

# Microbiology Lab Information Management and Visualization System

DESIGN DOCUMENT

Team 29

Client: Karrie Daniels

Adviser: Thomas Daniels

Brittany McPeek – Report/Documentation Manger

Benjamin Vogel – Team Manager

Rob Reinhard – Meeting Facilitator

Kyle Gansen – Meeting Scribe

Ben Alexander – Progress Manager

Samuel Jungman – Chief Engineer

Team Email: [sddec20-29@iastate.edu](mailto:sddec20-29@iastate.edu)

Team Website: <https://sddec20-29.sd.ece.iastate.edu>

Revised: 3/29/2020 / Version: 0.2

# Executive Summary

## Development Standards & Practices Used

- Detailed Use Case diagrams and User Stories
- Simple Design
- Pair Programming
- Heavy documentation in both code and manuals, updated regularly
- Single coding standard (Will use PEP-8)
- Often refactors
- Simple designs and simple APIs
- Don't write code you might need in the future, but don't need yet
- Construct and maintain a user manual for non-technical clients

## Summary of Requirements

- System should support importation of large amounts of data in a reasonable amount of time
  - Including from sources such as CSV and Excel
- The system should be able to abstract the data and place them into modifiable graphs
- The system should support exporting those graphs to be published into research papers
- The system should be able to maintain past data and support adding and modification with new data
- The system will be written in Python
- The system should be easy to understand and use by users with little-to-no background in programming
- The system should be maintainable by 1 to 2 IT personnel

## Applicable Courses from Iowa State University Curriculum

- COMS 309

- COMS 227
- COMS 228
- COMS 363

### New Skills/Knowledge acquired that was not taught in courses

- Graphing and graph visualization
- Python
- Data management/backup

# Table of Contents

1	Introduction	5
1.1	Acknowledgement	5
1.2	Problem and Project Statement	5
1.3	Operational Environment	5
1.4	Requirements	5
1.5	Intended Users and Uses	6
1.6	Assumptions and Limitations	6
1.7	Expected End Product and Deliverables	7
2.	Specifications and Analysis	9
2.1	Proposed Approach	9
2.2	Design Analysis	9
2.3	Development Process	10
2.4	Conceptual Sketch	10
3.	Statement of Work	11
3.1	Previous Work And Literature	11
3.2	Technology Considerations	11
3.3	Task Decomposition	12
3.4	Possible Risks And Risk Management	12
3.5	Project Proposed Milestones and Evaluation Criteria	13
3.6	Project Tracking Procedures	13
3.7	Expected Results and Validation	13
4.	Project Timeline, Estimated Resources, and Challenges	14
4.1	Project Timeline	14
4.2	Feasibility Assessment	15
4.3	Personnel Effort Requirements	15
4.4	Other Resource Requirements	16
4.5	Financial Requirements	16
5.	Testing and Implementation	18
5.1	Interface Specifications	18
5.2	Hardware and software	18
5.3	Functional Testing	18

5.4 Non-Functional Testing	18
5.5 Process	18
5.6 Results	18
6. Closing Material	20
6.1 Conclusion	20
6.2 References	20
6.3 Appendices	20

**List of figures/tables/symbols/definitions** (This should be the similar to the project plan)

Table 1: Deadlines - page 7

Table 2: Deliverables - page 7

Table 3: Personal Effort Requirements - page 15

Figure 1: Conceptual Sketch - page 12

Figure 2: Project Timeline – page 14

# 1 Introduction

## 1.1 ACKNOWLEDGEMENT

Our group would like to acknowledge and thank Thomas Daniels for his guidance, support, and technical advice throughout this project. Our group would also like to acknowledge and thank Karrie Daniels for providing us information and requirements for this project, as well as act as a sample user of our end product. We appreciate the time and commitment these two individuals have donated towards this project.

## 1.2 PROBLEM AND PROJECT STATEMENT

Many scientists and researchers dedicate large amounts of time towards organizing, maintaining, and visualizing the data they collect. The purpose of this project is to find a solution to this problem. The solution should be able to automate the process of organizing, maintaining, and visualizing data. It is important that scientists and researchers have more time to collect and analyze their data, especially in time-sensitive experiments; thus resolving the issue of organizing, maintaining, and visualizing their data will be beneficial to scientists and researchers.

Our group proposes creating an application that allows the user to import pre-existing data from Excel and manage and visualize the data. The application will allow users to select data elements and a type of graph/statistical analysis and visualize the resulting information in the form of a graph or another type of visual. The graphing utilities will allow the user to customize the appearance of the graphs to be created and will meet scientific publication standards. Additionally, the application will save backups of the data and allow the data and visuals to be exported and shared with another person. Our hope is to create an easy-to-use application that does not require too much maintenance and allows scientists and researchers an easier method to organize, maintain, and visualize their data.

## 1.3 OPERATIONAL ENVIRONMENT

Our end product will most likely be only software based. The software's environment is any computer that it runs on. Our end users will most likely be in a climate-controlled, indoor location which means that a computer will be able to operate without any problems.

## 1.4 REQUIREMENTS

Our client wants to be able to use this software to log and manage data about microbiology experiments. Here is a list of functional requirements that the client wants:

- Data import from Excel
- Data modification after import
- Custom data visualization based on specified data elements
- Data sets and graphs should be saved to the file system
- The data should backup to a separate location

- Function to export the data to be shared with someone else
- Creation of projects that contain multiple experiments
  - Collation of multiple graphs from similarly based experiments

Our non-functional requirements are:

- It should be possible for the system to be maintained by one or two people
- The system should be secure enough so that research data can't be seen by anyone else
- We must use libraries in Python for data visualization
- The data must be parsed after it is imported

## 1.5 INTENDED USERS AND USES

This project's end product is intended to be used by scientists and researchers for inputting, organizing, and visualizing large amounts of data efficiently and effectively. The visualizations created by the product should meet scientific publication standards so that these visualizations can be used in published reports, papers, etc. Non-technical persons should be able to utilize the end product with ease. Many of these users are not particularly 'tech-savvy' and are not accustomed to complicated and unintuitive software. This experience places high importance in the intuition and cleanliness of the UI of our software. Furthermore, this software will not have a real maintainer or IT team after it is delivered, so the project will need to be clean and bug free to match the lack of maintenance staff.

## 1.6 ASSUMPTIONS AND LIMITATIONS

Assumptions:

- The maximum number of users per instance of the product will be one
  - The client prefers the data to be modifiable by one person, with additions of simultaneous editors as a stretch goal. Given the size and complexity of the data, having simultaneous users modifying and graphing data from one set could provide a daunting task.
- The solution will not be distributed outside of Iowa State
  - Due to the specific solution the end product brings to the research department, the use of the product (at least in its early stages of development) will not be useful outside of the university.
- Python and the Python Interpreter will be used for the development
  - By the request of the client, using Python with the product should provide easy maintenance if needed by people who may be unfamiliar with the development of the product itself. This added with powerful graphing utilities can provide a solid solution to the client's problem
- The end user will require an instruction set about the end product
  - Due to the technical knowledge of the end users, a simplified instruction manual will be required to help convey the usages and information needed to operate the solution.

Limitations:

- The end product shall be able to be maintained by 1-2 IT workers on a minimal basis (Client Requirement)
- The end product shall cost the end user no more than \$200 (less than cost of current client’s solution)
- The end product shall be easy to run and navigate with little-to-no programming experience (Client Requirement)
- The end product shall be based on, and released for, the ISU research department (Geographical Constraint)

### 1.7 EXPECTED END PRODUCT AND DELIVERABLES

The final product will parse excel files and output graphs based on the data provided. The end-user will also be able to customize these graphs layout as well as select the variables inputted into the graphs so the project’s use extends outside of the provided data-set. Finally, the end product’s components will be decoupled to allow further customization if the needs of the end-product evolve following delivery (i.e. new types of data or graphs are needed).

The design of the project will be completed by the end of April, and the end product and documentation will be finalized by the end of December 2020. Documentation code will be ongoing as we meet each date. A look below shows a list of our deliverables over the course of the next two semesters.

Table 1: Projected deadlines for the project

ID	Date	Deliverable
D1	April 30th	Framework for each component of our project will have been created.
D2	September 30th	Prototype of each component of our project.
D3	October 15th	Functional Product
D4	November 15th	Final Product
D5	November 30th	Testing
D6	Ongoing	Documentation for above deliverables

Table 2: A description of the deliverables for the project

ID	Description
D1	By this point, “Hello World” programs will be built on the technological stack we have chosen. In addition, the preliminary frameworks for importation, graphing, exportation, and the GUI binding them together will have been created.
D2	Essential features of each component will have been implemented. Data can be imported from an Excel file. Then, you will be able to select data to create at least one type of graph. The graph will



	generate, and then you will be able to adjust the graph if needed. Finally, you will be able to export a high-quality image of the graph. In addition, data will be saved.
D3	Each of the core features will be fully implemented. Every type of graph requested will be available for creation, and the importation and exportation settings will be available. More types of files will be available for importation and exportation. In addition, backup snapshots will be implemented in case data is lost. Finally, the product will be able to calculate correlation and perform T-Tests and P-Tests from the graphs.
D4	Modifications requested by our clients looking at the functional product will be taken into account, and stretch-goals, if possible, will be implemented. Bugs that occur in normal boundaries will be squashed.
D5	While testing will occur as we develop the product to ensure proper functionality in normal circumstances, these two weeks the testing will be done in attempts to “break the product.” This way, our final product we created for the client will continue to properly function even in scenarios we didn’t account for.
D6	Two types of documentation will be provided: one for the end-user, and one documenting the code behind our end-product. The end-user documentation will consist of tutorials for the end-product. The code documentation will provide a detailed look at all classes and functions making up the back-end. The documentation will also outline how these classes and functions relate to each other. This will be demonstrated by providing “under-the-hood” looks at the tutorials in the end-user documentation. Finally, the documentation will show how classes and functions can be extended.

## 2. Specifications and Analysis

### 2.1 PROPOSED APPROACH

Our proposed approach is to create a local application using Python. The PEP 8 style guide for Python Code will be used during development. The user will interact with the program through a GUI likely to be made using the Tkinter Python library. The various functions of the application will then be split into components which interact with each other through interfaces. This should allow for independent development of the various components of the software. A figure showing how these components relate to each other is available in section 2.4.

The design addresses the functional and non-functional requirements from section 1.4 in the following ways:

#### Functional Requirements

- Data import and parsing from Excel will be handled through a data import component of the software. Python has a module “xlrd” which can be used to extract data from Excel spreadsheets.
- Data modification after import will be handled through a data holding component of the software. Python also has a module “xlutils” which can be used to modify an existing Excel sheet.
- Custom data visualization of the data will be handled through data visualization and graphing components of the software. The open source Python library “Plotly” is one potential library to use for these components.
- The graphs and data will be saved through a save files component of the software. Plotly allows static images of plots to be saved, and the data can be saved to new Excel files using the “xlwt” Python module.
- Data will be backed up to a separate location through a data backup component of the software. After data has been imported into the software, it can be written to a file in a separate location.
- A share files component of the software will allow for files created by the application to be uploaded to Google drive.

#### Non-Functional Requirements

- To ensure the system could be maintained by one or two people, a local application approach was chosen. Excluding a database from the design of the software eliminates the need for database maintenance.
- The research data will be secure due to the software being a local application.
- It was requested that Python libraries be used to visualize the data, and the Python library “Plotly” is robust enough to cover the functional requirements.
- The various Excel modules mentioned in the functional requirements allow for parsing the data after it has been imported.

### 2.2 DESIGN ANALYSIS

One of the strengths of the design in theory is the modularity of it. Breaking the software into separate components should allow for team members to work fairly independently of each other on the software. Another strength of the design is that it allows for new functionality to be added to the software without having to overhaul the design. New functions can be implemented as new components in the software.

One potential weakness is that all processing of the data will be done locally. The speed at which this occurs will depend on the processing power of the computer running the application, whereas a client-server application could allow for processing of data on the server side.

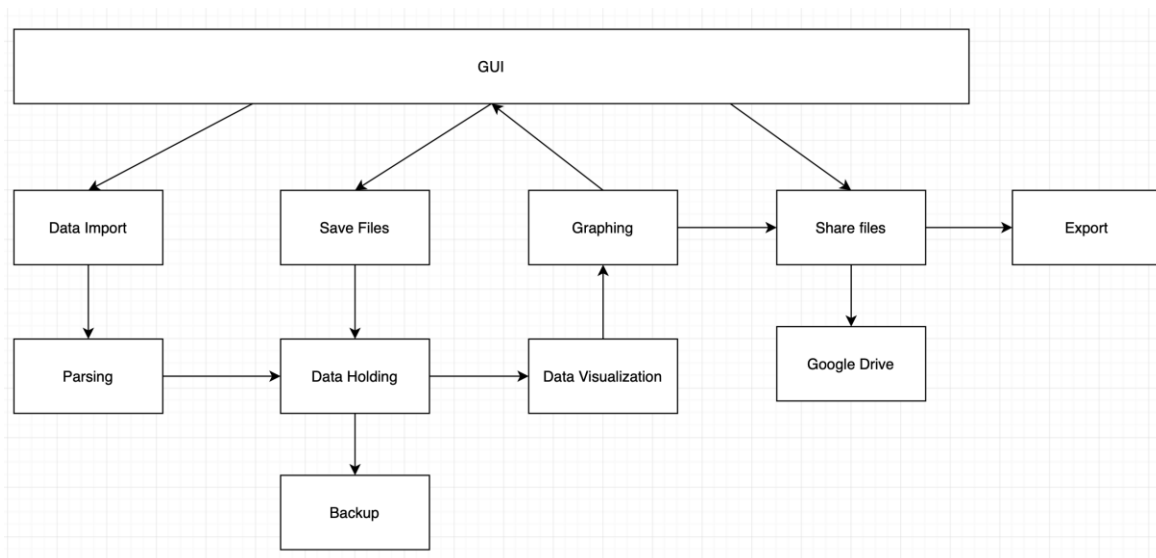
### 2.3 DEVELOPMENT PROCESS

The development processes will be an agile based development. We will use 1-2 week iterations to produce predefined goals. These goals will be written out in the form of story cards that will encapsulate use cases with diagrams to allow for easy to follow documentation and maintenance after the project is delivered. We will use a Kanban board of some kind to track the progress of story cards. The reasoning behind this process is that it will allow us constant feedback from our client as to the specifications and implementations of the project. This is a project that will require input as the design and intuition of the UI has high importance in the overall quality of the final deliverable.

### 2.4 CONCEPTUAL SKETCH

For each element in the diagram we would have an interface that abstracts each component and thus, simplifying how the elements interact. The benefits from using interfaces here is better collaboration. Each team member doesn't have to know everything about another component in order to use it, they can just use the interface.

Figure 1: Conceptual sketch of the application and interactions between modules



## 3. Statement of Work

### 3.1 PREVIOUS WORK AND LITERATURE

The previous work we are basing ourselves off of is the solution our client is currently using, GraphPad. This technology is designed specifically to take data and organize it similarly to a spreadsheet and then provide graphing utilities to help visualize the data. GraphPad itself does not operate data entry as a spreadsheet, but a more specialized version where they offer special data tables that can be catered to how the client wishes to organize.

One of the current major downsides to this technology is the price. GraphPad can be extremely expensive on a yearly subscription, especially when more than one person needs to have a license for it. The other downside that the client has leveraged to us is the lack of options for a robust suite of graphs. Currently, GraphPad only can visualize with bar graphs, which the client may use, but would rather work with more varied graphs such as scatter dot plot graphs and correlation graphs.

### 3.2 TECHNOLOGY CONSIDERATIONS

The possible solution we are bringing to the table would provide a free (or relatively inexpensive) variant to GraphPad. It would handle data in a similar way since that was valued by the client, but also utilize the power of Plotly (a Python library specialized with advanced graphing utilities) to provide a wider variety of graphs to use.

One strength of this solution would be the ability to use a free, open-source library on top of a free, open-source language and interpreter, which will be able to drive costs down on the development which can transfer over to the user. Python has a massive amount of packages that can be easily installed and incorporated into our project to help us not only provide a robust product for the client, but relieve a lot of the behind the scenes work for the developers (for example, importing and parsing large amounts of data). Another strength of this solution is due to the nature of Python and our structure, we will be able to make an end product that should be much easier to maintain and fix for people who are not as experienced in software development. By making the front end and the back end as accessible as possible for our end clients, we can ensure that the product will be able to live beyond the Senior Design class.

A downside to this possible solution is the nature of Python itself. Because Python runs on an interpreter, creating an EXE or application will be harder. There are packages and libraries to help with the process, but it just doesn't work as naturally as some other languages or solutions might. Another downside is the current solution is a local solution. The client, while making it a low-priority stretch goal, wanted to have simultaneous users. By keeping, modifying, and visualizing the data on a local machine, it will make a solution with simultaneous users difficult if not impossible.

### 3.3 TASK DECOMPOSITION

Our tasks can be decomposed into the modules similar to the structure of our solution in Figure 1:

1. GUI
  - a. Styling
    - i. Create a visually appealing front end that also shows all relevant data
    - ii. User should be able to edit styles to their own liking
  - b. Layout
    - i. Layout should be easily understandable by the end user
    - ii. Graphing should be the primary focus of the layout
2. Data Import
  - a. The client should be able to import CSV or EXCEL files to be analyzed and graphed
3. Parsing
  - a. The solution should be able to parse data from the files and sort them into data structures for better visualization options
4. Data Holding
  - a. Create dynamic data structures to contain the data for graphing uses
5. Saving Files
  - a. The user should be able to save a current file within the application
6. Backup
  - a. The system should store backups of the imported data in a hidden folder for data recovery
7. Data Visualization/Graphing
  - a. Create a system that integrates with Plotly to create and show graphs to the end user
  - b. The system should be able to handle the following types of graphs and computations: dots plots, scatter plots, bar graphs, t-test plots, p-value computation, and correlation computation
  - c. The graphs should be customizable in regards to the types of graphs, which variables are graphed, what color schemes are used, and what data point shapes are used
8. Share Files
  - a. Create an export tool that can either share to Google Docs through an external API or save it as a picture to their local machine.

### 3.4 POSSIBLE RISKS AND RISK MANAGEMENT

Unfamiliarity with both Python and some of the needed packages will become a possible risk for the group, though we are already taking steps to remedy the problem with research and practice before beginning on developing the actual solution.

In order to avoid most of the risks that come from developing this solution, we will be working closely to communicate to each other and the client in order to sort out misunderstandings and confusion as fast as possible. By doing so, we are able to mitigate risks faster and more effectively as they come up.

### 3.5 PROJECT PROPOSED MILESTONES AND EVALUATION CRITERIA

The key milestones in our proposed project are the framework, prototype, the functional product, and final product stages.

The framework stage is a key milestone because it will require us to build a theoretical model to implement. Since each of us are focusing on individual components of the project, the frameworks section is also critical because it will require us to have discussions on how data is passed between each component. The test for our frameworks are not programmatic, the test is that the frameworks we have built handle a consistent type of data and could work together.

The prototype stage is a key milestone because it will be the first time a product is delivered. By implementing core-features with the frameworks laid out in the previous stage, we will have our first demonstration of the product's components not only working, but working together. Testing will be done by passing in data through each component in a few key scenarios.

The functional product stage is a key milestone because it will be an implementation of all features we have been asked to deliver. Testing for these phases are done similar to the prototype stage. Using the same data, we test the features of that component, and then test the passing of the data to other components.

Finally, the final product stage is a key milestone because it will be a final tailoring of what we presented in the functional product stage for our client. Feedback will be taken into consideration, and stretch goals, if possible will be met. Testing is now for small adjustments, and similar to the functional product stage.

### 3.6 PROJECT TRACKING PROCEDURES

Our group is currently using Trello to manage and track progress for this and next semester. We also meet with the client regularly to discuss our progress and gain feedback so we know what portions of the project need correction and to narrow our next steps.

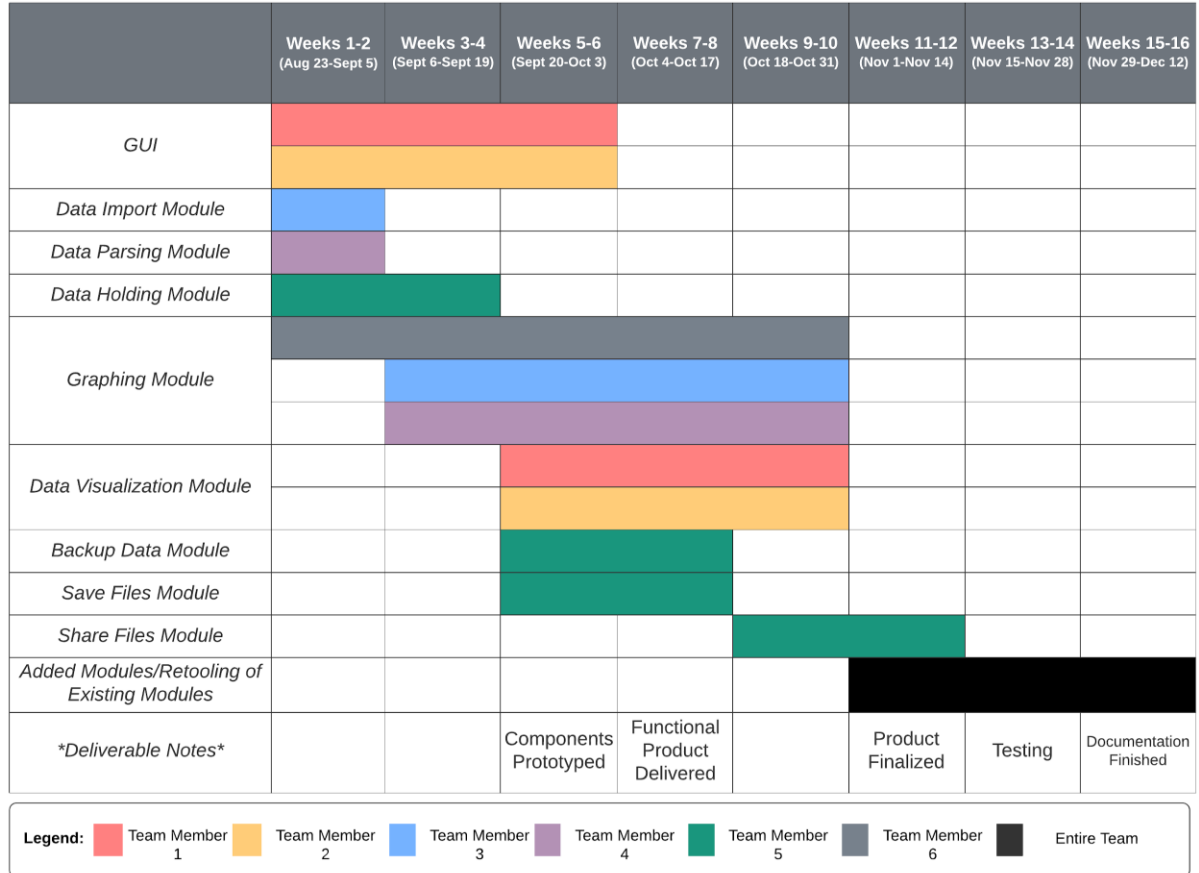
### 3.7 EXPECTED RESULTS AND VALIDATION

The desired outcome is a program that encompasses all of the tasks layed out in 3.3. Each of these tasks must behave properly within themselves as well as interact properly with one another. This can be verified at a **high level** using full integration tests in which individuals separate from our team will be asked to perform various tasks in a black box testing fashion that incorporate all of our key milestones and features that have been laid out in this document. It should be noted that this testing will occur only after rigorous Moq testing and unit testing on a low level as well as small scale integration testing at various points of development.

## 4. Project Timeline, Estimated Resources, and Challenges

### 4.1 PROJECT TIMELINE

Figure 2: Project Timeline



Work on the project on this proposed timeline begins with the GUI, data import module, data parsing module, data holding module, and graph module. These things set the groundwork for the project. For the future modules to be functional in the scope of the program, it is necessary that these get finished first. The GUI is vital for this project as it is the only way for a user to interact with the modules, so two team members are set aside for work on it for the first six weeks of the fall.

After significant progress has been made on the GUI and data can be imported into the program, the majority of the team is scheduled to work on the graph and data visualization modules. Because these are the key features of the programs, the graphing and data visualization modules will have five of the team's six members dedicated to working on them. After the product has been finalized, rigorous testing will be performed on the product. Testing will have already occurred throughout development, but this will be done to ensure the finished product is airtight.

## 4.2 FEASIBILITY ASSESSMENT

The project will be a standalone executable with its own files kept in the program location. It will encapsulate and allow for all of the main features previously discussed in this document. However, there are some foreseen challenges with this predicted project that have been laid out in the following paragraph as well as some preliminary plans to relive these challenges and improve the feasibility of this project.

The first main challenge will be the formatting and creation of the graphs themselves. Graphs will require some statistical knowledge and practice to be able to generate the graphs from the data. There will need to be some team research to accomplish this aspect of the project The second is the variety of Python libraries that will be needed to complete this project. Due to the immensity of Python and its libraries, our group has varying experience with Python and the libraries in particular and will need to conduct experiments and research to learn how to develop the project using these libraries. These present themselves as the main foreseen challenges of our project which is why this project seems to be well within the realm of feasible.

## 4.3 PERSONNEL EFFORT REQUIREMENTS

Table 3: Personal Effort Requirements

<b>Task</b>	<b>Description</b>	<b>Time Estimate</b>	<b>Difficulty Estimate (1 = Easy, 5 = very hard)</b>
GUI Styling	Create a desktop application that is aesthetically appealing and intuitive.	15 hours over 2 weeks	2
GUI Layout	The layout should be easy to navigate and the app functions easy to understand	15 hours over 2 weeks	2
Data Import	Import data into the app from an excel or csv file	10 hours over 2 weeks	1
Parsing	Analyze the imported data to find and sort relevant variables and measurements	25 hours over 3 weeks	3
Data Holding	The parsed data should be temporarily kept in the program to be used later	5 hours over 1 week	2



Task	Description	Time Estimate	Difficulty Estimate (1 = Easy, 5 = very hard)
Saving Files	Parsed data will be saved to the local storage in a csv file	10 hours over 2 weeks	1
Backup	Versions of program files will be backed up in the file system once every set time interval	15 hours over 2 weeks	3
Data Visualization	Use Plotly to create various kinds of graphs. User can customize data inputs, colors and shapes of data points	40 hours over 4-5 weeks	4
Share Files	Export files to Google Drive through API to share with other users	20 hours over 2 weeks	2

#### 4.4 OTHER RESOURCE REQUIREMENTS

This project will be completed entirely using Python and Python libraries such as Plotly. This means that the project will not require any physical materials or equipment, other than the computers we use to complete the project. The end product, an application, should not take up much space, and will mostly likely end up around 10MB. Furthermore, the application will not require a very powerful computer to run. This means no additional equipment will be required for our client to use the product since the client has computers available to them in their lab.

#### 4.5 FINANCIAL REQUIREMENTS

This project will only require the use of Python and some Python libraries. A Python install is completely free, and the libraries we will use, such as Plotly, are free and open-source. We also plan on using our own PCs to complete the project, so there is no additional cost in that regard. The client will not need to buy any special equipment to use our end product, as their computers in their lab will suffice. Therefore our project will not require any financial resources.



## 5. Testing and Implementation

Testing is an **extremely** important component of most projects, whether it involves a circuit, a process, or a software library

Although the tooling is usually significantly different, the testing process is typically quite similar regardless of CprE, EE, or SE themed project:

1. Define the needed types of tests (unit testing for modules, integrity testing for interfaces, user-study for functional and non-functional requirements)
2. Define the individual items to be tested
3. Define, design, and develop the actual test cases
4. Determine the anticipated test results for each test case
5. Perform the actual tests
6. Evaluate the actual test results
7. Make the necessary changes to the product being tested
8. Perform any necessary retesting
9. Document the entire testing process and its results

Include Functional and Non-Functional Testing, Modeling and Simulations, challenges you've determined.

### 5.1 INTERFACE SPECIFICATIONS

- Discuss any hardware/software interfacing that you are working on for testing your project

### 5.2 HARDWARE AND SOFTWARE

- Indicate any hardware and/or software used in the testing phase
- Provide brief, simple introductions for each to explain the usefulness of each

### 5.3 FUNCTIONAL TESTING

Examples include unit, integration, system, acceptance testing

### 5.4 NON-FUNCTIONAL TESTING

Testing for performance, security, usability, compatibility

### 5.5 PROCESS

- Explain how each method indicated in Section 2 was tested
- Flow diagram of the process if applicable (should be for most projects)

### 5.6 RESULTS

- List and explain any and all results obtained so far during the testing phase
  - - Include failures and successes

- – Explain what you learned and how you are planning to change it as you progress with your project
- – If you are including figures, please include captions and cite it in the text
  - This part will likely need to be refined in your 492 semester where the majority of the implementation and testing work will take place

-**Modeling and Simulation:** This could be logic analyzation, waveform outputs, block testing. 3D model renders, modeling graphs.

-List the **implementation Issues and Challenges.**

## 6. Closing Material

### 6.1 CONCLUSION

Summarize the work you have done so far. Briefly re-iterate your goals. Then, re-iterate the best plan of action (or solution) to achieving your goals and indicate why this surpasses all other possible solutions tested.

### 6.2 REFERENCES

This will likely be different than in project plan, since these will be technical references versus related work / market survey references. Do professional citation style(ex. IEEE).

### 6.3 APPENDICES

Any additional information that would be helpful to the evaluation of your design document.

If you have any large graphs, tables, or similar that does not directly pertain to the problem but helps support it, include that here. This would also be a good area to include hardware/software manuals used. May include CAD files, circuit schematics, layout etc. PCB testing issues etc. Software bugs etc.