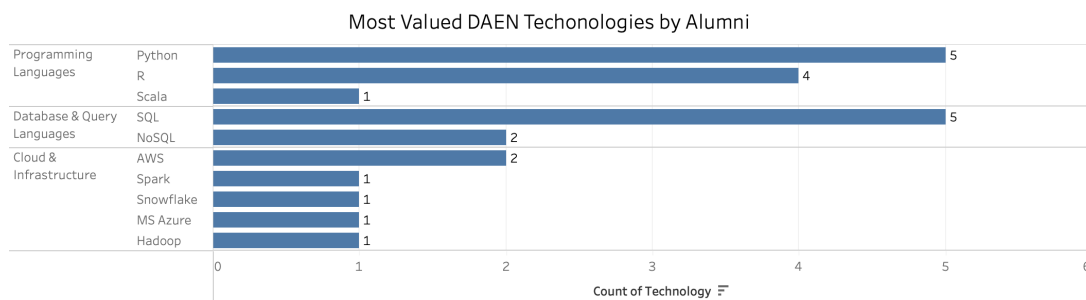


Figure 9: Valuable Technologies Acquired in DAEN Program

SQL and Python emerge as the most beneficial technologies learned, aligning well with industry demands as the most commonly used technologies in DAEN alumni careers.

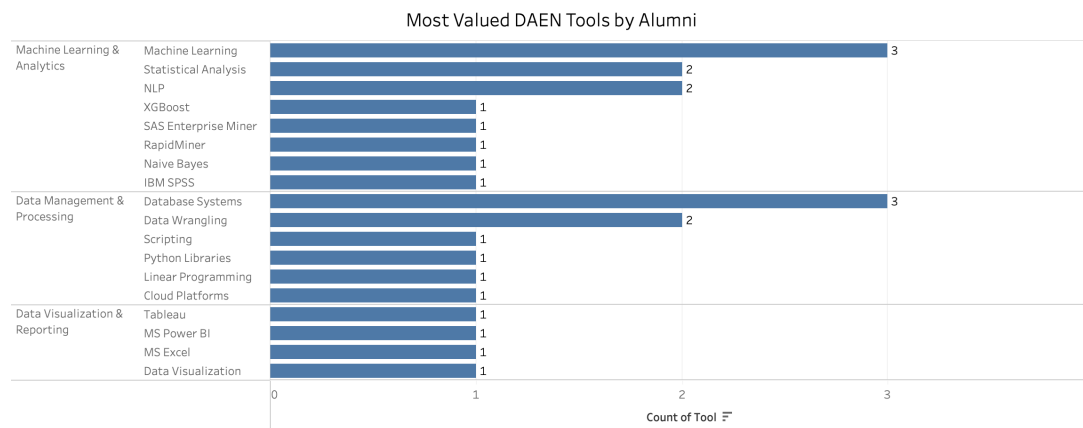


Figure 10: Valuable Tools Acquired in DAEN Program

Machine learning and database systems emerged as the most valuable tools acquired from the program.

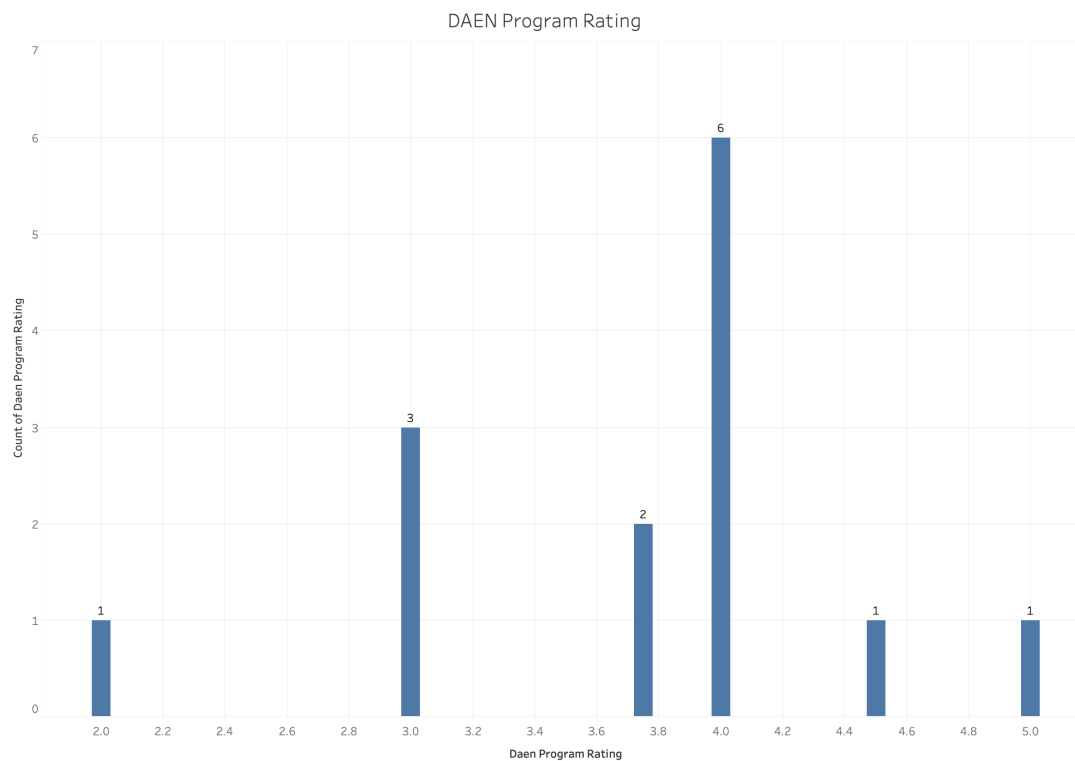


Figure 11: DAEN Alumni Program Ratings

The program received an average rating of 3.7 out of 5, with most alumni rating it 4 out of 5. This indicates relatively high satisfaction with their education while also suggesting areas for potential improvement.

Valuable Skills Acquired From DAEN Program

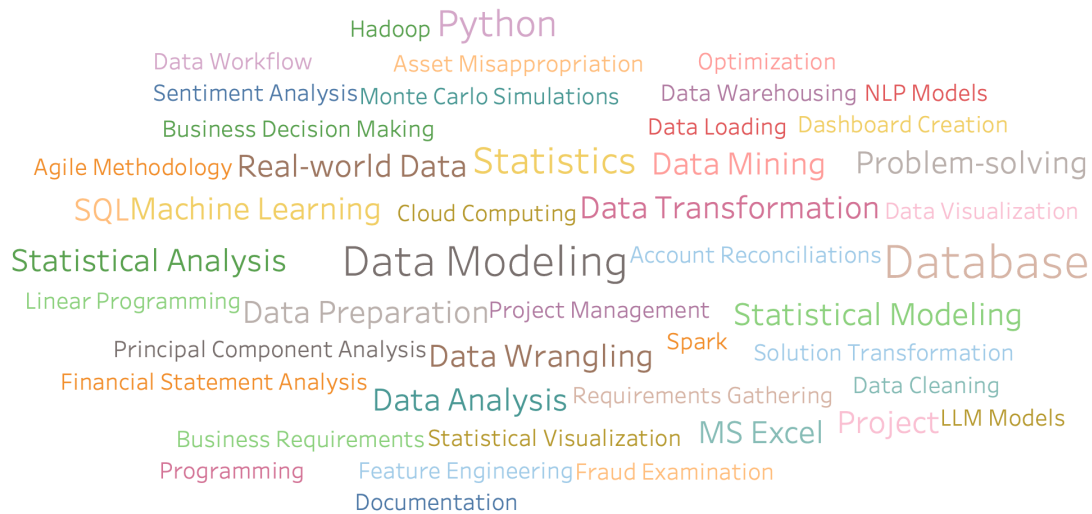


Figure 12: Valuable Skills Acquired in DAEN Program

Alumni particularly valued data modeling and database skills gained from the program. The emphasis on database systems and machine learning concepts appears to align well with industry needs.

DAEN Program Potential Improvements

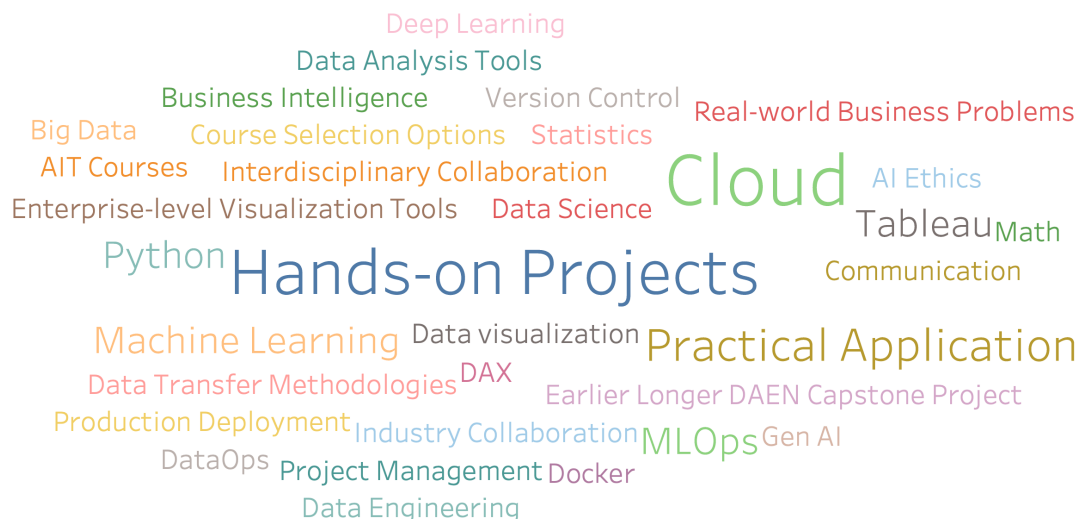


Figure 13: Potential Program Improvements Suggested by Alumni

Three main areas for improvement emerged from alumni feedback:

- Increased emphasis on cloud computing technologies, particularly AWS as many alumni use cloud in their roles and received limited exposure during their coursework
- More hands-on project opportunities

- Enhanced coverage of emerging technologies

These suggestions reflect the evolving needs of the industry and the gap between academic training and professional requirements.

Certifications Acquired in DAEN Alumni Careers

Azure Data Scientist
Tableau Desktop Specialist
Google Project Management AWS Data Scientist
Project Management LinkedIn Course Snowflake Snow Pro Core Certification
Apache Airflow Certification Construction Management Certification
Financial Modeling of Data from Wharton Online Agile Course
Integration Engineer Certification Snowflake Advanced Architect Certification
Team Communication LinkedIn Course ISME Security Awareness Training
DBT training for analytics engineer JIRA Course Indico Data
Snowflake Certification PMP certification
AWS Cloud Practitioner
UiPath

Figure 14: DAEN Alumni Completed Certifications Post-Graduation

The alumni reported diverse certifications post-graduation, reflecting individual career paths and employer requirements.

6 Findings

Key insights from the analysis are organized into several critical areas:

Program Demographics and Outcomes

- Concentrated graduate distribution between 2019-2023
- 100% MS degree completion rate among participants
- Average of 2 jobs held post-graduation
- Strong representation in data analyst roles

Technical Skill Utilization

- Dominant use of SQL and Python in professional roles
- High adoption of AWS cloud services
- Increasing importance of Snowflake and Databricks
- Significant focus on visualization tools (Power BI, Tableau)

Program Effectiveness

- Average program rating of 3.7 out of 5
- Strong satisfaction with database and ML coursework
- High value placed on hands-on project experience
- Notable impact of DAEN 690 capstone course

Areas for Enhancement

- Increased cloud computing coverage
- More emphasis on practical applications
- Integration of emerging technologies
- Enhanced focus on industry certification preparation

7 Next Steps

Program Development

- Establish a structured process for regularly collecting alumni feedback through surveys and interviews.
- Implement a comprehensive strategy for alumni outreach and engagement to maintain strong connections.
- Develop an automated pipeline using NLP tools to efficiently process and analyze data from various interview formats.
- Create a secure, centralized database to confidentially store and manage alumni information and feedback.

Curriculum Enhancement

- Expand coursework and hands-on training focused on popular cloud platforms, particularly AWS.
- Incorporate more real-world, applied projects to allow students to directly apply their technical skills.
- Update curriculum to include emerging tools and methodologies like Snowflake, Databricks, and MLOps.
- Tailor program content to help students prepare for in-demand industry certifications.

Alumni Engagement

- Implement periodic surveys and outreach efforts to continuously gather alumni insights.
- Connect current students with alumni mentors to foster professional development.
- Organize alumni events and conferences to facilitate networking and collaboration.
- Offer workshops, webinars, and resources to help alumni stay current with industry changes.

Continuously protect the confidentiality of all alumni data collected and stored throughout these initiatives.

8 Lessons Learned

Research Methodology

- Engaging alumni early after graduation is crucial for encouraging their participation in the research.
- Designing clear, standardized interview questions helps elicit more specific, actionable responses.
- Offering multiple participation methods like video calls and surveys improves overall engagement.
- Consistent formatting and recording of interview data enables more efficient processing and higher-quality analysis.

Data Collection

- Leveraging professional networking platforms like LinkedIn allows efficient identification of and outreach to alumni.
- Ensuring all interview recordings use a single, compatible format streamlines the data processing workflow.
- Implementing robust privacy safeguards encourages alumni to share candid feedback.
- Conducting persistent, polite follow-up outreach helps maintain alumni engagement.

Process Improvement

- Developing NLP-based data extraction scripts and workflows accelerates analysis by minimizing manual effort.
- Consistently mapping job titles, technologies, and skills to industry classifications leads to more meaningful insights.
- Maintaining detailed documentation of the research process ensures knowledge retention for future iterations.
- Incorporating input from stakeholders and subject matter experts iteratively refines the approach and findings.

A Background

This research, conducted through DAEN 698: Research Project, represents a significant departmental initiative led by DAEN student and DAEN program director. As the first systematic effort to collect and analyze alumni feedback, this study establishes a foundation for ongoing program assessment and enhancement. The insights gained from this pioneering research will stimulate further discussions and drive continued investment in understanding and improving program effectiveness.

B References

References

- [1] George Mason University. (2024). Data Analytics Engineering, MS. Retrieved from <https://catalog.gmu.edu/colleges-schools/engineering-computing/engineering/data-analytics-engineering-ms/>
- [2] Bergelson, I., Tracy, C., & Takacs, E. (2022). Best Practices for Reducing Bias in the Interview Process. *Current Urology Reports*, 23(11), 319-325.
- [3] Nikiforova, A. (2020). Definition and Evaluation of Data Quality: User-Oriented Data Object-Driven Approach to Data Quality Assessment. *Baltic Journal of Modern Computing*, 8(3), 391-432.

C Research Biases and Data Quality Considerations

In conducting this alumni interview research, we identified several key sources of potential bias and inaccuracy that could impact our findings. Despite implementing systematic controls, these factors should be considered when interpreting the results:

- **Selection Bias**

- Sample limited to LinkedIn-active alumni from 2019-2023
- Response rate of 14 out of 50 contacted alumni
- Potential over-representation of successful graduates

- **Data Accuracy**

- Memory recall variations depending on graduation year
- Standardization requirements for technical terminology

While these limitations exist, our methodological controls and transparent reporting help maintain the integrity of the findings while acknowledging areas for potential improvement in future research.