**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. You have the Inventory System, the IP Management System, and the Observability System. You are expected to ramp up a large security audit project, where the security team would like to run vulnerability scans on all of the active network assets. They were provided some data from the inventory system, but users have complained that the data seemed to be inaccurate or incomplete. Due to evolving processes and policies over time, it’s possible that the systems do not have all of the appropriate data. In order to proceed with the security project, it is necessary to determine the availability of assets across all systems.

**Research Question:** We can proceed with the security audit after validating the availabily of asset records in all three sources of truth.

**Hypothesis:** That an asset in the observability platform can be found in the inventory system, and the IPAM system over 75% of the time.

**Context:** The security audit project is intended to identify assets that have active vulnerabilities on them. KPI’s will need to be established to ensure that assets are being scanned in a timely fashion, and then remediated in a timely fashion. Before that can be started, we must validate that users will not spend most of their time trying to track down asset details. This situation warrants an investigation into the presence of assets across the three systems.

**Data:** We will need to collect asset data from the Observability system, and then compare it to the IPAM and Inventory system.

*The first data set will be the asset record from the Observability platform. This should have the device IP address, the device hostname, and the current device status. The second data set will be the asset record from the IP management system. This will contain the IP address, and the asset region. The region will be important, as the region determines which fix agent group will need to be engaged to remediate an identified vulnerability. The third data set will be the asset record from the Inventory system. This will contain the vendor and model information for the asset, as certain vulnerabilities will need vendor support to remediate.*

The datasets necessary for analysis will be generated through mock data creation. We will use python and specifically the Faker library to generate the initial asset list, as well as the subsequent data sets.

**Data Gathering:** As we will be generating the data, we will be putting it into a databse of some kind, probably sqlite. Then we will separately generate a python script that will use pandas to connect to the database and pull down the data. This will simulate extracting data from a database via python and sql. We will then generate the output files in CSV format.

**Data Analytics Tools and Techniques:** We will use Python, Pandas and the Data Wrangler extension in vscode.

**Justification of Tools/Techniques:** We don’t need to do an extensive analysis given that the dataset proves to be at least 75% complete. Data wrangler will demonstrate through basic profiling that the appropriate presence was attained. This can be screen captured and included in a power point that executive management will review to provide the go ahead.

**Application Type, if applicable (select one):**

mobile

web

stand-alone

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

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

**Database Management System, if applicable:** Postgres or sql lite

**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.**

**INSTRUCTOR’S NAME AND SIGNATURE:**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**INSTRUCTOR APPROVAL DATE:**

**Project Compliance with IRB (Y/N):**