The Ultimate C964 Capstone Guide

Introduction -- The following information is intended to help you better understand what the expectations are for your Computer Science capstone project. Once completed you'll provide to evaluations a functional piece of software and all required datasets. In addition, there are numerous pieces of documentation and artifacts (examples) you will need to include.

For assistance with your efforts be sure to contact your assigned course instructor (CI). If you don't have one, send an email to UGcapstoneIT@wgu.edu. CIs are glad to provide guidance on potential project ideas, course requirements, written work clarification, and review of documentation drafts. For assistance with the programming aspect of your project, contact a CI familiar with the language you are using.

Remember: All content you produce must be your original work. For written work, No more than a combined total of 30% of the submission An originality report is provided when you submit your task that can be used as a guide.

There are two tasks required to complete your capstone project. What follows is information that will help you complete those tasks. Please do not submit Task 2 until Task 1 has successfully passed.

Task 1

Task 1 is essentially getting your project approved and completing information about the nature of the project. Here are descriptions for the three forms you'll need to submit:

Note: You will see human subjects mention in a number of forms. This has to do with the use of humans for experimental purposes and typically does not relate to a Computer Science capstone. You are still required to complete that information.

Choosing a Topic – While there is quite a bit of flexibility for topic selection it's crucial that you choose a topic/project focused on the requirements listed in Task 2 under Section C. The data-based application you select must provide a business solution for a client (e.g. company, organization, school, etc…) that involves visual data analysis with a "future" component. You will need to present a solid case for how your application meets that business need in the documentation. Justification of your solution in the context you describe is a major element of success. This may be fictional but that focus transcends throughout the documentation. Contact a course instructor for further guidance on project/topic selection. Go over the following checklist to ensure you're covering the requirements (More information may be found in Task 2 Prompt C.):

- ✓ Implement **one** descriptive method and **one** non-descriptive (predictive or prescriptive) method.
- ✓ Use of collected or available datasets.
- ✓ Includes a decision-support functionality related to the non-descriptive method.

- ✓ Use methods and algorithms supporting data exploration and preparation to make data useable.
- ✓ Implementation of interactive queries to access specific data.
- ✓ Include a machine-learning method(s) and algorithm(s) to support the decision making functionality.
- ✓ Incorporate functionalities that will evaluate the accuracy of the application's outcome.
- ✓ Provide an industry-appropriate security features that protects access to the application of data.
- ✓ Include a tool(s) to monitor the applications functionality for maintenance purposes.
- ✓ Include a user-friendly, functional dashboard that includes *at least* **three** visualization types (not just images) for data exploration and inspection.

Here is a good site that explains some popular machine learning/analysis algorithms: https://www.kdnuggets.com/2016/08/10-algorithms-machine-learning-engineers.html

Topic Approval form – Provide information about your topic including Analysis, Design and Development, and Implementation. Information should be high-level and considered a general overview of what you plan to accomplish. You must complete all sections including the language and database you plan on using.

Submit this form as a Word document to your CI for a signature. Your CI will either return it with a request for more information or will sign it and return it to you as a .pdf.

Release form – Here is where you will indicate whether or not your project and related documentation has restricted or sensitive information. If the project does not contain that material, type your name and date on the form. If it does, you'll need to also have the Third Party release form signed and submitted as well. It is highly recommended that, if your project has restricted material, you find a way to redact it and/or autotomize the setting so you may avoid using the Third Party Release Letter template (Authorization to Use Restricted Information).

Please ensure that only ONE of the options in the Capstone Release Form is selected on the release form. When the second option is selected, the third-party letter of authorization (Authorization to Use Restricted Information) must be included in the submission.

IRB form – To complete this form be sure to take the Human Subjects quiz found on the Course of Study under the Course Information section. You'll receive the results of the quiz via email. Cut and paste those results in the IRB form under Appendix A. Complete the parts in the form outside the block the Course Instructor will complete. This includes justification of why the project complies with IRB standards for exemption status. Most capstone proposals are exempt from IRB review as they do not engage human subjects and omit WGU related information. If an IRB review is warranted, your Course Instructor will advise you or a comment in your Task 1 evaluation will indicate further actions might be needed,

Submit this form as a Word document to your CI for a signature. Your CI will either return it with a request for more information or will sign it and return it to you as a .pdf.

Submit Task 1: Include the signed approval form, signed IRB form, and the release form (and a signed third-party letter of authorization if your proposal uses or relies upon restricted information (confidential and/or proprietary information owned by a third-party).

A Course Instructor must sign the Topic Approval and IRB Forms. Without a signature in any of the PDF documents, the task will be returned.

Task 2

Prompt A

This prompt requires two documents:

1. Letter of Transmittal -- Create a letter of transmittal (a single-page cover letter) to the client's senior leadership (non-technical background). This needs to be brief and concise addressing the following:

- summary of the problem
- recommendation of the solution (data product and type)
- description of how the proposed solution benefits the client
- objectives of the project
- total funding requirement
- expertise of developer relevant to the solution

The letter of transmittal should include all elements typically found in artifacts of business communication, including a subject line, date, inside address, greeting, complimentary close, and signature.

2. Project Recommendation -- A project recommendation to convince those same senior, non-technical executives to implement the data product you are proposing to design to meet their business need. This is not a lengthy document and should have minimal technical jargon (or explanations if jargon or technical terms are used), but sufficient material to explain what will be accomplished. Around two to three pages should suffice. Remember, you're establishing the context for your project and what it will accomplish for the client. The summary should include *each* of the following and *be written in future tense*:

Problem Summary -- A summary of the problem that the application will solve. This should include a comprehensive description of the setting so it's clear how the solution meets the client's need(s). This is a scope statement complete with what will be included and incidental items that will not.

Application Benefits -- A description of how the application benefits the client and how the application will support a decision-making process in the context of the business need. Where does the product fit in filling an identified gap?

Application Description -- A description, using technical details, of the data solution application and how it aligns with the client's business priorities you're planning to solve.

Data Description -- A description of the data that will be used to construct the data product including its type (nominal, real numbers, etc...), origin, and structure. Identify the independent and dependent variables. Comment on the anomalies of the dataset and the limitations of the data collected.

Objective and Hypotheses -- The objectives represent the primary desired outcomes for the project and application. State what you want the application to achieve in relation to the goals. The hypotheses should be structured as a proposed explanation for a phenomenon(s) based on what the application will produce. State the "if" condition and the "then" outcome which is the prediction of general outcomes. You'll describe the validity of your hypothesis in Prompt D.

Methodology -- An outline of the project methodology. Choose an industry standard methodology (e.g. ADDIE, Agile, or SDLC) that you'll use to manage your project. Describe why that methodology is appropriate and then indicate what parts of the project will align with the methodology phases.

Funding Requirements-- The project's funding requirements. Taking into consideration environment, personnel, and any licensing or programming tools required what would be the overall cost involved. Be sure this amount matches the letter of transmittal.

Stakeholders Impact – Stakeholders are individuals or organizations with a vested interest in the project including being affected as a result of project execution or project completion. They may also influence the project's objectives and outcomes. Stakeholders may be internal, or external (e.g. customers), or members of a community. Describe the impact of your solution might have on the project stakeholders.

Data Precautions – Identify the data from the dataset(s) that are sensitive and/or protected. Then review the general guidelines related to working with that data. Discuss the general guidelines related to working with the sensitive data. Precautions may include data preparation, data analysis, data storage, access to data, and data dissemination. Explain the relevance between the precautions and application treatment of said data. If these considerations do not apply to the project and the data you are using, review the ethical and legal precautions commonly used when working with sensitive data. Obvious situations would be health care (HIPAA), education (FERPA), or payment (PCI DSS).

Developer's Expertise – Provide information (e.g. academic training and professional expertise) about the software developer and how he/she is qualified to complete the project in a timely fashion. Be sure to relate it to the solution you're proposing. This information may be real or hypothetical to fit the project topic and scenario you have created.

Prompt B

Write a proposal for the technically savvy IT professional leadership of your client. This will contain industry appropriate jargon and sufficient technical details to describe the proposed project and resulting application. Remember, you're establishing the technical context for your project and what it will accomplish for the client. This should be around 8 – 10 pages. The proposal should include *each* of the following and *be written in future tense*:

Problem Statement – Describe the problem or opportunity you are solving. This could relate to the need for accurate analysis or predictive representation of the data.

Customer Summary – Provide a description of the customers or users of the application and how this product will fulfill their needs. What is the environment where the application will be used and what special skill sets might be necessary?

Existing System Analysis – Summarize the technology environment prior to project initiation and the desired state of the environment upon project completion. This is a

systems analysis for the project deliverables. This might also include a review of the existing functional gaps in the data products you are replacing.

Data– Expand on the nature of the data you're using including: how it is to be collected, what techniques will be required to make the data usable, how you'll deal with undesirable exceptions or data anomalies (outliers, incomplete data). If appropriate, identify the additional data that are required for the stages of the data product lifecycle (design, development, maintenance, or others).

Project Methodology – Elaborate on your industry standard methodology by providing specific details about what aspects your project will be managed by each of the methodology phases. How will the methodology guide and support the data product design and development? Include levels of testing (e.g. unit, integration, system, and acceptance) as part of the methodology.

Project Outcomes – Describe the deliverables associated with the design and development of the application. List and describe those deliverables. Also, include examples to help clarify what specific type of artifacts will qualify. Use categories to such as Project Deliverables (In-process such as schedule, test plans, wireframes, or mockups) and Product Deliverables (Final outcomes such as functional GUI, maintenance tools, custom access, or specialized database.) to help clarify the logical distinction of the items you identify.

Implementation Plan -- Explain how the project will be implemented. This has to do with how the software application will be put into the production environment, not how it will be created. So, consider the following:

- Strategy for implementation
- Phases of rollout
- Details for levels of testing and final distribution
- Dependencies and milestones
- Deliverables – both tangible and intangible
- User testing -- if applicable

Evaluation Plan – Describe the methods for validating and verifying that the developed application meets the requirements and subsequently the needs of the client. Discuss your approach to quality assurance and what criteria that is based on. Consider industry standards, regulatory requirements, and organizational policies. What metrics will be used, how measured, and what measurement indicates success. What are the plans for the testing cycles and closures?

Resources and Costs – Elaborate/itemize the costs required to test and complete the project in a production environment. Be sure to include the following:

Programming Environment -- Provide a clear picture of what hardware and software are required to complete the project.

Environment Costs -- Provide an explanation of the costs associated with the application. Some might be startup, first-time costs while others might be a percentage of licensing costs. Environment costs are relatively minimal. The environment where the system resides in a shared environment where costs are shared by the organizations.

Human Resource Requirements -- What is the time and cost for the labor to complete the application?

Timeline and Milestones– Generate a projected timeline, including milestones, start and end dates, duration for *each* milestone, dependencies, and resources assigned to *each* task. Use a table to display your timeline material.

Prompt C

Design and develop a fully functional data product (application) addressing your identified business problem or organizational need. Your ability to explain how your application meets said need is essential so remember it's the overall context what your application is providing. To the best of your ability, include *each* of the following attributes:

Data Methods – provide **one** descriptive method that discerns relationships and characteristics of the past data in at least three forms of visualization. Also, provide one nondescriptive where a decision or trend could be inferred. The descriptive method should be in the domains of cluster or association analysis, and the others could include pruning algorithm, discriminate analysis, regression analysis (linear, logistic), Bayesian methods, neural network, or support vector machines. Be sure to include a justification of the methods selected in relationship to the goal(s) of the project.

Examples of acceptable descriptive methods:

- K-means clustering
- Hierarchical clustering
- Other clustering methods (max distance, min distance, etc.)
- PCA (all variables need to be numeric)
- MCA (all variables need to be factorized)
- FAMD
- Logistic regression with interpretations of estimated coefficients

Multi-linear regression with interpretations of estimated coefficients

Examples of appropriate non-descriptive methods:

- Logistic regression
- Decision trees
- Random forest
- Neural network
- Multi-linear regression (if it includes a strong justification – like ease of interpretability)

Datasets – The use of dataset(s) is a critical element and involves the gathering and measuring of information on targeted variables in a systematic fashion. This could be student collected (Please consider IRB ramifications.) or publicly accessible such as websites (e.g. Kaggle.com), governmental (e.g. Department of Labor), or software related (e.g. GetHub.com). Be sure to consider the methodology used including possible disadvantages and challenges.

Analytics– Using the given data, your application needs to enable decisions to be formulated or support for given trends to be provided.

Data Cleaning – if applicable, create a function that will make the data usable prior to actually being used by the application. Things such as featuring, parsing, cleaning, and wrangling the datasets.

Data Visualization – You need at least three real-time (e.g. using the GUI/dashboard) formats to visualize the data in a graphic format. Look at things like charting, mapping, color theory, plots, diagrams, or other methods (tables must include heat mapping). These, in conjunction with or as a part of your GUI, would enable users to explore or inspect the data characteristics.

Real-Time Queries – As part of your GUI enable users to access and manipulate data real-time including data maintenance. This does not deal with data "freshness" but with the query response time being in seconds.

Adaptive Element – if appropriate for the business need, provide the implementation of machine-learning methods and algorithms to enable the application to improve with

experience. Examples include learning associations, classification, statistical arbitrage, prediction, extraction, and regression.

Outcome Accuracy – provide functionalities that evaluate the accuracy of the information/outcomes given by the application. What are the parameters for valid output data and how will those be checked by the application?

Security Measures – include industry-appropriate security features that will control access to the data and/or, how the data is stored or transmitted. The security features should be appropriate for the data product and the sensitivity of the data it interacts with. For example, with a web application this requirement might be satisfied by implementing username/password authentication.

Product Health Monitoring – include functionality that will enable the application's "health" and reliability to be monitored. It should answer the questions, "Is the application performing correctly?" For example, the use of displays or logs to quickly discover, isolate and solve problems that can negatively affect the application's performance and accuracy.

Dashboard – include a user-friendly, functional dashboard that enables the query and display of the data, as well as other functionality described in this section. This could be stand-alone, Web-based, or a mobile application interface.

Prompt D

Create a post-implementation report that addresses the various aspect of the project, providing sufficient detail so a reader, unfamiliar with the project, can understand what you have accomplished. The use of graphics (including screenshots) is highly encouraged to help explain and substantiate the capabilities of your application. Be sure your report includes the following elements with an accurate Table of Contents and cover page. This should be written in past tense:

Project Purpose – reiterate the business or organization problem that this application solved. How did the application address the "vision" or expectations of the client? Summarizes the technical functionality and end-user requirements that were met.

Datasets – provide examples of the raw datasets used by the application. If applicable, also include the cleaned datasets along with a sample of the code (or executable file) used to clean the data. If the data was cleaned or manipulated in any way, provide explanations of the interventions. If not, state so and why.

Data Product Code – review the functionality of the code used to perform the analysis of the data and how it provided the necessary functionality of descriptive and predictive outputs. You will also need to submit the entire, functional source code.

Hypothesis verification – discuss if and why the initially established project hypothesis was accepted or rejected based on the use of the application. Did your assumptions about the phenomenon included (described) in the data prove true or not. Rejected hypothesis are okay but explain why that was so.

Effective Visualizations and Reporting – provide a description of how your visualizations and elements effectively told an accurate story about the data. This needs to include items such as how your application supported data preparation, data analysis, and data summary. It should also include how your display techniques clearly explained any phenomenon detection if appropriate.

Accuracy analysis – assess how accurate your application is at presenting the data and providing predictive outcomes. Provide an example of what the data showed and explain why those offer representative artifacts of the application's accuracy.

Application Testing – clearly describe the different levels of testing you used to confirm the functionality of your application. These could include unit, integration, system, and acceptance. Also, did you conduct any beta tests or usability tests for your project? Summarize those tests and explain how results were used to modify/calibrate the product during its development.
Application Files – provide a comprehensive inventory of the files required to execute your application. Include a clear description of the interdependencies and file hierarchy. Describe how those files will be organized for submission. Include those files in the submission.
User's Guide – include a brief manual concerning the installation and use of your application. Be sure to describe all steps necessary to establish an environment capable of running your application. Provide clear, concise steps of how a user would execute the application and produce the results you've described in your documentation. Carefully consider and describe the procedural aspect of the application including know areas where certain crucial steps can affect the performance of the application. You must ensure that anybody can run the application. Please include details on the technology context that is required for your application to properly execute.
Summation of Learning Experience – describe how your prior experience assisted you with the capstone project then review the additional professional development you had to pursue to complete the capstone. What assistance did you seek out from other individuals? How has the experience contributed to your concept of life-long learning?

Section E

Acknowledge sources, using The Big Four method. The **Big Four** includes Author, Date, Title, and Location of information (e.g. publisher, web address, journal.) As an example, some sites may not have an author. For this reason, please exercise your own judgment to delineate if enough information on the sources is presented to pass the aspect.

Big Four:

- *Author*
- *Date*
- *Type*
- *Location of information (publisher, URL, journal)*

To be sure your citations provide enough information it's highly recommended that you use APA-formatted in-text citations and references.

Section F

Demonstrate professional communication in the content and presentation of your submission. It is paramount that you provide a well-organized, professionally formatted set of documents. Remember, the evaluator will be looking at the overall presentation of this project. Ensure that

all material is accurate and provide the proper context concerning the nature of the application and how it functions. Clear, concise, and accurate material will go a long way in ensuring your submission is successful.