Data Analytics Capstone Topic Approval Form

The capstone challenges students to integrate skills and knowledge from several program domains into one project. The guidelines for this capstone course require you to demonstrate the application of academic and professional abilities developed as an undergraduate student in the BSDA program. It is highly recommended that the topic of your capstone be about resolving a current or perceived business problem. Your research topic should exemplify scholarship and research at the highest level and should be significant enough that it would help potential employers identify your abilities. It is also recommended that you use publicly available datasets for transparency and external validity.

This document is designed to help you clearly state the research question you will be exploring in your capstone project, the scope of your project, and your timeline in order to ensure that all of these align with your degree emphasis. Without clearly defining each of these areas, you will not have a complete and realistic overview of your project, and it cannot be accurately assessed whether your project will be acceptable for this capstone course.

If your project is one you have already completed at work or elsewhere, this document should be easy to complete. Many students do use a project they have already completed in the past. In that case, you will write the proposal as if the project has not been completed yet, and when you report on your project, you will use your complete after-implementation report. If you have not yet completed your project, this document can help ensure the scope is within the acceptable range for this capstone. An instructor must approve this form before you submit this task for evaluation. The task will not be evaluated without an instructor's signature. The instructor may ask for additional information before approving this form.

Before submitting this form for approval, please remove all italicized directions in the form.

Please only submit a Topic Approval Form that has been signed by a course instructor for evaluation.

Capstone Topic Approval Form

The purpose of this document is to help you clearly explain your capstone topic, project scope, and timeline. Identify each of the following areas so you will have a complete and realistic overview of your project. Your course instructor cannot approve your project topic without this information.

Note: You must fill out and submit this form. Space within each section will expand as needed.

Any costs associated with the development of the data analytics solution will be the responsibility of the student.

Student Name: Jacob Porter

Student ID: 000311459

Capstone Project Name: Network Device System Presence

Project Topic: You are a data engineer for a large telecommunications company called Lightspeed. You have three systems that are considered sources of truth for network asset data: the Inventory System, the IP Management System, and the Observability System. As part of a security audit initiative, there is a need to ensure the availability and completeness of asset records across all systems before vulnerability scans can be reliably conducted. Initial data provided by the inventory system has raised concerns regarding accuracy and completeness. Due to evolving processes and policies, these systems may not all contain the same asset information. The goal of this project is to systematically validate the presence and completeness of asset records across all three systems and develop a predictive model to identify risk factors for missingness.

Research Question: Can we accurately validate and predict the availability of asset records across the three sources of truth (Observability, Inventory, IPAM systems) to ensure readiness for a security audit?

> **Hypothesis:** An asset present in the observability platform will be found in both the inventory and IPAM systems over 75% of the time. Furthermore, we can build a predictive model that identifies which asset characteristics are most strongly associated with missing records.

Context: The security audit project aims to ensure that all critical assets are scanned for vulnerabilities and that remediation actions can be planned effectively. Before launching large-scale scanning, it is crucial to validate that asset records are consistent and present across all systems, minimizing wasted effort and risk from incomplete data. Inconsistent asset presence can hinder the audit's accuracy and the organization's security posture.

Data: Asset data will be **generated using Python and the Faker library** to simulate real-world system records, providing device IP, hostname, status, region, vendor, and model information.
Data will be stored and exchanged as CSV (text) files, making the project fully portable and transparent.

Three datasets will be created: one for each system (Observability, Inventory, IPAM), with overlapping and missing entries according to realistic scenarios.

Data Gathering: Asset records will be generated programmatically using Python and the Faker library.
All datasets will be written directly to CSV files. Python scripts (using Pandas) will extract, join, and analyze these CSV files.
This simulates the kind of file-based ETL and data engineering workflow often used in analytics projects.

Data Analytics Tools and Techniques:

- **Python** (Pandas, Scikit-learn, Faker)
- **VSCode Data Wrangler** for profiling and exploration
- **Supervised machine learning model** (e.g., logistic regression or random forest) to predict asset presence/completeness based on asset features
- **Descriptive statistics and basic data profiling** to summarize completeness and distribution

Justification of Tools/Techniques: A supervised machine learning model is appropriate for this project as it enables not only the validation of asset presence rates, but also the identification of key factors associated with missingness across systems. Model performance (e.g., accuracy, precision, recall, F1 score) will be used to rigorously evaluate the predictive task, in addition to descriptive profiling. This approach exceeds the minimum requirements for analytical rigor and is aligned with best practices in data analytics.

Application Type, if applicable (select one):

- ☐ mobile
- ☐ web
- ☐ stand-alone

Programming/Development Language(s), if applicable: Python

Operating System(s)/Platform(s), if applicable: MacOS/ Windows

Database Management System, if applicable: Primary: Demonstrate >75% presence of asset records across all three systems to support the security audit. **Secondary:** Provide a trained predictive model and insights into asset characteristics that are most at risk of being missing in key systems, supporting remediation and process improvement.

Project Outcomes: Given that the record set scores over 75% average presence across the 3 sources of truth, we can proceed with the security audit. If any additional beneficial analysis is available that points to the root cause of data issues, that can also be provided.

Projected Project End Date: 7/15/2025

Sources: None necessary at this time.

Human Subjects or Proprietary Information
Does your project involve the potential use of human subjects? (Y/N): No
Does your project involve the potential use of proprietary company information? (Y/N): No

---

STUDENT SIGNATURE

Jake Porter

By signing and submitting this form, you acknowledge that any cost associated with the development and execution of your data analytics solution will be your (the student) responsibility.

TO BE COMPLETED BY AN INSTRUCTOR

The capstone topic is approved by an instructor.

COURSE INSTRUCTOR SIGNATURE:

Jim Ashe, Ph.D. Mathematics
COURSE INSTRUCTOR APPROVAL DATE:

7/26/2025

Project Compliance with IRB (Y/N):y