COMP3850 Project Deliverable Certificate

Name of Deliverable	*Increment 2*
Date Submitted	*18/05/2023*
Project Group Number	*31*
Rubric stream being followed for this deliverable *Note: the feasibility study has the same rubric for all streams.*	*DATA SCIENCE Rubric*

We, the undersigned members of the above Project Group, collectively and individually certify that the above Project Deliverable, as submitted, **is entirely our own work**, other than where explicitly indicated in the deliverable documentation.

INITIALS	SURNAME	GIVEN NAME	STUDENT NUMBER	SIGNATURE *(IN-PERSON OR DIGITAL)*
TP	*Philip*	*Tabitha*	*46347763*	
CJ	*Johns*	*Cassandra*	*46462457*	*CassandraJ*
LY	*Yates*	*Lachlan*	*45984352*	*L.Yates*
AG	*Gardiner*	*Ava*	*46410961*	*AGardiner*
RA	*Ali*	*Rory*	*45901848*	*RoryA*

List of tasks completed for the deliverable and activities since last deliverable certificate with totals for each individual team member and whole team *(copy individual total row for each member and copy pages if more pages needed)*

Performed by *(Student Names)*	Duration *(hrs)*	Complexity *(L, M, H)*	Name of task	Checked by *(Initials)*
Tabitha Philip	.5	L	Obtaining feedback from sponsor regarding deliverables	CJ
Tabitha Philip	6	M	Creating visuals in Viya for findings report and justifications	CJ
Tabitha Philip	3	L	Revisions across D4 document	RA
Tabitha Philip	0.5	L	Editing and finalising documents.	RA
Cassandra Johns	.5	L	Feedback from sponsors	TP
Cassandra Johns	5	H	Encoding for the Viya platform including learning their coding language basics	TP
Cassandra Johns	6	H	Creating models for important factors to each category. Included learning about Gradient Boosted and Random Forests.	TP
Cassandra Johns	2	L	Described how to implement my python method of linear regression model.	TP
Cassandra Johns	3	L	Added and described my findings to the findings report	TP
Cassandra Johns	2	L	Imported findings into presentation demo	TP
Rory Ali	.5	L	Sponsor meetings and feedback incorporation	TP
Rory Ali	5	M	Creating visuals and points for sponsor driven findings report	TP
Rory Ali	5	M	Viya screenshots and process for increment 2 Viya guide	TP
Rory Ali	.5	L	Final checks and doc submission	TP
Ava Gardiner	.5	L	Obtaining feedback from sponsors regarding deliverables.	TP
Ava Gardiner	6	M	Creating visuals in SAS Viya for the Findings Report.	TP
Ava Gardiner	2	L	Revising D3 based on the feedback given.	TP
Ava Gardiner	1	L	Edited and formatted Document.	TP

Lachlan Yates	.5	L	Obtaining sponsor meetings and feedback.	TP
Lachlan Yates	2.5	L	Creating SAS Viya installation guide and intro to Scripts/Model execution.	TP
Lachlan Yates	2.5	M	Creating 'How to set up Models' section.	TP
Tabitha Philip total	10			
Cassandra Johns total	18.5			
Lachlan Yates total	5.5			
Ava Gardiner total	9.5			
Rory Ali total	11			
Team Total	54.5			

TABLE OF CONTENTS

REVISION TABLE

Revision Number	Heading	Date	Person	Change to Previous Version (what/why)
1	**Version Control**	24/04/23	Tabitha Philip	Feedback from Deliverable two indicates that a considerable amount of work can be done to improve the section including clarity around tracking changes within the project and how this is observed throughout the project's lifespan.
2	**Handover Requirements**	24/04/23	Tabitha Philip	This section was requested as per the university deliverable three. The handover requirements allow for a seamless transition to delivering the project to another team to carry on the work where necessary.
3	**Standards**	25/04/23	Tabitha Philip	Feedback from deliverable 2 suggests the project should look at standards that may apply to the project. When broken down into smaller sections the project can observe a number of standards around data and information security.
4	**Project Scope**	23/04/23	Tabitha Philip	Revised to include four different deliverables instead of the initial three and change the coding language used to Python instead of R.
5	**Data Collections and Captures**	23/04/23	Tabitha Philip	Revised to account for the international organisation databases we may interact with throughout the project

				in order to support the team's findings.
6	**Data Quality**	23/04/23	Tabitha Philip	Revised to account for tailored data sanity checks that are applicable for our data visualisation project.
7	**Data Types and Visualisations**	23/04/23	Tabitha Philip	Added to Account for the type of data interacted with within the project and paint an initial picture of what our final visualisations may consist of.
8	**Storage**	23/04/23	Tabitha Philip	Revised to account for software platform types utilised and storage methods used.
9	**Modelling Constraints**	23/04/23	Tabitha Philip	Added to account for lack of section as indicated by D2 feedback and voice nature of the project.
10	**Deployment**	23/04/23	Tabitha Philip	Revised to four sponsor deliverables and emphasises project type (visualisation emphasis)
11	**Insights from 'Correlation between features and SSI'**	15/05/23	Tabitha P	To address the gap in data visualisation around deriving insights from the findings produced. These findings are preliminary to our final version and will analyse additional factors to the specific countries in further detail.
12	**Insights from Figure 2**	15/05/23	Tabitha P	To address the gap in data visualisation around deriving insights from the

				findings produced. These findings are preliminary to our final version and will analyse additional factors to the specific countries in further detail.
13	**Insights from Figures 3-6**	15/05/23	Tabitha P	To address the gap in data visualisation around deriving insights from the findings produced. These findings are preliminary to our final version and will analyse additional factors to the specific countries in further detail.
14	**Insights from Figure 7**	15/05/23	Tabitha P	To address the gap in data visualisation around deriving insights from the findings produced. These findings are preliminary to our final version and will analyse additional factors to the specific countries in further detail.
15	**Insights from Figures 8 and 9**	15/05/23	Tabitha P	To address the gap in data visualisation around deriving insights from the findings produced. These findings are preliminary to our final version and will analyse additional factors to the specific countries in further detail.
16	**Insights from Figures 10- 12**	15/05/23	Tabitha P	To address the gap in data visualisation around deriving insights from the findings produced. These findings are preliminary to our final version and will analyse additional factors to the specific countries in further detail.
17	**Solution Architecture**	18/05/2023	Ava G	Feedback from Deliverable 3 suggests that the Solution Architecture section should

				showcase a data collection and analysis pipeline in the form of a diagram. Therefore, I have created a diagram on LucidChart that illustrates the high-level data collection and analysis pipeline of the project, showcasing the key steps involved in gathering, preparing, and analysing the data for the Zoe Empowers program.
18	**Model Evaluation**	18/05/23	Tabitha P	To account for the quantitative and qualitative aspects of the model evaluation, making note of the number of visuals produced per section and the feedback from SAS on two occasions. The addition of the most recent meeting has been added to strengthen justification of quality.

PROJECT PLAN

1. INTRODUCTION

Effective project planning, execution, and management are essential for the success of any project. With this in mind, DataSynergy has developed a comprehensive project plan to guide our efforts in measuring the effectiveness of the Zoe Empowers program using advanced analytics.

Our primary objective is to gain valuable insights into the program's impact, identify obstacles hindering its success, and provide actionable recommendations to improve its efficacy. DataSynergy will analyse and visualise data from Rwanda and Kenya datasets, using the ZOE_SELF_SUFFICIENCY_INDEX to understand the program's impact and effectiveness. Using advanced analytics, we will create 3 key documents: A Finding Report, an Analysis and Recommendations Report, and a Presentation. These reports will provide us with insightful information about the program's various aspects and aid in pinpointing areas that require improvement.

To ensure the success of the project, our project plan will cover the following critical areas:

- **Project Organisation:** We will define the project structure, roles, and responsibilities of DataSynergy. We will outline the reporting hierarchy, decision-making processes, and communication channels for the project to ensure everyone involved in the project is aware of their responsibilities and can work together efficiently.
- **Risk Management:** Our project plan will include an analysis of potential risks that could impact the project's success, along with mitigation strategies to address those risks.We will outline the risk management approach, the risk register, and the contingency plan to ensure that we are prepared for any unforeseen challenges that may arise.
- **Project Resources:** We will define the resources required to complete the project successfully, including human resources, technology, equipment, and materials. We will also outline the procurement plan, including the process for selecting and acquiring necessary resources.
- **Project Schedule:** Our project plan will include a detailed project schedule, including a timeline of all activities, deliverables, and milestones. We will also identify dependencies and critical paths to ensure that the project stays on track and within the planned timeline.
- **Tracking:** We will establish a system for tracking progress and performance against the project plan. We will include regular reporting mechanisms to keep SAS informed of project status, progress, and potential issues. We will also outline the process for making changes to the project plan as needed.

By addressing these critical aspects of project planning, DataSynergy will ensure that the project is executed successfully, achieves its objectives, and delivers high-quality results that meet SAS Institute Australia and Zoe Empowers' expectations.

2. PROJECT ORGANISATION

2.1. GROUP ORGANISATION

The success of our project depends on the organisation and management of the team involved. In this project, the team consists of five members who each play an essential role in ensuring the project's success.

PROJECT MANAGER: TABITHA

- Defining project goals and objectives.
- Creating and maintaining a project plan and schedule.
- Monitoring project progress and ensuring that it stays on track.
- Developing data models and visualisations to support decision-making.
- Communicating project updates and status reports to SAS.
- Providing technical support for data-related tools and technologies.
- Managing project risks and issues.
- Facilitating team meetings and ensuring that team members are collaborating effectively.
- Ensuring that the project is delivered on time and to the required quality standards.

PROJECT OFFICER/CHANGE OFFICER: AVA

- Supporting the development of the project plan and schedule.
- Coordinating project activities and ensuring that they are completed on time.
- Facilitating communication between team members and SAS.
- Helping to manage project risks and issues.
- Supporting the implementation of project changes.
- Ensuring that project documentation is up-to-date and accessible to all team members.

PROJECT OFFICER/DATA CONSULTANT: RORY

- Gathering and analysing data related to the project.
- Developing data models and visualisations to support decision-making.
- Ensuring that data quality standards are met.
- Providing technical support for data-related tools and technologies.
- Collaborating with other team members to identify opportunities for using data to improve project outcomes.

DATA SPECIALIST: CASSANDRA

- Collecting, analysing, and interpreting data to provide valuable insights for informed decision making.
- Ensuring the accuracy and quality of data to effectively communicate her findings and recommendations to our partner organisation.
- Mitigate risks associated with data.

CHANGE OFFICER/BUSINESS ANALYST: LACHLAN

- Conducting impact assessments to understand the potential impact of project changes on our partner organisation and the project.
- Developing change management plans and strategies.
- Support the implementation of change management activities.
- Providing business analysis support to the project team.
- Collaborating with other team members to identify opportunities for improving project outcomes.

Each team member has specific responsibilities, but we will also work collaboratively to achieve the project's goals. Effective communication, collaboration, and coordination are essential for the successful completion of this project.

2.2. CLIENT ORGANISATION

DataSynergy's partner organisation, SAS Institute Australia (SAS) is a global leader in business analytics software and services, driven by a strong commitment to helping clients make informed decisions through the power of data and analytics. The SAS team consists of three members who each play a critical role in guiding, supporting and supervising DataSynergy.

- **LUCY BIASI: ACADEMIC PROGRAM MANAGER (ANZ)**
  Lucy is responsible for overseeing the project team and ensuring that all project deliverables are completed on time and meets their expectations. As the Project Lead, Lucy will be the primary point of contact for our team and will be responsible for managing project timelines, coordinating meetings, and ensuring that all team members are aligned on project objectives. She will work closely with our team to ensure that we are meeting their needs and that we are delivering high-quality results.

- **JORDAN MOWLAI - TECHNICAL CONSULTANT**
  Jordan will be responsible for ensuring that the project is aligned with the overall strategy and goals of the organisation, and that it is completed on time and meets expectations. Jordan's expertise in project management and technical skills will be invaluable in guiding the team and providing technical direction throughout the project.

- **CHRIS GIBSON - TECHNICAL CONSULTANT**
  Chris has a deep understanding of the industry and extensive experience in working with similar projects. Chris will be providing guidance and support to the team and will help ensure that the project is executed successfully. His technical expertise and experience in managing complex projects will help ensure that the project meets the high standards of quality and excellence that are expected by SAS.

SAS is a committed partner organisation in this project and is dedicated to ensuring its success. Lucy, Jordan and Chris all bring a wealth of expertise and resources to the project and are eager to collaborate and support DataSynergy to deliver valuable insights into the impact of Zoe Empowers' program.

3. RISK MANAGEMENT

In the following section, DataSynergy will address potential risks that may arise throughout our project. We will present a detailed risk matrix that outlines the nature of the risk, description, likelihood of occurrence, impact and corresponding risk level. Additionally, we will provide effective strategies that DataSynergy can employ to mitigate these risks.

3.1. RISK IDENTIFICATION

To determine risks that may affect the completion of the project, the group allocated a significant amount of time during our in-person meeting in week 3. We compiled a rough list of potential risks and hazards that may present themselves throughout the project's duration. One thing that was prevalent to everyone was that due to the nature of the task, which is not heavily based on the process of building a software like many of the other pace tasks, it was difficult to refine the rough risks into a concise list while still conducting a substantial risk analysis process. Many topics and risks associated with aspects of development, such as prototyping, beta rollouts and software testing do not necessarily apply to the nature of the task, which includes the heavy use of the sponsored 3rd party software which we have no developmental control over and also the

deliverable structure of the project submitted in a number of different analytical reports, as opposed to submissions of built from scratch software versions.

With completing a risk register ease of understanding is paramount. Upon seeing the widespread use of a risk register that is split into risk categories, it was decided to proceed with this template to ensure that the sponsors and any others who are already used to this identification process would be able to understand our categorisation process with no trouble. The categories decided that would hold the most weight with our project were: human, operational, technical and external.

Definitions are as follows:

Human Risk	Risks arising from human behaviour, including negligence, errors, or malicious intent, which can result in negative consequences for the project. (E.g. An individual leaking sensitive information to unauthorised parties due to carelessness or lack of awareness about data security policies.)
Operational Risk	Risks arising from internal processes or procedures, including those related to data collection, management, and analysis, that can impact project outcomes. (E.g. Data is not collected or processed accurately, resulting in flawed analyses and incorrect conclusion)
Technical Risk	Risks related to the use of technology, including hardware, software, and infrastructure, which can result in system failures or other negative impacts on project operations. (E.g. Software systems experience a critical error or malfunction, which causes data loss, system downtime, or other negative impacts on the organisation's operations.)
External/Residual Risk	Risks that are outside the control of the project team or sponsors, such as those arising from unforeseen events or factors in the external environment, which can have a significant and unexpected impact on the project. (E.g. Black Swan event)

3.2. RISK REGISTER

The risk register currently being used by the team. This list is subject to change as the project evolves.

Risk	Number	Probability (L/M/H)	Impact (L/M/H)	Cause/Description
Poor communication, loss of contact with team	H1	L	H	Team member is absent without explanation for meeting or submission
Team member leaves project before completion (drops unit)	H2	L	H	One member leaves the unit for unknown reasons
Team member conflict which affects working environment	H3	L	M	Members clash over specific differences in ideas or personalities
Deadlines missed	O1	L	H	Team cannot keep up with assigned deadlines for project deliverables.
Report versions repeatedly rejected by sponsor	O2	L	M	Team cannot provide the sponsor with a suitable report thesis or structure.
Scheduling clash or duplicates produced	O3	L	M	Proper workload not divided or scheduled properly leading to team confused who is doing what, may end with group members doing the same task
SAS Viya goes through period of long outage	T1	M	M	3rd party SAS software used to create most data models experiences an unforeseen outage
Data given by Zoe Empowers changes or updates	E1	L	H	Zoe empowers conducts a revision of their own datasets changing them significantly, rendering work already done obsolete
Zoe Empowers is put under legal scrutiny or shut down	E2	L	H	The charity itself was suddenly put under scrutiny for factors out of the group's control. This will stop work completely

SAS and/or Zoe Empowers finds no benefit from completed work	E3	L	L	More of a morale issue, means that our work while accepted was not at the standard that the sponsor was expecting.

The team developed actionable responses to all of these main threats. The team also has as previously mentioned a process and complete agency to add to the risk register if new risks arise and also add to the threat mitigation table too, which will be updated and sent to all parties involved.

Risk	**Response**
H1	All minutes from every meeting are taken and posted within group Discord and tagged to everyone. This will notify everyone as long as they use their phone/desktop. These notes can then be referred back to if forgotten and also used to fill in anyone who missed a meeting. From a submission standpoint as a group we set a deadline for projects 2-3 days before the deliverable is due to make sure all parts are done and are available to review. This allows us to see clearly if a part is not done or needs revising well before the actual due date and hence allows the rest of the group to cover a part that might not be done in an extreme scenario.
H2	A member leaving mid project is a huge obstacle for the team to overcome. This is something that will immediately be taken to the unit convenor and sponsor to await further recommendations on how to proceed.
H3	Heated team member interactions can be dissolved by taking appropriate mediation actions to diffuse any tension while also discussing the situation with team leaders. In the case that a group member is consistently not contributing and has gained ire from all group members this will be escalated to the unit convenor.
O1	As mentioned in H1, all deadlines are well planned out in advance to help avoid official deadlines. Through the use of project management software such as Kanban visuals and gantt charts all group members know their assigned tasks and deadlines well in advance so any concerns of work not that wont be completed will be brought up well before the due date. If the workload becomes seriously too much for the group then the sponsor will be contacted to resolve work expectation issues.
O2	Sponsor is being kept informed of all ideas and processes being currently undertaken and worked on. This gives them the chance to advise if a certain aspect of a deliverable or idea is not going to work well before submission. This allows the group to be flexible and adjust to their needs, hence avoiding any disappointment in what is submitted.
O3	All work is divided properly in team meetings which is all written and posted to group discord and other workspaces. Project management software as well and gantt charts created by team leaders ensure everyone is aware of their parts in each deliverable.
T1	Upon any technical difficulties with the online SAS Viya for learners platform sponsor is immediately notified who then raises the issue with their respective team. Members are also all responsible for saving their own word and in particular downloading any important graphs and saving them within shared group workspaces.

E1	If data from Zoe Empowers changes, sponsors will contact us to inform and also advise on how to proceed properly. The team has already arranged a meeting with the SAS representative for the data collected so it is likely that this risk can be put to them directly, as well as ask for an appropriate guideline to follow if they do choose to update the data.
E2	In this scenario it is completely out of our control and we will wait for instructions from the sponsors and also university
E3	This can be avoided through strong teamwork, brainstorming and keeping the sponsors involved with all work and decisions made.

We believe that risks can be properly monitored and controlled by maintaining a proactive approach, while also thinking of new ways to take initiative with each deliverable. For example, setting advanced deadlines can help resolve deadlines and group work issues proactively. New frameworks will be developed as we start to work better as a group also. While these processes are in place, all human and operational risks will be adequately monitored. Technical and external risks monitoring and control is achieved by staying in constant contact with the sponsor. This is done by including them in all group in-progress documents, maintaining constant email communication, and holding bi-weekly meetings where we prepare questions ahead of time to make the most of the meeting time.

4. RESOURCE MANAGEMENT

For communication means the team has multiple points of contact with the main hub of communication and work sharing on a dedicated Discord server, which is organised into different channels for work and topic distribution. Messenger chats are also used for quick conversations and also a group Github for further version control of documents or any external code that is used and iterated on. In person meetings are booked bi-weekly on campus in a library meeting room to coincide with sponsor meeting times.

For the project all hardware requirements are covered by everyone's own devices. Software needed for the project is mainly the SAS Viya for learners platform, which we all have access to. Some members will use Rstudio for more advanced breakdown of certain data in a more familiar coding environment. Project management is being managed through software such as Projectlibre for work breakdown structures and Trello for general task management. Shared work is all done on Google Docs with the final presentation likely to be done using Canva.

For any questions regarding the project, we contact SAS directly using this email address academic@oz.sas.com.

5. CHANGE MANAGEMENT

5.1. MANAGING REQUIREMENT AND SCOPE CHANGE

The management of project requirements and scope change are achieved across an array of steps. Effective change management is crucial to ensuring the project is able to ongoingly adapt to change while remaining on course for development and completion.

Actively Anticipate Change	Team members agree that change may be necessary throughout the project. Staying open and adapting to new situations is key to managing change effectively.
Communicate Need for Change	Clear verbal or online communication between the team explaining the need for change and its impact on the project.
Team Involvement	All members are made aware of changing processes to the project. Members are also encouraged to share their ideas and perspectives on how we can manage change successfully.
Plan Development	The development of plans specific to change management e.g. reassignment of roles and responsibilities, setting new goals, adjustment of schedules, and other necessary changes.
Monitor Progress	Close monitoring of project progression will ensure change is handled and implemented effectively. Weekly meetings and frequent check-ins to assess the impact and management of change will ensure necessary adjustments are made.

5.2. CHANGES TO DOCUMENTS, CODE AND DATA

In this section, we will explore how DataSynergy plans to manage changes to documents, code and data, and what tools/methods are in place to facilitate effective change management.

5.2.1. VERSION CONTROL (REVISED)

Continuously changing environments of a project can prove challenging without the implementation of proper change management strategies. The team has decided to implement a number of tools to mitigate any risk associated with continuous change including lack of communications and awareness. The use of a dedicated project Github Repository to store deliverable documents, generated data visualisations, and data code allows the team to stay on top of version control through the constant observation of track changes and uploads. Github allows any number of team members to modify any document at a given time and see what changes have been made.

Google Docs is another tool the team utilises to create, edit and modify documents. The software as a service (SaaS) platform allows multiple team members to modify the same document in real time. The version history tool is particularly useful in observing track changes done by each team member and the ability to mention, tag and comment enables effective communications across the board.

To further enhance communications around version control team members are encouraged to voice concerns around dedicated sections within deliverables and work together collaboratively to rectify any issues. These communications tend to happen within out weekly meetings and are recorded within the weekly minutes for future reference and discussion.

5.2.2. TRACKING

The tracking of the project is accomplished using the visual project management tool Trello, real-time voice, video, text app Discord, and social media app Facebook messenger. Each enables team members to successfully distribute project information, documents and resources with one

another. Further, the use of weekly and fortnightly meetings between team members and sponsors, afford the ability to effectively track the project. Implementation of tracking tools, controls, and processes are essential to successful project progression and eventual completion.

5.2.3. TRACKING PROGRESS

Effective tracking of progress over the course of the project is crucial. Our use of tracking tools, such as Trello and Discord, were purposely chosen as they fall within the agile approach of our project, or the Kanban methodology. These tools help members visualise and understand our work better while simultaneously keeping on the same page as one another (Dan Radigan, 2019).

- **Trello:** The software is used as a project management visualisation tool that allows us to visually recognize tasks and subtasks needing to be completed. The visual aspect of the softwares interface helps team members more easily understand when deliverables are due, benefiting our ability to meet appointed deadlines. The interface encompasses project resources, to-do lists, completed tasks, questions for meetings, and weekly reports, ensuring the project is on schedule and meeting goals.
- **Discord:** The discord server created serves predominantly as a communication channel, however it effectively breaks areas of our project down into text channels. This break-down of text channels within the server ensures specific tracking of reports, resources, schedules, assessments calendars, meetings and general communication, directly contributing to the project progress.
- **Weekly/Fortnightly Meetings:** In-person and/or online meetings on a weekly basis between team members help keep track of ongoing change and progress of the project. Fortnightly meetings via zoom with SAS allows for a show of progress to the sponsor, and further space to ask questions and receive feedback, affording SAS the ability to track project progress, and further reinforce progress to members.

5.2.4. COMMUNICATION METHOD

Communication is an ongoing occurrence between both team members and SAS sponsors. To ensure communication is held to standard, it is carried out across weekly team meetings, fortnightly meetings with SAS sponsors, discord communication channels, and Facebook messenger. Communication was determined a key team value to guide us to achieve project goals, therefore utilisation of both appropriate and effective forms of communication management saw us adopt a diverse set of communicative tools. Communication methods adopted by the team each serve a particular purpose. There is no one particular communication stream that allows us to effectively communicate, hence the adoption of multiple methods was deemed appropriate. While Facebook messenger serves as a familiar communication method, it is not effective in sharing information and resources with members. Hence the adoption of Discord proved appropriate to navigate this weakness.

5.2.5. COMMUNICATION METHODS

Communication Type	Description
Weekly Team Meetings (online & face-to-face)	The scheduled weekly meeting is carried out amongst team members either via discord voice call or face-to-face meetings in a reserved study space at Macquarie university library. The weekly meeting allows team members to ensure project goals are remaining within scope, deadlines are being reinforced, need for change is communicated, and

	development of plans can occur, each allowing us to collectively monitor project progress.
Fortnightly Meetings with Sponsor	Scheduled fortnightly meetings via video-call on Zoom take place between team members and SAS sponsors. The meeting ensures ongoing communication between the team and sponsor, allowing the team to ensure the project is continuing to align with sponsor expectations, wants and needs.
Discord	A discord server is used by all team members as a communication channel to gather and share project information, resources, reports, weekly minutes, and assessment schedules.
Facebook Messenger	Used as a means of both group and individual communication. The app is of familiarity to all team members, where discord is not the go to stream of communication for some members. This allows ease of direct communication amongst the team.
iLearn Private Group Forum	Forum to post breakdowns of weekly meetings tracking weekly activities, progress, and team member details.
Emails	Emails between team members are used predominantly to pass on responses from SAS regarding questions and work we have sent them. This helps keep track of valuable responses and information without over complicating tracking of emails.

6. QUALITY MANAGEMENT

6.1. QUALITY CONTROL

Quality control processes are used as a detective means to identify faults within the project. These work to ensure that team members possess the ability to effectively navigate faults and move forwards ensuring project quality is improved or maintained.

1. **Deliverable/Feedback Review**: Monitoring feedback and comments on our deliverables allow us to understand its faults and make adjustments. This ensures the project is rid of these faults and project quality is upheld.
2. **Weekly reports:** Allows the team to identify what they found challenging and what they did well. This reflection helps identify faults within team members' operations throughout the week, and work to navigate these faults for the week ahead.
3. **Error Checking/Editing**: Group members have the ability to leave comments on reports and documents providing recommended improvements or editing fixes. Members providing checks are to do so for work other than their own.
4. **Sponsor Feedback:** Feedback from sponsors provides an additional set of eyes to identify errors in the project, or areas that are lacking in quality or information. Sponsor feedback also ensures our project is aligning with sponsors expectations and needs.
5. **Ongoing Collaboration with SAS:** Ensuring that we reach out via email to our sponsors throughout the week and not just wait for fortnightly meetings. Continually checking

project changes and requirements are being met with the sponsors will ensure we can ongoingly identify faults and make changes that align with SAS requirements.

6. **One-on-one meetings with project manager:** Group members have the option to catch up with the project manager. Members can voice challenges or issues they may be facing and the project manager and member can work collaboratively to navigate and improve said challenges.

6.2. QUALITY ASSURANCE

Quality assurance processes are employed as preventative measures. The following processes work to prevent faults and errors before they occur or heavily impact the project.

- **Project Manager Review/Audit:** Our project manager ensures proactive management of project progression, risks, and tracking. Enforcing measures of facilitation of team collaboration, ensuring documentation is up-to-date, and maintaining project plans and schedules, ensures measures are preventative and not reactive.
- **Deadlines/Checklists**: Visual checklists on Trello provide the team the ability to visualise tasks to be completed and upcoming deadlines. This ensures quality is upheld whereby we can ensure nothing is missing in reports, and we can move onto other tasks.
- **Weekly Reviews:** This encompasses both weekly meetings and reflections and works as an assurance method. Regular check-ins in with members and having individual reflection each week ensures operations are up-to-date and work isn't lagging.

7. SCHEDULE

7.1. GANTT CHART

The following Gantt Chart in Appendix A, illustrates the breakdown of deliverables across the semester and is in line with the due dates of the dictated assessments schedule. The deliverables intended to be submitted to the university will have the starting letter D, whereas deliverables intended for the Sponsor will hold the initials SD (Sponsor Deliverable). The project management tool 'ProjectLibre', was utilised to carry out the product of the Gantt Chart on a Macintosh operating system.

7.2. WORK BREAKDOWN / NETWORK DIAGRAM

The Work Breakdown Structure (WBS) in Appendix B is produced from the Gantt chart on the same software and reproduces activities in a finer detail breaking down larger tasks into subtasks. The following WBS is to be interpreted with the efforts of all individuals involved in each task and not one sole individual doing the whole task. Group discussions, brainstorming and feedback sessions are used as an accountability activity in order to encourage collaboration in all aspects of the project. For the purpose of this deliverable a network diagram of the WBS has been produced due to the limited page count.

7.3. COST AND TIME

This project has no cost breakdown or allocation as PACE is a voluntary, industry-placement program. Meaning that no additional cost is incurred to the client and not monetary value is received by the project team. The sponsor has however mentioned that they have set time allocations for engaging with this PACE project throughout their day to day activities and budget. The student team however does not engage in this.

The total time it will take to complete the project is 97 days, with an expected 25.5 hours per deliverable (rough averaging from first deliverable) totaling 229.5 hours for the project.

7.4. RESOURCE ALLOCATION AND ASSUMPTIONS

The following Resource Allocation as observed in Appendix C, has been derived from the Gantt chart and Work Breakdown. Kindly note that the software Project Libre tends to alter hours worked according to percentage assigned to resource and time to task and recalculates at each adjustment, therefore please observe the percentages and not the hours.

It is assumed that all resources will be available throughout the duration of the project with alternatives on standby. For example an alternative to the library study pods would be a local cafe or external library. Another example is that we have alternative data analytical platforms for use as well.

In previous meetings with the sponsor the team has expressed our personal experience with the SAS Viya platform and have voiced concerns about it being slow and sometimes not working. The SAS team were more than happy to report the feedback to the development team in order to adjust the platform for better use.

8. ASSUMPTIONS

- It is assumed that access to necessary data to measure the effectiveness of Zoe Empowers program is available, accessible, and in a usable format.
- It is assumed that the data is of sufficient quality and consistency to support accurate and meaningful analysis.
- It is assumed that the necessary data analytics tools and technologies are available and can be used to analyse and visualise the data effectively.
- It is assumed that SAS will be engaged throughout the project and will provide feedback and support as needed.
- It is assumed that the project will be completed within the defined timeline, and any potential delays or obstacles will be addressed promptly to ensure timely completion.
- It is assumed that the permittance of available external systems will be of direct beneficial assistance to conducting research for the project.
- It is assumed that predictive modelling falls within the scope of the project where forecasting predictive outcomes are essential to successful analysis and research.
- It is assumed that weekly sprint meetings between the team and fortnightly meetings with sponsors will greatly benefit the progress of the project.
- It is assumed that the agile methodology will allow the team to effectively complete the project while adapting to change both efficiently and effectively as we collectively progress.
- It is assumed that the team will remain within the scope of the project plan, working to reduce the risk of scope creeping, to ensure that we deliver a solution that meets MVP.

9. HANDOVER REQUIREMENTS

9.1. GENERAL OVERVIEW

The purpose of the handover requirement is to explain what the sponsor (SAS) will be receiving from the team at Data Synergy. The requirements from SAS are situated outside of the university project, and are specific to what SAS have asked us to provide them, that being a findings report, an analysis report, a recommendations report, and a presentation. The presentation will first be given to SAS, followed by the remaining reports.

9.2. KEY EVENTS

An array of meetings have ensued over the course of the semester between the team and SAS. Meetings have been used to validate ongoing work, receive feedback, and communicate the wants and needs of the sponsor. Remaining key events are two meetings with SAS, the team presentation and the handover of reports..

- Wednesday, May 3 (4:30pm-5pm), meeting with SAS and Team 32.
- Tuesday, May 16 (9:30am-10am), meeting with SAS to review increment 2 and get assistance.
- Thursday, June 1 (5pm-7pm), on-campus presentation to SAS.
- Thursday, June 1 (5pm-7pm), handover of findings report, analysis report, and recommendation report to SAS.

9.3. KEY ACTIVITIES

The key activities of the team prove numerous. Where some activities apply to university work, and others to the sponsor, all activities must be completed. Without university work, sponsor needs cannot be met, and without sponsor needs being met, university work cannot be completed. Remaining key activities are diverse and include weekly team meetings, fortnightly meetings with the sponsor, analysis and research on the Viya for learners platform, analysis and research on Python/R studio, collectively writing and submitting deliverables 3-8, and the findings, analysis, and recommendations reports. As stated, each key activity must be completed should the team successfully meet sponsor wants and needs.

9.4. KEY RESPONSIBILITIES

Each team member has key responsibilities to ensure successful completion of the project. These responsibilities have been developed based on the skills of members, and directly contribute to providing SAS with a findings report, analysis report, and recommendations report.

Team Member	**Key Responsibilities**
Tabitha Philip (Project Manager)	- Creating, maintaining, and developing a project plan and schedule. - Communicating project updates and status reports to SAS. - Managing project risks and issues.
Cassandra Johns (Data Specialist)	- Collecting, analysing, and interpreting

	data. - Ensuring accuracy and quality of data to communicate to SAS. - Mitigate risks associated with data.
Lachlan Yates (Change Officer/Business Analyst)	- Developing change management plans and strategies. - Support implementation of change management activities. - Provide business analysis support to the project team.
Ava Gardiner (Project Officer/Change Officer)	- Supporting the development of project plan and schedule. - Implementation of project changes. - Ensuring documentation is up-to-date and completed on time.
Rory Ali (Project Officer/Date Consultant)	- Gathering and analysing data related to the project. - Developing data models and visualisations. - Provide technical support for data-related tools and strategies.

The specific key responsibilities of team members are tailored to ensuring SAS wants are met. Each responsibility forms a part within each report needing completion. Team members must ensure they uphold their key responsibilities. Doing so will ensure we are able to cover all needs of the sponsor and complete the university project.

9.5. KEY DOCUMENTS

Findings Report

The findings report will include detailed data visualisations and exploration of raw data found in the Kenya, Rwanda, and SSI data sets. The report will showcase understanding of the demographics of individuals and the environment that they live in. The report will further discuss the data methods deployed by the team to support our findings and breakdown our key findings.

Analysis Report

The analysis report will be developed to identify gaps/vulnerabilities that may exist within the Zoe Empowers program. The gaps/vulnerabilities identified will be further used in the recommendations report. The report will provide a greater understanding of the links between the particular factors and why they might be causing a positive/negative impact. The report works as a deeper dive into the findings report.

Recommendations Report

The recommendations report will be the final report given to SAS. The report will provide suggestions for improving the Zoe Empowers program. This will include ways the program can further improve the lives of individuals even further. Suggestions will be derived from the gaps and vulnerabilities identified in the analysis report.

Presentation

The presentation is scheduled for Thursday, June 1, 2023. It will be presented to SAS at a slot within a two hour time frame (5pm-7pm), and is the team opportunity to visually communicate the mixture of our findings report, analysis report, and recommendations report and pinpointing various aspects of the Zoe empowers program requiring improvement. The presentation will follow a similar structure as below, keeping in mind that it is subject to change.

Presentation Structure	
INTRODUCTION	- Discuss focal point (Does Zoe Empowers improve the lives of participants?) - What is SSI? - What factors influence the SSI?
FINDINGS	- What factors influencing SSI can be observed throughout the programs lifetime + its participants - Graphs, visualisations
ANALYSIS	- Show factors affecting SSI (why do they affect SSI? Global resources) - What gaps can be observed from the findings (why do these gaps present themselves? Global resource)
RECOMMENDATIONS	- What Zoe Empowers can do to address this and improve the life long experience of its participants? - Taking the “gaps analysis” to influence the above point ^. - Reaffirm recommendations using data from - global organisations - UN sustainment goals. - Why should Zoe Empowers observe these (benefits). Because it observes a global authority with global efforts towards the SSI, reaffirming efforts. - Contributes to the long term sustainment of an individual's self-sufficiency. - Inturn assists in the contribution to

	global efforts for self-sufficiency.

10. STANDARDS (ADDED)

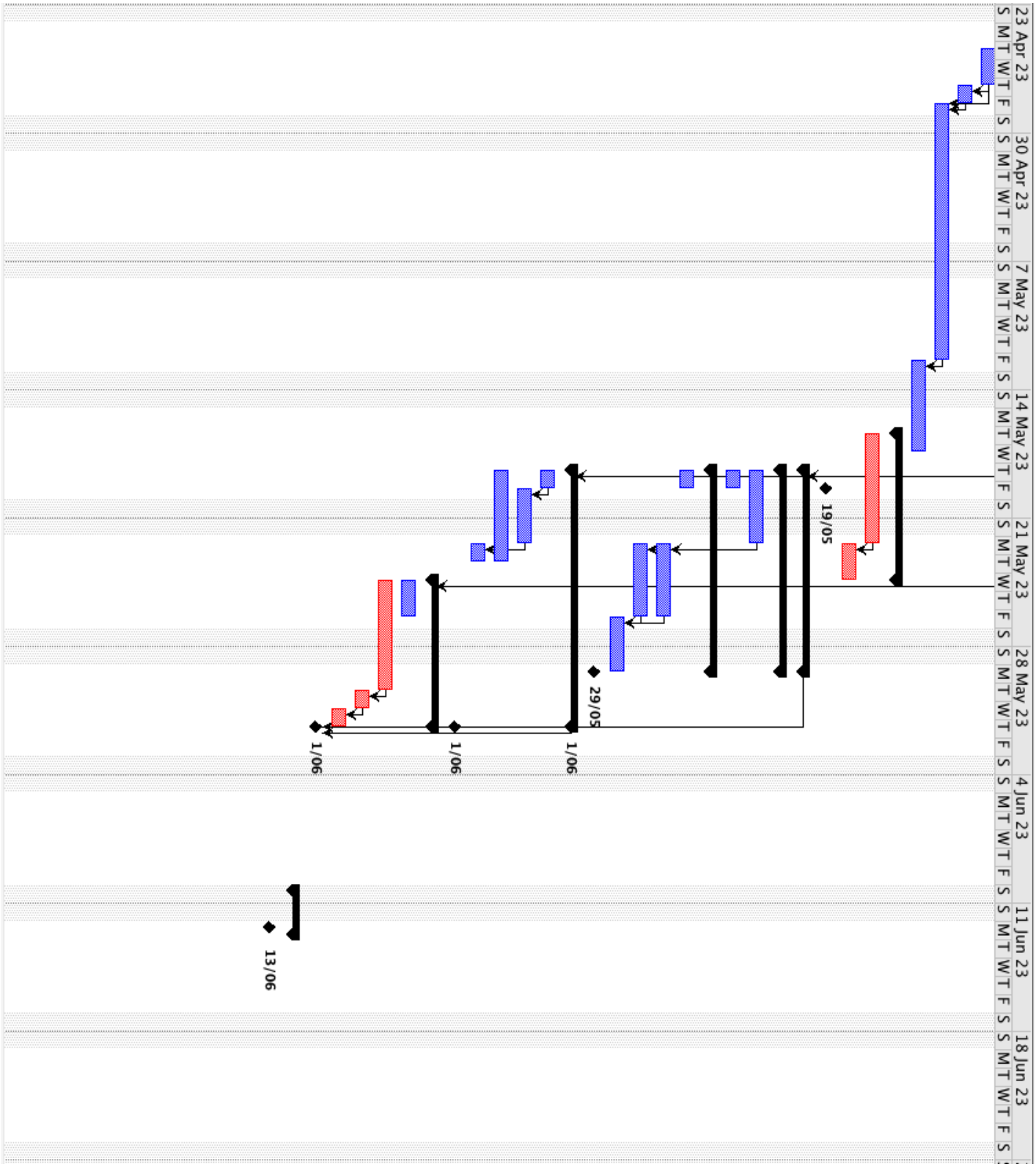
DataSynergy is dedicated to creating the highest quality work to the best of our ability. In doing so, the team adheres to a number of ethical and global standards when it comes to the handling and management of personally identifiable information and sensitive (personal-privacy) data. The following standards are observed in the teams day to day practice:

a) Australian Computer Society (ASC) Code of Conduct
b) International Standards Organisation (ISO) 8000-1 : Data Quality
    i) Ensuring data is from a reputable source and is up to standard before investigation.
c) ISO 2700 Family of Standards : Information Security Standards
    i) To protect the data being handled.
d) National Institute of Standards and Technology 800-122 : Guide to Protecting the Confidentiality of Personally Identifiable Information (PII)
    i) To protect and treat the information handled accordingly.

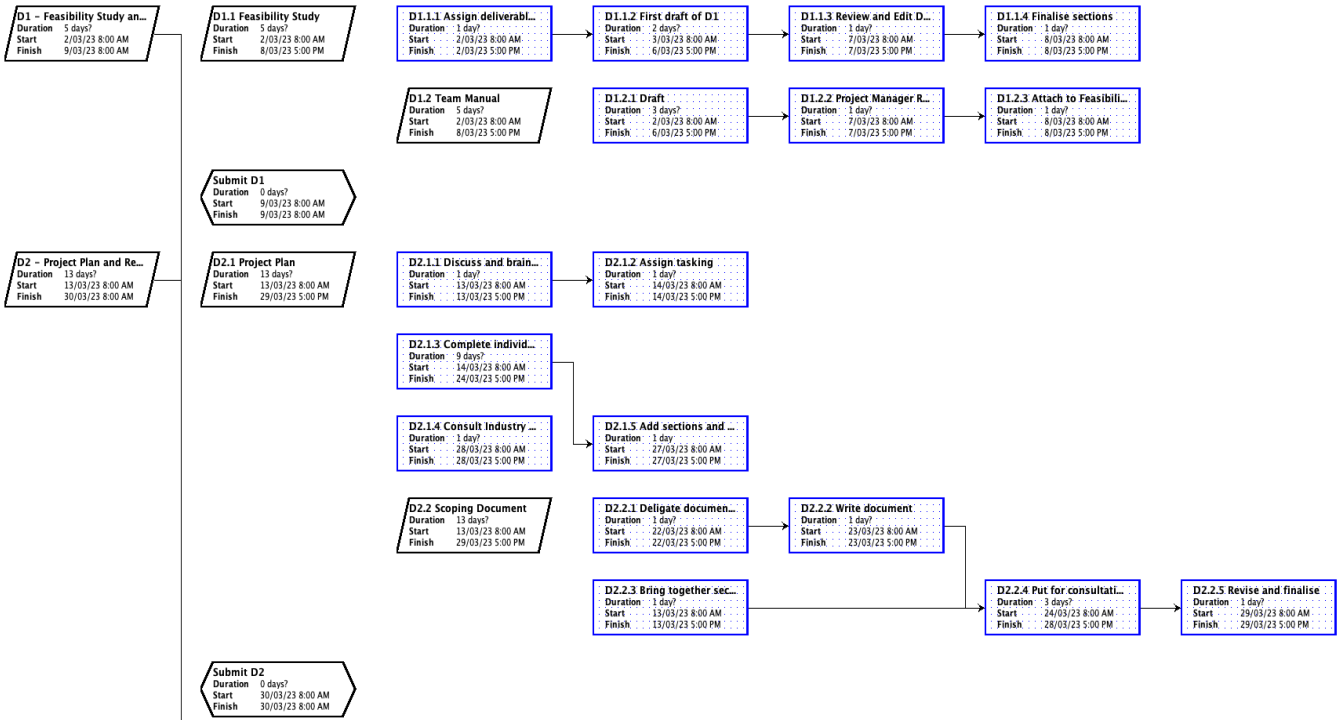
11. APPENDICES

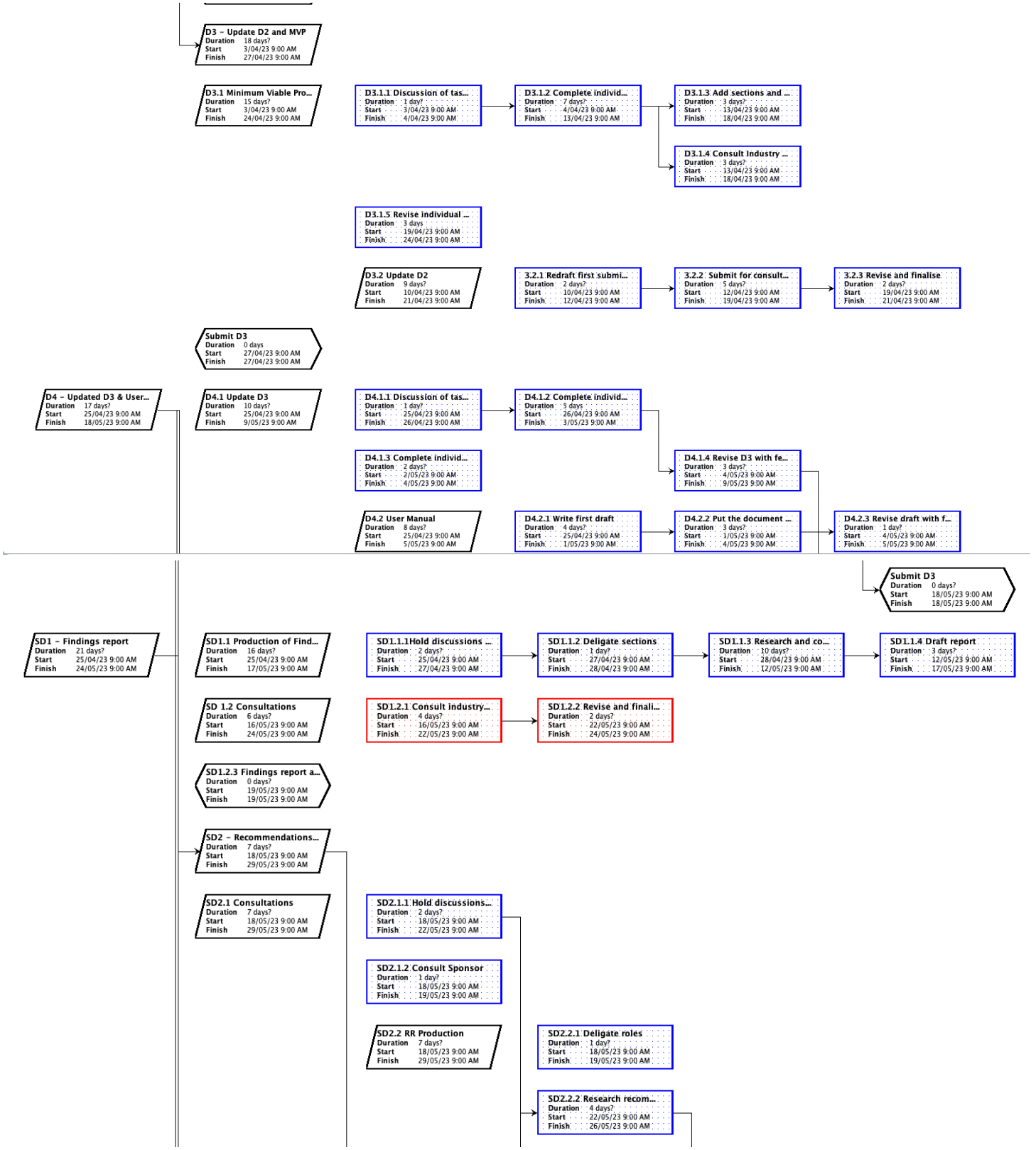
a) Appendix A - Gantt chart
b) Appendix B - Work Breakdown Structure
c) Appendix C - Resource Allocation

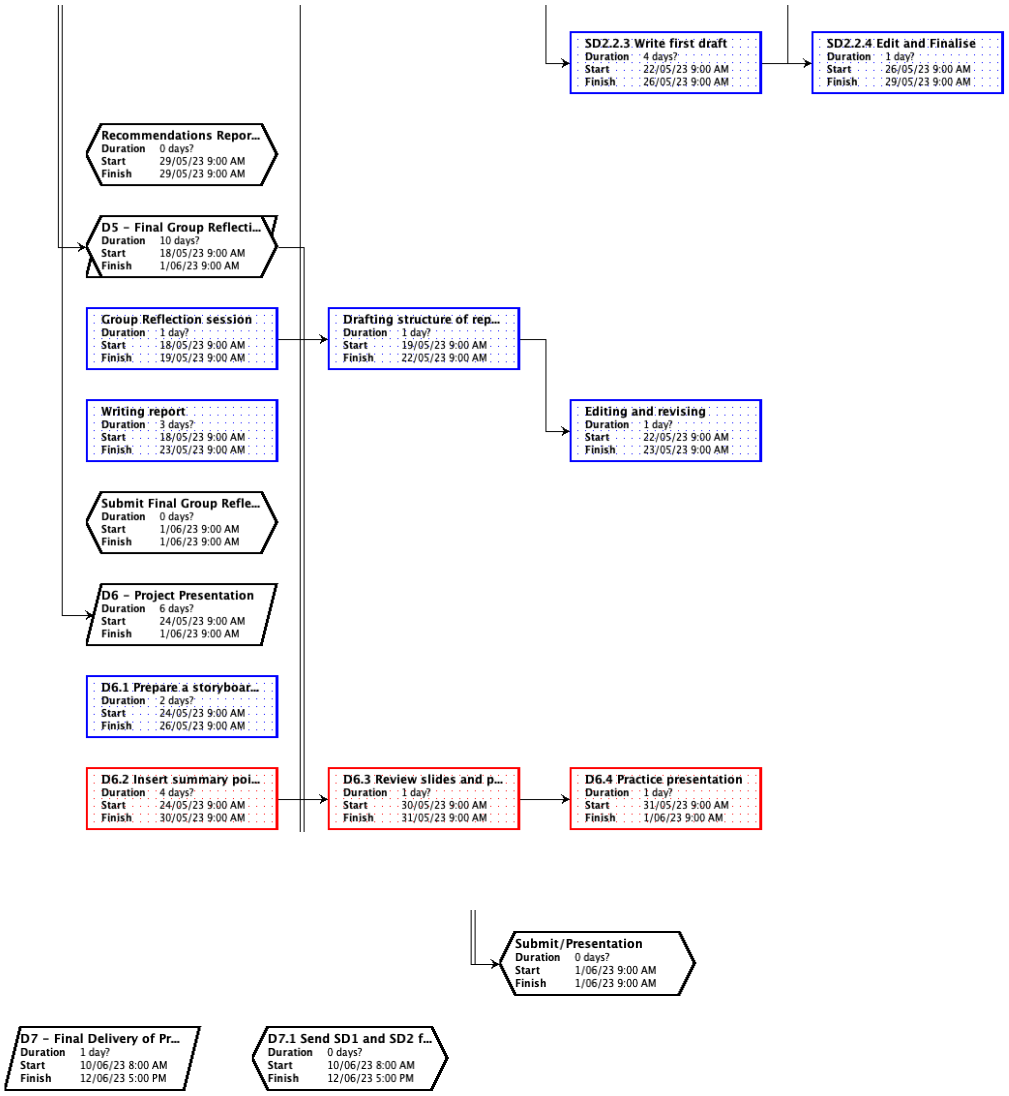
Appendix A - Gantt Chart



Appendix B


D1 - Feasibility Study an...
Duration 5 days?
Start 2/03/23 8:00 AM
Finish 9/03/23 8:00 AM
D1.1 Feasibility Study
Duration 5 days?
Start 2/03/23 8:00 AM
Finish 8/03/23 5:00 PM
D1.1.1 Assign deliverabl...
Duration 1 day?
Start 2/03/23 8:00 AM
Finish 2/03/23 5:00 PM
D1.1.2 First draft of D1
Duration 2 days?
Start 3/03/23 8:00 AM
Finish 6/03/23 5:00 PM
D1.1.3 Review and Edit D...
Duration 1 day?
Start 7/03/23 8:00 AM
Finish 7/03/23 5:00 PM
D1.1.4 Finalise sections
Duration 1 day?
Start 8/03/23 8:00 AM
Finish 8/03/23 5:00 PM
D1.2 Team Manual
Duration 5 days?
Start 2/03/23 8:00 AM
Finish 8/03/23 5:00 PM
D1.2.1 Draft
Duration 3 days?
Start 2/03/23 8:00 AM
Finish 6/03/23 5:00 PM
D1.2.2 Project Manager R...
Duration 1 day?
Start 7/03/23 8:00 AM
Finish 7/03/23 5:00 PM
D1.2.3 Attach to Feasibili...
Duration 1 day?
Start 8/03/23 8:00 AM
Finish 8/03/23 5:00 PM
Submit D1
Duration 0 days?
Start 9/03/23 8:00 AM
Finish 9/03/23 8:00 AM
D2 - Project Plan and Re...
Duration 13 days?
Start 13/03/23 8:00 AM
Finish 30/03/23 8:00 AM
D2.1 Project Plan
Duration 13 days?
Start 13/03/23 8:00 AM
Finish 29/03/23 5:00 PM
D2.1.1 Discuss and brain...
Duration 1 day?
Start 13/03/23 8:00 AM
Finish 13/03/23 5:00 PM
D2.1.2 Assign tasking
Duration 1 day?
Start 14/03/23 8:00 AM
Finish 14/03/23 5:00 PM
D2.1.3 Complete individ...
Duration 9 days?
Start 14/03/23 8:00 AM
Finish 24/03/23 5:00 PM
D2.1.4 Consult Industry ...
Duration 1 day?
Start 28/03/23 8:00 AM
Finish 28/03/23 5:00 PM
D2.1.5 Add sections and ...
Duration 1 day
Start 27/03/23 8:00 AM
Finish 27/03/23 5:00 PM
D2.2 Scoping Document
Duration 13 days?
Start 13/03/23 8:00 AM
Finish 29/03/23 5:00 PM
D2.2.1 Deligate documen...
Duration 1 day?
Start 22/03/23 8:00 AM
Finish 22/03/23 5:00 PM
D2.2.2 Write document
Duration 1 day?
Start 23/03/23 8:00 AM
Finish 23/03/23 5:00 PM
D2.2.3 Bring together sec...
Duration 1 day?
Start 13/03/23 8:00 AM
Finish 13/03/23 5:00 PM
D2.2.4 Put for consultati...
Duration 3 days?
Start 24/03/23 8:00 AM
Finish 28/03/23 5:00 PM
D2.2.5 Revise and finalise
Duration 1 day?
Start 29/03/23 8:00 AM
Finish 29/03/23 5:00 PM
Submit D2
Duration 0 days?
Start 30/03/23 8:00 AM
Finish 30/03/23 8:00 AM


D3 – Update D2 and MVP
Duration 18 days?
Start 3/04/23 9:00 AM
Finish 27/04/23 9:00 AM
D3.1 Minimum Viable Pro...
Duration 15 days?
Start 3/04/23 9:00 AM
Finish 24/04/23 9:00 AM
D3.1.1 Discussion of tas...
Duration 1 day?
Start 3/04/23 9:00 AM
Finish 4/04/23 9:00 AM
D3.1.2 Complete individ...
Duration 7 days?
Start 4/04/23 9:00 AM
Finish 13/04/23 9:00 AM
D3.1.3 Add sections and ...
Duration 3 days?
Start 13/04/23 9:00 AM
Finish 18/04/23 9:00 AM
D3.1.4 Consult Industry ...
Duration 3 days?
Start 13/04/23 9:00 AM
Finish 18/04/23 9:00 AM
D3.1.5 Revise individual ...
Duration 3 days
Start 19/04/23 9:00 AM
Finish 24/04/23 9:00 AM
D3.2 Update D2
Duration 9 days?
Start 10/04/23 9:00 AM
Finish 21/04/23 9:00 AM
3.2.1 Redraft first submi...
Duration 2 days?
Start 10/04/23 9:00 AM
Finish 12/04/23 9:00 AM
3.2.2 Submit for consult...
Duration 5 days?
Start 12/04/23 9:00 AM
Finish 19/04/23 9:00 AM
3.2.3 Revise and finalise
Duration 2 days?
Start 19/04/23 9:00 AM
Finish 21/04/23 9:00 AM
Submit D3
Duration 0 days
Start 27/04/23 9:00 AM
Finish 27/04/23 9:00 AM
D4 – Updated D3 & User...
Duration 17 days?
Start 25/04/23 9:00 AM
Finish 18/05/23 9:00 AM
D4.1 Update D3
Duration 10 days?
Start 25/04/23 9:00 AM
Finish 9/05/23 9:00 AM
D4.1.1 Discussion of tas...
Duration 1 day?
Start 25/04/23 9:00 AM
Finish 26/04/23 9:00 AM
D4.1.2 Complete individ...
Duration 5 days
Start 26/04/23 9:00 AM
Finish 3/05/23 9:00 AM
D4.1.3 Complete individ...
Duration 2 days?
Start 2/05/23 9:00 AM
Finish 4/05/23 9:00 AM
D4.1.4 Revise D3 with fe...
Duration 3 days?
Start 4/05/23 9:00 AM
Finish 9/05/23 9:00 AM
D4.2 User Manual
Duration 8 days?
Start 25/04/23 9:00 AM
Finish 5/05/23 9:00 AM
D4.2.1 Write first draft
Duration 4 days?
Start 25/04/23 9:00 AM
Finish 1/05/23 9:00 AM
D4.2.2 Put the document ...
Duration 3 days?
Start 1/05/23 9:00 AM
Finish 4/05/23 9:00 AM
D4.2.3 Revise draft with f...
Duration 1 day?
Start 4/05/23 9:00 AM
Finish 5/05/23 9:00 AM
Submit D3
Duration 0 days?
Start 18/05/23 9:00 AM
Finish 18/05/23 9:00 AM
SD1 – Findings report
Duration 21 days?
Start 25/04/23 9:00 AM
Finish 24/05/23 9:00 AM
SD1.1 Production of Find...
Duration 16 days?
Start 25/04/23 9:00 AM
Finish 17/05/23 9:00 AM
SD1.1.1Hold discussions ...
Duration 2 days?
Start 25/04/23 9:00 AM
Finish 27/04/23 9:00 AM
SD1.1.2 Deligate sections
Duration 1 day?
Start 27/04/23 9:00 AM
Finish 28/04/23 9:00 AM
SD1.1.3 Research and co...
Duration 10 days?
Start 28/04/23 9:00 AM
Finish 12/05/23 9:00 AM
SD1.1.4 Draft report
Duration 3 days?
Start 12/05/23 9:00 AM
Finish 17/05/23 9:00 AM
SD 1.2 Consultations
Duration 6 days?
Start 16/05/23 9:00 AM
Finish 24/05/23 9:00 AM
SD1.2.1 Consult industry...
Duration 4 days?
Start 16/05/23 9:00 AM
Finish 22/05/23 9:00 AM
SD1.2.2 Revise and finali...
Duration 2 days?
Start 22/05/23 9:00 AM
Finish 24/05/23 9:00 AM
SD1.2.3 Findings report a...
Duration 0 days?
Start 19/05/23 9:00 AM
Finish 19/05/23 9:00 AM
SD2 – Recommendations...
Duration 7 days?
Start 18/05/23 9:00 AM
Finish 29/05/23 9:00 AM
SD2.1 Consultations
Duration 7 days?
Start 18/05/23 9:00 AM
Finish 29/05/23 9:00 AM
SD2.1.1 Hold discussions...
Duration 2 days?
Start 18/05/23 9:00 AM
Finish 22/05/23 9:00 AM
SD2.1.2 Consult Sponsor
Duration 1 day?
Start 18/05/23 9:00 AM
Finish 19/05/23 9:00 AM
SD2.2 RR Production
Duration 7 days?
Start 18/05/23 9:00 AM
Finish 29/05/23 9:00 AM
SD2.2.1 Deligate roles
Duration 1 day?
Start 18/05/23 9:00 AM
Finish 19/05/23 9:00 AM
SD2.2.2 Research recom...
Duration 4 days?
Start 22/05/23 9:00 AM
Finish 26/05/23 9:00 AM


SD2.2.3 Write first draft
Duration 4 days?
Start 22/05/23 9:00 AM
Finish 26/05/23 9:00 AM
SD2.2.4 Edit and Finalise
Duration 1 day?
Start 26/05/23 9:00 AM
Finish 29/05/23 9:00 AM
Recommendations Repor...
Duration 0 days?
Start 29/05/23 9:00 AM
Finish 29/05/23 9:00 AM
D5 - Final Group Reflecti...
Duration 10 days?
Start 18/05/23 9:00 AM
Finish 1/06/23 9:00 AM
Group Reflection session
Duration 1 day?
Start 18/05/23 9:00 AM
Finish 19/05/23 9:00 AM
Drafting structure of rep...
Duration 1 day?
Start 19/05/23 9:00 AM
Finish 22/05/23 9:00 AM
Writing report
Duration 3 days?
Start 18/05/23 9:00 AM
Finish 23/05/23 9:00 AM
Editing and revising
Duration 1 day?
Start 22/05/23 9:00 AM
Finish 23/05/23 9:00 AM
Submit Final Group Refle...
Duration 0 days?
Start 1/06/23 9:00 AM
Finish 1/06/23 9:00 AM
D6 - Project Presentation
Duration 6 days?
Start 24/05/23 9:00 AM
Finish 1/06/23 9:00 AM
D6.1 Prepare a storyboar...
Duration 2 days?
Start 24/05/23 9:00 AM
Finish 26/05/23 9:00 AM
D6.2 Insert summary poi...
Duration 4 days?
Start 24/05/23 9:00 AM
Finish 30/05/23 9:00 AM
D6.3 Review slides and p...
Duration 1 day?
Start 30/05/23 9:00 AM
Finish 31/05/23 9:00 AM
D6.4 Practice presentation
Duration 1 day?
Start 31/05/23 9:00 AM
Finish 1/06/23 9:00 AM
Submit/Presentation
Duration 0 days?
Start 1/06/23 9:00 AM
Finish 1/06/23 9:00 AM
D7 - Final Delivery of Pr...
Duration 1 day?
Start 10/06/23 8:00 AM
Finish 12/06/23 5:00 PM
D7.1 Send SD1 and SD2 f...
Duration 0 days?
Start 10/06/23 8:00 AM
Finish 12/06/23 5:00 PM

Appendix C - Resource Graph (Please observe the %)

1	**⊟D1 – Feasibility Study and Team Manual**	**100%**	**52.919 hours**	**8 days?**	**2/03/23 8:00 AM**	**13/03/23 5:00 PM**
2	**⊟D1.1 Feasibility Study**	**100%**	**52.919 hours**	**8 days?**	**2/03/23 8:00 AM**	**13/03/23 5:00 PM**
3	D1.1.1 Assign deliverable responsibilities	100%	8 hours	1 day?	2/03/23 8:00 AM	3/03/23 5:00 PM
	PROJECT MANAGER	*100%*	*8 hours*	*1 day*	*3/03/23 8:00 AM*	*3/03/23 5:00 PM*
4	D1.1.2 First draft of D1	24%	7.779 hours	4 days?	6/03/23 8:00 AM	9/03/23 5:00 PM
	PROJECT OFFICER/CHANGE OFFICER	*20%*	*1.986 hours*	*1.241 days*	*7/03/23 8:00 AM*	*8/03/23 9:55 AM*
	PROJECT MANAGER	*20%*	*4.8 hours*	*3 days*	*7/03/23 8:00 AM*	*9/03/23 5:00 PM*
	DATA SPECIALIST	*20%*	*0.827 hours*	*0.517 days*	*7/03/23 8:00 AM*	*7/03/23 1:08 PM*
	PROJECT OFFICER/DATA CONSULTANT	*20%*	*0 hours*	*0 days*	*7/03/23 8:00 AM*	*7/03/23 8:00 AM*
	CHANGE OFFICER/BUSINESS ANALYST	*20%*	*0.165 hours*	*0.103 days*	*7/03/23 8:00 AM*	*7/03/23 8:49 AM*
5	D1.1.3 Review and Edit D1	100%	8 hours	1 day?	10/03/23 8:00 AM	10/03/23 5:00 PM
6	D1.1.4 Finalise sections	100%	8 hours	1 day?	13/03/23 8:00 AM	13/03/23 5:00 PM
7	**⊟D1.2 Team Manual**	**100%**	**21.14 hours**	**2.75 days?**	**2/03/23 8:00 AM**	**6/03/23 3:00 PM**
8	D1.2.1 Draft	96%	7.7 hours	1 day?	2/03/23 8:00 AM	2/03/23 5:00 PM
	PROJECT OFFICER/CHANGE OFFICER	*70%*	*5.6 hours*	*1 day*	*2/03/23 8:00 AM*	*2/03/23 5:00 PM*
	PROJECT MANAGER	*10%*	*0.7 hours*	*0.875 days*	*2/03/23 8:00 AM*	*2/03/23 4:00 PM*
	PROJECT OFFICER/DATA CONSULTANT	*20%*	*1.4 hours*	*0.875 days*	*2/03/23 8:00 AM*	*2/03/23 4:00 PM*
	DATA SPECIALIST	*0%*	*0 hours*	*0 days*	*2/03/23 8:00 AM*	*2/03/23 8:00 AM*
9	D1.2.2 Project Manager Review and team edit	100%	8 hours	1 day?	3/03/23 8:00 AM	3/03/23 5:00 PM
	PROJECT MANAGER	*100%*	*8 hours*	*1 day*	*3/03/23 8:00 AM*	*3/03/23 5:00 PM*
10	D1.2.3 Attach to Feasibility study and clean up	91%	5.44 hours	0.75 days?	6/03/23 8:00 AM	6/03/23 3:00 PM
	PROJECT OFFICER/CHANGE OFFICER	*40%*	*2.4 hours*	*0.75 days*	*6/03/23 8:00 AM*	*6/03/23 3:00 PM*
	PROJECT MANAGER	*40%*	*2.4 hours*	*0.75 days*	*6/03/23 8:00 AM*	*6/03/23 3:00 PM*
	PROJECT OFFICER/DATA CONSULTANT	*10%*	*0.32 hours*	*0.4 days*	*6/03/23 8:00 AM*	*6/03/23 11:12 AM*
	DATA SPECIALIST	*10%*	*0.32 hours*	*0.4 days*	*6/03/23 8:00 AM*	*6/03/23 11:12 AM*
11	Submit D1	100%	0 hours	0 days?	9/03/23 8:00 AM	9/03/23 8:00 AM

12	**⊟D2 – Project Plan and Requirements/Scoping Document**	**100%**	**126.598 hours**	**13 days?**	**13/03/23 8:00 AM**	**30/03/23 8:00 AM**
13	**⊟D2.1 Project Plan**	**100%**	**126.598 hours**	**11.625 da...**	**13/03/23 8:00 AM**	**28/03/23 2:00 PM**
14	D2.1.1 Discuss and brainstrom task	84%	4.2 hours	0.625 days?	13/03/23 8:00 AM	13/03/23 2:00 PM
	PROJECT MANAGER	*20%*	*1 hour*	*0.625 days*	*13/03/23 8:00 AM*	*13/03/23 2:00 PM*
	PROJECT OFFICER/CHANGE OFFICER	*20%*	*0.8 hours*	*0.5 days*	*13/03/23 8:00 AM*	*13/03/23 1:00 PM*
	PROJECT OFFICER/DATA CONSULTANT	*20%*	*0.8 hours*	*0.5 days*	*13/03/23 8:00 AM*	*13/03/23 1:00 PM*
	DATA SPECIALIST	*20%*	*0.8 hours*	*0.5 days*	*13/03/23 8:00 AM*	*13/03/23 1:00 PM*
	CHANGE OFFICER/BUSINESS ANALYST	*20%*	*0.8 hours*	*0.5 days*	*13/03/23 8:00 AM*	*13/03/23 1:00 PM*
15	D2.1.2 Assign tasking	100%	4 hours	0.5 days?	13/03/23 2:00 PM	14/03/23 9:00 AM
	PROJECT MANAGER	*100%*	*4 hours*	*0.5 days*	*13/03/23 2:00 PM*	*14/03/23 9:00 AM*
16	D2.1.3 Complete individual sections	100%	72 hours	9 days?	14/03/23 8:00 AM	24/03/23 5:00 PM
	PROJECT MANAGER	*20%*	*14.4 hours*	*9 days*	*14/03/23 8:00 AM*	*24/03/23 5:00 PM*
	PROJECT OFFICER/CHANGE OFFICER	*20%*	*14.4 hours*	*9 days*	*14/03/23 8:00 AM*	*24/03/23 5:00 PM*
	PROJECT OFFICER/DATA CONSULTANT	*20%*	*14.4 hours*	*9 days*	*14/03/23 8:00 AM*	*24/03/23 5:00 PM*
	DATA SPECIALIST	*20%*	*14.4 hours*	*9 days*	*14/03/23 8:00 AM*	*24/03/23 5:00 PM*
	CHANGE OFFICER/BUSINESS ANALYST	*20%*	*14.4 hours*	*9 days*	*14/03/23 8:00 AM*	*24/03/23 5:00 PM*
17	D2.1.4 Consult Industry partners	100%	5 hours	0.625 days?	28/03/23 8:00 AM	28/03/23 2:00 PM
	PROJECT MANAGER	*20%*	*1 hour*	*0.625 days*	*28/03/23 8:00 AM*	*28/03/23 2:00 PM*
	PROJECT OFFICER/CHANGE OFFICER	*20%*	*1 hour*	*0.625 days*	*28/03/23 8:00 AM*	*28/03/23 2:00 PM*
	PROJECT OFFICER/DATA CONSULTANT	*20%*	*1 hour*	*0.625 days*	*28/03/23 8:00 AM*	*28/03/23 2:00 PM*
	DATA SPECIALIST	*20%*	*1 hour*	*0.625 days*	*28/03/23 8:00 AM*	*28/03/23 2:00 PM*
	CHANGE OFFICER/BUSINESS ANALYST	*20%*	*1 hour*	*0.625 days*	*28/03/23 8:00 AM*	*28/03/23 2:00 PM*
18	D2.1.5 Add sections and edit	67%	1.598 hours	0.297 days	27/03/23 8:00 AM	27/03/23 10:22 AM
	PROJECT MANAGER	*18%*	*0.076 hours*	*0.053 days*	*27/03/23 8:00 AM*	*27/03/23 8:25 AM*
	PROJECT OFFICER/CHANGE OFFICER	*30%*	*0.381 hours*	*0.159 days*	*27/03/23 8:00 AM*	*27/03/23 9:16 AM*
	PROJECT OFFICER/DATA CONSULTANT	*18%*	*0.381 hours*	*0.264 days*	*27/03/23 8:00 AM*	*27/03/23 10:06 AM*
	DATA SPECIALIST	*18%*	*0.381 hours*	*0.264 days*	*27/03/23 8:00 AM*	*27/03/23 10:06 AM*
	CHANGE OFFICER/BUSINESS ANALYST	*16%*	*0.381 hours*	*0.297 days*	*27/03/23 8:00 AM*	*27/03/23 10:22 AM*

19	**⊟D2.2 Scoping Document**	**100%**	**39.8 hours**	**10.975 da...**	**13/03/23 8:00 AM**	**27/03/23 4:48 PM**
20	D2.2.1 Deligate document sections	100%	1 hour	0.125 days?	22/03/23 8:00 AM	22/03/23 9:00 AM
	PROJECT MANAGER	*100%*	*1 hour*	*0.125 days*	*22/03/23 8:00 AM*	*22/03/23 9:00 AM*
21	D2.2.2 Write document	100%	10 hours	1.25 days?	22/03/23 9:00 AM	23/03/23 11:00 AM
	PROJECT MANAGER	*30%*	*3 hours*	*1.25 days*	*22/03/23 9:00 AM*	*23/03/23 11:00 AM*
	DATA SPECIALIST	*70%*	*7 hours*	*1.25 days*	*22/03/23 9:00 AM*	*23/03/23 11:00 AM*
22	D2.2.3 Bring together sections and edit	100%	8 hours	1 day?	13/03/23 8:00 AM	13/03/23 5:00 PM
	DATA SPECIALIST	*50%*	*4 hours*	*1 day*	*13/03/23 8:00 AM*	*13/03/23 5:00 PM*
	PROJECT MANAGER	*50%*	*4 hours*	*1 day*	*13/03/23 8:00 AM*	*13/03/23 5:00 PM*
23	D2.2.4 Put for consultation	100%	16 hours	2 days?	23/03/23 11:00 AM	27/03/23 11:00 AM
	PROJECT MANAGER	*50%*	*8 hours*	*2 days*	*23/03/23 11:00 AM*	*27/03/23 11:00 AM*
	PROJECT OFFICER/DATA CONSULTANT	*0%*	*0 hours*	*0 days*	*23/03/23 11:00 AM*	*23/03/23 11:00 AM*
	DATA SPECIALIST	*50%*	*8 hours*	*2 days*	*23/03/23 11:00 AM*	*27/03/23 11:00 AM*
24	D2.2.5 Revise and finalise	100%	4.8 hours	0.6 days?	27/03/23 11:00 AM	27/03/23 4:48 PM
	PROJECT MANAGER	*60%*	*2.88 hours*	*0.6 days*	*27/03/23 11:00 AM*	*27/03/23 4:48 PM*
	DATA SPECIALIST	*40%*	*1.92 hours*	*0.6 days*	*27/03/23 11:00 AM*	*27/03/23 4:48 PM*
25	Submit D2	100%	0 hours	0 days?	30/03/23 8:00 AM	30/03/23 8:00 AM
26	**⊟D3 – Update D2 and MVP**	**100%**	**163.367 hours**	**18 days?**	**3/04/23 9:00 AM**	**27/04/23 9:00 AM**
27	**⊟D3.1 Minimum Viable Product**	**100%**	**163.367 hours**	**15.625 da...**	**3/04/23 9:00 AM**	**24/04/23 3:00 PM**
28	D3.1.1 Discussion of task and deligating parts	100%	1 hour	0.125 days?	3/04/23 9:00 AM	3/04/23 10:00 AM
	PROJECT MANAGER	*100%*	*1 hour*	*0.125 days*	*3/04/23 9:00 AM*	*3/04/23 10:00 AM*
29	D3.1.2 Complete individual sections	24%	24.3 hours	12.5 days?	3/04/23 10:00 AM	19/04/23 3:00 PM
	CHANGE OFFICER/BUSINESS ANALYST	*60%*	*10.8 hours*	*2.25 days*	*3/04/23 10:00 AM*	*5/04/23 1:00 PM*
	DATA SPECIALIST	*15%*	*3 hours*	*2.5 days*	*3/04/23 10:00 AM*	*5/04/23 3:00 PM*
	PROJECT OFFICER/DATA CONSULTANT	*5%*	*5 hours*	*12.5 days*	*3/04/23 10:00 AM*	*19/04/23 3:00 PM*
	PROJECT OFFICER/CHANGE OFFICER	*15%*	*3 hours*	*2.5 days*	*3/04/23 10:00 AM*	*5/04/23 3:00 PM*
	PROJECT MANAGER	*5%*	*2.5 hours*	*6.25 days*	*3/04/23 10:00 AM*	*11/04/23 1:00 PM*
30	D3.1.3 Add sections and edit	100%	14.4 hours	1.8 days?	19/04/23 3:00 PM	21/04/23 1:24 PM
	PROJECT OFFICER/CHANGE OFFICER	*60%*	*8.64 hours*	*1.8 days*	*19/04/23 3:00 PM*	*21/04/23 1:24 PM*
	PROJECT MANAGER	*40%*	*5.76 hours*	*1.8 days*	*19/04/23 3:00 PM*	*21/04/23 1:24 PM*
31	D3.1.4 Consult Industry Partners	100%	24 hours	3 days?	19/04/23 3:00 PM	24/04/23 3:00 PM
	CHANGE OFFICER/BUSINESS ANALYST	*100%*	*24 hours*	*3 days*	*19/04/23 3:00 PM*	*24/04/23 3:00 PM*
32	D3.1.5 Revise individual part with feedback	100%	24 hours	3 days	19/04/23 9:00 AM	24/04/23 9:00 AM
	PROJECT OFFICER/DATA CONSULTANT	*50%*	*12 hours*	*3 days*	*19/04/23 9:00 AM*	*24/04/23 9:00 AM*
	CHANGE OFFICER/BUSINESS ANALYST	*50%*	*12 hours*	*3 days*	*19/04/23 9:00 AM*	*24/04/23 9:00 AM*
33	**⊟D3.2 Update D2**	**100%**	**75.667 hours**	**9.667 day...**	**10/04/23 9:00 AM**	**21/04/23 3:20 PM**
34	3.2.1 Redraft first submission	92%	19.667 hours	2.667 days?	10/04/23 9:00 AM	12/04/23 3:20 PM
	PROJECT OFFICER/CHANGE OFFICER	*50%*	*10.667 hours*	*2.667 days*	*10/04/23 9:00 AM*	*12/04/23 3:20 PM*
	PROJECT MANAGER	*50%*	*9 hours*	*2.25 days*	*10/04/23 9:00 AM*	*12/04/23 11:00 AM*
35	3.2.2 Submit for consultation to Sponsors	100%	40 hours	5 days?	12/04/23 3:20 PM	19/04/23 3:20 PM
	PROJECT MANAGER	*100%*	*40 hours*	*5 days*	*12/04/23 3:20 PM*	*19/04/23 3:20 PM*
36	3.2.3 Revise and finalise	100%	16 hours	2 days?	19/04/23 3:20 PM	21/04/23 3:20 PM
	PROJECT MANAGER	*100%*	*16 hours*	*2 days*	*19/04/23 3:20 PM*	*21/04/23 3:20 PM*
37	Submit D3	100%	0 hours	0 days	27/04/23 9:00 AM	27/04/23 9:00 AM
38	**⊟D4 – Updated D3 & User/Training Manual**	**100%**	**132.418 hours**	**21.518 da...**	**25/04/23 9:00 AM**	**24/05/23 2:08 PM**
39	**⊟D4.1 Update D3**	**100%**	**132.418 hours**	**21.518 da...**	**25/04/23 9:00 AM**	**24/05/23 2:08 PM**
40	D4.1.1 Discussion of task and deligating parts	100%	8 hours	1 day?	25/04/23 9:00 AM	26/04/23 9:00 AM
	PROJECT MANAGER	*100%*	*8 hours*	*1 day*	*25/04/23 9:00 AM*	*26/04/23 9:00 AM*
41	D4.1.2 Complete individual sections	24%	35.085 hours	18.518 d...	26/04/23 9:00 AM	22/05/23 2:08 PM
	CHANGE OFFICER/BUSINESS ANALYST	*40%*	*7.407 hours*	*2.315 days*	*26/04/23 9:00 AM*	*28/04/23 11:31 AM*
	PROJECT OFFICER/DATA CONSULTANT	*40%*	*2.963 hours*	*0.926 days*	*26/04/23 9:00 AM*	*27/04/23 8:24 AM*
	DATA SPECIALIST	*5%*	*7.407 hours*	*18.518 d...*	*26/04/23 9:00 AM*	*22/05/23 2:08 PM*
	PROJECT OFFICER/CHANGE OFFICER	*5%*	*7.407 hours*	*18.518 d...*	*26/04/23 9:00 AM*	*22/05/23 2:08 PM*
	PROJECT MANAGER	*10%*	*9.9 hours*	*12.375 d...*	*26/04/23 9:00 AM*	*12/05/23 1:00 PM*
42	D4.1.3 Submit for consultation	100%	16 hours	2 days?	2/05/23 9:00 AM	4/05/23 9:00 AM
	CHANGE OFFICER/BUSINESS ANALYST	*100%*	*16 hours*	*2 days*	*2/05/23 9:00 AM*	*4/05/23 9:00 AM*
43	D4.1.4 Revise D3 with feedback	100%	16 hours	2 days?	22/05/23 2:08 PM	24/05/23 2:08 PM
	CHANGE OFFICER/BUSINESS ANALYST	*100%*	*8 hours*	*1 day*	*22/05/23 2:08 PM*	*23/05/23 2:08 PM*
	PROJECT OFFICER/DATA CONSULTANT	*50%*	*8 hours*	*2 days*	*22/05/23 2:08 PM*	*24/05/23 2:08 PM*
44	**⊟D4.2 User Manual**	**100%**	**57.333 hours**	**7.333 day...**	**25/04/23 9:00 AM**	**4/05/23 11:40 AM**
45	D4.2.1 Write first draft	95%	25.333 hours	3.333 days?	25/04/23 9:00 AM	28/04/23 11:40 AM
	PROJECT OFFICER/DATA CONSULTANT	*20%*	*4 hours*	*2.5 days*	*25/04/23 9:00 AM*	*27/04/23 2:00 PM*
	PROJECT OFFICER/CHANGE OFFICER	*40%*	*10.667 hours*	*3.333 days*	*25/04/23 9:00 AM*	*28/04/23 11:40 AM*
	PROJECT MANAGER	*40%*	*10.667 hours*	*3.333 days*	*25/04/23 9:00 AM*	*28/04/23 11:40 AM*
46	D4.2.2 Put the document out for consultation	100%	24 hours	3 days?	28/04/23 11:40 AM	3/05/23 11:40 AM
	PROJECT MANAGER	*100%*	*24 hours*	*3 days*	*28/04/23 11:40 AM*	*3/05/23 11:40 AM*
47	D4.2.3 Revise draft with feeback and finalise	100%	8 hours	1 day?	3/05/23 11:40 AM	4/05/23 11:40 AM
	PROJECT OFFICER/CHANGE OFFICER	*100%*	*8 hours*	*1 day*	*3/05/23 11:40 AM*	*4/05/23 11:40 AM*
48	Submit D3	100%	0 hours	0 days?	24/05/23 2:08 PM	24/05/23 2:08 PM

49	**⊟SD1 – Findings report**	100%	118.4 hours	39 days?	**25/04/23 9:00 AM**	**19/06/23 9:00 AM**
50	**⊟SD1.1 Production of Findings report**	100%	70.4 hours	9.625 day...	**25/04/23 9:00 AM**	**8/05/23 3:00 PM**
51	SD1.1.1Hold discussions and brainstorming session	100%	16 hours	2 days?	25/04/23 9:00 AM	27/04/23 9:00 AM
	PROJECT MANAGER	*20%*	*3.2 hours*	*2 days*	*25/04/23 9:00 AM*	*27/04/23 9:00 AM*
	PROJECT OFFICER/CHANGE OFFICER	*20%*	*3.2 hours*	*2 days*	*25/04/23 9:00 AM*	*27/04/23 9:00 AM*
	PROJECT OFFICER/DATA CONSULTANT	*20%*	*3.2 hours*	*2 days*	*25/04/23 9:00 AM*	*27/04/23 9:00 AM*
	DATA SPECIALIST	*20%*	*3.2 hours*	*2 days*	*25/04/23 9:00 AM*	*27/04/23 9:00 AM*
	CHANGE OFFICER/BUSINESS ANALYST	*20%*	*3.2 hours*	*2 days*	*25/04/23 9:00 AM*	*27/04/23 9:00 AM*
52	SD1.1.2 Deligate sections	100%	1 hour	0.125 days?	27/04/23 9:00 AM	27/04/23 10:00 AM
	PROJECT MANAGER	*100%*	*1 hour*	*0.125 days*	*27/04/23 9:00 AM*	*27/04/23 10:00 AM*
53	SD1.1.3 Research and compute data	100%	30 hours	3.75 days?	27/04/23 10:00 AM	2/05/23 5:00 PM
	DATA SPECIALIST	*50%*	*15 hours*	*3.75 days*	*27/04/23 10:00 AM*	*2/05/23 5:00 PM*
	CHANGE OFFICER/BUSINESS ANALYST	*10%*	*3 hours*	*3.75 days*	*27/04/23 10:00 AM*	*2/05/23 5:00 PM*
	PROJECT OFFICER/DATA CONSULTANT	*20%*	*6 hours*	*3.75 days*	*27/04/23 10:00 AM*	*2/05/23 5:00 PM*
	PROJECT OFFICER/CHANGE OFFICER	*10%*	*3 hours*	*3.75 days*	*27/04/23 10:00 AM*	*2/05/23 5:00 PM*
	PROJECT MANAGER	*10%*	*3 hours*	*3.75 days*	*27/04/23 10:00 AM*	*2/05/23 5:00 PM*
54	SD1.1.4 Draft report	78%	23.4 hours	3.75 days?	3/05/23 8:00 AM	8/05/23 3:00 PM
	CHANGE OFFICER/BUSINESS ANALYST	*12%*	*3.6 hours*	*3.75 days*	*3/05/23 8:00 AM*	*8/05/23 3:00 PM*
	DATA SPECIALIST	*30%*	*6 hours*	*2.5 days*	*3/05/23 8:00 AM*	*5/05/23 1:00 PM*
	PROJECT OFFICER/DATA CONSULTANT	*12%*	*1.2 hours*	*1.25 days*	*3/05/23 8:00 AM*	*4/05/23 10:00 AM*
	PROJECT OFFICER/CHANGE OFFICER	*30%*	*9 hours*	*3.75 days*	*3/05/23 8:00 AM*	*8/05/23 3:00 PM*
	PROJECT MANAGER	*12%*	*3.6 hours*	*3.75 days*	*3/05/23 8:00 AM*	*8/05/23 3:00 PM*
55	**⊟SD 1.2 Consultations**	**100%**	**48 hours**	**24 days?**	**16/05/23 9:00 AM**	**19/06/23 9:00 AM**
56	SD1.2.1 Consult industry partners	100%	32 hours	4 days?	16/05/23 9:00 AM	22/05/23 9:00 AM
	DATA SPECIALIST	*100%*	*32 hours*	*4 days*	*16/05/23 9:00 AM*	*22/05/23 9:00 AM*
57	SD1.2.2 Revise and finalise Findings Report	10%	16 hours	20 days?	22/05/23 9:00 AM	19/06/23 9:00 AM
	PROJECT OFFICER/CHANGE OFFICER	*2%*	*3.2 hours*	*20 days*	*22/05/23 9:00 AM*	*19/06/23 9:00 AM*
	PROJECT OFFICER/DATA CONSULTANT	*60%*	*3.2 hours*	*0.667 days*	*22/05/23 9:00 AM*	*22/05/23 3:20 PM*
	PROJECT MANAGER	*2%*	*3.2 hours*	*16 days*	*22/05/23 9:00 AM*	*13/06/23 9:00 AM*
	DATA SPECIALIST	*30%*	*3.2 hours*	*1.333 days*	*22/05/23 9:00 AM*	*23/05/23 11:40 AM*
	CHANGE OFFICER/BUSINESS ANALYST	*5%*	*3.2 hours*	*8 days*	*22/05/23 9:00 AM*	*1/06/23 9:00 AM*
58	SD1.2.3 Findings report added to GitHub	100%	0 hours	0 days?	19/05/23 9:00 AM	19/05/23 9:00 AM
	PROJECT OFFICER/CHANGE OFFICER	*100%*	*0 hours*	*0 days*	*19/05/23 9:00 AM*	*19/05/23 9:00 AM*

59	**⊟SD2 – Recommendations Report**	100%	96.999 hours	12.302 da...	**24/05/23 2:08 PM**	**9/06/23 4:33 PM**
60	**⊟SD2.1 Consultations**	100%	96.999 hours	12.302 da...	**24/05/23 2:08 PM**	**9/06/23 4:33 PM**
61	SD2.1.1 Hold discussions and brainstorm	100%	16 hours	2 days?	24/05/23 2:08 PM	26/05/23 2:08 PM
	PROJECT MANAGER	*20%*	*3.2 hours*	*2 days*	*24/05/23 2:08 PM*	*26/05/23 2:08 PM*
	PROJECT OFFICER/CHANGE OFFICER	*20%*	*3.2 hours*	*2 days*	*24/05/23 2:08 PM*	*26/05/23 2:08 PM*
	PROJECT OFFICER/DATA CONSULTANT	*20%*	*3.2 hours*	*2 days*	*24/05/23 2:08 PM*	*26/05/23 2:08 PM*
	DATA SPECIALIST	*20%*	*3.2 hours*	*2 days*	*24/05/23 2:08 PM*	*26/05/23 2:08 PM*
	CHANGE OFFICER/BUSINESS ANALYST	*20%*	*3.2 hours*	*2 days*	*24/05/23 2:08 PM*	*26/05/23 2:08 PM*
62	SD2.1.2 Consult Sponsor	100%	8 hours	1 day?	24/05/23 2:08 PM	25/05/23 2:08 PM
	PROJECT OFFICER/DATA CONSULTANT	*100%*	*8 hours*	*1 day*	*24/05/23 2:08 PM*	*25/05/23 2:08 PM*
63	**⊟SD2.2 RR Production**	100%	72.999 hours	12.302 da...	**24/05/23 2:08 PM**	**9/06/23 4:33 PM**
64	SD2.2.1 Deligate roles	100%	1 hour	0.125 days?	24/05/23 2:08 PM	24/05/23 3:08 PM
	PROJECT MANAGER	*100%*	*1 hour*	*0.125 days*	*24/05/23 2:08 PM*	*24/05/23 3:08 PM*
65	SD2.2.2 Research recommendations	50%	32 hours	8 days?	26/05/23 2:08 PM	7/06/23 2:08 PM
	PROJECT OFFICER/DATA CONSULTANT	*50%*	*6.4 hours*	*1.6 days*	*26/05/23 2:08 PM*	*30/05/23 9:56 AM*
	CHANGE OFFICER/BUSINESS ANALYST	*10%*	*6.4 hours*	*8 days*	*26/05/23 2:08 PM*	*7/06/23 2:08 PM*
	DATA SPECIALIST	*20%*	*6.4 hours*	*4 days*	*26/05/23 2:08 PM*	*1/06/23 2:08 PM*
	PROJECT OFFICER/CHANGE OFFICER	*10%*	*6.4 hours*	*8 days*	*26/05/23 2:08 PM*	*7/06/23 2:08 PM*
	PROJECT MANAGER	*10%*	*6.4 hours*	*8 days*	*26/05/23 2:08 PM*	*7/06/23 2:08 PM*
66	SD2.2.3 Write first draft	43%	31.999 hours	9.302 days?	26/05/23 2:08 PM	8/06/23 4:33 PM
	PROJECT OFFICER/CHANGE OFFICER	*30%*	*7.442 hours*	*3.101 days*	*26/05/23 2:08 PM*	*31/05/23 2:57 PM*
	PROJECT MANAGER	*15%*	*7.442 hours*	*6.201 days*	*26/05/23 2:08 PM*	*5/06/23 3:45 PM*
	PROJECT OFFICER/DATA CONSULTANT	*10%*	*7.442 hours*	*9.302 days*	*26/05/23 2:08 PM*	*8/06/23 4:33 PM*
	DATA SPECIALIST	*15%*	*7.442 hours*	*6.201 days*	*26/05/23 2:08 PM*	*5/06/23 3:45 PM*
	CHANGE OFFICER/BUSINESS ANALYST	*30%*	*2.232 hours*	*0.93 days*	*26/05/23 2:08 PM*	*29/05/23 1:35 PM*
67	SD2.2.4 Edit and Finalise	100%	8 hours	1 day?	8/06/23 4:33 PM	9/06/23 4:33 PM
	PROJECT OFFICER/CHANGE OFFICER	*100%*	*8 hours*	*1 day*	*8/06/23 4:33 PM*	*9/06/23 4:33 PM*
68	Recommendations Report added to GitHub	100%	0 hours	0 days?	29/05/23 9:00 AM	29/05/23 9:00 AM

69	**D5 – Final Group Reflection**	100%	18.198 hours	5.482 day...	**24/05/23 2:08 PM**	**1/06/23 9:00 AM**
70	Group Reflection session	100%	8 hours	1 day?	24/05/23 2:08 PM	25/05/23 2:08 PM
	PROJECT MANAGER	*20%*	*1.6 hours*	*1 day*	*24/05/23 2:08 PM*	*25/05/23 2:08 PM*
	PROJECT OFFICER/CHANGE OFFICER	*20%*	*1.6 hours*	*1 day*	*24/05/23 2:08 PM*	*25/05/23 2:08 PM*
	PROJECT OFFICER/DATA CONSULTANT	*20%*	*1.6 hours*	*1 day*	*24/05/23 2:08 PM*	*25/05/23 2:08 PM*
	DATA SPECIALIST	*20%*	*1.6 hours*	*1 day*	*24/05/23 2:08 PM*	*25/05/23 2:08 PM*
	CHANGE OFFICER/BUSINESS ANALYST	*20%*	*1.6 hours*	*1 day*	*24/05/23 2:08 PM*	*25/05/23 2:08 PM*
71	Drafting structure of report	84%	1.598 hours	0.238 days?	25/05/23 2:08 PM	25/05/23 4:02 PM
	PROJECT MANAGER	*20%*	*0.076 hours*	*0.048 days*	*25/05/23 2:08 PM*	*25/05/23 2:31 PM*
	PROJECT OFFICER/CHANGE OFFICER	*20%*	*0.38 hours*	*0.238 days*	*25/05/23 2:08 PM*	*25/05/23 4:02 PM*
	PROJECT OFFICER/DATA CONSULTANT	*20%*	*0.381 hours*	*0.238 days*	*25/05/23 2:08 PM*	*25/05/23 4:02 PM*
	DATA SPECIALIST	*20%*	*0.381 hours*	*0.238 days*	*25/05/23 2:08 PM*	*25/05/23 4:02 PM*
	CHANGE OFFICER/BUSINESS ANALYST	*20%*	*0.381 hours*	*0.238 days*	*25/05/23 2:08 PM*	*25/05/23 4:02 PM*
72	Writing report	90%	0.6 hours	0.083 days?	24/05/23 2:08 PM	24/05/23 2:48 PM
	CHANGE OFFICER/BUSINESS ANALYST	*5%*	*0 hours*	*0 days*	*24/05/23 2:08 PM*	*24/05/23 2:08 PM*
	PROJECT OFFICER/CHANGE OFFICER	*5%*	*0 hours*	*0 days*	*24/05/23 2:08 PM*	*24/05/23 2:08 PM*
	PROJECT MANAGER	*90%*	*0.6 hours*	*0.083 days*	*24/05/23 2:08 PM*	*24/05/23 2:48 PM*
73	Editing and revising	100%	8 hours	1 day?	25/05/23 4:02 PM	26/05/23 4:02 PM
	PROJECT OFFICER/DATA CONSULTANT	*100%*	*8 hours*	*1 day*	*25/05/23 4:02 PM*	*26/05/23 4:02 PM*
74	Submit Final Group Reflection	100%	0 hours	0 days?	1/06/23 9:00 AM	1/06/23 9:00 AM
75	**D6 – Project Presentation**	100%	87.999 hours	10.839 da...	**19/06/23 9:00 AM**	**3/07/23 4:42 PM**
76	D6.1 Prepare a storyboard/ slide deck template	100%	24 hours	3 days?	19/06/23 9:00 AM	22/06/23 9:00 AM
	CHANGE OFFICER/BUSINESS ANALYST	*100%*	*24 hours*	*3 days*	*19/06/23 9:00 AM*	*22/06/23 9:00 AM*
77	D6.2 Insert summary points per report and process outlined	100%	32 hours	4 days?	19/06/23 9:00 AM	23/06/23 9:00 AM
	PROJECT OFFICER/CHANGE OFFICER	*100%*	*32 hours*	*4 days*	*19/06/23 9:00 AM*	*23/06/23 9:00 AM*
78	D6.3 Review slides and prepare speaking notes	50%	8 hours	2 days?	23/06/23 9:00 AM	27/06/23 9:00 AM
	PROJECT MANAGER	*50%*	*1.6 hours*	*0.4 days*	*23/06/23 9:00 AM*	*23/06/23 1:12 PM*
	PROJECT OFFICER/CHANGE OFFICER	*15%*	*1.6 hours*	*1.333 days*	*23/06/23 9:00 AM*	*26/06/23 11:40 AM*
	PROJECT OFFICER/DATA CONSULTANT	*10%*	*1.6 hours*	*2 days*	*23/06/23 9:00 AM*	*27/06/23 9:00 AM*
	DATA SPECIALIST	*10%*	*1.6 hours*	*2 days*	*23/06/23 9:00 AM*	*27/06/23 9:00 AM*
	CHANGE OFFICER/BUSINESS ANALYST	*15%*	*1.6 hours*	*1.333 days*	*23/06/23 9:00 AM*	*26/06/23 11:40 AM*
79	D6.4 Practice presentation	62%	23.999 hours	4.839 days?	27/06/23 9:00 AM	3/07/23 4:42 PM
	PROJECT MANAGER	*30%*	*11.613 hours*	*4.839 days*	*27/06/23 9:00 AM*	*3/07/23 4:42 PM*
	DATA SPECIALIST	*10%*	*3.871 hours*	*4.839 days*	*27/06/23 9:00 AM*	*3/07/23 4:42 PM*
	PROJECT OFFICER/CHANGE OFFICER	*10%*	*3.871 hours*	*4.839 days*	*27/06/23 9:00 AM*	*3/07/23 4:42 PM*
	PROJECT OFFICER/DATA CONSULTANT	*10%*	*3.871 hours*	*4.839 days*	*27/06/23 9:00 AM*	*3/07/23 4:42 PM*
	CHANGE OFFICER/BUSINESS ANALYST	*20%*	*0.774 hours*	*0.484 days*	*27/06/23 9:00 AM*	*27/06/23 1:52 PM*
80	Submit/Presentation	100%	0 hours	0 days?	19/06/23 9:00 AM	19/06/23 9:00 AM
	PROJECT MANAGER	*100%*	*0 hours*	*0 days*	*19/06/23 9:00 AM*	*19/06/23 9:00 AM*

81	**D7 – Final Delivery of Products**	100%	1 hour	0.125 day...	**10/06/23 8:00 AM**	**12/06/23 9:00 AM**
82	D7.1 Send SD1 and SD2 from GitHub to Delivery Partner	100%	1 hour	0.125 days?	10/06/23 8:00 AM	12/06/23 9:00 AM
	PROJECT OFFICER/CHANGE OFFICER	*100%*	*1 hour*	*0.125 days*	*10/06/23 8:00 AM*	*12/06/23 9:00 AM*

SRS/SCOPING

1. OVERVIEW

1.1. INTENDED AUDIENCE

The intended audience for our project will include the key project stakeholders (SAS Institute and Zoe Empowers) and where appropriate, interested parties hoping to gain insight into the impacts the Zoe Empowers (ZE) project has on both wider and local communities in developing nations.

1.2. PROJECT SCOPE

The aim of this project is to analyse the data provided by SAS for their partnered charity organisation, Zoe Empowers. The project will use data analytics methods and tools, such as Viya (SAS-approved cloud platform) and Python, to extract key insights regarding the program and its participants' engagement/success before, during, and after the program. DataSynergy will produce three reports: Findings, Analysis and Recommendations. The Findings Report will provide detailed data visualisations and exploration of the raw data provided by SAS. The Analysis Report will identify any vulnerabilities/gaps that may exist within the program structure/activities, based on key information derived from the Findings Report. The Recommendations report will provide suggestions for improving the effectiveness of the program based on the vulnerabilities/gaps identified in the Analysis report. Lastly, DataSynergy will deliver a presentation to SAS that provides insightful information regarding the three reports and pinpoint various aspects of the Zoe Empowers program that may require improvement. It is important to note that this project will not deliver any solutions to SAS, regarding any aspects of the Zoe Empowers program. Additionally, it will not project future data outcomes, but instead, will focus on analysing the current efforts of the program and how they can be improved for future intakes.

2. DATA UNDERSTANDING

2.1. INITIAL DATA SOURCES

The data that has been provided to us by SAS has been deemed sufficient to execute the entirety of the project. The sponsor has provided three data files in a csv format that describe different aspects of the Zoe Empowers' work.

Firstly, two of the three data sets are split into two respective countries, Kenya and Rwanda, with roughly 400 rows and 500 rows respectively. Both these datasets are in the exact same format with perfectly corresponding headings, which has been deemed beneficial for the project team as the alternative of joining two tables would be trivial. The data sets identify participants per row with a unique identifier that does not overlap between datasets. There is no personally identifying information (PII) about individual participants outside what is useful to the study, maintaining confidence in PII. The data that is included for the participant ranges from basic classifiers like, gender, religion, age and level of education, none of which are personally identifiable. Information that specifically pertains to the study includes points of interest and growth that Zoe Empowers has identified for us. The years in which the participants have answered the questionnaire ranges from 2015 to 2018.

There is a detailed list of 56 columns including categories that carry simple information whereas some columns proved unhelpful due to the amount of filler data that exists. The following identified information has been sighted as ‘helpful’ to the project,including:

- Nutrition
    - How much food participants have and what types
    - How to prepare food
- Finance Status
    - Type of financial position participants are in
    - Their savings;
    - or investment in farming
- Program Structure/ Provisions
    - What provisions were received by ZE
    - What stage in training they’re up to in the program;
    - Recently on boarded or recently graduated from the program.
- Knowledge around rights

It is important to note that there is a variety of categorical and continuous types of data items within the data set. It is also quite full of N/A values that have to be managed. The data set is quite extensive and will provide a large amount of interesting points to go over and investigate.

The last data set is possibly the most important out of the three, it contains the Self Sufficiency Index (SSI) for all the participants. The Self Sufficiency Index is created by Zoe Empowers and SAS to quantify how self-sufficient a participant is at the point of doing the questionnaire. The data set contains all the same unique identifiers for both Kenya and Rwanda, which is ideal for joining the data with only less than 10 participants not matching up. It also contains 9 other columns, the index itself which is a summation of the other eight. The other eight categories are Food Security and Nutrition, Housing, Community Connections, Health and Hygiene, Child Rights, Education, Economy / IGA and Spiritual Strength. These all are a 4 point value range from 0 to 3, which also means that the self sufficiency index ranges from 0 to 24. There are some places in this data set that do have null values (which are sometimes represented with a ‘.’) but all the Self Sufficiency Indexes are filled, so these can be simply reverse engineered if required.

2.2. DATA COLLECTIONS AND CAPTURES

The data that is provided is the basis for our entire project. It is presumably direct from the source and appears to have already been managed for us. There is a large complexity to this data that could lead to many different investigations and angles to tackle this project. However, there is always room for more data to be included. Extra outside information that could improve the contextual understanding of what is happening in Kenya and Rwanda could prove valuable, especially since the data is from 2015 - 2018 and the pandemic that has since ensued would have impacted the participants of the Zoe Empowers. The data we would look for is likely to be provided from reputable sources such as government/international agencies like WHO’s global health observatory (GHO) databases and the UN’s Sustainable Development Goals (SDG) global database. However it is unlikely that we would need to collect/organise more data than what we have been provided with. Data sighted from either WHO or the UN will provide supplementary information to assist in drawing analysis and conclusions to the project's work. Examples of reference data can include WHO’s GHO indicator ‘Joint Child Malnutrition estimation’ and the UN’s SDG ‘Sustainable cities and communities’ CVS data files.

2.3. DATA QUALITY

At first glance, the data is of high quality.

- It is structured very effectively with lots of features pertaining to an individual participant.
- There is no present evidence of duplication within the data.
- It is simple to understand.
- There are only three files, each lining up almost perfectly with each other
    - Appending Kenya and Rwanda with the same column names.
    - Joining the Self-Sufficiency Index on the unique id.

Upon investigating the data and its quality, two notable issues presented themselves and were dealt with accordingly using tailored data sanity checks.

Not Applicable (N/A) Data Checks:

- There is a tendency to have NAs that are concentrated in a few direct columns such as 'Vocational Training' in the Kenya and Rwanda datasets and 'Education' in the Self Sufficiency Index dataset.
- This was rectified by placing the values in excel and using short functions and filtering to remove N/A values.

Unaccounted for category of data:

- There is no data for when the participants have graduated from Zoe Empowers and practically (a single row out of around 900) no data on truly before the participants begin work with Zoe Empowers. This will make the discoveries around the impact of participating in Zoe Empowers difficult to properly understand.

Duplication Data:

- Additionally be used to identify the distinct values within indexes and attempt de-duplication of values.
- Since no de-duplication was generated by excel, it is safe to assume that there are no obvious duplicates. Joining the three files is more technical, software such as Viya or Python has the tools available to append/join files and also produce a report on successfulness.

3. DATA PREPARATION

Due to the high quality of the data, there does not seem to be a large amount of data preparation required compared to most data analysis projects. That does not mean there is nothing to do. Firstly there will need to be some cleaning performed on the datasets. Next for particular models

the data needs to be transformed into a suitable format. Lastly there would be a great benefit to gain from appropriately appending and joining data sets as already described.

3.1. DATA TYPES AND VISUALISATION

Across the project the team will interact with data that is collated to yes (Y,1) and no (N,0) values and encoded satisfaction/response ratings (0 least satisfied to 4 most satisfied) in regards to questions of answer and response for example 'Level of Family Education' as 0 = No education completed,1 = Primary School, 2 = Secondary and 3 = University. All this data is encoded into numerical values and are both analysed and visualised using the AI within the Viya platform. The visualisations are done through a series of 'objects' within the Viya AI tool including and not limited to:

- Decision Trees - answer specific survey questions along with other factors.
- Bar Charts - presents numerical importance and responses.
- Box Plots - demonstrates the minimum, median and maximum pertaining to data.
- Scatter plots - demonstrates distribution of data.
- Heat Maps - demonstrates distribution of data.
- Pie Charts - shows number of responses/ input for a particular category.
- Dual-axis line/chart - demonstrates distribution of data and linear attributes.

All visualisations can be linked to each other using Viya's features and produce results that are tailored for specific groups, factors and categories providing deeper insight into the data.

3.2. CLEANING

The first step in cleaning the data would be managing the null/NA values. Across the board there are several points where the null values appear and need to be managed. Firstly the SSI data set has a pattern of missing values specifically in one index. Luckily for us, the SSI column is simply a summation of the others, when only one column is missing it is possible to calculate exactly what the value was supposed to be. The Rwanda data set in particular has a large amount of null values that will be more difficult to find the answer to. The amount of features that the data has is quite large, this is a beneficial property as some columns can be extrapolated, this technique is called Hot-Deck Imputation. For example, "Taught vocational training" would be dependent on if they've completed or even started "Vocational/Skill Training". This can be applied across several different sections of the data but doesn't solve all the issues. Unfortunately due to the nature of some indexes deletion is the only choice, this is aimed at the indexes that have categorical data that provide all possible information. It would be possible to encode a numerical representation and then take an average to fill the missing value, but considering the nature that this is participants lives and that some of these columns are primarily missing data, deletion is the better option. Specifically, keeping to pairwise deletion as mentioned before, there are many features and to perform listwise deletions and remove all data points that contain null values would result in a far greater loss of information than simply removing them when needed. This does make the approach in modelling more complex as cleaning is required almost per model, but the information retained from this approach is required for a majority of analysis to be performed.

3.3. TRANSFORMATION

Several techniques would be helpful for transforming the data into a better format to aid different models. The model that would benefit the most from transformations would be the multiple regression models. Firstly, the model can handle both categorical and continuous data, but is more

sensitive to continuous/numerical data. To help build a more accurate model first we should encode categorical data into a numerical format, change indexes that have Y and N values to 1 and 0s, and encode the multiple choice questions into their respective values. For example, the index “Membership status” has four options, Not Yet a Member, Member less than 3 months, Member 1yr - 2yr and Recent Graduate (less than three months) can be encoded to values 0 to 3. It would be a time consuming process to translate all indexes from the start, so again it is likely to be done similar to pairwise deletion, on demand. The next transformation that the multiple regression model would benefit from is normalisation. This involves reducing the data from a full range to a range between 0 and 1 for all values. This provides more accurate results for both simple linear regression and multiple regression models as it is easier to calculate the relevant significance scales between the predictor variables and their effect on the response variable. Additionally, normalisation increases readability to explorative models and visualisation techniques as the scales would be equal across every axis. If all the techniques were applied to the data from the beginning, the format that would be left would be all values encoded numerically and then normalised. However, both encoding and normalisation applies a level of abstraction to the data, keeping a record of the original data to cross check assumptions and understanding would be key to progressing effectively.

3.4. STORAGE

The storage systems for managing this project are pretty simple. The files are not too excessive and can be stored both locally and on a **cloud system** like google drive. The benefit of having **cloud storage** is the ability to see exactly the same item as a team member, with version control not considered a particular concern if the team keeps prime documents separated. Another noteworthy point is that these raw files are stored on SAS’s **service software** Viya, a **cloud-based**, data visualisation AI platform that retains an original copy of the data at all times. Throughout the project, the direct coding produced on software platforms will be pushed onto **GitHub** for version control and access by team members. It is important to keep all the places that we store the data files private as we are managing the data that was provided to us by SAS and we ourselves do not hold ownership. On top of that an Intellectual Property agreement has been signed with SAS regarding managing the data sets and the reports that we are going to produce for them.

The team needs to be mindful that as we start analysing the data sets we will generate new sets, some larger as we combine information and some smaller as we subject the data to categorisations. It’s an extremely helpful thing to stay ahead of the growth of complexity and communicate the different storage systems and the current location of where everything is kept.

4. MODELLING

There is a defiant split between two types of models during the lifetime of a data science project, exploration and evaluation. Firstly exploration models are about learning what is in the data. They delve into the raw data points and summarise information, identify trends/relationships and build questions about what we can analyse. On the other hand evaluation models are those that build more statistically significant information and demonstrate quantitatively how data relates to produce information. They work in tandem, exploration models help produce evaluation modes to define our understanding better. Evaluation models can produce more questions themselves,

especially if they fail to produce results, which push us to scope outwards and generate more exploration models to “rephrase” our questions.

4.1. EXPLORATION MODELS

These models are the beginning of extrapolating information from the data. They are used to present overviews of individual variables and interaction between them. They’re doing the work to provide context from random data points, turn it from a clean excel spreadsheet into something that can be understood at a quick glance. An even more important role that they serve is helping to prepare data for the evaluation models by showing facts such as outliers.

The most basic and beginning part to any data exploration is the classic histogram visualisation. The histogram shows the frequency distribution of a single variable. This makes for a simple way to begin understanding how the data is behaving in this project. It will guide us in knowing if the data is normally distributed or not. With correct bin selection it will also help us identify outliers early. The first place this would be deployed is directly on the variable of most interest, the SSI, this index distributed according to the Central Limit Theorem, is a considerable number of participants for the Theorem to take hold. If the data does not hold up to the Central Limit Theorem and doesn’t approach a normal distribution, then straight away the histogram is producing interesting questions for deeper explorations. Once a variable of interest has been identified, more quantitative techniques are required. To identify normal distribution of the data, a Normal Q-Q plot can be employed, if the distribution lines up with the diagonal, then the assumption of normality can be confirmed for more evaluative models and/or hypothesis testing techniques. Furthermore, if needing to identify outliers a box plot will visually show the quantitative measurement of inter quartile range of a variable, any points that are outside the expected range should be removed before employing models such as multiple regression as they can impact accuracy.

An interesting model for judging the significance of predictor variables and the variable of interest is a decision tree. This model classifies the impact of both categorical and continuous variables onto the chosen variable of interest. It shows the path of significance from the root node through the predictor variable to the variable of interest quantitatively, showing what are the significant predictors earlier in the path. It will identify more interesting variables in the data and several decision trees can show us different trends. It is also designed as a classifier model, helping to provide another way to classify other rows in the data that might have not been entered correctly into the data set, however it is a supervised model and must be trained with a steady hand to attempt to avoid overfitting and biases. If in the earlier stages it was difficult to clear missing data points a well trained decision tree can help us fill the classification of particular variables. These also point us in the right direction for more robust statistical models about variable impacts such as multi linear regression. Although before deployment we need to acknowledge that the decision trees are not a great model for complex data, which our data set for it. It should not be blindly applied to the entire or large sections of the dataset. When we have questions that focus on a particular few variables this is when decision trees are best employed. They should prove helpful in identifying the way that the more economically focused data points impact on the Self Sufficiency Index and lead us towards more research, if required.

4.2. EVALUATION MODELS

Explorative models are helpful visualisation tools that teach us about the shapes of and between variables in data. They generate questions and intrigue about the nature of the data that we desire

to analyse, a blurb to the more vigorous component of statistical analysis that quantifies and evaluates the data. Different models have to be used to acquire statistical significance.

4.2.1. SIMPLE LINEAR REGRESSION

A model that is likely to show a great deal of interest in evaluating the effectiveness of Zoe Empower's methodology is linear regression. Simple linear regression is useful for quantifying the relationship between two variables. Simple analysis between the predictor, time spent in Zoe Empower's and the effect, participants' SSI, should have a relationship. Linear regression models will help quantify how much of an effect the predictor will have on the predicted variable, if at all. Additionally an advantage of the model is that it can describe the level of uncertainty within the model and even demonstrate a visual representation of it with a confidence interval. More statistically rigorous methods such as finding the p value of the predictor and quantifying the relationship through an R squared measurement will also be deployed. Linear regression also has some constraints, it assumes linearity, which is to say that it assumes the relationship between variables will be linear, it does not have the ability to measure different relationships such as polynomials. However we predict that the participants' Self Sufficiency Index will increase with time, as Zoe Empowers appears to be an effective charity organisation and our statistically rigorous measurements will help identify if this prediction is reality. Unfortunately, linear regression is susceptible to outliers in the data, they can have a dramatic effect on the models effectiveness and accuracy. To manage this, the explorative models like histograms and boxplots will help us identify the outliers in specific categories and remove them before applying the linear regression model.

4.2.2. MODELLING CONSTRAINTS

A significant constraint that was identified from the data provided was the lack of response from individuals before the program and those post program. This missing information doesn't allow DataSyngery to paint an extensive picture of the long-term impact or prior impact to the program but constrains output to only to newly onboarded participants, current participants (which is substantially smaller) and newly graduated (within three months) participants. This data therefore constrains the team to producing immediate data visualisations to the years of 2015 to 2018. No modelling assessments can be provided due to the nature of the project being data visualisation dominant and the use of an AI cloud software that produces output without demonstrating the process in between otherwise known as a black box approach. However we can evaluate qualitative metrics and quantitative metrics regarding the visualisation aspect of the project. The qualitative aspect will focus on quantifying intangible components of the project and relating these to global organisations such as the UN and WHO's global statistics and databases (as previously mentioned) to cross reference relevancy and quality of the project's output. Quantitative measures include the cumulative values of the encoded responses from participants demonstrating patterns and pave the way for in depth analysis with supporting numerical evidence.

5. EVALUATION

The main focus on the project is to determine the effectiveness of the Zoe Empowers' program as a whole. With a general nudge towards investigating the demographics and macroeconomic factors that influence the participants self-sufficiency. This leads us to a more open exploration of the data once we have found our main points about effectiveness. There has been other research conducted on this dataset and the effectiveness of Zoe Empowers' program and the participants' self sufficiency. An example is an executive summary by Kenneth Hinze, a sociologist-demographer that retired in 2001 from being a research-active Professor at Louisiana State University in Shrevepo. This summary is a rigorous statistical analysis of the SSI and the

validity of the claim that Zoe Empowers' program does indeed create improvement. The reports we will be creating will take the context that there has been research already conducted, but will not be comparing our findings with those that have already been reported on.
After analysis has been completed we will then need to create recommendations based on our findings. Our goal is to make insightful recommendations on improving specific attributes of the program, singling out what data points make a significant impact on the participants. It is likely that we will also find some recommendations about their data acquisition processes, but that is almost always a given for even data recommendation reports. There will never be enough data in the world for analysis. If these recommendations are of high enough quality and spark enough interest, there may be a slim chance that these findings will be shown directly to Zoe Empowers. Other than that, it is not likely there will be any steps to take after this project has demonstrated the information found.

6. DEPLOYMENT

The project will have four main deliverables to the sponsor: The Findings Report, Analysis Report, Recommendations Report and a Presentation. The Presentation will be the first deliverable to the Sponsor SAS Institute, followed by the three reports handed in together. These deliverables are in accordance with the dictated assessments schedule from the university and the specified Sponsor deliverable brief .

The project deliverables will be monitored throughout its development, via a series of university deliverables which will act as "checkpoints" for the "monitoring and maintenance" across documentation. The deliverables have been scheduled into a Gantt chart to track timing of tasks and are visually displayed in a project Trello board. **Because this project's deliverables are not systems based there is no need for a systematic approach for any system evaluation.** Feedback regarding each report and its contents will be briefly revised at the fortnightly sponsor meetings to clarify any points the sponsor wishes to address and ensure the team is on track.

In terms of training, a brief summary of instructions will be provided in each delivered document and a quick rundown of these instructions will be presented during the final presentation.

7. SAS Feedback and Team Response/Action

Meeting date and time: Tuesday 28th March at 9:30 am - 10 am.

Feedback received

- Scope section although good for academic submission doesn't necessarily meet industry standards, it can be shortened to what will be delivered rather than how. Keeping it short and concise is the best industry approach.
- SAS highlighted that they were looking for a distinct analysis section from the findings to be included.

Follow-up feedback from Jordan Mowlai SAS Representative:
As discussed in the catch-up here are my points regarding the scoping document –

Business project scope documents are clear and concise on the project scope and deliverables. It is meant for non-technical individuals to get a better understanding of what is required and the constraints involved.

Usually the project scope is within the range of just a couple of pages, I understand that you have a marking rubric to go off and thus I will talk against this. Overall I think it is a very comprehensive document and I can see a lot of thought has gone into it. All of the main points along the rubric have been addressed, except two 1. Integration and 2. Formats. **As our project these two are not required it might be best to simply mention that these are not relevant to the project just in case they mark you down for this.**

Another time to address is just general documentation work, grammatical errors, formatting, language, etc.

Again, unfortunately you are constrained by the rubric as it's not exactly what a scoping document is usually set up but overall looks very well done.

Team response/action points

- Revise scope to semi meet industry standards and academic standards (*meeting halfway*)
- Will divide the findings/analysis report into 2 distinct sections with findings and analysis.

INCREMENT ONE

1. ANALYSIS & DESIGN DOCUMENT

1.1. FEATURE ENGINEERING

We are tasked with evaluating the effectiveness of the Zoe Empower's program at improving the lives of their participants. We have split this into two main sections.

- Assess the improvement of overall SSI for participants.
- Break down the aspects of the SSI and learn what exactly is being improved and subsequent trends within those factors.

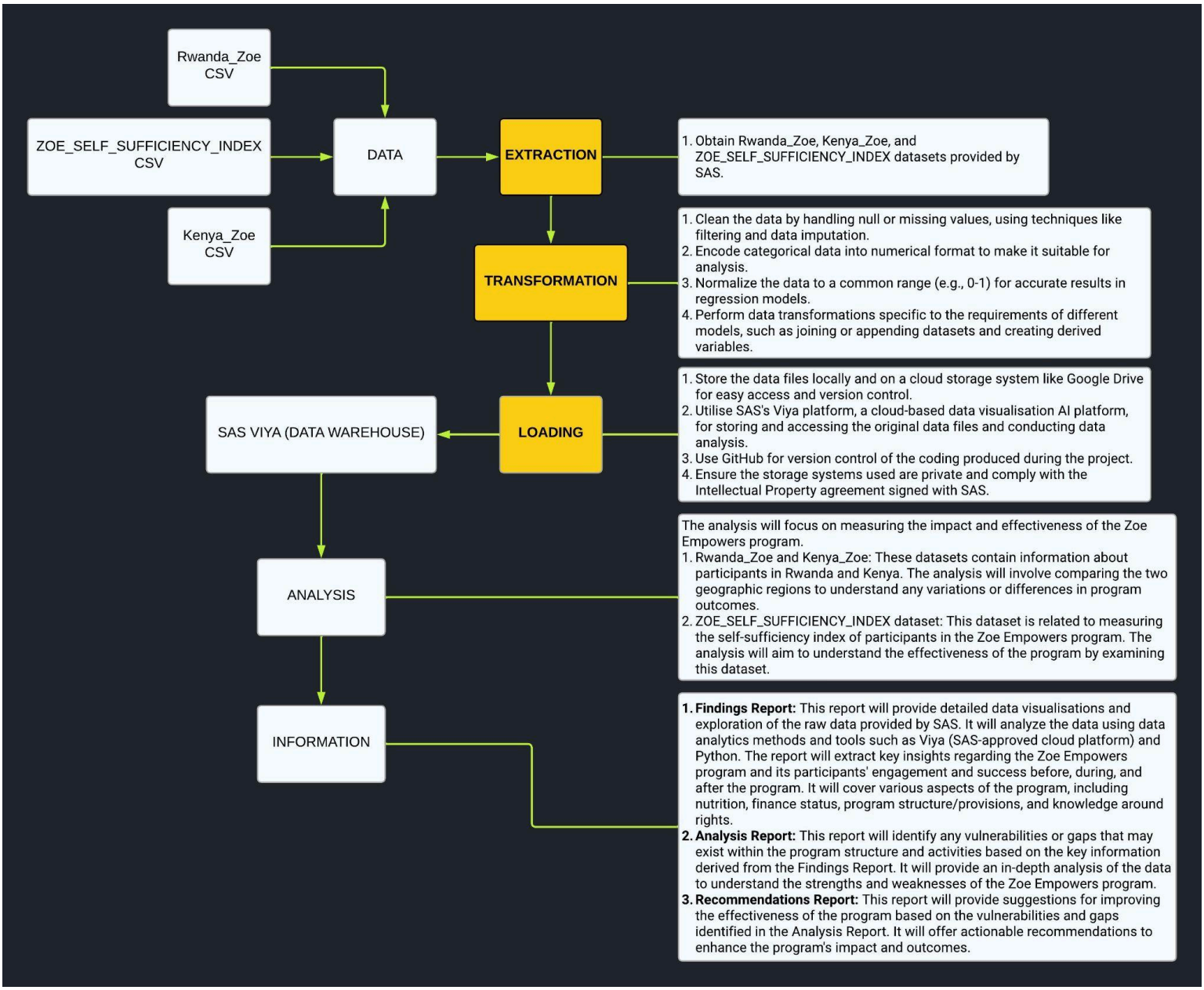
For the first section we are able to identify immediately two factors within the data that will demonstrate the relationship of SSI and participation in the program. The data has two very specific columns that will allow us to investigate this relationship, the overall SSI index itself, and the membership status of the participants. Once the datasets have been correctly appended and joined it would be simple to start analysing their relationship with each other. We are hoping to find that the SSI will be very related to the membership status showing that the amount of time spent in the program increases self-sufficiency and improves the lives of those in the program. The range of self sufficiency goes from 0 to 24. We're hoping that these will be able to be "bucketed" by the membership status and then show improvement in each stage of the membership. For example, someone who has just started the program will have an SSI around 2-8 and someone who has recently finished all three years to be around 20-24.

The breakdown of the SSI is less easy to identify. The first step however, is easy. The SSI is made from 8 categories (listed in the data description area). These categories are different indexes that summate to the total SSI. Breaking it into those 8 categories is extremely simple. However the analysis of those categories becomes the task of our project. The visualisation tools provided to use through SAS's Viya platform allow us to start learning about the different trends and relationships that occur within those categories (see the MVP for more details). The approach to identifying these trends and relationships will be an iterative process, there are tools that give us summarizations of indexes and how they contribute to different relationships. From there we will have to delve deeper into understanding those summaries and gaining insights into the way features interact with each other and these interactions can help identify if the holistic approach of Zoe Empowers is effective and/or even carried out properly. We are expecting areas such as access to food and financial security to be the most influential parts of the SSI. There won't be many findings that have statistical validations in this section.

1.2. SOLUTION ARCHITECTURE (REVISED)

1.2.1. ETL DIAGRAM (ADDED)

The following ETL (Extract, Transform, Load) diagram is to illustrate the high-level data collection and analysis pipeline of the project, showcasing the steps involved in gathering, preparing, and analysing the data for the Zoe Empowers program.



1.2.2. DATA INGESTION

Rwanda_Zoe, Kenya_Zoe and ZOE_SELF_SUFFICIENCY_INDEX are provided to us via SAS Viya to complete the entire project.

1.2.3. DATA PROCESSING

The SAS Analytics team has cleaned and prepared the data prior to the project. However, there were still some issues that needed immediate attention. For instance, the ZOE_SELF_SUFFICIENCY_INDEX dataset had a missing value in one index. Nonetheless, it was still possible to calculate the missing value, given that the Self-Sufficiency Index column was a summation of the others. On the other hand, the Rwanda dataset had more significant null

values that proved to be more challenging to deal with. With several features in the data, the Hot-Deck Imputation technique was used to extrapolate some columns. However, this technique only addressed some of the issues, and in some cases, deletion was the only option, especially for indexes with categorical data that provided all the necessary information. The pairwise deletion approach was used to remove null values only when necessary since listwise deletions would result in a more significant loss of information. To assist different models, the data had to be transformed into a better format using various techniques. Multiple regression models would benefit the most from transformations, such as encoding categorical data into a numerical format, converting Y and N values to 1 and 0s, and encoding multiple-choice questions into their respective values. For example, the index “Membership status” has four options, Not Yet a Member, Member less than 3 months, Member 1yr - 2yr and Recent Graduate (less than three months) are encoded to values 0 to 3. Normalising the data was also crucial, reducing the data to a 0 to 1 range for all values, making it easier to calculate the relevant significance scales between the predictor variables and their effect on the response variable. However, both encoding and normalisation applies a level of abstraction to the data, keeping a record of the original data to cross check assumptions and understanding would be key to progressing effectively.

1.2.4. DATA STORAGE

Our data is securely stored on SAS Viya, which is a cloud-based in-memory analytics engine. Furthermore, we keep all of our coding centralised on Github, providing a streamlined location for storing and organising code. This makes it simple to track changes and revert to previous versions if needed.

1.2.5. DATA ANALYSIS

Analysing the data provided is crucial to identify areas where participants are making progress and areas where they may need more support to achieve self-sufficiency. By breaking down the SSI into eight categories (listed in the data description), we can identify which areas are most important for participants to focus on to improve their overall SSI score. This information can then be used to tailor interventions and programs to better meet the needs of participants. For example, analysing the relationship between the SSI and membership status can help identify whether being a member of the program is associated with greater improvements in self-sufficiency compared to non-members.

1.2.6. DATA VISUALISATION

Across the project the team will interact with data that is collated to yes (Y,1) and no (N,0) values and encoded satisfaction/response ratings (0 least satisfied to 4 most satisfied) in regards to questions of answer and response for example ‘Level of Family Education’ as 0 = No education completed,1 = Primary School, 2 = Secondary and 3 = University. All this data is encoded into numerical values and are both analysed and visualised using the AI within the SAS Viya platform.

The visualisations are done through a series of ‘objects’ within the SAS Viya AI tool, including and not limited to:

- Decision Trees - answer specific survey questions along with other factors
- Bar Charts - presents numerical importance and responses
- Box Plots - demonstrates the minimum, median and maximum pertaining to data
- Scatter plots - demonstrates distribution of data
- Heat Maps - demonstrates distribution of data

- Pie Charts - shows number of responses/ input for a particular category
- Dual-axis line/chart - demonstrates distribution of data and linear attributes

All visualisations can be linked to each other using Viya's features and produce results that are tailored for specific groups, factors and categories providing deeper insight into the data.

1.2.7. MODEL DEPLOYMENT

Not Applicable: Our project aims to gain valuable insights into the program's impact, identify obstacles hindering its success, and provide actionable recommendations to improve Zoe Empowers' efficacy. SAS has asked DataSynergy to follow the analytical lifecycle and explore and visualise the Zoe Empowers data using their software, SAS Viya. To achieve this, we are using a range of visualisation methods to analyse and visualise data from Rwanda and Kenya datasets, using the ZOE_SELF_SUFFICIENCY_INDEX to understand the program's impact and effectiveness. Using advanced analytics, we will create 3 key documents: A Finding Report, an Analysis and Recommendations Report. These reports will provide us with insightful information about the program's various aspects and aid in pinpointing areas that require improvement. Therefore, it is important to note that this project **will not deliver any solutions** to SAS, regarding any aspects of the Zoe Empowers program. Additionally, **it will not project future data outcomes**, but instead, will focus on analysing the current efforts of the program and how they can be improved for future intakes.

1.3. ALGORITHM/MODEL METHODS

There are two parts to our solution that have different reasons for being resolved the way they are. We will be modelling the effect of participation in Zoe Empower's program to the participant's self sufficiency and we will be dissecting the SSI into different factors, gaining an understanding of what directly is improving.

Firstly, our section analysing the effect of membership on the SSI will be using linear regression. It is designed to show the relationship between two variables, the predictor variable of the membership and response variable of the SSI. Linear regression allows us to have metrics that describe the accuracy of this relationship, the P-Value of coefficients, the R-Squared metric and the MSE. The simplicity of the model makes it perfect for our simple goal. Other models, even just simply multiple linear regression, convolute the description of the relationship between membership and SSI.

Linear regression requires specifically numerical data. This means that the categorical variable of membership must be encoded first. This is how it is being encoded:

- Not yet a member: 0
- Member less than 3 months: 1
- Member 1 yr - 2 yr: 2
- Recent Graduate (less than 3 months): 3

Linear regression is also affected by outliers, there is only a single record with a status of 0 (Not yet a member). This would be affecting the model and is removed. Further outliers will be looked at. Additionally, linear regression does not have to be describing a perfectly linear relationship, it data transformations can be applied to change the behaviour of the relationship being described. For example applying a logarithmic transformation to the data. Different transformations need to be explored but the starting ideas are untransformed and logarithmically transformed.

Secondly, the deconstruction of the SSI into factors will not be directly modelled. Any modelling done is rudimentary and intended to be visualisations in a presentation, for example a decision tree can be used as a classification model, but in our case it will be used to understand the causes and effects of variables. Visualisations will be used to describe trends and relationships within the data in a less statistically rigorous way.

1.4. DETAILED DATA DESCRIPTION

Our project uses SAS Viya, a cloud-based analytics engine known for its powerful capabilities in supporting the entire analytics lifecycle. SAS Viya is fully equipped to handle everything quickly, from processing and discovering data to deploying it. With SAS Viya, we can access Rwanda_Zoe, Kenya_Zoe, and ZOE_SELF_SUFFICIENCY_INDEX from the SAS library and integrate, transform, and modify them all within one software environment. Additionally, the interactive self-service visualisation tools, powered by AI, allow us to search for relationships, trends, and patterns, which allows us to showcase how Zoe Empowers is helping young participants to move from hopelessness and crushing poverty to meeting their needs across eight major life areas.

The Rwanda_Zoe and Kenya_Zoe dataset have been curated by the Data for Good program at SAS Analytics, who have meticulously prepared and cleaned the data collected from the 2015-2019 Impact Surveys conducted by Zoe Empowers. These surveys were conducted at three points in the three-year program: at intake before the youth receives any benefits, at the midpoint, and at graduation. Zoe Empowers chooses the empowerment groups to be surveyed from the total list of empowerment groups using an Excel randomiser with the following subgroups: countries, program year, and group start date. Zoe Empowers survey's 20-25% of the groups (a lower percentage for larger programs) with a minimum of three groups chosen at each level in each country. One survey is given to the head of the household (the youth acting as "parent" in a youth headed household; usually the oldest or most capable sibling in the family unit) in the groups randomly selected. The surveys gather information on various aspects related to the program, including food security and nutrition, income generation, safe housing, health and hygiene, child rights, community connections, education, and spiritual strength. In addition, the surveys include background data on each participant and their empowerment group to accurately measure the program's impact at the family/household level, individual level, and community level.

Both Rwanda_Zoe and Kenya_Zoe datasets are in the exact same format and each dataset identifies participants per row with unique identifiers. The datasets include a mix of categorical and continuous data items, which are quite extensive and provide interesting insight to compare the Zoe Empower's program in Rwanda and Kenya. The data can identify any differences in programs effectiveness and determine what factors may be contributing to these differences. For example, we can compare the percentage of participants in each country who have received vocational or skills training, and the percentage who have started their own business as a result. We can also compare the number of meals participants are able to afford per day, the level of household income, and the proportion of participants who report feeling safer and more secure as a result of the program. Additionally we can compare the percentage of participants in each country who have access to medical care and clean water, and the percentage who report attending community events or feeling empowered to seek help when needed. By analysing these and other data points, we can identify potential differences in the programs effectiveness between the two countries, and work to address any barriers to success to improve the overall impact of the Zoe Empowers program.

However, the ZOE_SELF_SUFFICIENCY_INDEX is the most important dataset out of the three because it contains the Self-Sufficiency Index (SSI) for all the participants, which was calculated by SAS Analytics and Zoe Empowers to quantify how self-sufficient a participant is at the point of doing the questionnaire. The Self-Sufficiency Index is measured based on the households head's answers to four questions: number of meals eaten per day, eat enough to be satisfied, adequacy of housing, and sufficient income for household necessities. The dataset contains 9 columns, the index itself which is a summation of the other eight, which range from 0 to 24. The other eight categories are Food Security and Nutrition, Housing, Community Connections, Health and Hygiene, Child Rights, Education, Economy / IGA and Spiritual Strength. These eight categories are all 4qSSI (from 0-3) best characterised as near 0 denotes EXTREMELY VULNERABLE, near 1 means VULNERABLE, near 2 means SELF-SUSTAINING, and near 3 denotes FLOURISHING.

The Self-Sufficiency Index provided by the ZOE_SELF_SUFFICIENY_INDEX dataset allows us to quantitatively measure the success of the Zoe Empowers program in improving self-sufficiency of participants. By joining the three dataset together to compare SSI scores of participants in Rwanda and Kenya, we can determine which areas of the program are working well and which areas may need improvement. For example, if participants in Rwanda have lower SSI scores in the Education category compared to participants in Kenya, this could indicate a need for targeted intervention to improve access to education in Rwanda. Similarly, if participants in Kenya have lower SSI scores in the Health and Hygiene category, this could indicate a need for increased access to medical care and health education in Kenya. By analysing and interpreting the data in these three datasets, we can provide meaningful recommendations to Zoe Empowers to help them achieve their goal of empowering vulnerable youth to become self-sufficient and thrive in their communities.

2. MVP/PROTOTYPE DOCUMENT (REVISED)

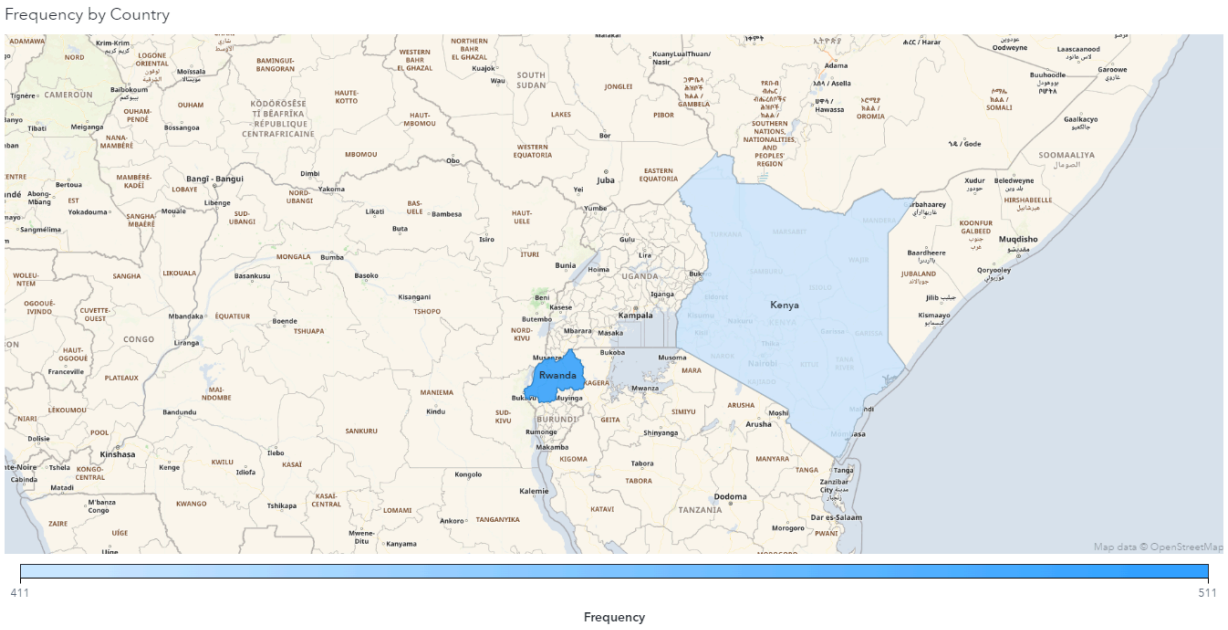
2.1. BACKGROUND

The effectiveness of the Zoe Empowers program is shown through the summation and collation of the Self Sufficiency Index (SSI). The SSI of an individual is assigned by a calculation of different attributes that have predetermined values and whose sum total is the index. These attributes are collected through their response to a survey that is sent to all participating members of the Zoe Empowers Program.

2.2. PURPOSE

The purpose of this document is to present a minimum standard that meets the baseline requirements of the Sponsor and to demonstrate early visualisations regarding what definitive factors or influences have the largest effect on the Self Sufficiency Index. By identifying these factors and impacts it enables the team to make more effective and impactful recommendations to SAS about the Zoe Empowers program and what may need to be changed or improved to ensure the longevity of an individual's ability to sustain their SSI. The dataset provided by the Sponsor gives a sample of 922 participants split across Rwanda and Kenya and is dated from 2015-2018. The initial findings were then presented to SAS on the 18th of March 2023 in order to engage feedback including some guidance on where to take the project as the team nears the final deliverables.

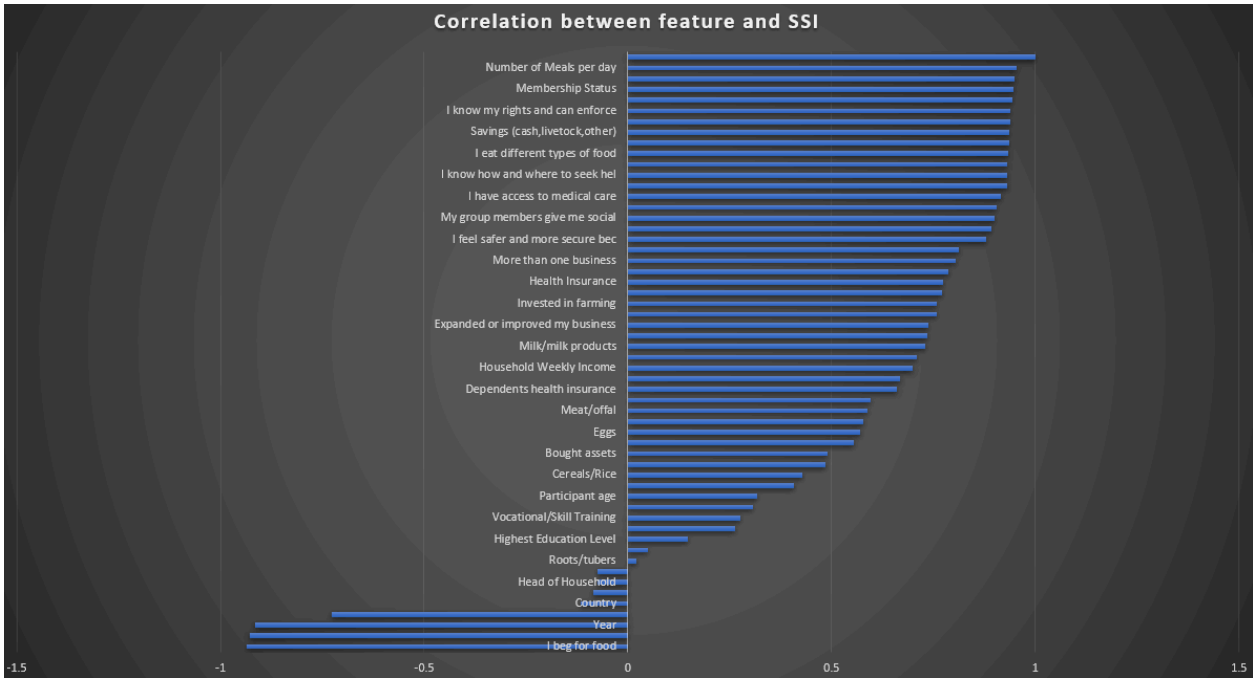
2.3. INITIAL VISUALISATION



(Frequency by country)

Insights from 'Frequency by Country

The following visualisation demonstrates a higher concentration of participants within the Rwandan region compared to Kenya. Using the frequency legend we are able to see that the participants number for Rwanda is in the high 500s and for Kenya it is in the lower 400s, this concentration corresponds to the gradient of blue used with darker being a higher number of participants and vice versa.



(Figure 1 Correlation between features and SSI)

Preliminary data extraction makes certain correlations clear and allows for more focused research into which factors have the heaviest influencers on the index. It was clear that access to food and how much the individual is eating seems to have the strongest correlation with the SSI. This represents a good starting point for further models using the SAS Viya platform.

Insights from 'Correlation between features and SSI' (ADDED)

From the following bidirectional bar chart we are able to see the factors that strongly influence the SSI within the program and those that do not. Out of the twenty eight listed factors four of those factors have a negative correlation to the SSI these include the 'Head of Household', 'Country', 'Year' and 'I beg for food. This may indicate that holding a status position within a participant's home has no impact on their ability to self suffice. The country and year negatively pertaining to the SSI can indicate the program operates consistently within the two countries and does not allocate great energy or resources over the other giving participants from both countries and all years equal opportunity to attain greater self sufficiency. Because the program has a strong emphasis on orphans being able to attain basic nutrition and meals the factor 'I beg for food' has the strongest negative correlation indicating methods surrounding food and nutrition provisions within the overall program, are effective and efficient in raising an individual's SSI.

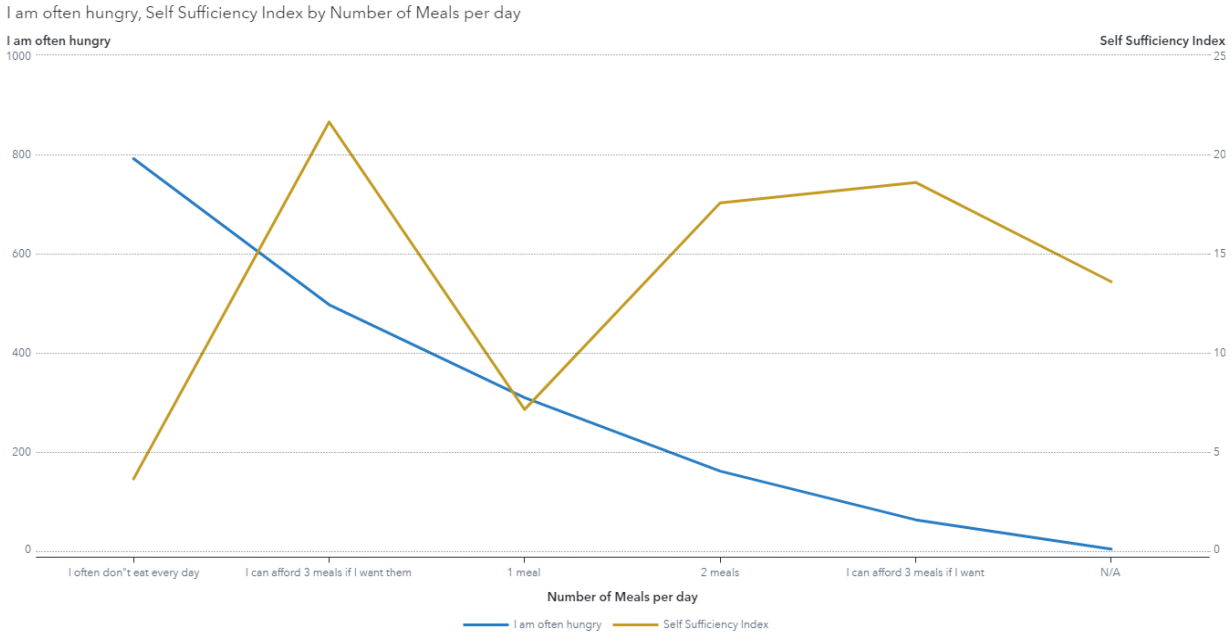


Figure 2 (SSI and meal frequency)

The dual axis bar chart (Fig2) presented provides further evidence to this by showing the correlation between the number of meals an individual is consuming and their overall SSI as a significant.

Figure 2 Insights (ADDED)

The figure demonstrates that where individuals are able to consume an average and constant amount of three meals a day they are more likely to sustain a higher self sufficiency above twenty. This information may be obvious to some however in the cases of participants of Rwanda and Kenya, poverty stricken individuals find it harder to obtain nutritious sources of meals and sustain a balanced diet from the availability of food options. This increase of SSI and meals consumed can be used to aid corresponding findings of resources and base income as it is likely that an individual who obtains a strong and steady income is more likely to self suffice and consume the three meals a day. The yellow line indicating self sufficiency is still strong at the two meal mark which also demonstrates an increase of food intake by participants compared to the initial self sufficiency being below a five rating.

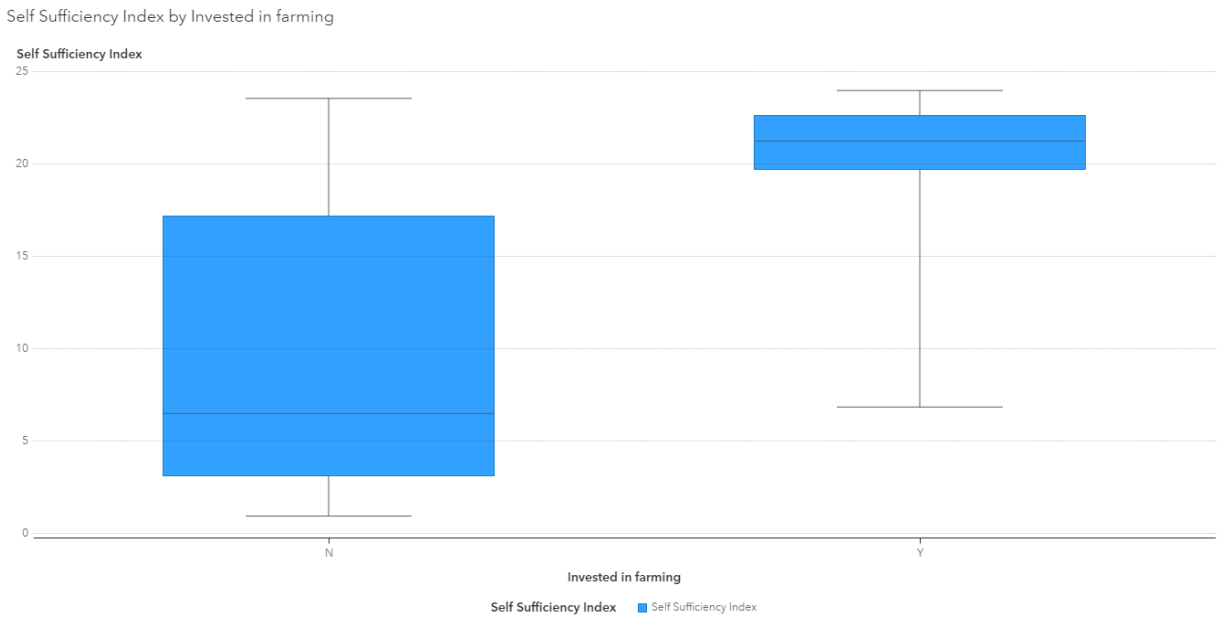


Figure 3 (SSI and participants choice to farm)

The education provided within the program encouraged the purchase of livestock and also provided training into leveraging their animals to help create business. This aspect of the charity is also heavily reflected with the increase of SSI results (fig 3). These herds of livestock can also be used for the families own food source, thus prompting them to vote higher in the food availability statistics in turn boosting their SSI.

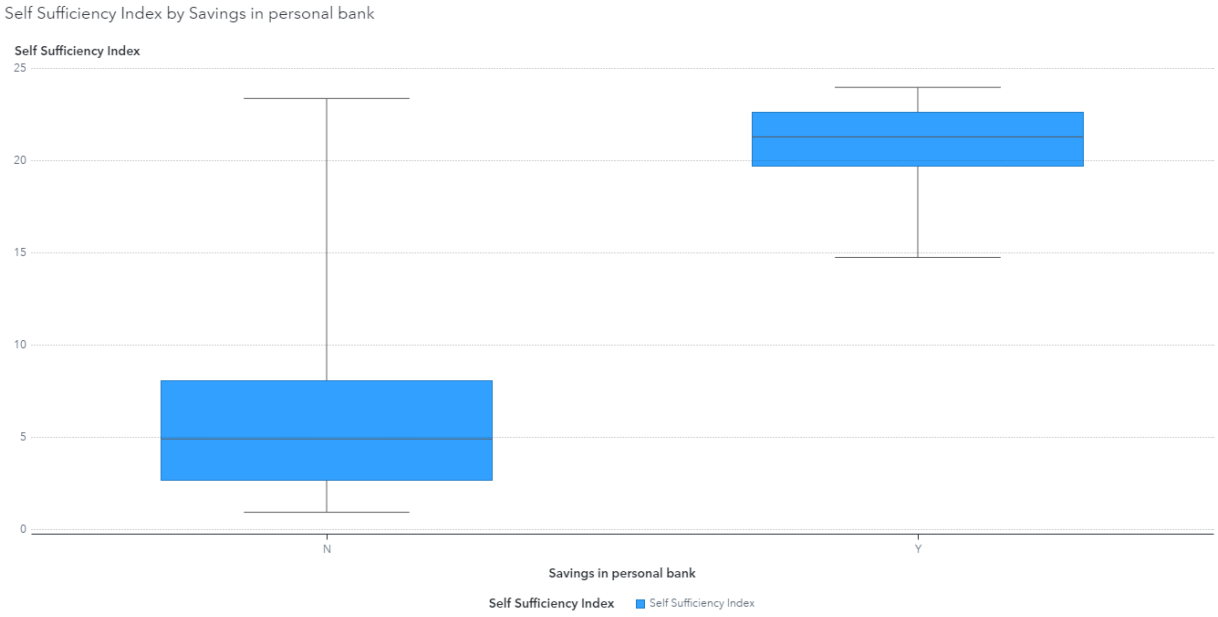


Figure 4 (SSI and Savings in personal bank)

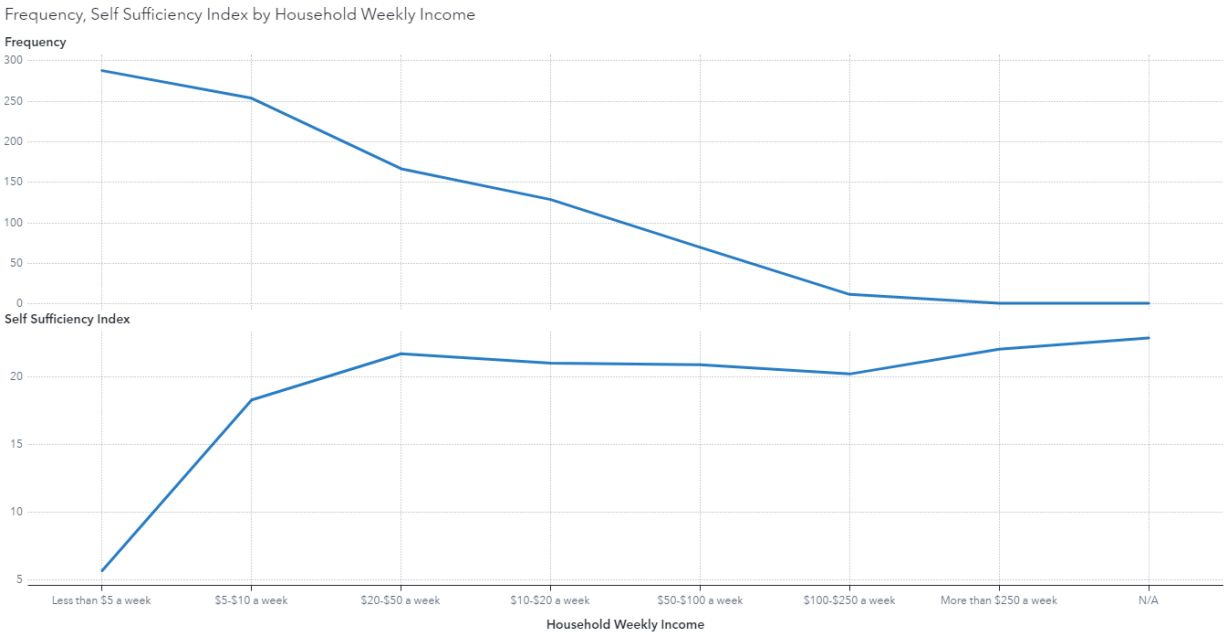


Figure 5 (SSI and frequency of different incomes)

Cash savings and more importantly earning potential was also noted as having a strong correlation to the SSI (fig 4); it was important to visualise its specific effects as well as taking into account the specific advantages that possessing multiple businesses can give the participants. The Bar chart above (fig 5) shows the SSI correlation, the decision tree below (fig 6) proves that participants with multiple businesses are making up those higher earner submissions and therefore has a direct correlation with one's SSI.

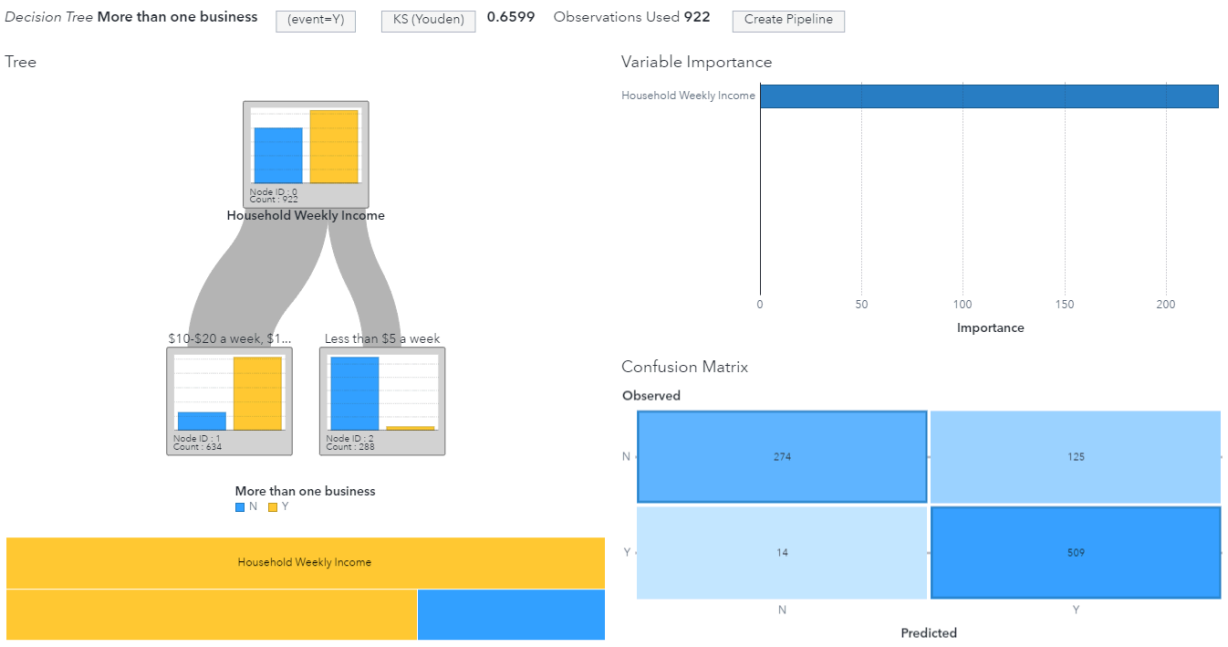
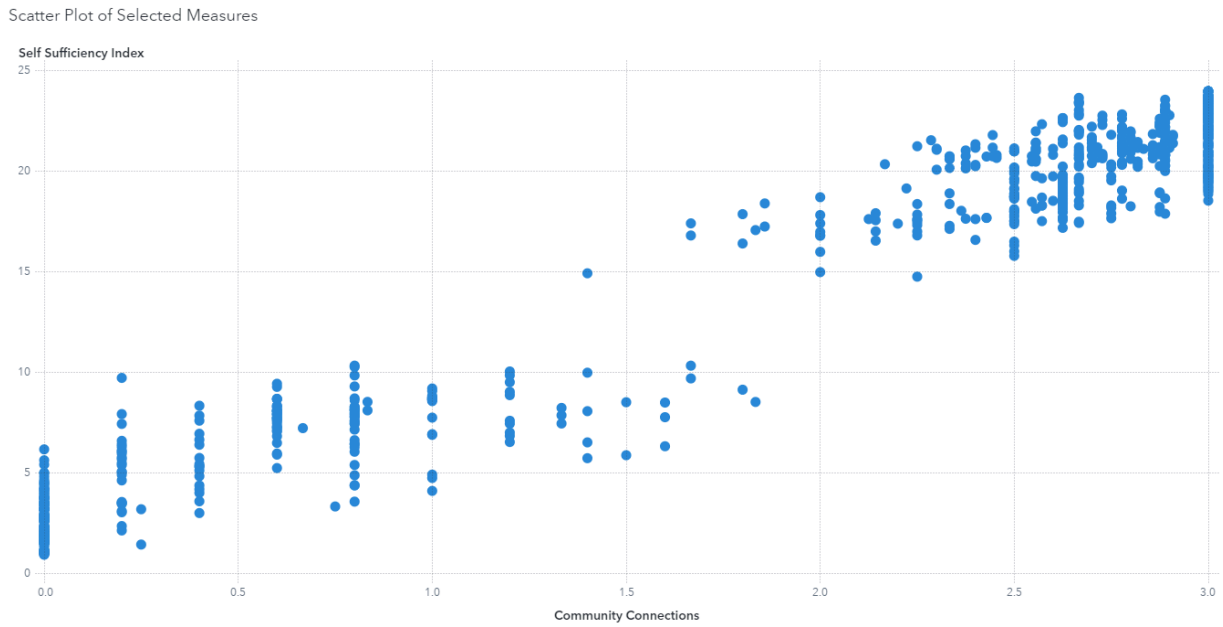


Figure 6 (Decision tree comparing numbers of business to income)

The program also places heavy emphasis on the establishment of thriving community interactions. Through the establishment of these businesses communities themselves become more supportive of their members. This is evident through the data (fig 7), with individuals also expressing that their sense of community has a direct impact on their lives as is shown in a community score of over half drastically increasing the average SSI of respondents.

Figures 3-6 Insights (ADDED)

Using the collective figures we are able to paint a picture of strong dependency among factors of personal income (income and savings), overall household income and investments into farming which contribute to the SSI and having positive impacts to an individual. These findings also demonstrate the effectiveness of Zoe Empowers as a program teaching the orphans to manage their finances and obtain farming investments that can be further used for business (selling crops) or their own consumption. Furthermore the collective findings can be used to aid the case that Zoe Empowers must continue positive enforcement around financial education for its participants as it has a proven record across a number of categories to improve self sufficiency.



(Figure 7 Community Connections and SSI)

Figure 7 Insights (ADDED)

The following scatter plot expresses the correlation between participants having community connections and their SSI index. Where individuals engage in their respective community through events, group activities and business they were able to benefit from support attained, safety within groups and work support.

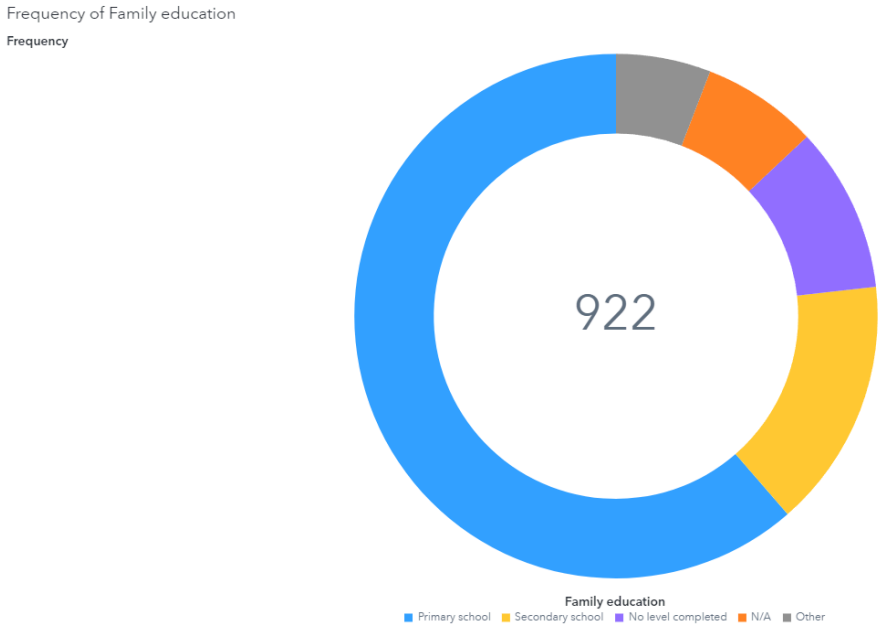


Figure 8 (Levels of education completed)

The Zoe Empowers program additionally stresses the importance of proper education for young children, as this is something that is sorely lacking within the remote communities they encountered.While over half of the participants have completed Primary school (Fig 8), any continuation into secondary or tertiary studies is severely lacking.

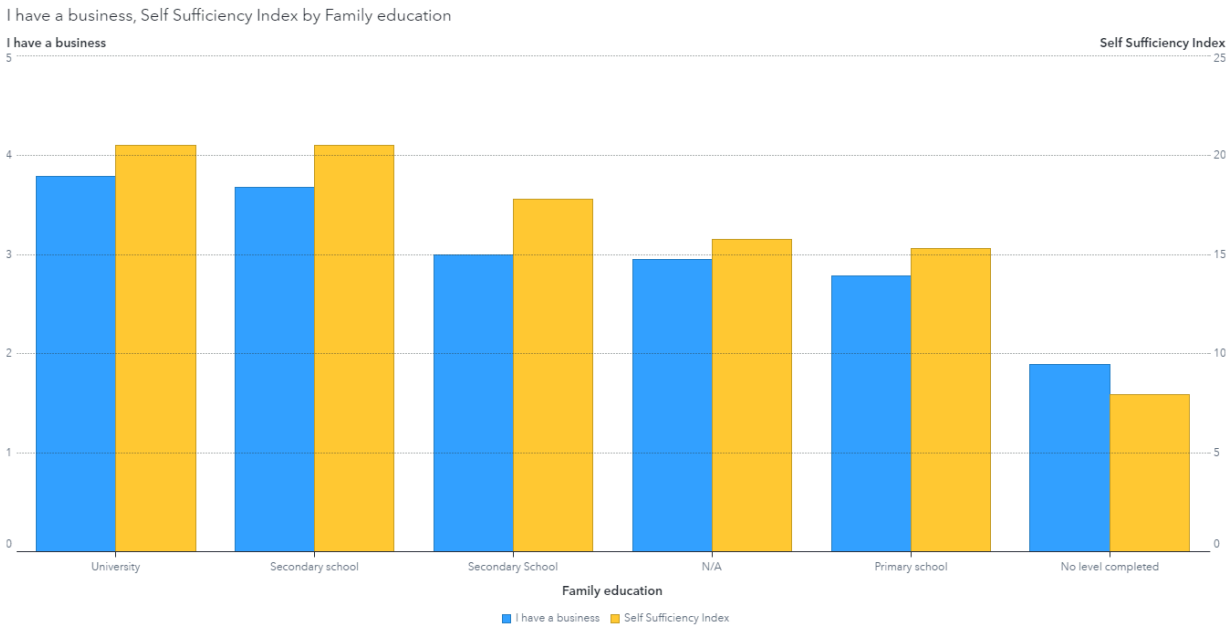


Figure 9 (Business score and SSI by level of education completed)

This is a part of the program that should definitely be encouraged more, seeing the higher overall business and SSI scores from the small percentage of people who have chosen to continue their studies (fig 9). Using the Viya platform also an interactive report was created to help better show which factors influence the SSI.

Figure 8 and 9 Insights (ADDED)

Observing the collective insights we are able to see the impacts of family education to an individual's SSI. Family education is an area that dissects each member of the family's history with education which in turn influences their perceptions of their other members' education. Where the members are from non or lower levels of education other members may be affected negatively due to the support to pursue their own higher education. Members who have completed tertiary and secondary education are seen to have an SSI above twenty however, that being states of the 922 participants involved in the family education survey over half of the respondents sit in the primary school level of family education. This in turn produces the SSI below 15. It can be noted that Zoe Empowers must place greater emphasis on the education components of the program not only for the orphans but for their extended family as they bring awareness

to the benefits of further education and opening opportunities for further education benefits the great Zoe Empowers community.

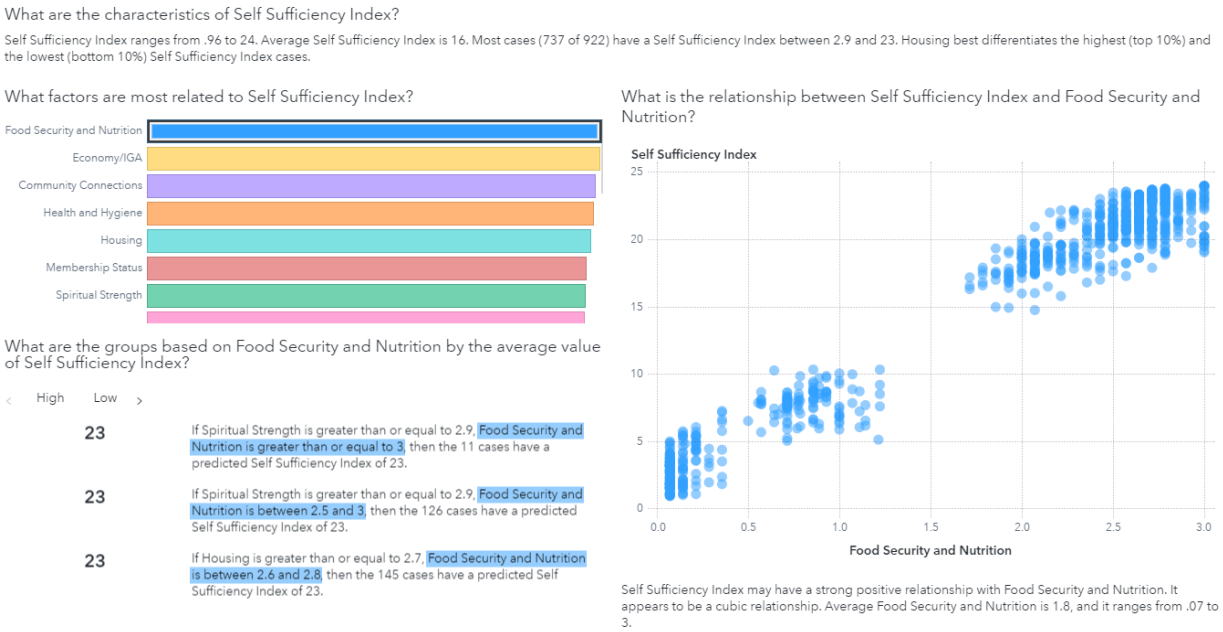


Figure 10 (SSI relationship with food and nutrition)

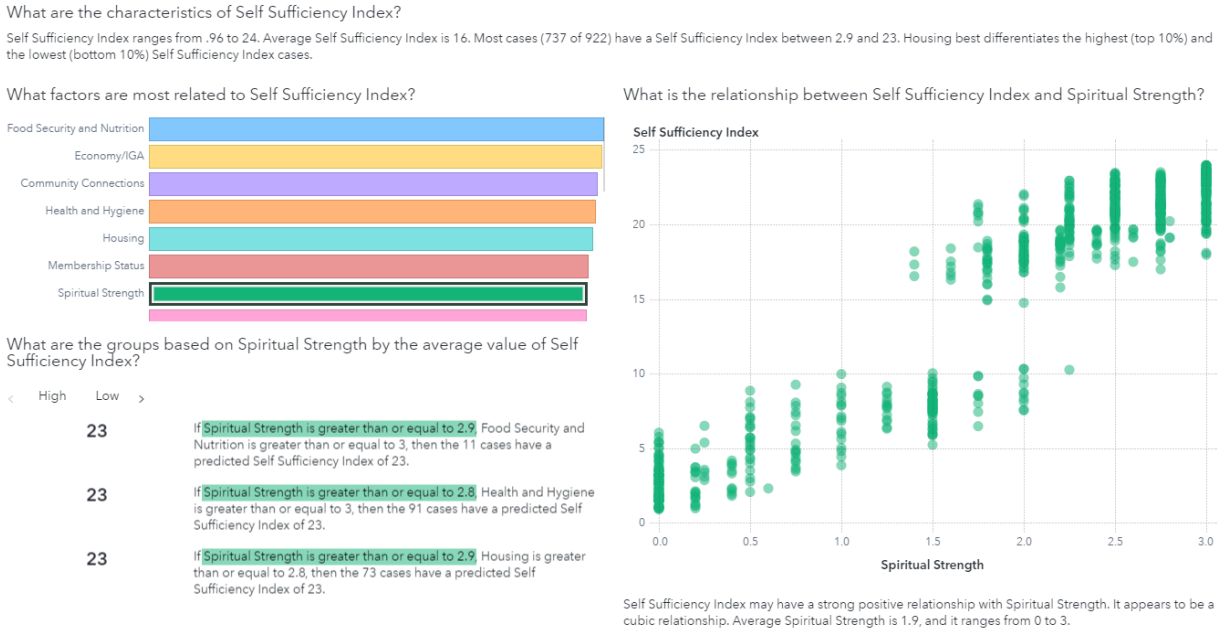


Figure 11 (SSI relationship with Spiritual strength)

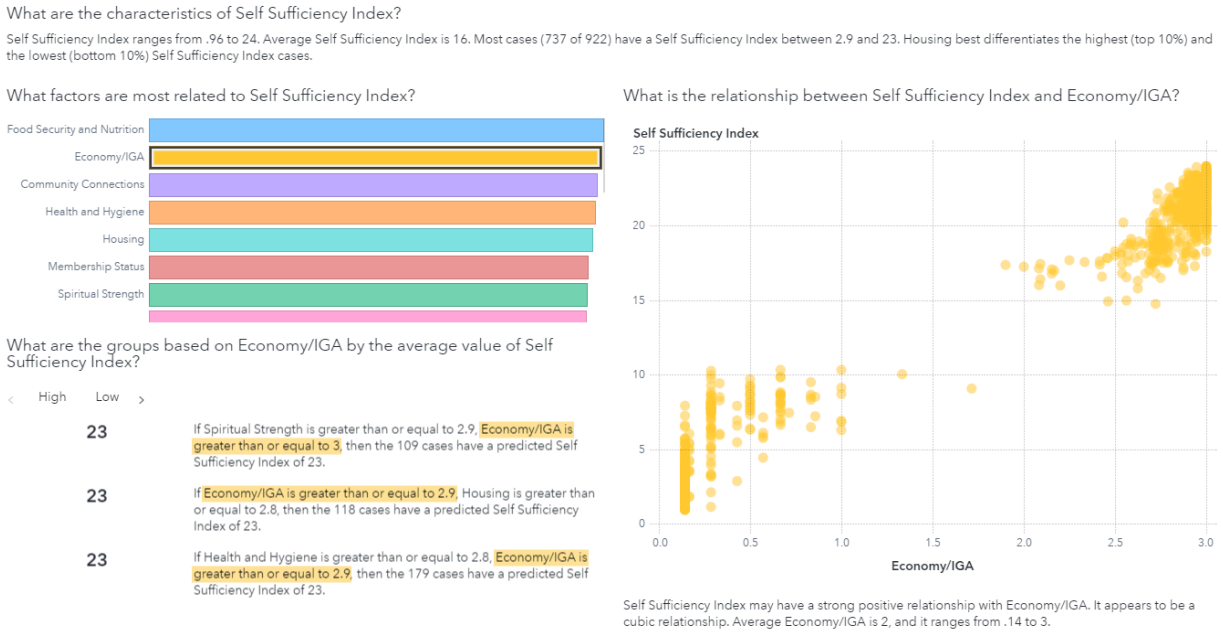


Figure 12 (SSI relationship with economy)

Figures 10- 12 Insights (ADDED)

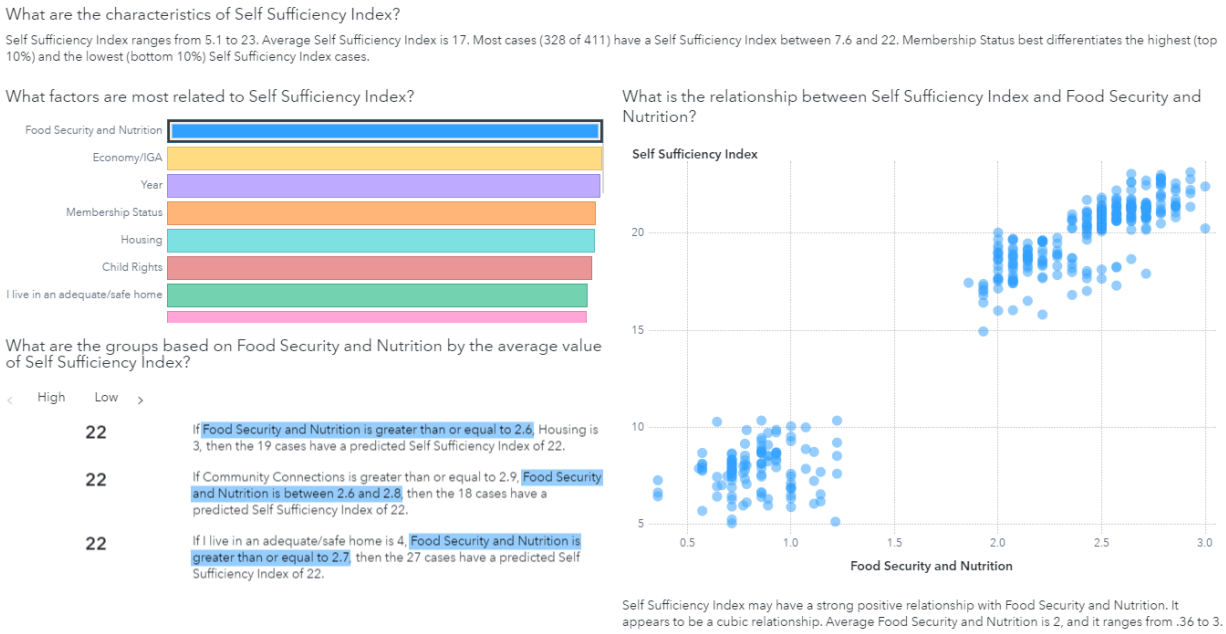
The following snapshots of collated findings use SSI and visualize its relationship with economic, food and nutrition and spiritual strength factors. We are able to see the trending patterns and impacts each measure has on the SSI. The Viya analytics tool also generates an average SSI across participants being sixteen but when observed with middle to higher ranged answer responses with each category the average SSI increases by nine points to twenty three. These findings indicate that the Zoe Empowers program has great benefit to its participants increasing their ability to attain food, economic value and spiritual strength.

This report condenses our findings into a format which can be viewed by anyone with a SAS account, hence making it a great small deliverable to return to the sponsor as a summary. This report will automatically rearrange and change supporting graphs depending on which factor the user has selected.

2.4. Sponsor Meeting, Feedback and Response to Feedback

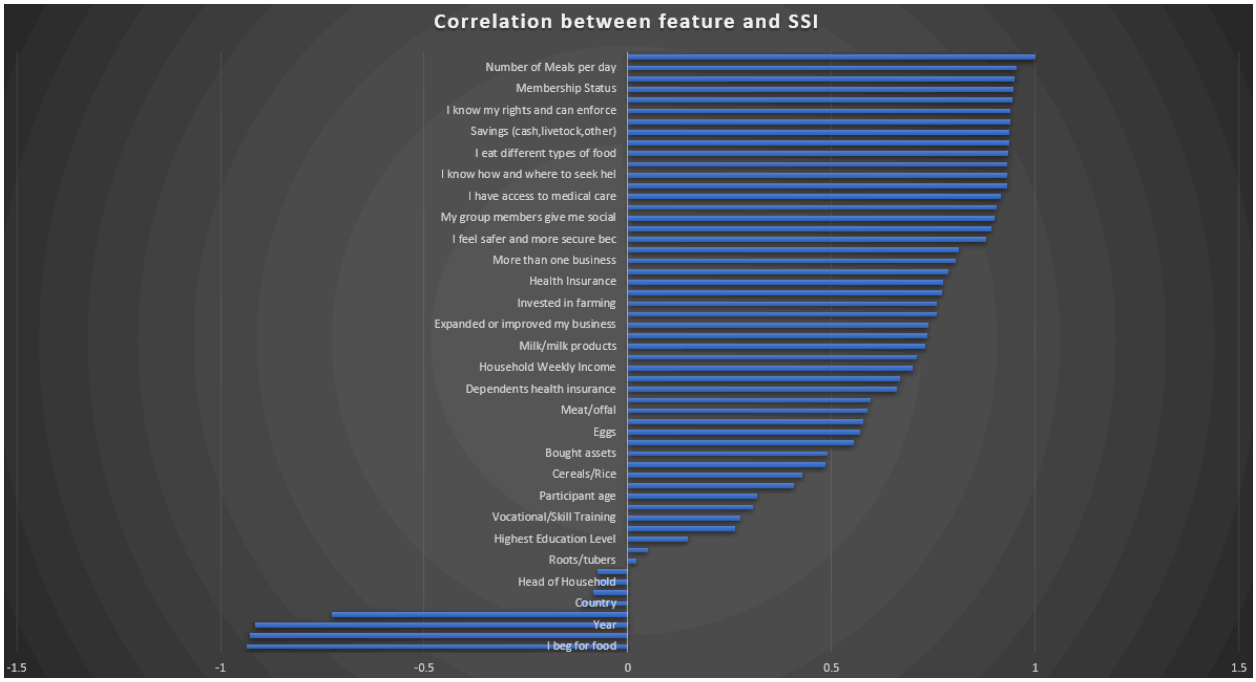
Feedback from our sponsor (SAS Institute) was given during a thirty-minute meeting on April 18, 2023 from 9:30 am to 10:00 am. Team members met in-person at Macquarie University library, making contact with SAS via a scheduled fortnightly meeting through Zoom. The course of the meeting saw all team members in attendance and actively demonstrating our findings to SAS for feedback.

The presentation firstly saw members showcase screenshots to SAS, through share screen on Zoom, of graphs and analysis conducted on the Viya for learners platform. The screenshots were visual findings from our early use of the Viya platform, used to broaden our understanding of the impacts of Zoe Empowers program on Kenya and Rwanda. Our skill level of Viya was at beginners level and our feedback from SAS on what we had found so far was important. The feedback received from Jordan (Consultant at SAS), was really positive. Jordan made note that the graphs looked visually appealing and were easy to read, further backed up by Chris (Intern at SAS). Jordan made further comments that the findings were relevant to what SAS were looking for, and to keep persisting with creating more graphs on the platform as the more graphs and analysis we create, the better our reports and findings will be.



(Graphs presented by Rory to SAS)

The second part of the meeting saw Cassandra (data specialist) demonstrate her findings from R Studio and Python. Each has been utilized as a second means to congregating and analyzing data for the project and to provide another angle to our findings. SAS were really happy that Cassandra was still utilizing R studio and python as a second resource to Viya. This allowed us to provide SAS with insights, analysis and learnings on resources outside of Viya that some sponsors were unfamiliar with.



(Correlation presented by Cassandra to SAS)

The end of the meeting was used to receive feedback from our sponsor on our presented findings, and also directly ask sponsors questions the team had prepared before the commencement of the meeting. Our team developed succinct questions we believed would provide the most effective responses, allowing us to move forward in the project and correct mistakes and misconceptions we were currently facing. The responses provided from SAS we found to be reassuring, helpful, and motivating.

Questions presented to SAS are as follows:

1. How similar to the 'Zoe Empowers Report Summary Example' would you like our reports to be? For example, are you expecting us to develop and produce machine learning/predictive models to understand which metrics most impact self sufficiency, or would you prefer the 'easier to read' graphs and charts, or a blend of both? Are we aiming to build predictive models? Are we aiming to make statistical inferences about the data? Are we aiming to make a visual presentation to explore data?
   Sponsor feedback: Jordan "So at the end of the day we want from you guys a findings report, analysis report and a recommendations report. Because the question gives you guys more freedom and is more open, feel free to use whatever means of analysis but try to stick to the Viya for learners for more visualised based data analysis and building. We don't want you to necessarily build predictive models. If you have the skills to do that, find that helps with creating your report, go for it. But focus more on Viya because we understand that not all you guys have a data science background and Viya will help you with that. VIya will give you guys the ability to create visual findings for your presentation and reports too."

2. Rory's graphs focus on Food Security and Nutrition. Are there any particular areas you would like us to include or focus on in particular or are you happy with a more diverse range of focal areas from the SSI?
   Sponsor feedback: Jordan "Keep exploring the data sets in Viya and collating a broad range of datasets. Because the question gives you guys a bit more freedom, it's up to you guys what you feel is important to focus on for your analysis and findings reports, but what you've collected so far is good, keep it up."
3. A student team of 6 members from UNSW was assembled and assigned to identify trends and patterns in data collected about orphans in Kenya and Rwanda using data analysis, visualization skills, and tools. They were asked to come up with solutions based on the identified patterns. If you know the project, are you able to inform us on their performance during the project, how they were marked, what they did well and what they could've done better?
   Sponsor feedback: Lucy "Jordan wasn't a part of the SAS team yet but I recall they did quite well. Definitely feel free to use their works for inspiration but of course don't blatantly copy it."
   Jordan "If you do use their work make sure to reference it. But regardless of what inspiration you guys draw from other works, what you've produced so far is really solid and on the right track."
4. Based on what we have provided you with today, are you happy with our progress and do you believe we are heading in the right direction as of now? If not, can you give us some pointers on where we can improve or further look into/research?
   Sponsor feedback: Jordan "Yes you guys are doing a great job so far. I can see you guys are beginning to really utilise Viya for learners and are starting to develop a wide range of analysis to help further your understanding of the effectiveness of the Zoe Empowers program."
5. Is it true that Viya is shut down on the third weekend of every month?
   Sponsor feedback: Lucy and Jordan "That's a great question! We believe it is taken down and operated on from the United States so it should only be down in the early hours of Saturday morning, hopefully having very little impact on your ability to access Viya and complete work."

The feedback and response from SAS was found to be really insightful and reassuring. We were able to visually communicate and present early prototype findings to the sponsor, to which feedback on what we are doing well, and where we can further improve was received. Furthermore, asking direct questions also ensured that the team was moving in the right direction. Overall, the team concluded that SAS was very happy with our minimum standard prototype presentation and questions, and ensured to us that we are meeting the baseline requirements of what SAS expects of us.

3. TESTING DOCUMENT

3.1. MODEL EVALUATION (REVISED)

There will only be one section of our presentation that will involve evaluation: The effect of participating in the program has on the participants' SSI. Is there a significant effect on the SSI that is produced from the different time spent in the program? There is only one real model that can show this trend: Linear regression. The participants' time in the program either does significantly improve over the duration of their time in the program or it doesn't. Simple linear regression will allow us to make effective and statistically sound conclusions about the relationship from the data provided to us. The three metrics used to assess this relationship:

1. P values of the coefficient and slope
2. R-squared metric

3. Mean Squared Error

Positive changes to the model will be reflected by the improvements of these metrics. For example, applying a logarithmic transformation my improve the model by:

1. Reducing the size of the P values
2. Increasing the R-Square metric
3. Reducing the mean squared error

There is not much to be gained from trying different models; there is only a strong interest in if participation increases SSI or not.

The rest of the presentation is about the exploration of the data and will likely not include different models. Therefore, there will not be any other models to evaluate.

In terms of quantitative and qualitative measures around the models delivered to SAS, the quantitative aspect will be handled by producing three to five visualisations per factor area of the SSI (this includes Health and Hygiene, Education, Housing, Community, Spiritual Strength, Food Security and Nutrition, Child Rights and Economy). This number of visualisations will produce a total output of twenty four to forty visuals. This range of visuals will also allow for greater emphasis on research focusing the analysis on specific findings, uncovering keysights and strengthening the final recommendations furthermore increasing the quality of output. Qualitative measures have been discussed with the sponsor on two notable occasions

- Wednesday 3rd of May 2023
- Tuesday 16th of May 2023

On both occasions SAS was very receptive to the team's initial data visualisations including quality and number of discoveries produced. The first meeting involved SAS's Senior Systems Engineer Jonothan Barlow who commented on producing findings beyond the basic two dimensional bar graphs and replicating insights using a number of data objects. The team has since reviewed this and produced a demonstration findings to SAS again at the meeting after (16th), the quality of visualisation including correlation matrices, bar charts, linear regression graphs, dot plots and parallel bar charts received positive feedback and was appreciated by the sponsor. Comments on ordering and expanding data legends were the only comments made and were noted for future improvement.

3.2. PERFORMANCE EVALUATION RESULTS

So far the tests for the simple linear regression has been of two types:

1. Encoded but not transformed data .
2. Encoded and logarithmically transformed membership status.

The results are shown below with statistical and error metrics, as well as visually demonstrated:

3.2.1. PERFORMANCE VALUES

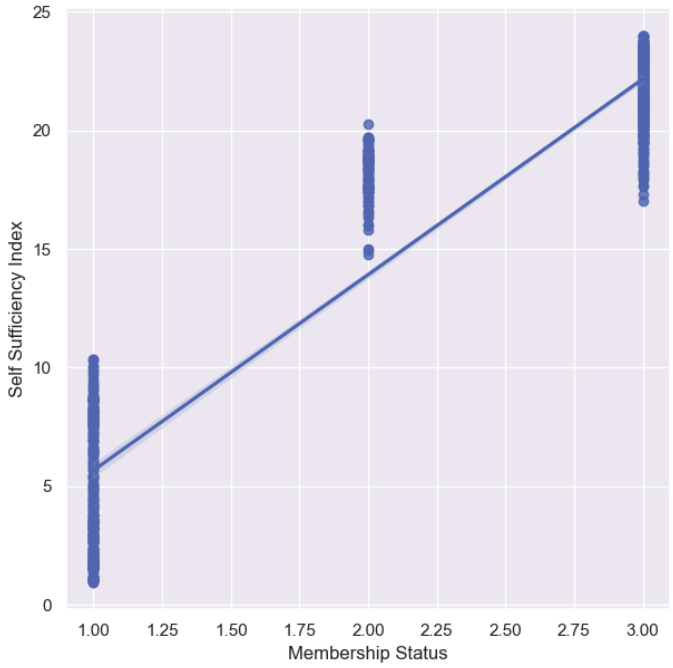
	P - Value	**R-Squared**	**MSE**

Simple Linear	< 0.00…	0.89	6.04
Log Transformed	< 0.00…	0.93	4.49

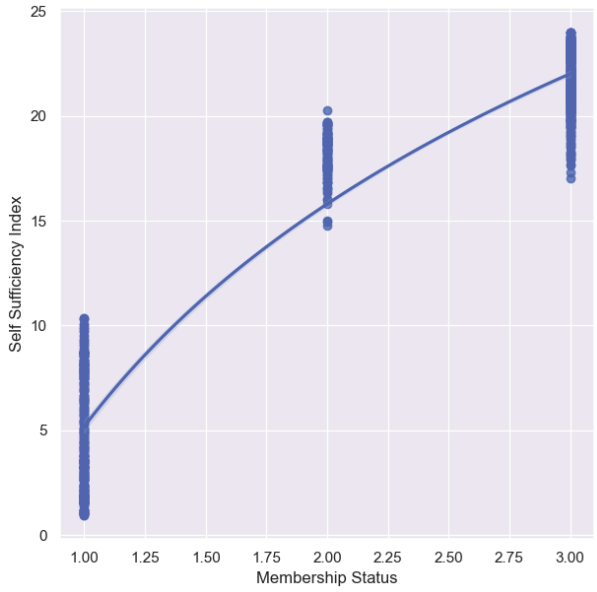
Note: *The range of Y values are between 0 - 24 for the MSE*

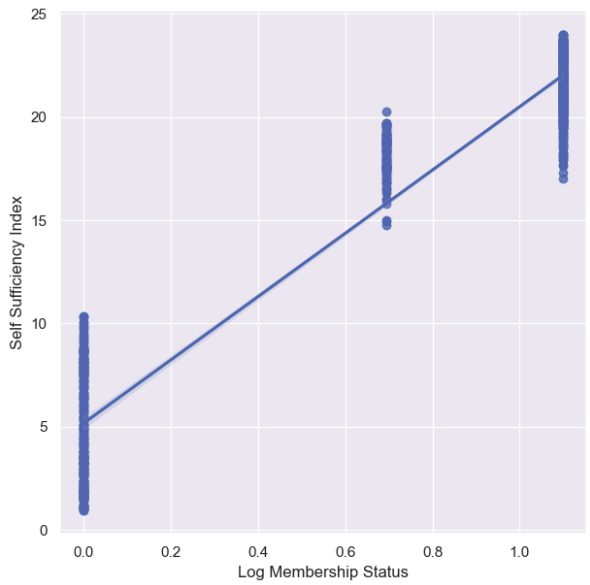
3.2.2. PERFORMANCE GRAPHS

Simple Linear Graph



Log Transformed Graphs (Equivalent)





Note: *The confidence intervals are shown, but they are small enough that it is difficult to see.*

The results show that the membership status shows a very significant factor to the improvement of SSI. The log transformed relationship was slightly better performing than the untransformed inputs with a higher R Squared value and a lower MSE.

We plan on trying different transformations on both the Membership Status and the SSI, but we suspect that there will not be a result that will perform better than the log transformed linear regression.

For deliverable four we would have experimented with more transformations and improved the visualisation to be more appealing and follow the style of the final presentation. Although the style improvements may only happen after deliverable four and before handover as we refine the presentation. Additionally we would have produced more analytical visualisations but these are unlikely to be evaluated in the same way the linear regression model has been.

INCREMENT TWO

1. SCRIPTS/MODEL EXECUTION

1.1. SAS VIYA FOR LEARNERS

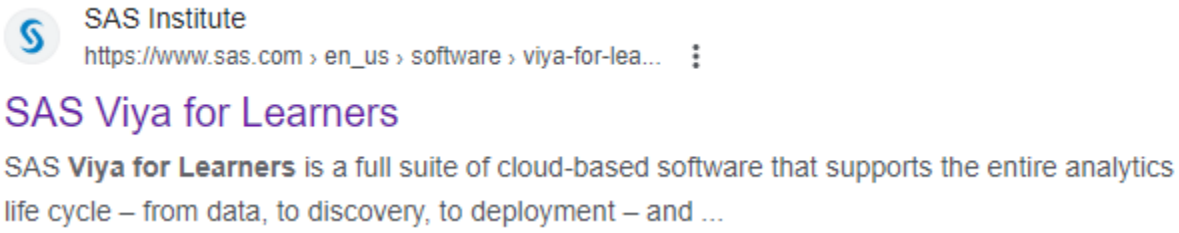
SAS Viya works as a cloud-based software where users of the platform are fully supported across the data analytics life cycle. Targeted users are educators and students, Viya providing the ability to manage data, develop models, and deploy insights. Viya equips audiences with analytics software that opens the door to teaching and learning analytics, however, being able to deploy this analysis requires in-depth knowledge of the platform.

The following scripts and model execution documentation will breakdown the process of SAS Viya. Following installation of the platform, a careful breakdown on how to prepare chosen data for particular analysis, how to visually explore and manipulate chosen data sets, and how to create, manage, and deploy models.

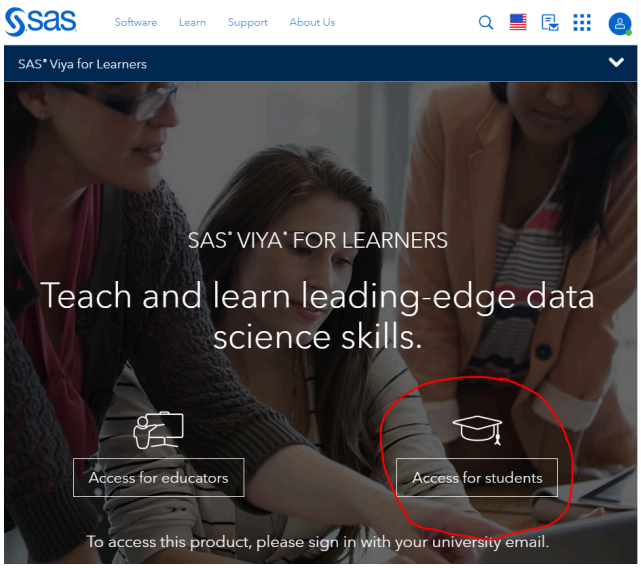
1.1.1. INSTALLATION GUIDE

Free access to Viya is available for educators and students. The platform can also be bought on Azure Marketplace, but to ensure appropriate explanation of scripts and model execution documentation on Viya for learners, installation of the free version will be explained.

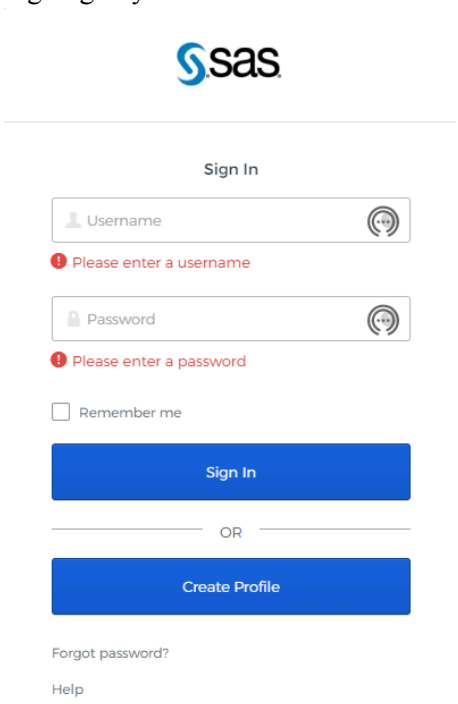
Step 1 - Internet search on Chrome, Firefox, or Safari 'SAS Viya for Learners', clicking on the following link.



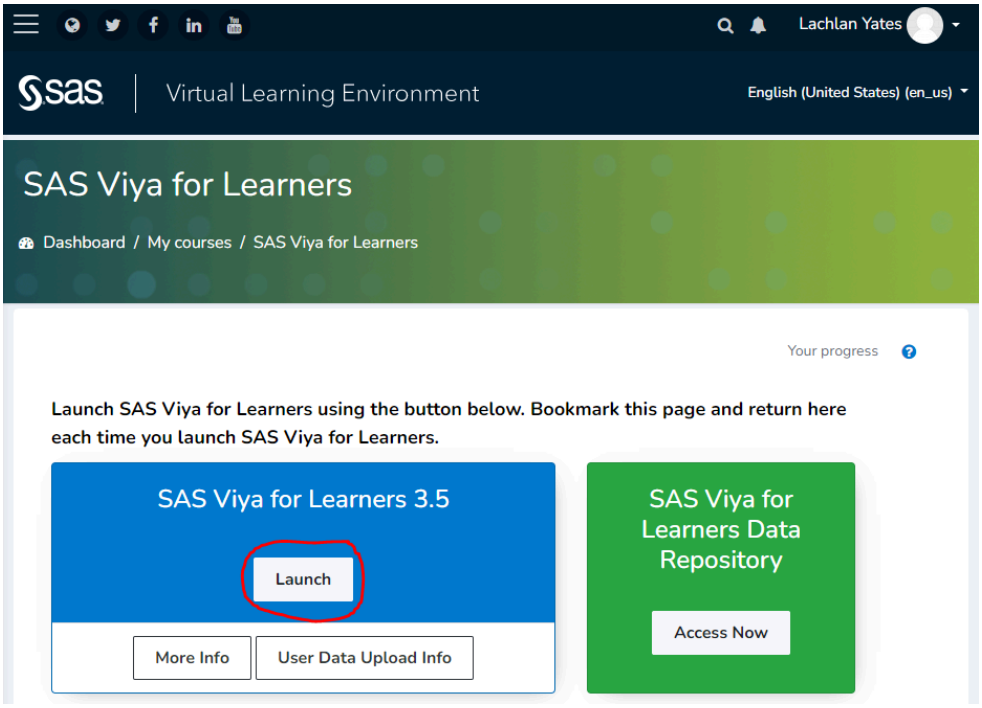
Step 2 - Once on the website, register for student access from the main website, clicking on 'Access for students'.



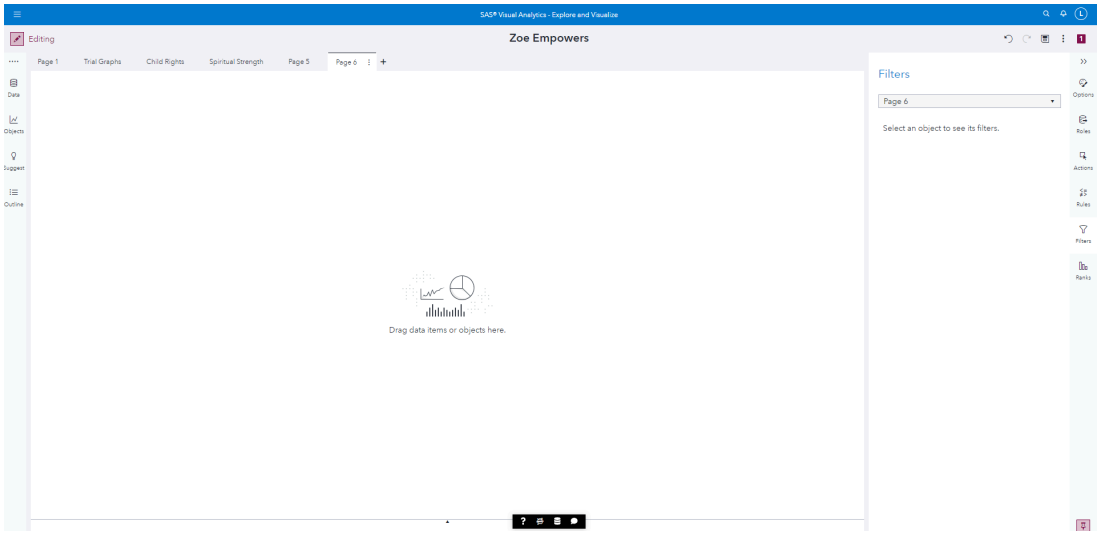
Step 3 - Create an account and sign in under a student account with the appropriate institution. After creating an account and signing in you will be directed to another webpage.



Step 4 - Once your account is created and you have signed in, the following webpage will appear. Click on 'Launch' to access the most recent version of Viya, being 'SAS Viya for Learners 3.5'. As SAS is a cloud-based platform, there is no direct 'downloading; of the platform, hence the 'Launch' option. It is recommended that once launched, users favourite the link to their personal SAS Viya drive.



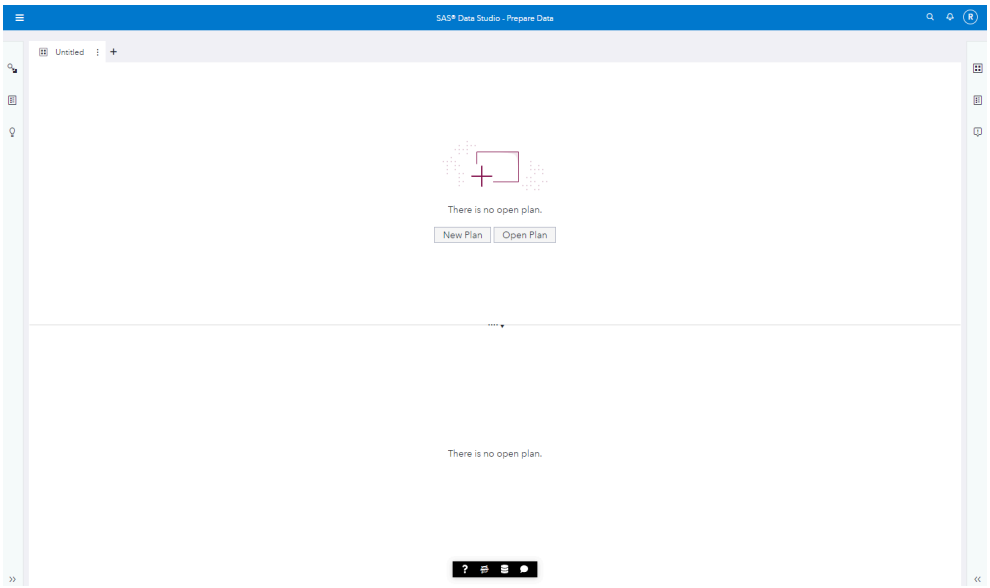
Step 5 - Once launched, users will be directed to their personal SAS Viya drive. It is recommended that once launched and directed to the drive, users favourite this webpage to easily access their drive in the future.



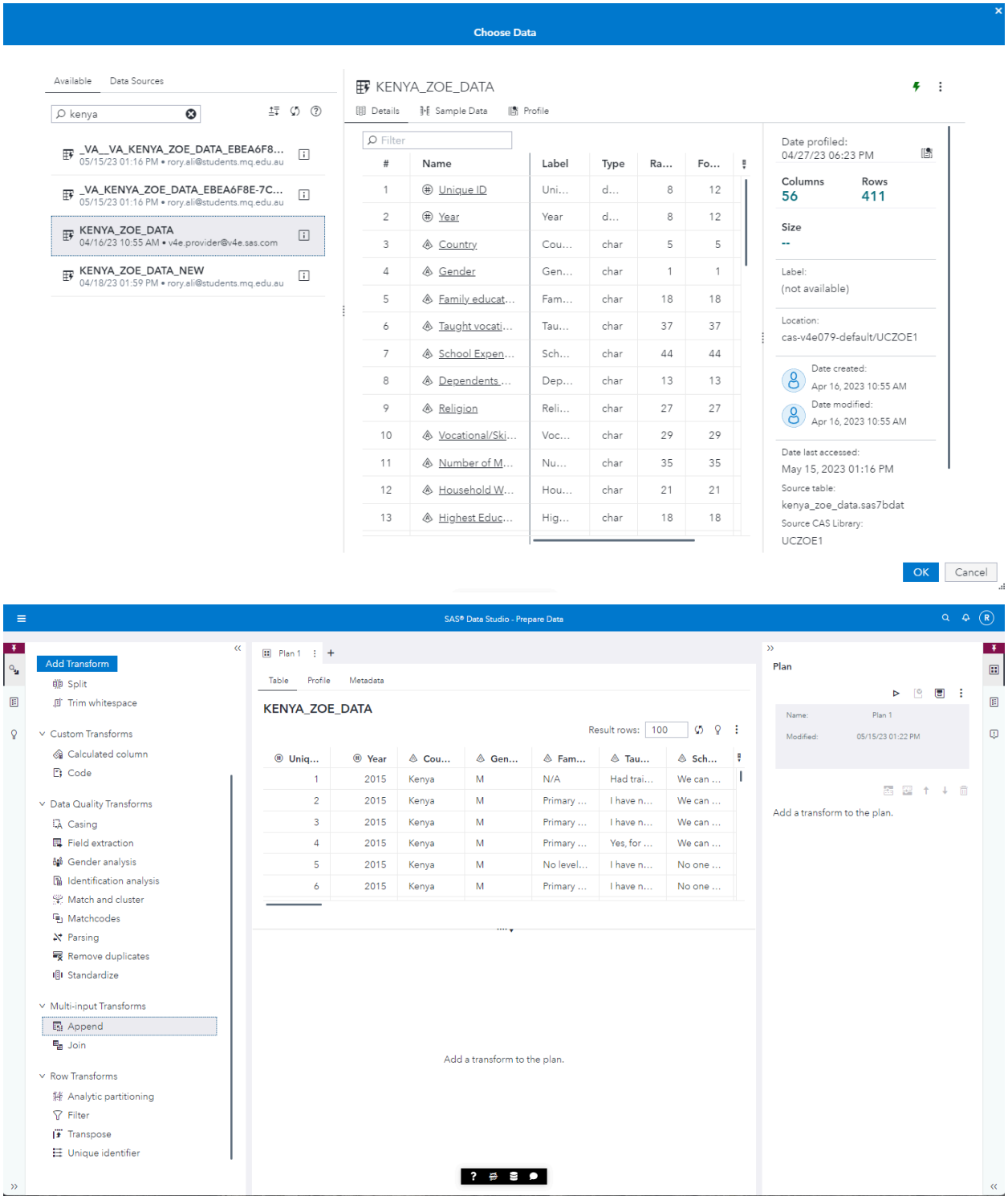
Step 6 - SAS Viya is now installed. Users are now able to proceed in appending and joining datasets.

1.1.2. HOW TO JOIN DATA TO CREATE VISUALISATIONS

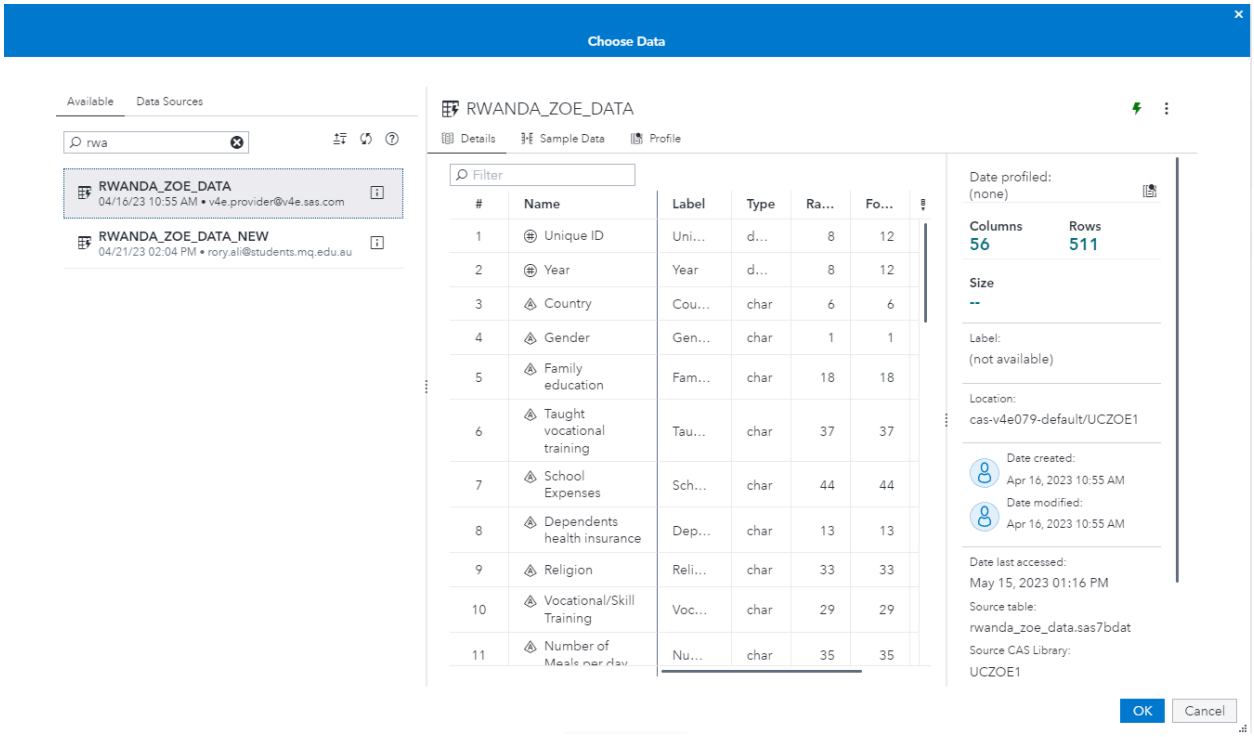
Within the Viya for learners platform, in order to start any data manipulation or visualisation you must first proceed with appending and joining the provided datasets. The following images outline the appropriate steps that need to be taken within the Viya for learners platform in order to properly complete these crucial steps so the data can be properly utilised. Furthermore, manipulated data used for our own project will be displayed and demonstrated as examples to provide real insight into Viya usage.

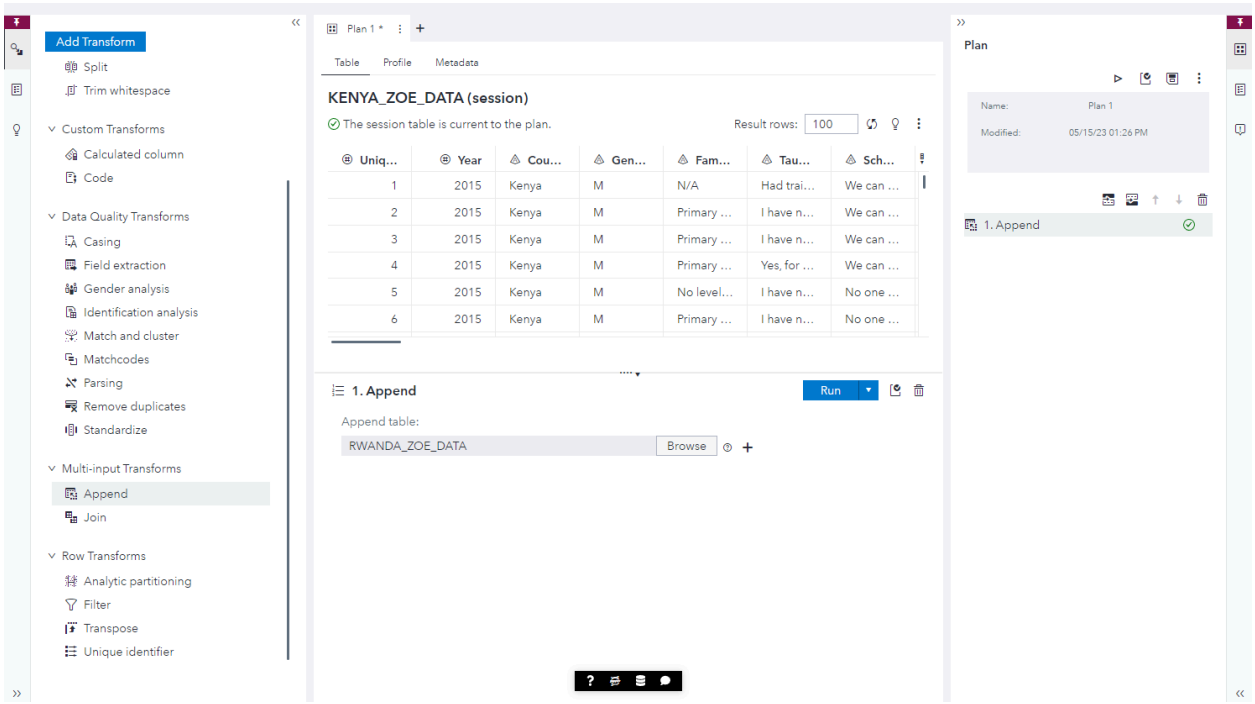


Within the menu the prepare data tab needs to be selected to be taken to the data preparation page.



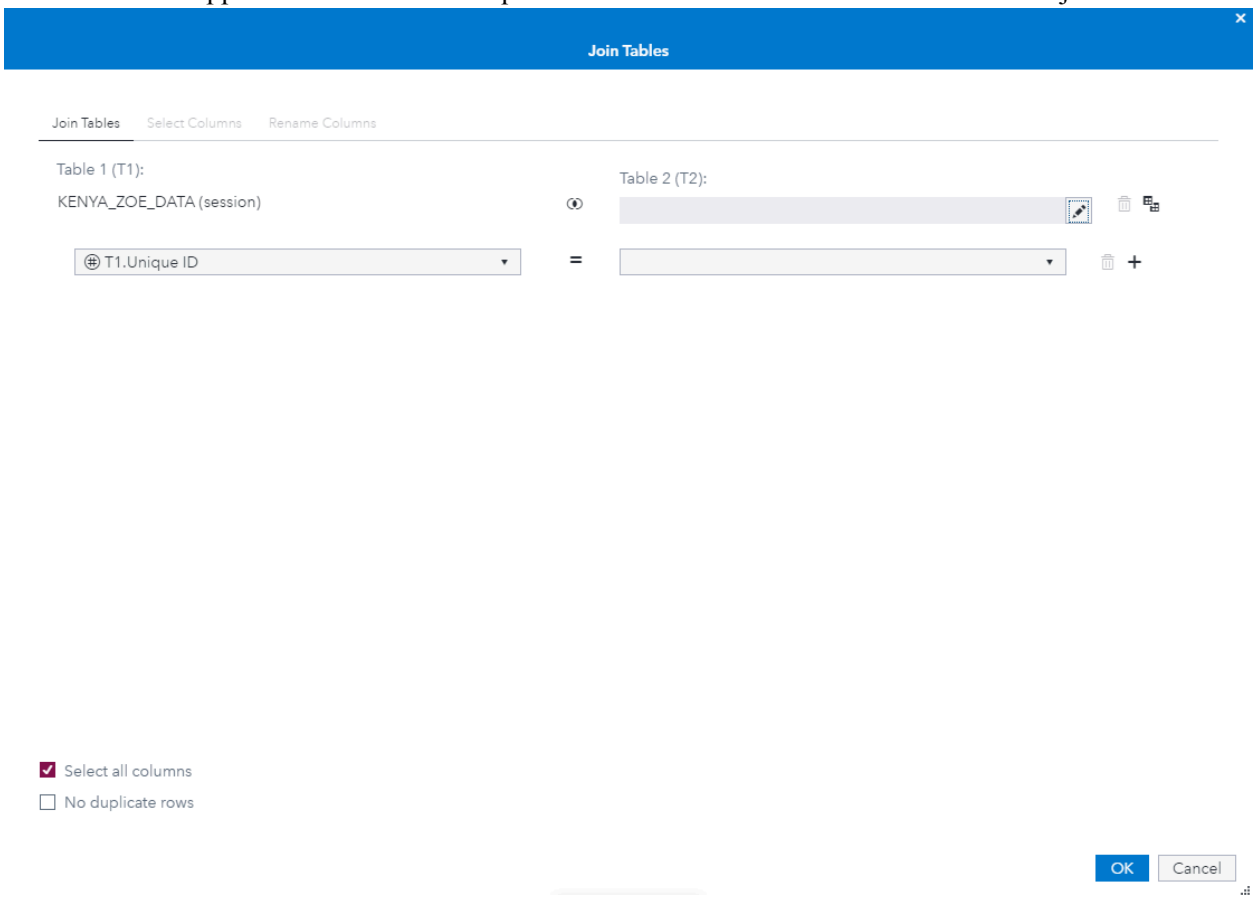
Once “New Plan” is selected the specific base dataset must be searched for. For this specific task “kenya_zoe_data” will be used. Once selected this will become the base dataset shown within the interface. Following this the user must select the “append” tab.



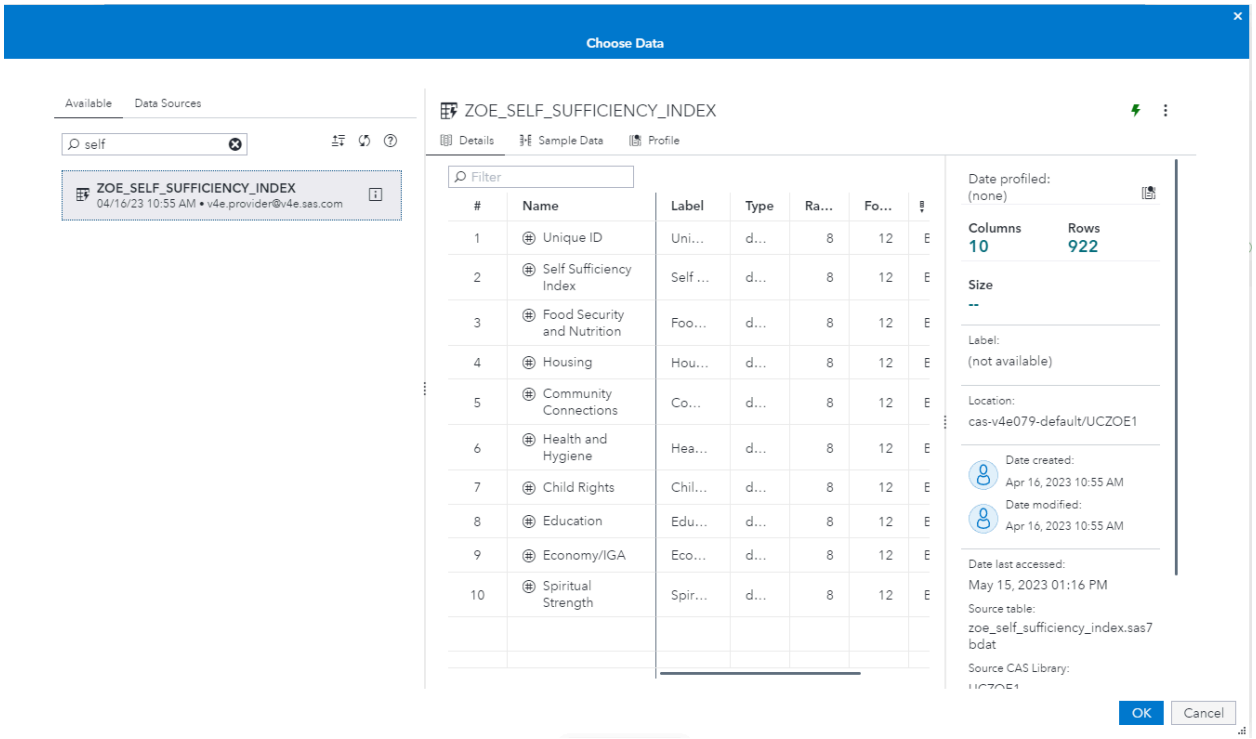


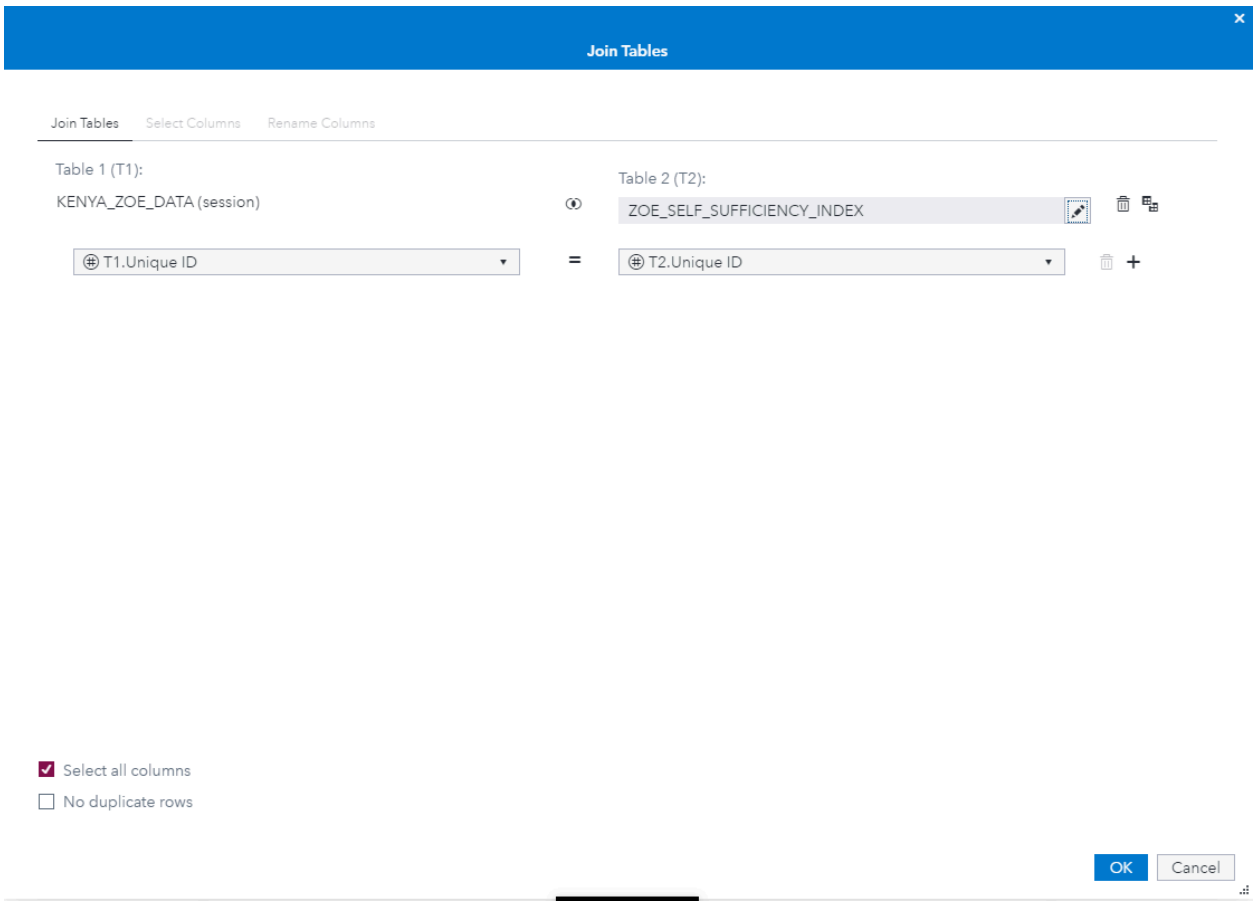
Once selected, the dataset needing to be appended to the kenya set is the matching rwanda data. It will show up as Rwanda_Zoe_Data when searched. Once selected and ok is pressed it will appear under the kenya session as shown above. The user then needs to proceed and press the run button for the append to

be complete which will add the green tick next to the append plan on the right of the screen. After the append is complete to begin the join sequence the user must select the “join” feature under the previously used append button. The subsequent screen will ask the user for which table to join:

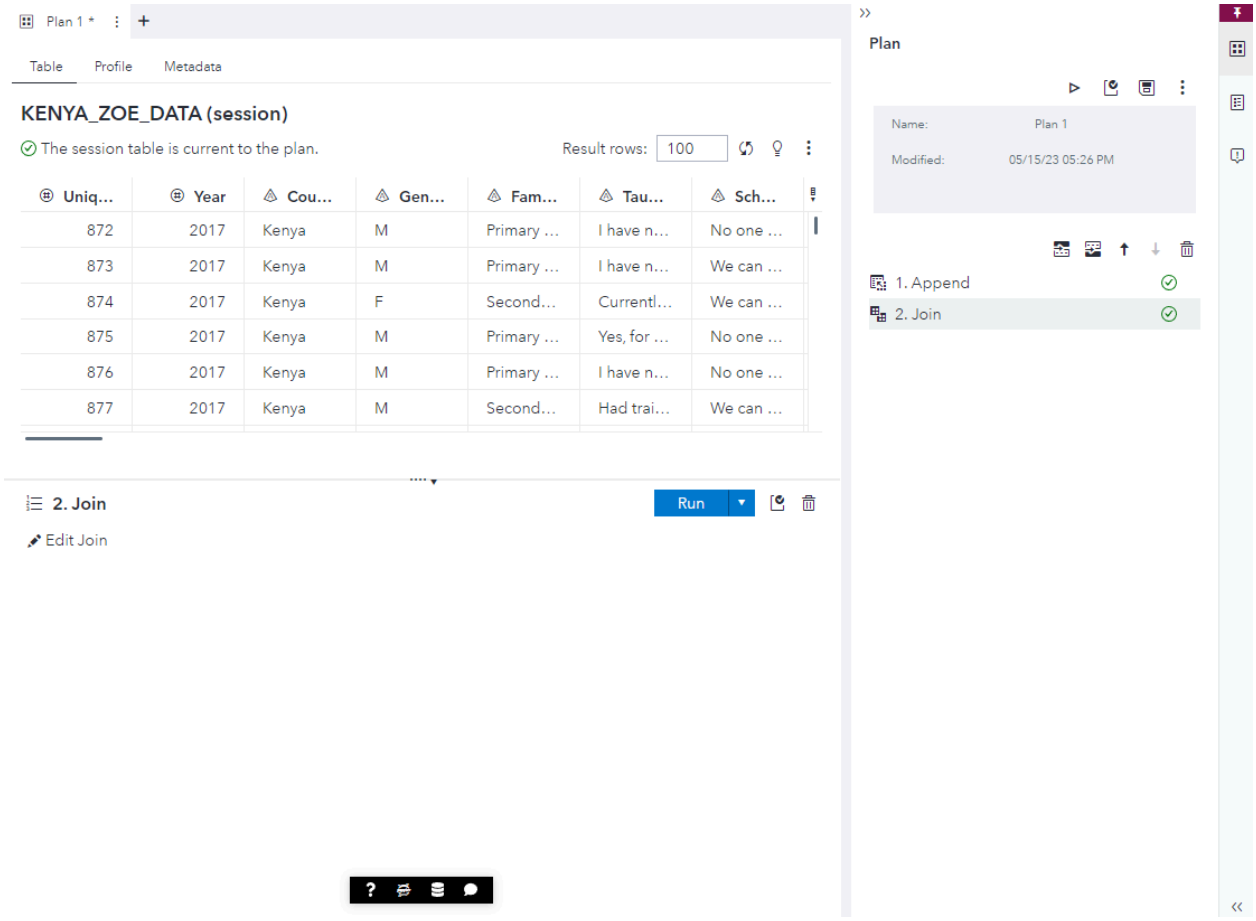


The bar under table two is where the user will search for “Zoe_self_sufficiency_index”

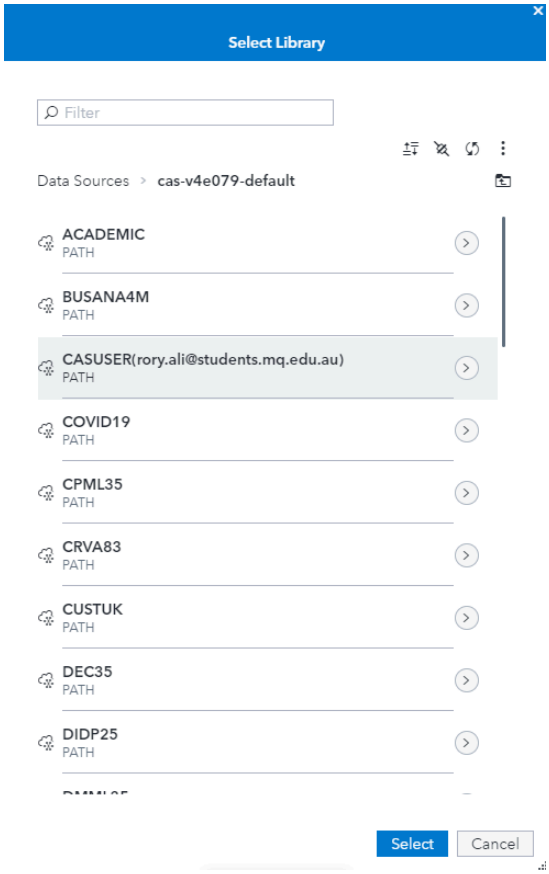




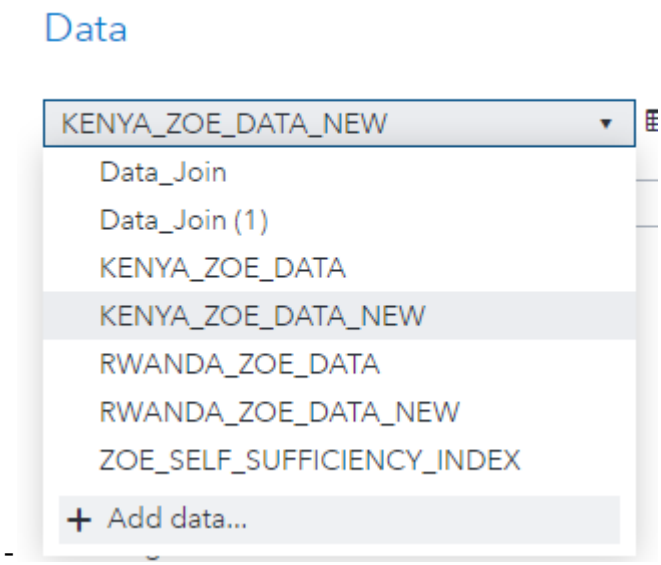
Once selected and run using the same button that was used to run the append the user should see a green tick next to both the append and join prompts.

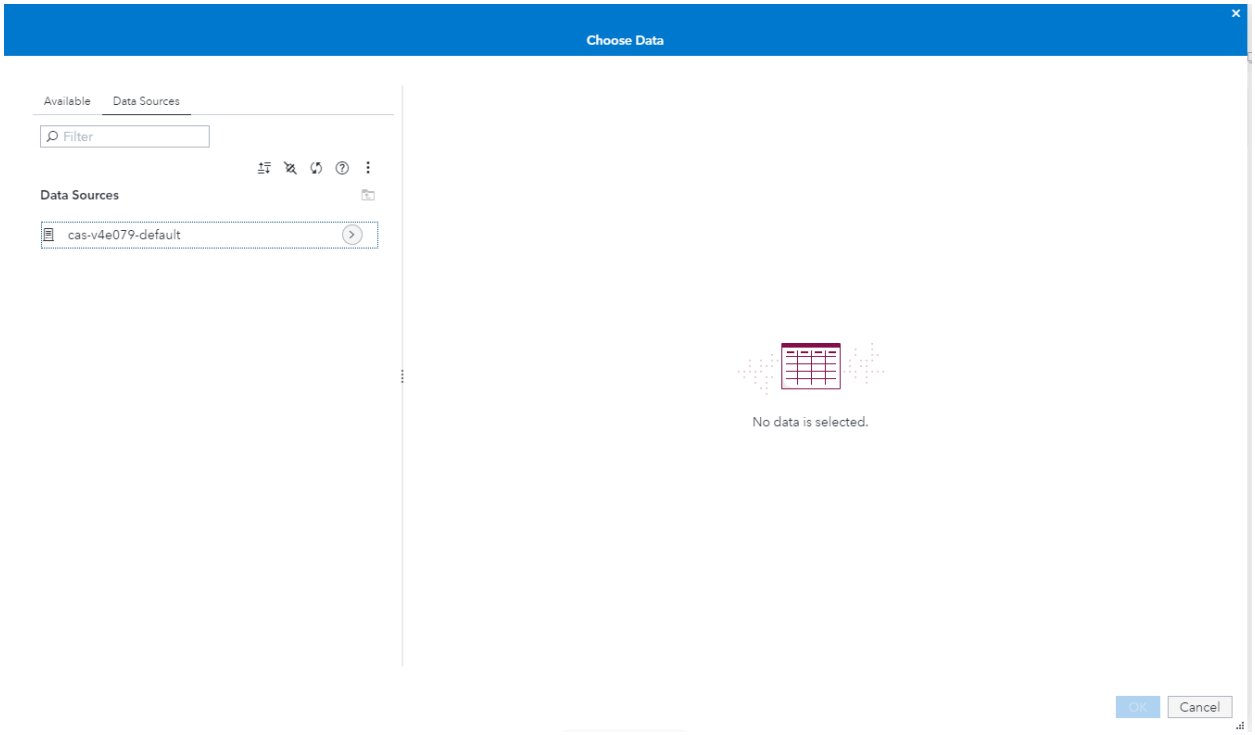


This completed table will need to be saved before it can be used for any visualisation work. This will be done through the save tab just above the plan name. The User will then need to save it within their personal SAS Library (CASUSER).

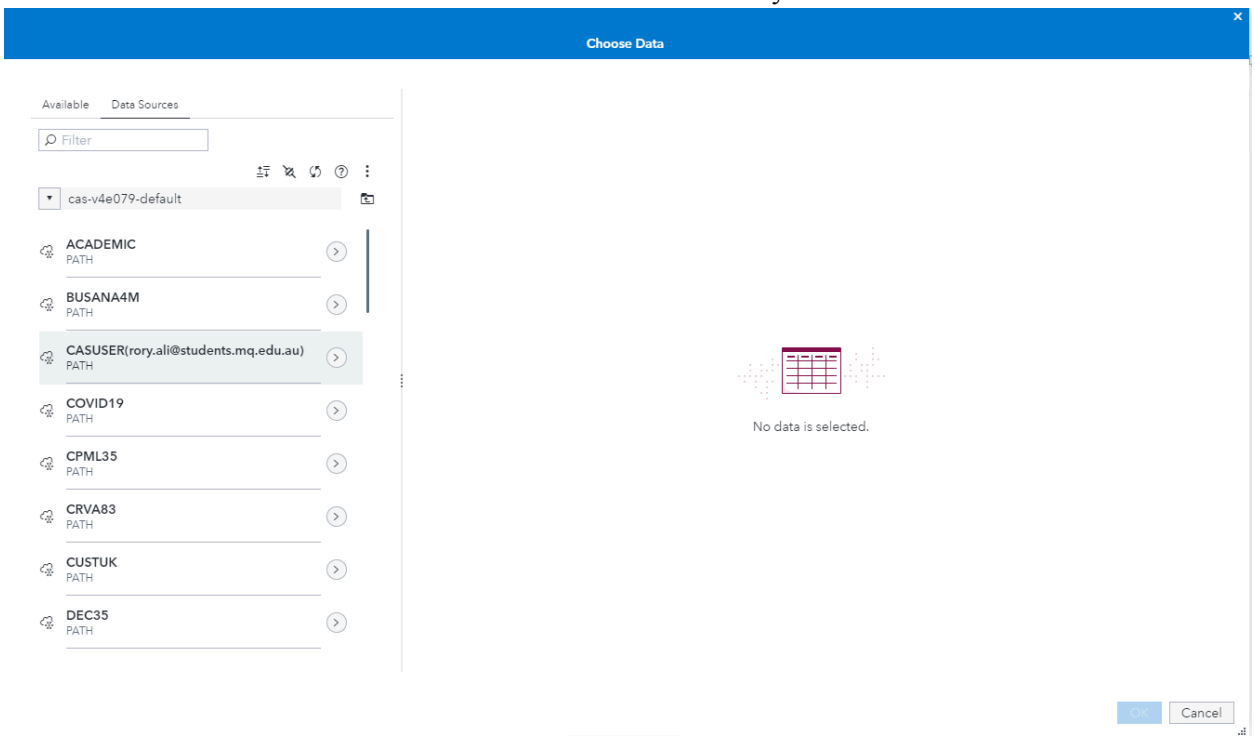


Once saved to make the data accessible within the report it will need to be added though the explore and visualise page by pressing the add data function.

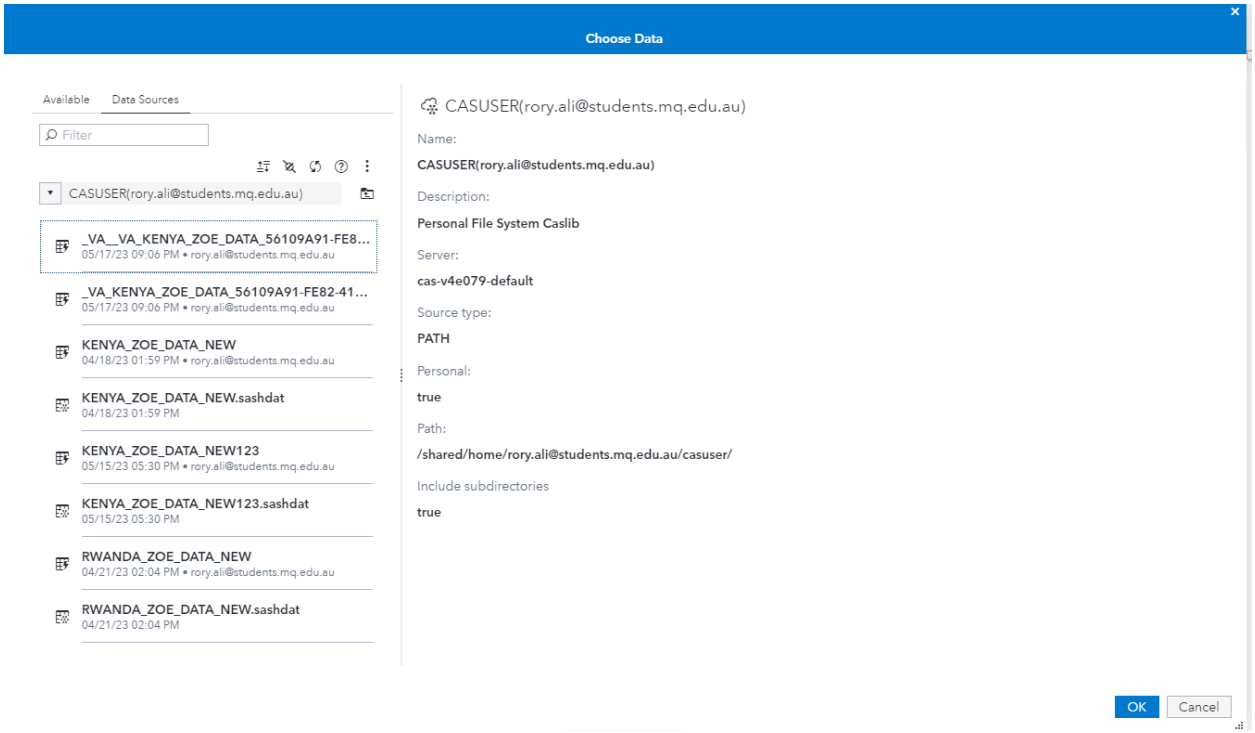




Proceed to the data sources tab and find the same user library that the new data was saved into.



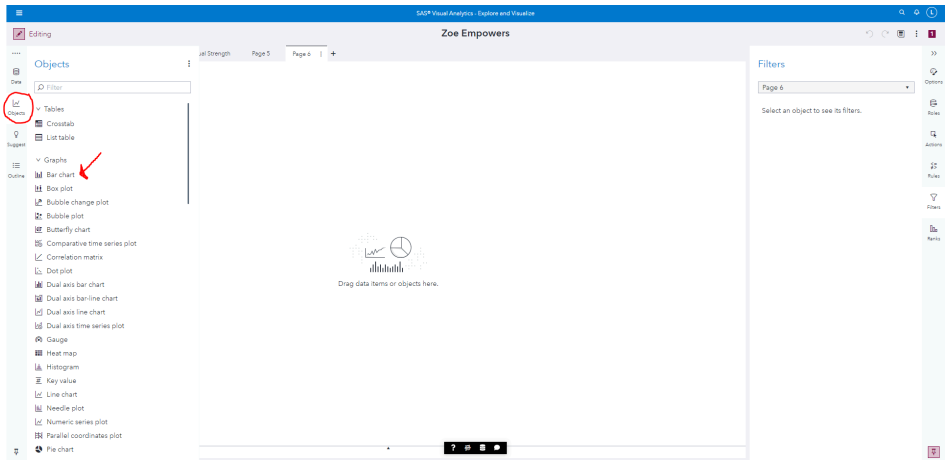
If all steps are done correctly then this folder will show your dataset as well as any other new ones that have been appended, joined or modified and saved.



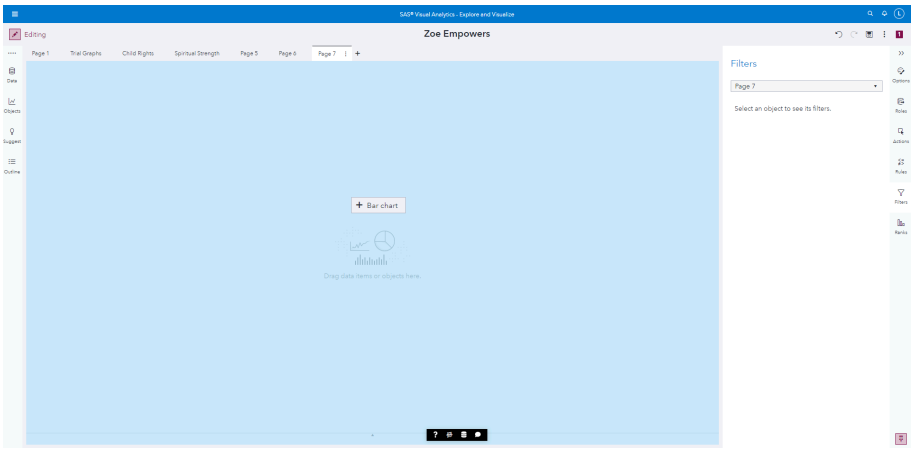
Select the desired dataset and it will load into the data list ready for manipulation within the SAS platform.

1.1.3. HOW TO SET UP MODELS?

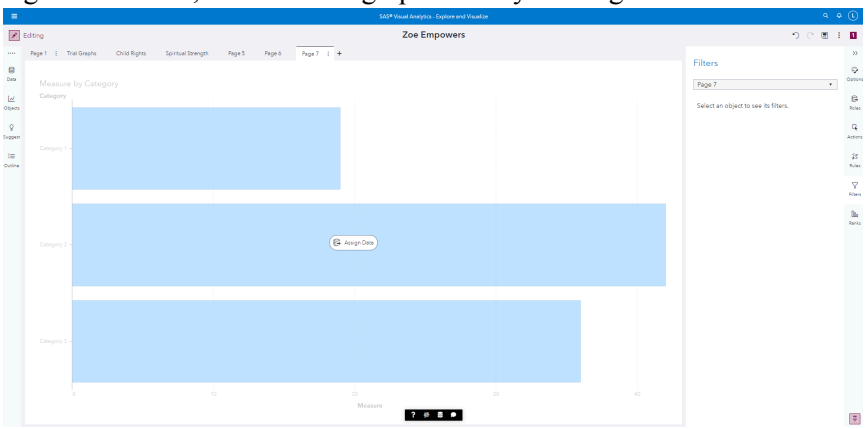
The following steps showcase how to successfully set up a model. On the left hand side of the screen, click on ‘Objects’. This will reveal an array of possible graphs to choose from when setting up a particular model. For this demonstration, we will be using a bar graph.



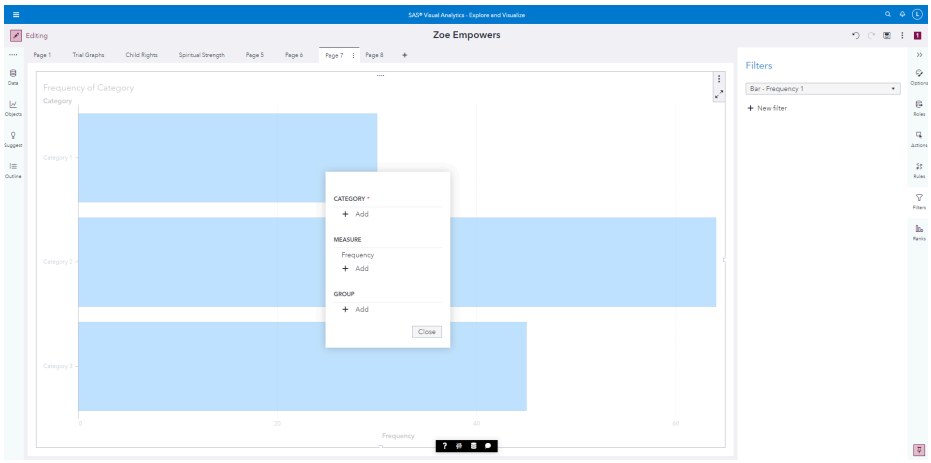
When clicking on the bar graph, hold in left click, and drag the bar graph across to the middle of the screen where “Drag data items or objects here’.’appears, to which the screen will be highlighted blue. Once highlighted blue, release the left click.



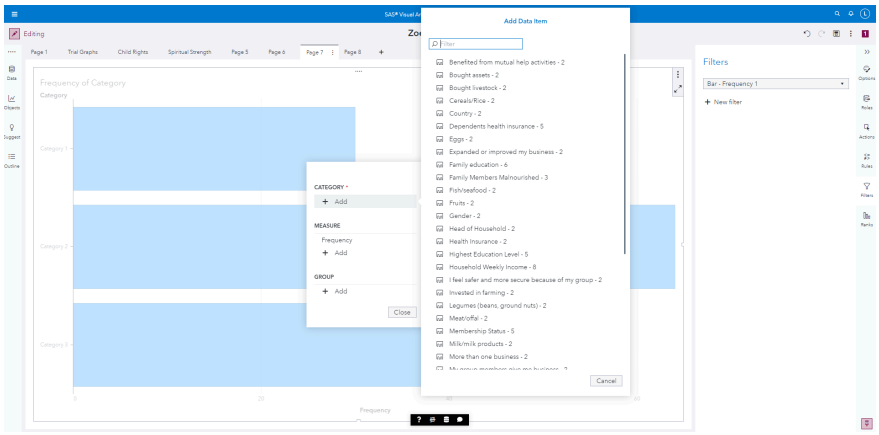
After releasing the left click, the bar chart/graph is ready to assign data to create a desired model.



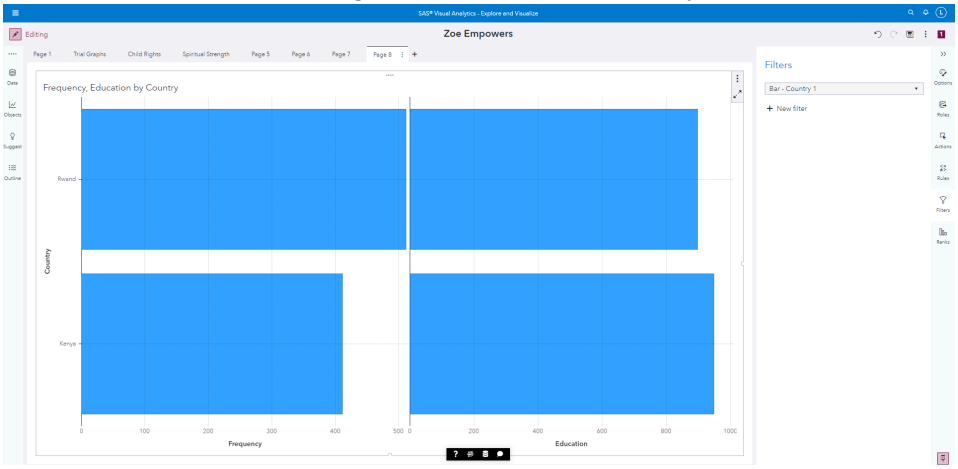
Click on assign data to add in a chosen category, measure, and group.



Click on the '+ Add' option under 'Category', 'Measure', and 'Group', assigning data of interest that will be used in your bar chart.



Once data has been assigned, click on close to view your data model.

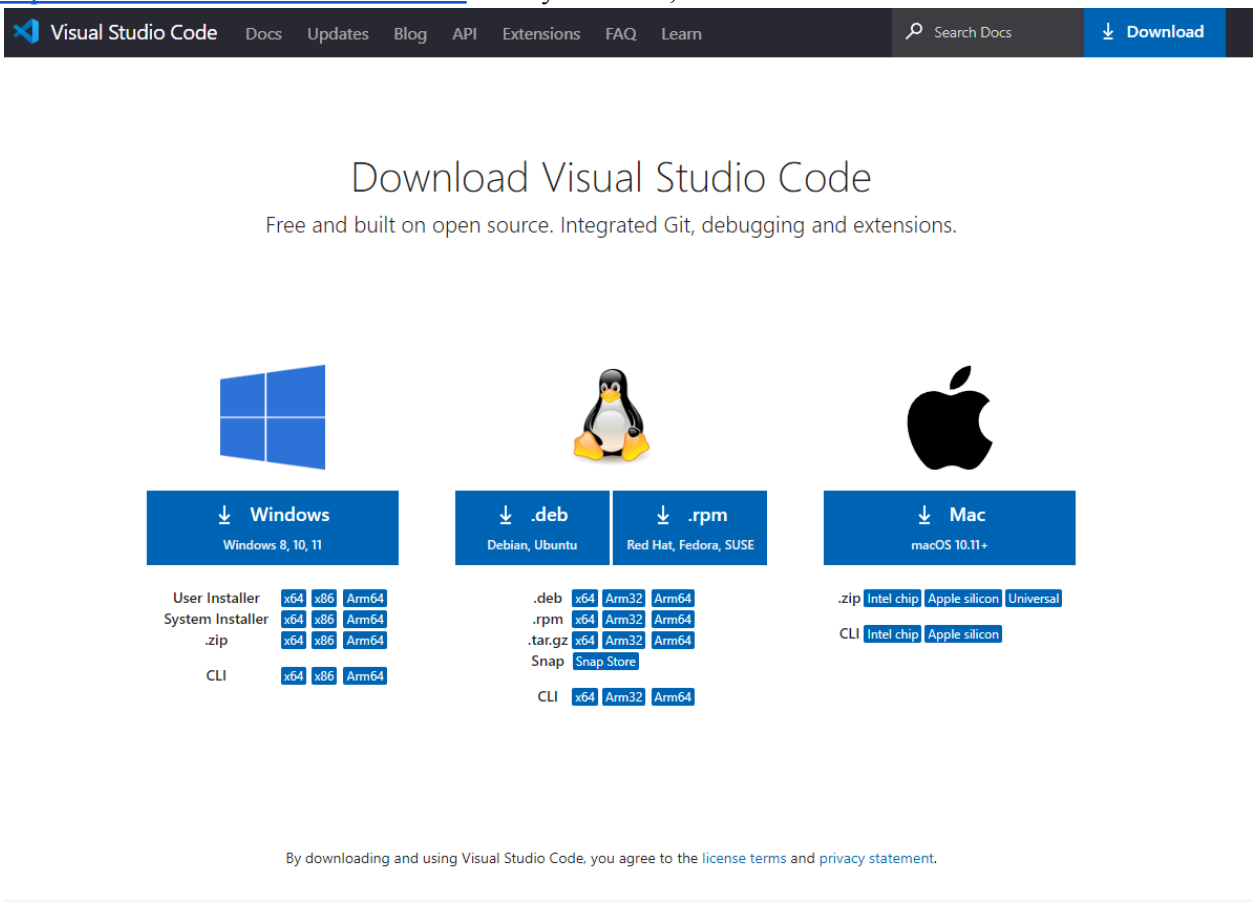


The user has now successfully set up their model and is now able to utilise their model.

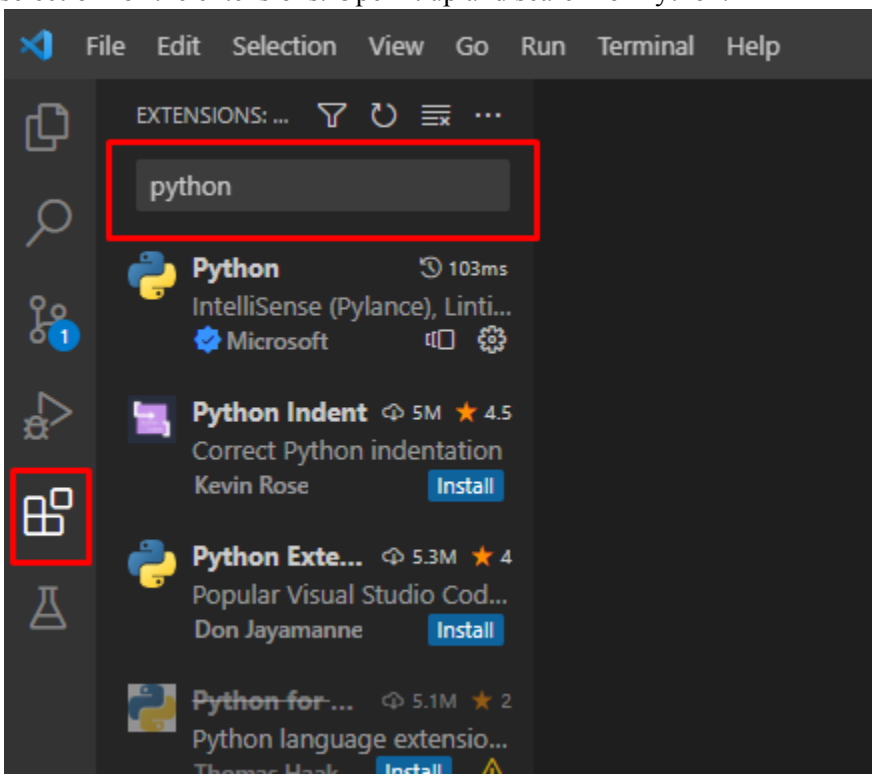
1.2. PYTHON

1.2.1. SETTING UP YOUR IDE

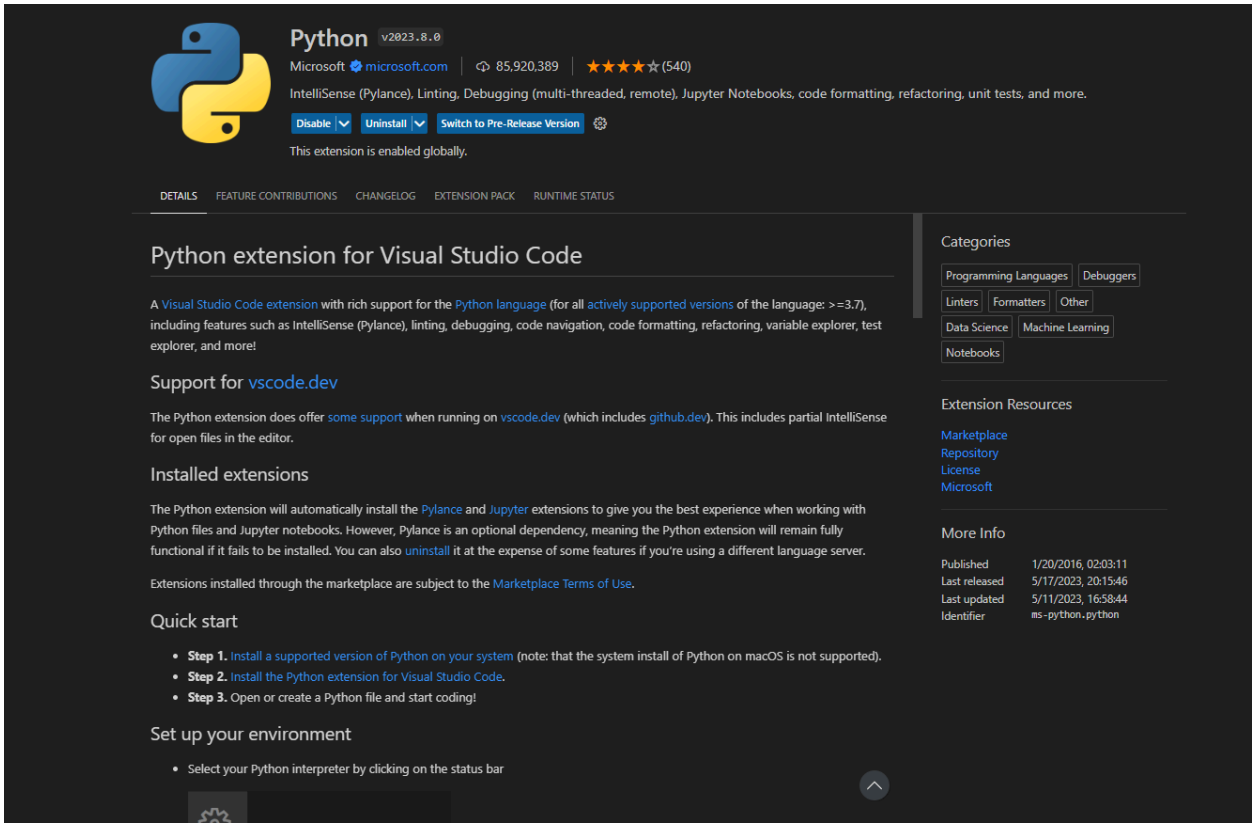
To get started we have used Visual Studio Code to do the python programming. You can download it from https://code.visualstudio.com/download. Pick your distro, download and install



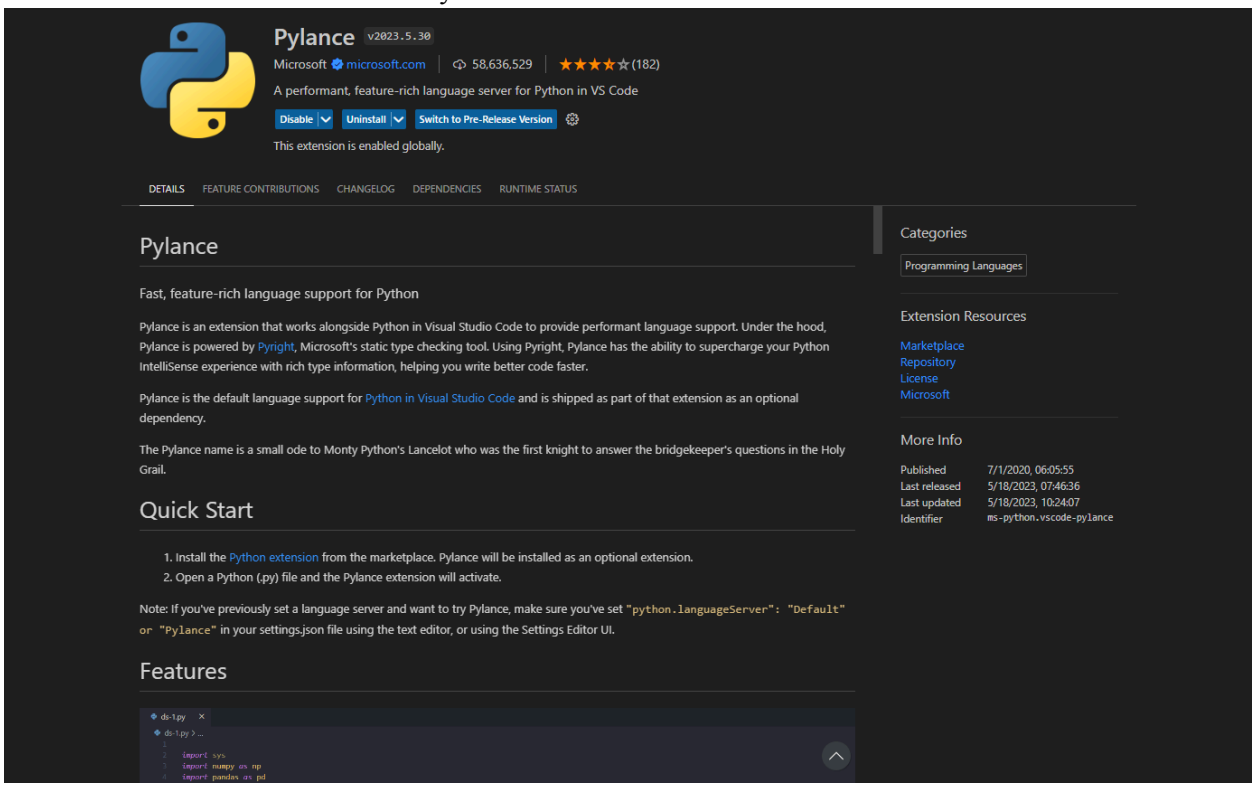
Step 1 - Importing the dependencies in VS code
The power of VS code is its extensive extensions. To start with it is quite the shell of an IDE but after installing these dependencies it will be a powerful tool. On the right hand side there is a selection for the extensions. Open it up and search for Python.



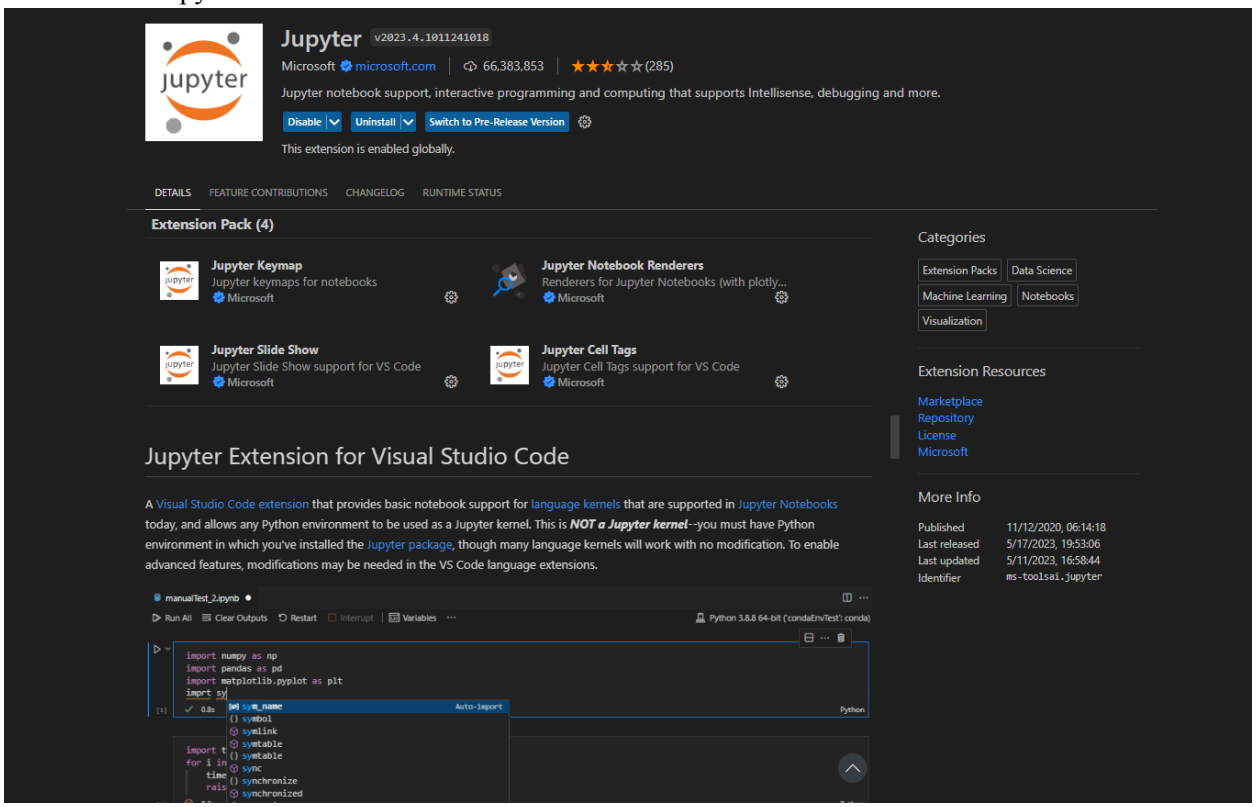
Install.



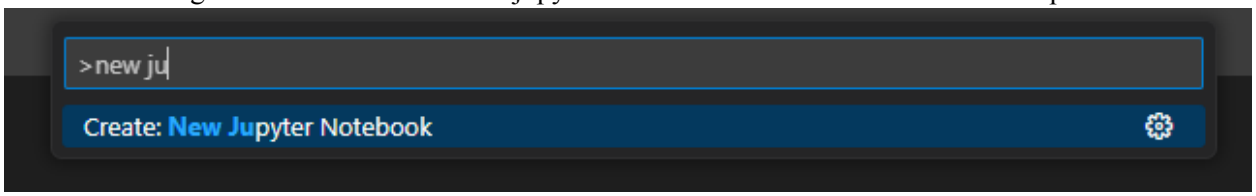
Next do the same to collect PyLance



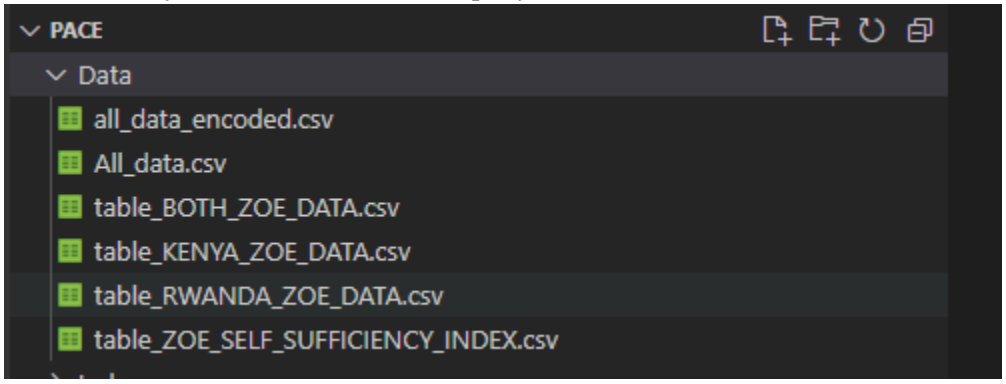
And Jupyter.



Now to get started we create a new jupyter notebook from the search bar at the top of VS code.

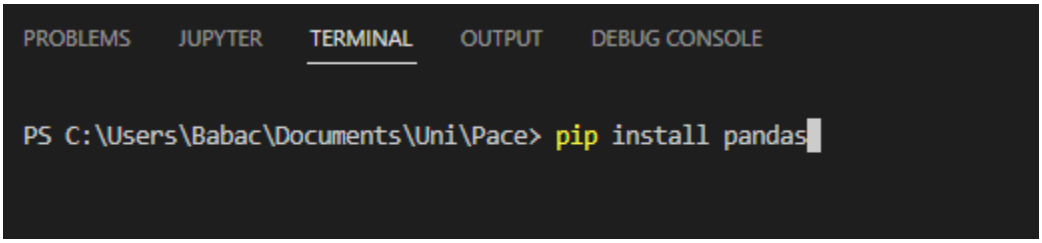


Finally, create a data folder and put your files in there.



1.2.2. PIP INSTALL

We are going to be using some more dependencies, but these are for python itself. Firstly, put this into your terminal to install pandas. If you are having problems please go to their documentation. https://pandas.pydata.org/docs/getting_started/install.html



Pandas is made for reading and manipulating data files, like CSVs.
There are many required instals. If you are having problems, click on them for a link to their installation documentation.

- Numpy: A library for managing arrays and manipulating them mathematically/functionally.
- Seaborn: Visualisations based on the matplotlib that have been simplified.
- Sklearn: A tool for predictive analysis based on Numpy and Matplotlib
- Matplotlib: For making statistical and analytical visualisations.
- Statsmodels: For deeper statistical analysis of models used.

1.2.3. PYTHON IMPORTS

In a new cell, we need to import these libraries into our project.

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split
from sklearn.metrics import mean_squared_error, r2_score
```

1.2.4. PANDAS TO OPEN A CSV FILE

To get started on actually analysing our data we open them in the script using pandas.read_csv. Then we merge them on the column that is equivalent between them.

```
both_regions = pd.read_csv("data/table_BOTH_ZOE_DATA.csv")
scores = pd.read_csv("data/table_ZOE_SELF_SUFFICIENCY_INDEX.csv")
main = pd.merge(scores, both_regions, on="Unique ID")
main.head()
```

	Unique ID	Self Sufficiency Index	Food Security and Nutrition	Housing	Community Connections	Health and Hygiene	Child Rights	Education	Economy/IGA	Spiritual Strength	...	I beg for food	I eat enough food each day so th	I live in an adequate/safe home	I have access to medical care	I use clean or boiled water	I know my rights and can enforce	I know how and where to seek hel	I attend community events (meeti	I have a business	I have livestock or crops
0	1	21.836310	2.500000	2.50	2.666667	2.857143	3.000000	3	2.812500	2.50	...	1	4	3	4	4	4	4	4	4	4
1	2	21.817460	2.571429	2.50	2.444444	2.857143	3.000000	3	2.944444	2.50	...	1	4	3	4	4	4	4	4	4	4
2	3	21.107143	2.500000	2.75	2.857143	2.166667	2.166667	3	2.916667	2.75	...	1	4	4	3	4	3	3	4	4	4
3	4	20.252976	2.357143	2.50	2.666667	2.000000	2.166667	3	2.812500	2.75	...	1	3	3	3	3	3	3	4	4	3
4	5	17.658333	2.500000	2.50	2.375000	2.333333	2.333333	.	2.866667	2.75	...	1	4	3	3	3	3	4	3	4	3

5 rows × 65 columns

The head function shows the top five results of the data. Just to help make sure we have what we're expecting.

1.2.5. ENCODING

For our goal of getting the linear regression model, we need to clean and encode some sections. This quick clean function converts all the Y & Ns to boolean values and there was a specific issue with some full stop that did result in a zero.

```
def clean_bool(df):
    cDf = df
    for col in cDf.columns:
        cDf.loc[cDf[col] == 'Y', col] = 1
        cDf.loc[cDf[col] == 'N', col] = 0
        cDf.loc[cDf[col] == "            .", col] = 0
    return cDf
main = clean_bool(main)
```

Additionally the membership status needs to be encoded, to be able to run a linear regression model, all values are required to be numeric.

```
main.loc[main["Membership Status"] == "Not yet a member", "Membership Status"] = 0
main.loc[main["Membership Status"] == "Member less than 3 months", "Membership Status"] = 1
main.loc[main["Membership Status"] == "Member 1 yr - 2 yr", "Membership Status"] = 2
main.loc[main["Membership Status"] == "Recent Gradute (less than 3 months)", "Membership Status"] = 3
main = main[main["Membership Status"] != 0]
```

On Top of that, the membership status "not yet a member" only appears once out of the 900 rows, it is an outlier that needs to be removed.

1.2.6. SK LEARN

To prepare the data we have to remove all rows that are NA and force the status into a float.

```
l_data = main

l_data = l_data[l_data["Membership Status"].notna()]
l_data["Membership Status"] = l_data["Membership Status"].astype(float)
```

Then using numpy we take the logarithm of all the membership status and acquire our x and y for the linear regression model.

```
l_data["Log Membership Status"] = np.log(l_data["Membership Status"])
```

```
x = l_data["Log Membership Status"]
x = np.array(x.dropna()).reshape(-1, 1)
y = l_data["Self Sufficiency Index"]
```

Now building the model we use the provided model from Sklearn.
Then split the data into training and testing sets fit the model on the training set.
Then we can collect the results of this training through the model's information.

```
regression_model = LinearRegression()

x_train, x_test, y_train, y_test = train_test_split(x, y)
regression_model.fit(x_train, y_train)

y_pred = regression_model.predict(x_test)

print("Resulting Equation: y = ", regression_model.coef_, "* X + ", regression_model.intercept_)
print("Mean squared error: %.2f" % mean_squared_error(y_test, y_pred))
print("Coefficient of determination: %.2f" % r2_score(y_test, y_pred))
```

```
Resulting Equation: y =  [15.37169515] * X +  5.190646216537669
Mean squared error: 4.34
Coefficient of determination: 0.93
```

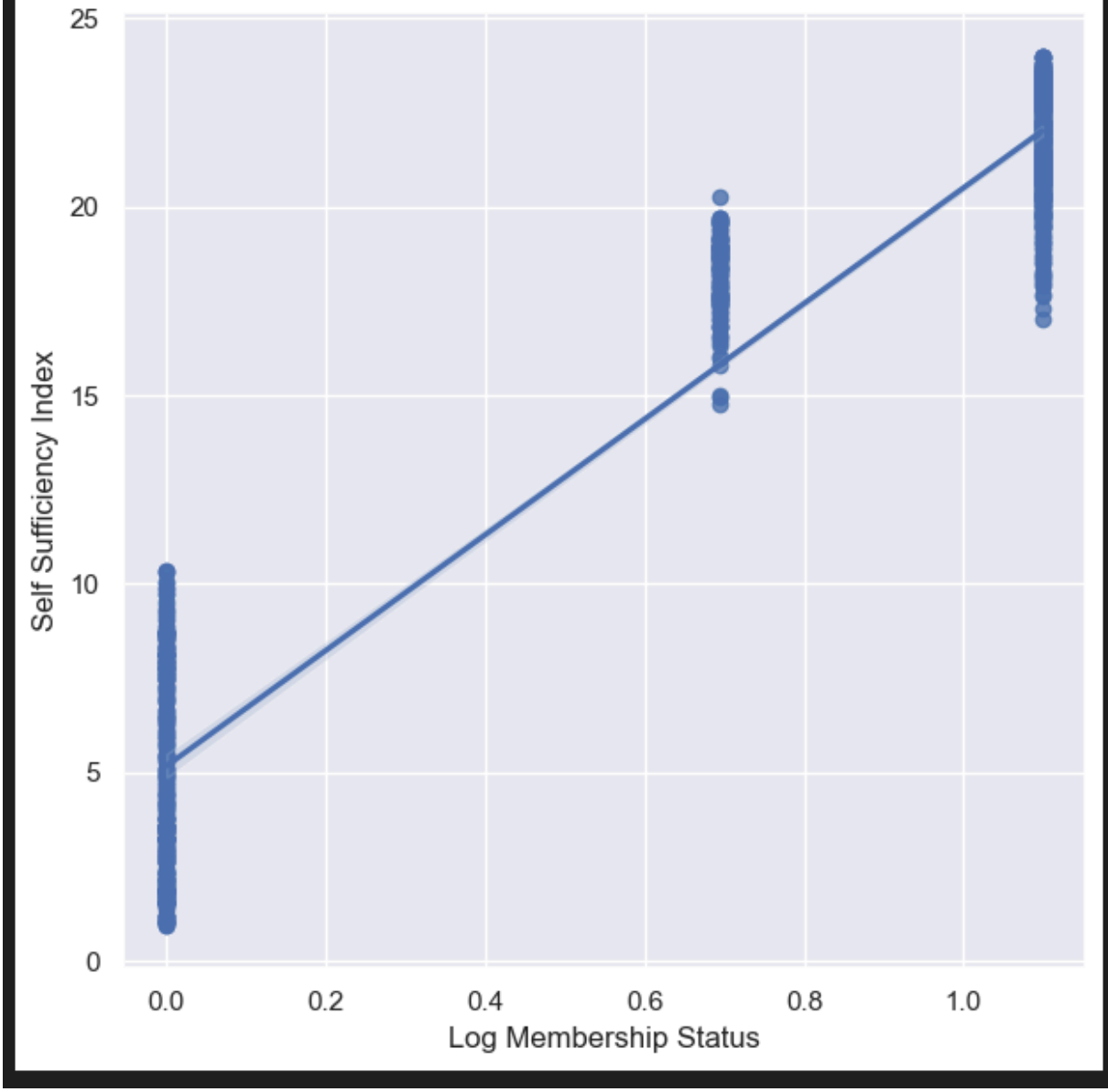
However this is very arbitrary and a visualisation will help us see what is actually happening.

1.2.7. SEABORN

The regplot for sea born will show us the linear regression that the SKLearn produces.

```
sns.regplot(x="Log Membership Status", y="Self Sufficiency Index", data=l_data)
```

```
<Axes: xlabel='Log Membership Status', ylabel='Self Sufficiency Index'>
```



1.2.8. STATSMODEL

Lastly using the statsmodel we can get even more in depth statistical information about the model that we are attempting to work with.

```
import statsmodels.api as sm
x = l_data["Log Membership Status"]
y = l_data["Self Sufficiency Index"]
model = sm.OLS(y, x).fit()
print(model.summary())
```

```
                                OLS Regression Results
==============================================================================================
Dep. Variable:     Self Sufficiency Index   R-squared (uncentered):                   0.955
Model:                                OLS   Adj. R-squared (uncentered):              0.955
Method:                     Least Squares   F-statistic:                          1.936e+04
Date:                    Thu, 18 May 2023   Prob (F-statistic):                        0.00
Time:                            10:56:27   Log-Likelihood:                         -2502.1
No. Observations:                     916   AIC:                                      5006.
Df Residuals:                         915   BIC:                                      5011.
Df Model:                               1
Covariance Type:                nonrobust
==============================================================================================
                             coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------------------
Log Membership Status    20.2849      0.146    139.132      0.000      19.999      20.571
==============================================================================
Omnibus:                       50.664   Durbin-Watson:                   0.176
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               53.874
Skew:                           0.564   Prob(JB):                     2.00e-12
Kurtosis:                       2.629   Cond. No.                         1.00
==============================================================================
```