Medicines & Healthcare products
Regulatory Agency

10 South Colonnade
Canary Wharf
London
E14 4PU
United Kingdom
gov.uk/mhra

End Point Assessment Report

Level 4 Data Analyst pathway

Shaheer Hussain

NOTICE : THIS DOCUMENT HAS A 3500 WORD COUNT, THIS EXCLUDES THE WRITING IN THE PROJECT MAPPING, APPENDIX AREA WHICH INCLUDES REFERENCES AND FIGURE NOTES IN ITALICS. WORD COUNT STARTS FROM PAGE 19 AND ENDS PAGE 27.

Operates data systems in compliance with all organisational and legislative requirements including principles of Privacy by Design	
	Project Mapping
S1 Use data systems securely to meet requirements and in line with organisational procedures and legislation, including principles of Privacy by Design. [Figure 6] [Figure 7] [Figure 10] [Figure 23]	When working on this LRT dashboard I have made sure that only I as well as the stakeholders have access to this dashboard and the data it relies on, this is to prevent any sort of unauthorised access to the data or dashboard. I have ensured that I work on this dashboard in a secure area (I am working from home and take precautions like locking my screen when leaving my desk) and not in areas like the train where I can lose personal belongings, or personal items can be stolen. I also make sure not to store any data on any external storage devices to reduce any potential of losing data which can result in a data breach/security risk. These are all measures I will have to take to meet requirements and be in line with organisational procedures and legislation. I will conduct security risk assessments to identify potential security vulnerabilities with the dashboard and the way it works.

Outlines and applies the principles of data analysis lifecycle to the steps of data analysis	
	Project Mapping
K3 Principles of the data analysis life cycle and the steps involved in carrying out routine data analysis tasks.	To show my knowledge of the principles of data analysis life cycle I have created a timeline for the 6 weeks I will have to develop this dashboard, these 6 weeks will be split into section of:

		• Identifying problems & understanding business requirements • Data collection • Data cleansing and preparation • Data modelling • Dashboard design and analysis – Visualisation • Deployment I have explained how I have completed each section of the data analysis life cycle for this project.
S2 Implement the stages of the data analysis lifecycle. [Figure 1] [Figure 2] [Figure 3] [Figure 6] [Figure 7] [Figure 9] [Figure 10] [Figure 12] [Figure 13] [Figure 14] [Figure 15] [Figure 16] [Figure 19] [Figure 20] [Figure 21] [Figure 22] [Figure 23] [Figure 24]		To implement the stages of the data analysis lifecycle I have started by communicating with the Head of Laboratory resources to understand the specific requirements and objectives for this order management dashboard, this means that I have established and understood the requirements and challenges to face. To go further into our conversation, I have talked about where the stakeholders store their data, what kind of data they store and the format as this will enable them to identify the sources of relevant data which will then allow me to think about how it can be Ingested into PowerBI.I chose to use both the SharePoint and Excel workbook connectors built into PowerBI. The next task is to cleanse the data and prepare it for the dashboard, this includes removing irrelevant data such as null columns and ensuring its quality is suitable for analysis. This also includes tasks such as removing top rows, promoting headers, changing data types and creating custom columns to meet requirements. In PowerBI the next stage to tackle will be modelling the data, this will include linking up the fact and dimension tables to make sure the relevant data is related. I then designed a dashboard interface, using simple tables and visualisations to explore the data and identify useful insights into the order management processes, I have made sure to create an iterative loop of feedback from stakeholders via Microsoft Teams,this is because im taking a more agile approach which will give me many opportunities to cross check whether I am meeting the

		stakeholder's requirements and what I can improve on. Finally, I have deployed the final dashboards. I have evaluated each stage of the data analysis life cycle by analysing what needs to be done to meet the requirements of the dashboard and what I have done, if my work in a stage has not been up to standard, I have evaluated further on what I need to do to meet the requirement. I also evaluate the stages by asking whether the work in each stage will contribute to my objectives.

Describe the principles of data including open, public, administrative and research data and how they relate to the data used within the project		
		Project Mapping
K4 Principles of data, including open and public data, administrative data, and research data.		Here are some principles of data : Accuracy Data should be precise and free from errors and mistakes Regular validation and verification processes should be implemented to maintain data integrity Integrity Data should be reliable and trustworthy. Mechanisms to prevent unauthorised alteration and ensure consistency over its lifecycle are essential Confidentiality Sensitive data must be protected against unauthorised access to prevent breach Strong encryption, access controls, and privacy policies are critical to maintaining confidentiality Availability Data should be readily accessible when needed for authorised users Backup and recovery plans along with robust infrastructure will ensure more data availability Compliance Data management practices should adhere to relevant laws, regulations, and standards Regular audits and compliance checks are necessary to meet legal and regulatory requirements Transparency

	Data collection, usage, and management processes should be clear and understandable Organisations should be open about how they handle data, promoting trust and accountability Security Protect data from threats and breaches using appropriate security measures Implement firewalls, intrusion detection systems, and secure coding practices to safeguard data Minimisation Only collect and retain data that is necessary for the intended purpose Avoid excessive data collection and regularly review data retention policies I will be working on administrative data as the purpose of my dashboard is to manage laboratory functions including equipment maintenance and acquisition of equipment so this would be restricted to internal staff responsible for the administration and upkeep of the laboratory.

Demonstrates a reasoned application of the principles of data classification. Explains where any flexibilities in application have been applied and their purpose	
	Project Mapping
S3 Apply principles of data classification within data analysis activity, flexing approach as necessary. [Figure 9] [Figure 10] [Figure 23]	I have applied the principles of data classification by identifying data based on its sensitivity and importance, for example the order management dashboard does not have much personal data like employee names and other such details however, it will contain a lot of confidential information that isn't public since it will list orders of equipment including prices which means finance will be involved. I have assessed the sensitivity of the data

	provided to me and taken into consideration privacy regulations and organisational policies i.e. the dashboard will not be shared with anyone other than the head of laboratory resources due to the confidential nature of the data involved. The MHRA has its own data classification framework so I will apply this classification criteria to the data involved in my analysis, this will allow me to flex my approach so if I'm working with highly sensitive data, I will consider and adopt strict security measures like encryption and monitor the effectiveness of this classification to observe any opportunities for improvement. I have also made use of SharePoint as a connection to the data and storage for the data, this is because it makes use of TLS/SSL encryption, advanced threat protection and role based access controls meaning I can make sure that only the people who need access to the data have access.

Identifies quality risks in data analysis and outlines methods to mitigate, escalate and/or resolve them	
	Project Mapping
K8 Quality risks inherent in data and how to mitigate/resolve these. [Figure 18] [Figure 25] [Figure 26] [Figure 29] [Figure 30] [Figure 31] [Figure 36]	As a result of the Laboratory resources team using a spreadsheet for their asset management, I already understand that there will be issues with data quality, for example there can be issues with completeness as some cells may be missing data, the data could be inaccurate due to human error or the some of the data may not even be relevant. I will be cleansing the data to make sure that I only have relevant data that is complete, and I will assess the data to identify any potentially inaccurate data. In other words, I will be conducting root cause analysis to identify and understand the underlying factors behind any data quality issues. I will commit to improving the data

	quality by monitoring the data as well as soliciting feedback from stakeholders. Here is an overview on some factors like consistency, timeliness and integrity: Consistency : Inconsistencies in data may occur when different sources present conflicting information or when there are no protocols for data entry. Another situation where inconsistency can arise is when updates or modifications to the data are not uniformly applied across all systems or documents. To mitigate the risks associated with consistency issues it is essential to establish data governance policies and procedures to ensure that all data sources follow the standards. Conducting audits and checks can help detect inconsistencies allowing for a swift resolution. Timeliness : Issues related to timeliness occur when data is not updated or entered promptly resulting in outdated information that may not accurately reflect the circumstances. This can be especially problematic in paced environments where information changes rapidly such as markets or healthcare settings. To address risks effectively it is crucial to establish protocols for data updates and introduce automated processes wherever feasible to reduce delays. Setting expectations regarding the frequency of data updates and

	enforcing accountability can also play a role in ensuring effective management of timely data practices. Integrity : Data integrity concerns the accuracy and reliability of data throughout its existence. Risks to data integrity can stem from sources, including access, data corruption or malicious tampering. Ensuring security measures, like access controls, encryption and data validation checks is crucial for protecting the integrity of data. It's also important to have backups and plans for disaster recovery to reduce the risks linked to data loss or corruption. Additionally fostering a culture of data ethics and integrity, within the company can encourage employees to practice handling of data.
S6 Identify and escalate quality risks in data analysis with suggested mitigation/resolutions as appropriate. [Figure 25] [Figure 26] [Figure 29] [Figure 30] [Figure 31] [Figure 36]	Similar to K8, I will be conducting data quality assessment and root cause analysis, this will allow me to prioritise any potential data quality risks and escalate them to the stakeholders, I plan to escalate these quality risks via Teams meetings as this will make it easy to present the possible issues and discuss the list of mitigation/resolutions, I plan to be completely transparent about mitigation/resolutions so the stakeholder understands how they will affect the dashboard and this will act as a pipeline for stakeholder feedback.

Outlines and applies the principles for defining customer requirements and implements findings in data analytics planning and outputs	
	Project Mapping

K9 Principal approaches to defining customer requirements for data analysis. [Figure 3] [Figure 16]		I have made use of communication methods like messages and emails as well as setting up meetings to draw out requirements from the stakeholders however meetings will be the most preferable as it will be the easiest to discuss the requirements for the dashboard, I have communicated with the main stakeholder whilst considering that they are not the most technical stakeholder, this means I have spent time explaining my process of creating the dashboard, the features and capabilities of the dashboard and how to utilise this dashboard effectively. I have documented stakeholder requirements using notes and proper documentation to make sure I can keep on track with fulfilling these requirements as well as recording meetings with the stakeholder (As long as consent to record is given to me which is incredibly important, if I need to record a meeting I like to ask everyone in the call and anyone who will join us if I can record the call/meeting.). I make use of mockups to make the stakeholder further understand how the dashboard may work which will give me the opportunity to have a constant stream of feedback to improve the dashboard's final design. I intend to create an iterative feedback loop to constantly gather feedback directly from stakeholders to make sure I am meeting their specific needs.
S7 Undertake customer requirements analysis and implement findings in data analytics planning and outputs. [Figure 1] [Figure 2] [Figure 3] [Figure 11] [Figure 12] [Figure 13] [Figure 16]		I have demonstrated this skill by engaging with the key stakeholders to understand their specific requirements for this dashboard and what their main objective is. This process included clarifying any misunderstanding or answering the stakeholders' questions about the dashboard. In our conversation I identified the key metrics and visualisations needed to display the information the stakeholder needs. This allowed me to plan the techniques I would use for the analysis of data. During the development of this dashboard I made it a priority to incorporate feedback from the key stakeholders to make sure that the final product sufficiently meets the requirements. After generating the insights and interpreting the results of the

[Figure 26] [Figure 32] [Figure 33] [Figure 34] [Figure 35]	visualisations, I communicated the findings with the key stakeholders in a clear and concise manner, this included training the stakeholders on how to effectively use the dashboard.

Demonstrates how data from different sources is combined and prepared for data analysis setting out how they identified the risks and challenges inherent in combining data within the project.	
	Project Mapping
S8 Identify data sources and the risks, challenges to combination within data analysis activity. [Figure 6] [Figure 7] [Figure 18]	I have identified the main data source for this project to be an excel spreadsheet stored in a SharePoint, this means that there can be a lot of potential data quality risks regarding the completeness of data, the accuracy of data and even the relevancy of data in the spreadsheet. This can present some challenges as the data may need to be cleansed a lot or a little depending on how the Laboratory resources team has formatted their asset management system. I will conduct a risk/challenge analysis for combining the data from the different potential data sources in this project and investigate points like compatibility and reliability.

Describes the tools and methods used by their organisation for data analysis and identifies which were used within the project with reasoning for the choices made to achieve the best outcome	
	Project Mapping

K11 Approaches to organisational tools and methods for data analysis. [Figure 6] [Figure 7] [Figure 17] [Figure 27]		At MHRA, PowerBI, SharePoint and Microsoft Excel are commonly used tools for data storage and visualisation so I made use of these tools to develop the LRT dashboard, this is because PowerBI is very capable for the objective of the dashboard and supports a huge variety of data sources including excel which is what the LRT asset management system is saved as, SharePoint will be a good area of storage for the excel sheet and I can use PowerBI to directly connect to SharePoint so we can conduct automatic refreshes on PowerBI WEB. I have included a tool comparison between Powerbi and excel as excel was a potential tool I could use for this project.
S15 Select and apply the most appropriate data tools to achieve the best outcome. [Figure 17] [Figure 27]		Selecting the most appropriate data tool will require me to understand the requirements of the project, the main outcome of this project is to have a dashboard for order management, the tools we use at the MHRA to ingest data and design dashboards include PowerBI and MicroStrategy however PowerBI alone will fulfil the requirements of this project by allowing me to design a dashboard with the key metrics and insights the stakeholders desire, PowerBI is also incredibly versatile when it comes to ingesting data as it supports hundreds of different data sources and allows for easy exploration of data. I concluded that excel would not be appropriate as its not as easy to automate data refresh and the visualisations I can create are much more advanced on powerbi which means there's more opportunity, I also have more experience on PowerBI that excel and microstrategy.

Analyses data sets taking account of different data structures and database designs		
		Project Mapping
S4 Analyse data sets taking account of different data structures and database designs.		In this project I may have to make tables relational by establishing the cardinality between these tables in PowerBI's Model view; this kind of data

[Figure 24]		structure will enforce referential integrity, preventing any sort of inconsistency within the data. This will also ensure that the data is normalised which will reduce data redundancy and improve data integrity. The data for this dashboard will be taken from a spreadsheet stored in a SharePoint, this means that I will be dealing with a free database design which could pose challenges depending on how LRT has structured their asses management system However I will have to transform and load the data to reach my objectives for the dashboard. I will analyse the data sets with different data structures by identifying common variables across the data sets and merge or join the datasets if this is needed.

Outlines the choice of organisational data architecture		
		Project Mapping
K12 Organisational data architecture.		For this project the data will be coming from an excel sheet stored in a SharePoint (SharePoint is used because it is a very secure platform,Microsoft strictly follows the stringent regulations like GDPR so they are trustworthy but on top of that they make use of SSL/TLS encryption to prevent man in the middle attacks and physical interception attacks on servers, they also make use of role based access controls to prevent internal attacks) and entered data in PowerBI however, I will be working from an offline copy of the asset management excel system to prevent any issues, The data will be collected from this excel sheet and transformed/loaded in PowerBI to assess the quality of the data and keep

	what is relevant to the project. I will have to model the relationships between data tables in the modelling view of PowerBI, I have to manage user access to this report and keep in contact with the stakeholder to ensure that the dashboard is in continuous improvement and moving towards the stakeholders needs. I have uploaded my dashboard into a workspace in Powerbi-service where I can manage the access to the dashboard, I will only provide access to the dashboard for the specific stakeholders that I converse with on this project and other LRT staff that they ask me to grant permissions to. Generally in the MHRA we store reports and dashboard in workspaces on powerbi-service so that we can have sections for all the different types of reports, for example the “Communications and engagement” reports will be separate from the “Finance” reports.

Communicates and collaborates with all relevant stakeholders and adapts communication style to meet audience and situational requirements	
	Project Mapping
S12 Collaborate and communicate with a range of internal and external stakeholders using appropriate styles and behaviours to suit the audience. [Figure 1] [Figure 2] [Figure 36] [Figure 26]	The first step I have taken to complete this is to identify the key stakeholders which in this case will be the Head of Laboratory resources and my managers. With each stakeholder I understand their expertise and level of understanding of the project and the tools I will use in order to gauge the appropriate communication methods, for example, the head of laboratory resources is not be an extremely technical stakeholder so I have to explain how to use a PowerBI dashboard, in terms of communication type, I stick to Microsoft Teams meetings, messages and emails as this is the communication standard in the MHRA. Communicating effectively with the

	key stakeholders will then enable me to collaborate with them, allowing for stakeholders and I to discuss ideas and the scope of the project.

Describes how they work independently and collaboratively detailing their impact on the work of others	
	Project Mapping
B3 Works independently and collaboratively. [Figure 1] [Figure 2]	In this project I have worked independently to ingest the data into Power BI, cleanse and prepare the data for suitable use, model the data to make sure the relevant data is all related and create a final dashboard design, However, I have also been working collaboratively with the head of Laboratory resources to gather requirements for the dashboard, collecting the data and refining the design of the dashboard as these tasks will require the input of the key stakeholder and our collaboration to make a final product suited for their needs and objectives.

Acts independently to establish logical and analytical solutions such as exploring new data sets or resolving issues within the data	
	Project Mapping
B4 Logical and analytical. [Figure 11] [Figure 12] [Figure 13] [Figure 14]	I have displayed logical and analytical behaviour by tackling this project in a structured timeline following the data analysis lifecycle with an agile methodology to promote continuous improvement and stakeholder involvement, I have produced a dashboard design with visualisations that are necessary to the stakeholder and make sense for the stakeholder as the main objective is to provide useful insights whilst also considering critical design principles like reducing cognitive load and making the interface as easy to use as possible.

Distinction

Evaluates the outcomes of data analysis and suggests alternative tools/methods which would be of benefit to all stakeholders.		
		Project Mapping
K8 Quality risks inherent in data and how to mitigate/resolve these. [Figure 25] [Figure 29] [Figure 30] [Figure 31] [Figure 36]		I will work hard to identify potential quality risks that are present in the dataset such as data incompleteness or inaccuracy / duplication, this will be followed up by a risk assessment in analyse the impact of these quality risks on the end product of my data analysis, depending on the risk I will develop mitigation strategies to handle these risks and potentially improve the data quality, this will be done through methods like data validation to identify and correct errors, establishing data standards or even cleansing the data itself. This will be followed by continuous monitoring to make sure the data quality is within the standard of what is expected.
K11 Approaches to organisational tools and methods for data analysis. [Figure 17] [Figure 27]		I understand that tools like PowerBI and Excel are widely used in the Data science department in the MHRA and I have been quite familiar with PowerBI, I will assess the tools used in my department and choose what tools will be most suitable for the project based on effectiveness and limitations, I will create POCs or proof of concepts using the identified tool and methods and demonstrate how these tool(s) can improve the data analysis outcomes and provide the necessary insights to the stakeholders.

B4 Logical and analytical. [Figure 17]	In assessing the results of data analysis, for the dashboard it's important that I approach the task with an analytical mindset. This includes evaluating how well the dashboard meets its goals, like providing staff with visibility on equipment dates. I can also use my abilities to detect any inconsistencies or irregularities in the data that may need exploration or refinement of analysis techniques. When recommending tools or methods that could benefit all parties involved I would take into account aspects like scalability, ease of use, compatibility with existing systems and potential cost savings. This might involve delving into analytics methods incorporating data sources or utilising other visualisation tools to enhance how effective and user friendly the dashboard is for employees in LRT

Analyses the requirements of the customer to produce a data analysis plan which provides an optimum solution	
	Project Mapping
S7 Undertake customer requirements analysis and implement findings in data analytics planning and outputs. [Figure 1] [Figure 2] [Figure 3] [Figure 11] [Figure 12] [Figure 13] [Figure 16] [Figure 26]	I have conducted meetings and discussions with the key stakeholders to identify the requirements whilst making notes of these requirements, this will serve as the reference for me to develop the optimum solution, With my information about the requirements I will be able to create a data analysis plan by asking for further information about where data is stored, what data is stored, how the LRT Asset management system works and what they're trying to achieve, this will help me further understand the requirements of the dashboard and give me the ability to start on the project. From here I will be able to observe the structure of the asset management system and plan my data analysis plan around how I will transform/load the data to make sure that I am using data that has integrity and is relevant, I can also prepare the data for visualisation which will be key for the dashboard. This

		will be taking place whilst being in contact with the stakeholders for feedback.

Demonstrates the use of data sets with different data structures and database designs to solve problems or improve the accuracy or efficiency of data analysis.		
		Project Mapping
S4 Analyse data sets taking account of different data structures and database designs. [Figure 18]		I have completed this skill by taking data from excel sheets that are stored locally and online (SharePoint) as well as creating data in PowerBI I have worked to identify and address potential analytical challenges with these datasets which will include different database designs and structures, I made plans, outlining my objectives, methodologies and tools to analyse the dataset whilst considering any sort of unique characteristics of the dataset. This will then allow me to prepare / transform the data where necessary, extract the insights from the data and feed back to the stakeholder.

Situation

Introduction

The Medicines and healthcare products regulatory agency (MHRA) comprises many divisions of departments, One of which is called “Science, Research and Innovations” and this division has a team called Laboratory Resources Team or LRT. One of the main jobs for employees in LRT is to maintain the records for laboratory equipment that needs to be serviced and this ranges from equipment like beakers to equipment considered to be “Financial assets” which means they are over [redacted] in value. This job is important not only to make sure that the team does not have compliance issues but also because if these equipment are not managed properly then they may be susceptible to breaking which is just a worst case scenario as LRT deals with a great volume of orders for lab equipment and maintenance, in fact since April 2023 they have completed [redacted] orders worth [redacted] million pounds.

Challenges for LRT

The problem is that LRT or the Laboratory resources team at the MHRA uses an order management system what heavily relies upon manual processes and the team holds all of their data on excel spreadsheets which is definitely a problem as it can make all of the data hard to understand as you are viewing a sheet of various numbers, it can also be a problem in terms of scalability, data integrity and automation. Within the LRT order management system it is very easy for LRT to make mistakes and miss out on equipment service due dates which can cost a lot of money and result in unnecessary purchases. It is also very difficult for LRT to manage open orders of equipment; this can result in extra purchases that can result in lost

funds, and this can influence the budget of LRT which subsequently can have an effect on projects that LRT invests in. The specific impact of these issues is that LRT will be less efficient as there will be inevitable human errors being made and this makes the work of LRT less effective as they won't be correctly managing the equipment of the laboratories and the orders coming in.

Opportunities

This is very important as the LRT deals with a great volume of orders for lab equipment and maintenance, in fact since April 2023 they have completed 3200 orders worth 5.5 million pounds. Building a dashboard to help LRT to define which orders need to be renewed or reviewed will both aid LRT with keeping track of their assets but it will also help the finance department with calculating in terms of understanding how much money is actually needed for maintenance and much more. As mentioned earlier, tracking orders will be incredibly relevant as it can prevent LRT spending funds unnecessarily which will be important when they will need to order equipment that they actually need which can be very expensive, if money is being wasted then it will be harder for LRT to pay for these equipment which subsequently makes it harder for laboratory staff to conduct the work they need to do for example on bacteria. If LRT continues with this standalone manual excel solution, the human errors that naturally occur can result in increased cost if an employee marks a piece of equipment down as needing more orders to the labs when there isn't an actual need for this equipment. Implementing a Power-bi solution which provides data visibility will improve efficiency as the dashboard will be automated in how it fetches data and it will provide insight on what equipment needs servicing or what orders are actually open without the stakeholders having to look at a confusing mess of data, this enables them to understand the situation more faster and more importantly, they will understand more of the data, allowing them to make more informed decisions.

Aims & Objectives

The main aim of this project was to develop and implement an effective solution to improve the Laboratory resources team's visibility of their data, I also aimed to develop a solution that will provide financial benefits or prevent unnecessary spending (Unnecessary spending is spending finances on equipment that is not

needed or spending on servicing equipment that have already been serviced, preventing this will present my opportunity to save money for LRT and I plan to use simple visualisation techniques to improve data visibility and prevent this unnecessary spending.) on licences/equipment for the Laboratory resources team and improve insights on the status of equipment orders e.g. whether equipment orders were still open or had closed. To successfully achieve these aims I must complete the following objectives:

- Objective 1 – My first objective is to design a dashboard to identify product licences that need to be maintained or managed for the current month. I will be able to quantify this objective by observing whether my dashboard accurately shows all the licences that are due, overdue, or completely outdated. A successful dashboard would provide insight into what product licences need to be renewed for the current month as well as the next, this will align with LRT's as well as MHRA's objective to work more "leaner" meaning to work more efficiently. I plan to achieve this in the final week of my project timeline which would be between the 1st of August to the 5th of August, giving me approximately 1 week to execute my visualisation concepts to Identify these Product licences.
- Objective 2 – My second objective is to design my dashboard whilst considering and implementing key concepts of user interface and UI/UX heuristics like "System and real-world matching" as well as "user control and freedom". The desired outcome of this objective is that the dashboard is easy to understand and utilise effectively so Laboratory resources team members do not make human error when it comes to managing equipment licences or orders and as a result LRT will save money which can potentially be used to fund expenditure on more equipment that will have an actual positive impact in other workflows in LRT like testing. I will be able to quantify this by effectively communicating with stakeholders regarding their opinion and preferences of the UI/UX as well as observing whether the dashboard is easy for them to understand and utilise effectively.I plan to achieve this in the final week of my of my project timeline which would be between the 1st of August to the 5th of August by allowing the stakeholders to test my dashboard and observing the results of using my dashboard in terms of how much money or time has been saved.
- Objective 3 – My third objective is to Provide useful Insights around the data on the order management side to help the stakeholder easily understand what orders open and what orders are closed as well as quantifying these categories to give the stakeholders a better overview of all the orders, I will be able to quantify this objective by observing how my the orders side of my dashboard

has affected the workflow of LRT in asset management and by gathering feedback by stakeholders on the dashboard. I plan to deliver on this objective in the final week of my of my project timeline which would be between the 1st of August to the 5th of August by handing over my dashboard to the stakeholders for testing and feedback, this will allow the stakeholders to view their data in a visualised manner, providing them with insight into what the current overview is with equipment and licences.

Task

My role as a Data analyst means I am responsible for the sourcing of appropriate data, cleaning and transforming this data to include only what is needed, modelling relationships to relate data from different data sources and visualising this data to give insights. However, focussing on this project my role was to create a solution (for the stakeholders at the Laboratory resources team) to the lack of data visibility around the LRT's equipment servicing and equipment orders. I chose to take an Agile methodology as I will be able to continuously improve the report and gather feedback from the stakeholder to make sure the insights I am providing are meaningful. This will also help me deliver a higher quality report and avoid any risks of developing a report that is not what the stakeholder wants. For this project I chose to use the following tools which can be seen in Figure 4.

Action

Identifying problems & understanding business requirements

The problem is that LRT or the Laboratory resources team at the MHRA uses an order management system what heavily relies upon manual processes and the team holds all of their data on excel spreadsheets which is definitely a problem as it can make all of the data hard to understand as you are viewing a sheet of various numbers, it can also be a problem in terms of scalability, data integrity and automation. Within the LRT order management system (which is currently just a spreadsheet) there is a column to show when certain pieces of lab equipment are about to expire, and the LRT need to be on track to renewing these pieces of equipment however, it is very easy for LRT to make mistakes and miss out on these due dates which can cost a lot of money and result in unnecessary purchases

To understand the requirements of this project I made use of Microsoft Teams to contact the stakeholders about setting up meeting to talk about what's needed for this dashboard as well as any questions I had, This can be seen in Figure 1 and Figure 2, Figure 3 even show some of the notes i typed out during meeting to remind myself of requirements.

Figure 1 refers to the LRT meeting catchup invite, this is used to discuss the dashboard regarding any progress and improvements that could be made.
Figure 2 refers to some conversation between me and one stakeholder from LRT about the requirements for the dashboard.
Figure 3 refers to personal notes I made to myself on MS Teams in the meeting to discuss the requirements for the LRT dashboard.

Data collection

LRT has chosen to store all of their records for equipment service dates in an excel sheet which is stored in a SharePoint site (Figure 7), this is what I used to connect to the data in PowerBI as this program has a SharePoint connector (Figure 6).I decided to focus on the service maintenance portion of this project first as the data was more complete than the orders spreadsheet however this was provided to me via Microsoft

teams so I could connect to the equipment orders dataset via the excel workbook in PowerBI.

Figure 6 refers to the data source settings for the dashboard
Figure 7 refers to the actual SharePoint source for the dashboard

Data processing

To cleanse and transform the data to meet stakeholder requirements, I removed rows of null data which have been caused by the way LRT employees have structured the original excel sheet (They left blank space at the top of the workbook), I had to promote the headers to get the actual columns to be recognised by powerBI, if this is not done then the column names will be recognised as part of the raw data itself, I had removed a lot of null columns in each query to keep a focus on the columns needed and I filtered the existing Service due date to exclude null values as any rows where service due date is null would not provide significant data.

I had also created two custom columns, one being a alteration of the service due date which is created for potential future needs and another to display whether the current service due date is “OUTDATED”, “OVERDUE”, “DUE” or “DUE LATER” and this was the major requirement for this portion of the project.I later changed the data types for these custom columns to date and text data types.I have included a before and after to display the difference of these applied steps in power query editor (Figure 9),(Figure 10).

Figure 9 refers to the data before cleansing and transformation
Figure 10 refers to the data after cleansing and transformation

Data analysis

I created a page to display a data overview (Figure 13),(Figure 14), this shows a breakdown of all the equipment in all categories and their service status, this will allow the stakeholders to have an instantaneous view of the current situation around equipment maintenance in every category.Figure 14 shows three boxes at the top of the screen for Landing page, page 1 and page 2. There are functioning buttons I created to aid the stakeholder in navigating the pages of my dashboard, the landing page redirects the user to the homepage (Figure 11) which allows them to navigate to other pages easily, the page 1 and page 2 buttons switch between views in the data overview page.

For each of the queries (Figure 8) I created a page to display the equipment and their service status, I included a bar chart to show an overview on the number of equipment that is labelled as “DUE”, “DUE LATER”, “OUTDATED” and “OVERDUE”. I

had also included a slicer so the user can filter by service status, an example of this page has been included for the Miscellaneous equipment tab (Figure 12).

The orders side of this project was much more simple as I had just created a page to show the open orders from the year 2023 to 2024 (Figure 19), this should show all of the orders that have not been closed off in the specified years, I had also created a page to track all the orders from the year 2024 to 2025 (Figure 20), Finally this was followed up with a data overview page just like the equipment side.

When developing my dashboards I had considered key UX or User experience principles like usability, performance, consistency, customisation and effective visualisation **(Faisal, 2024)**. In an effort to create a user centric design, I considered usability by designing my interface to be easy to use in order to minimise the learning curve and maximise the users productivity, this was done by making use of a very simple layout to clearly display a table or visualisation alongside their filter options.

I considered performance or feedback/interaction by implementing custom buttons in my dashboard, allowing the user to easily switch between important pages without having to waste time scrolling through the tabs in PowerBI, this facilitates a smooth flow of interaction, the filters i have made use of are also incredibly responsive and this was done to prevent user frustration when using the dashboard but also this will increase user productivity and will make the users more willing to learn how to use the dashboard.

I implemented consistency within my design as I maintained the same simple layout within many of the pages of my dashboards, these pages all resemble Figure 12 and this was done to aid the user to feel more comfortable with the dashboards and help them tackle the learning curve of using the dashboard. I was able to do this by copying pages and reformatting visualisations to suit different types of equipment or orders of equipment.

Finally I considered and implemented both customisation alongside the most effective visualisations, I gave the users the power to customise the visualisations with filters to allow them to suit the visuals to their needs, allowing them to have the opportunity to gain much deeper insight into the LRT data. I made the choice that making use of bar charts and column charts would be the most effective visuals to use as they are incredibly simple to understand, aiding in reducing cognitive load but also it is suitable as i have used these visuals to compare categorical context e.g. the equipment that have services that are due now, due later or overdue. The implementation of these principles can be seen in Figure 12 and Figure 28.

Figure 8 refers to a list of queries for the Equipment service management portion of LRT Dashboard
Figure 11 refers to the LRT Dashboard Landing Page

Figure 12 refers to the equipment service management page for the Miscellaneous Laboratory equipment
Figure 13 refers to the data overview page which had been requested later on in the project by the stakeholders
Figure 14 refers to the other side of the data overview page related to figure 13
Figure 19 refers to the Open Orders for the years 2023 to 2024 page
Figure 20 refers to the Order tracking for orders within 2024 to 2025
Figure 28 refers to the Visualisation comparison for the LRT Dashboards

Data modelling

I have created a new table in PowerBI to hold all of the service statuses, I have created a many to one relationship from all of the main queries to this service status table, this way i have a clever way of filtering all of the data to be fixed where the equipment is “DUE”, “DUE LATER”, “OVERDUE” and “OUTDATED” in terms of equipment maintenance. This can be seen in Figure 24.

Figure 24 refers to the ERD / Entity relation diagram

Deployment

When first developing the LRT dashboard I was working of a local offline copy of the data but now I have established a live connection to the data, However the user will still need to press the refresh button in PowerBI to see the most up to date data, to get rid of this manual process and introduce automation into this project I deployed the dashboard into PowerBI Service (Figure 21), this allows me and the users to have a online version of the report in case PowerBI desktop is not available, this also allows me to tweak some advanced settings like setting up a refresh schedule which can be seen in Figure 22.Figure 22 shows that I have setup the dashboard to refresh every morning at 9 am so that the users are getting the most up to date data being displayed to them in the visualisations, this pairs quite nicely with the refresh visualisations that can be seen in Figure 11 as this reconfirms that the user is viewing the most recent data.

Figure 11 refers to the LRT Dashboard Landing Page
Figure 21 refers to the deployment to PowerBI Service / Online and accessible to users.
Figure 22 refers to the Refresh setting suggesting live connection and up to date data.

Result / Review

The result of all this work on this dashboard is an effective solution to improve the Laboratory resources team's visibility of their data and a solution that will provide financial benefits or prevent unnecessary spending on equipment or servicing.

I have designed a dashboard to identify product licences that need to be maintained or managed for the current month and this is quantified by the custom column i had created called "Service Status" which can be seen in Figure 12.This has been extremely successful as the custom column allows the user to instantly observe what equipment is due for servicing, what equipment needs to be serviced next month and equipment that is overdue or completely outdated when it comes to servicing.I had even implemented some conditional formatting to make it much easier to read what the status is for equipment and with the smooth navigation system i have implemented (Figure 12, Figure 13, Figure 14) in the form of navigation buttons that are labelled and react when the user hovers over them with the mouse, it only makes it easier for the user to see the current status of equipment that need to be maintained.

Figure 12 refers to the equipment service management page for the Miscellaneous Laboratory equipment
Figure 13 refers to the data overview page which had been requested later on in the project by the stakeholders
Figure 14 refers to the other side of the data overview page related to figure 13

Appendix

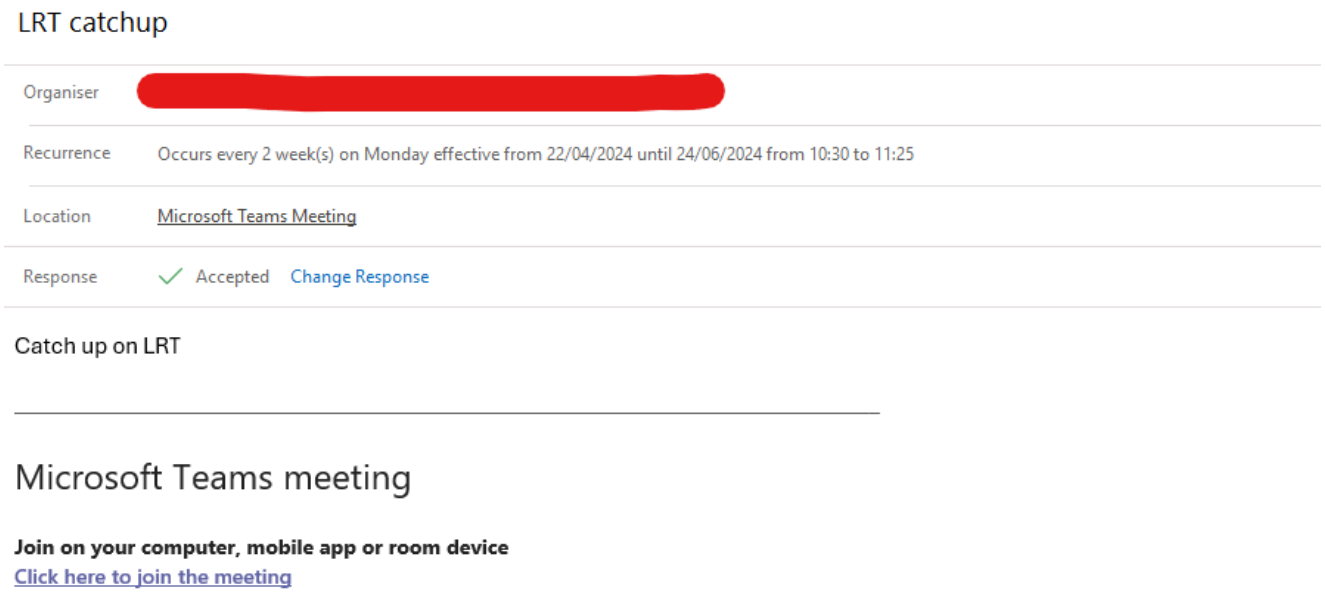


Figure 1 - LRT meeting catchup invite, this is used to discuss the dashboard regarding any progress and improvements that could be made.

Figure 2 - This is some conversation between me and one stakeholder from LRT about the requirements for the dashboard.

Figure 3 - Some personal notes I made to myself on MS Teams in the meeting to discuss the requirements for the LRT dashboard.

Tool Used	What it was used for	Reason for selection
PowerBI	import the data, transform and cleanse the data, create visualisations and a dashboard to provide insights	This is a widely used tool within the agency and it is very intuitive which makes it easy for a beginner to create useful dashboards
PowerBIService	Used to demonstrate the report to stakeholders and give them an opportunity to handle the dashboard themselves as well as setting refresh schedules	This is a widely used tool within the agency and it is very intuitive which makes it easy for a beginner to create useful dashboards
SharePoint	Used to store data in a shared location so multiple people can access the same data at once	It is a very secure tool and is easy to use, it also has its very on connector in PowerBI desktop making it a very compatible tool for PowerBI
Microsoft Teams	Communication and meetings	Another heavily used tool in the MHRA but it's very effective when trying to get a message across very quickly which makes communication around the project very easy.

Figure 4 - Tools used to create this LRT Dashboard

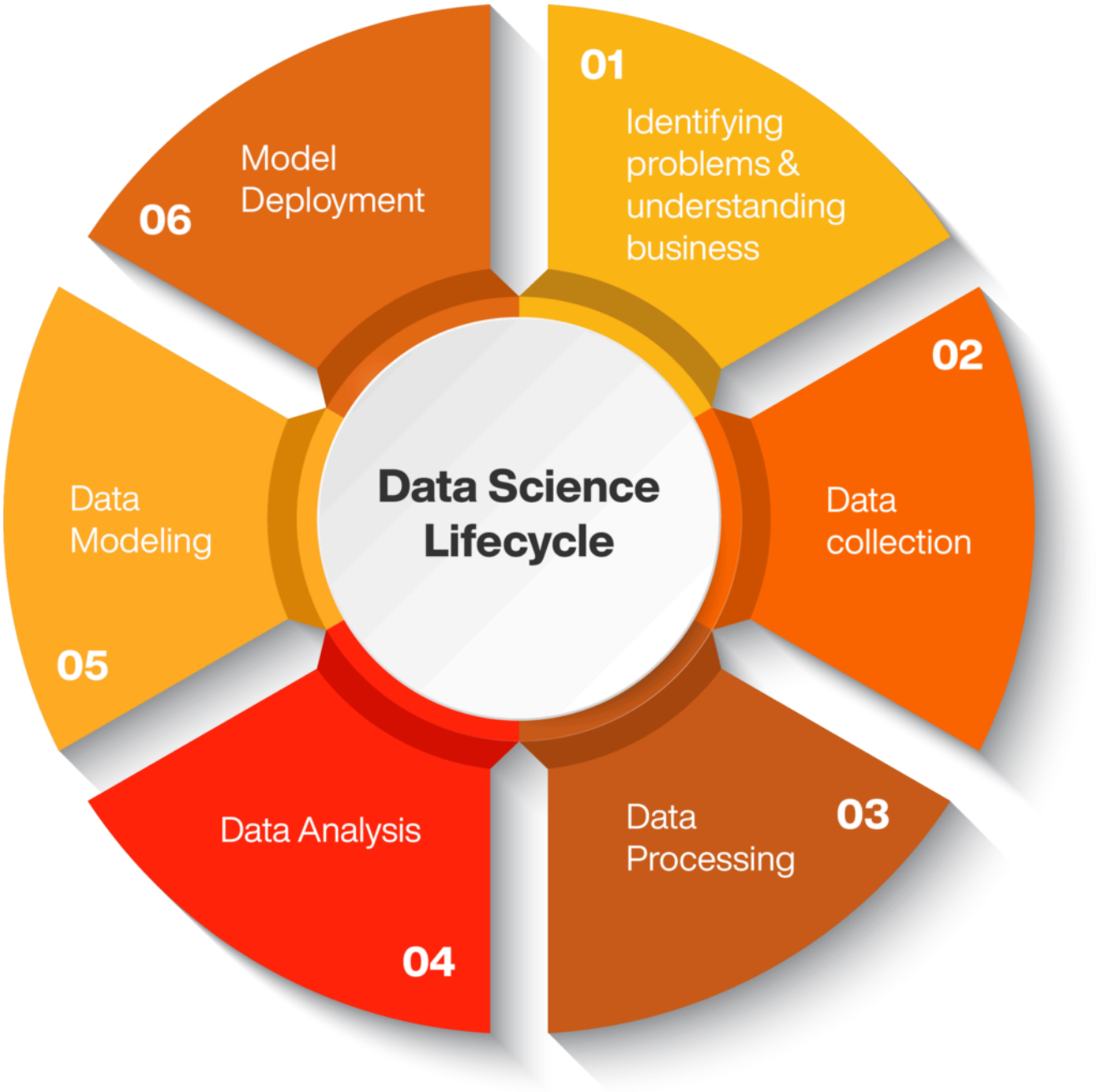


Figure 5 - Data analysis lifecycle **(onlinemanipaleditorialteam, 2022)**

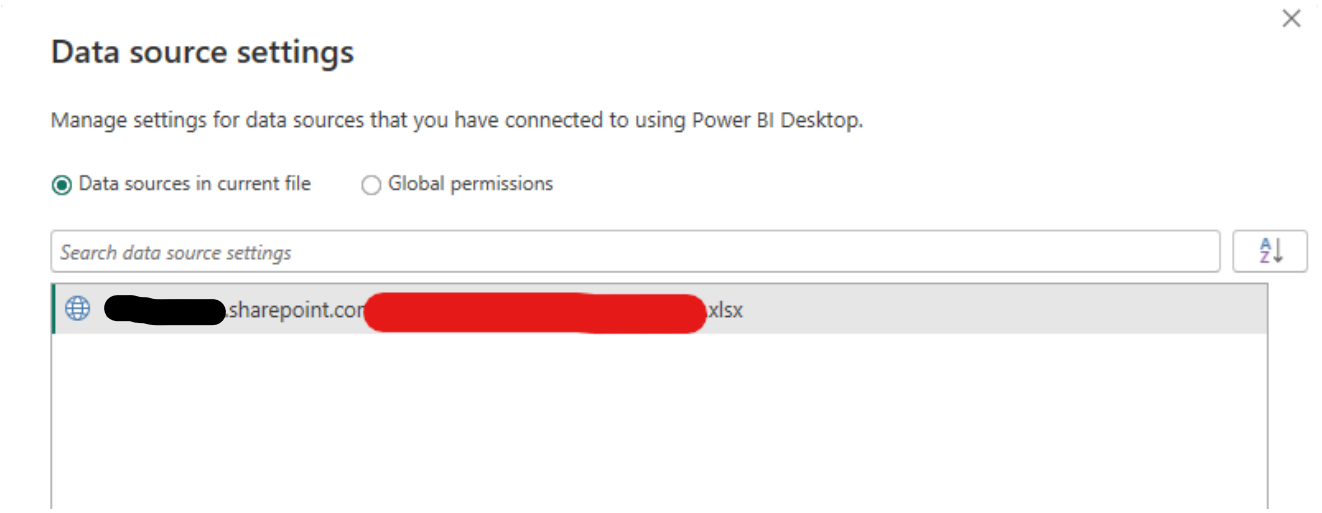


Figure 6 - Data source setting for Dashboard

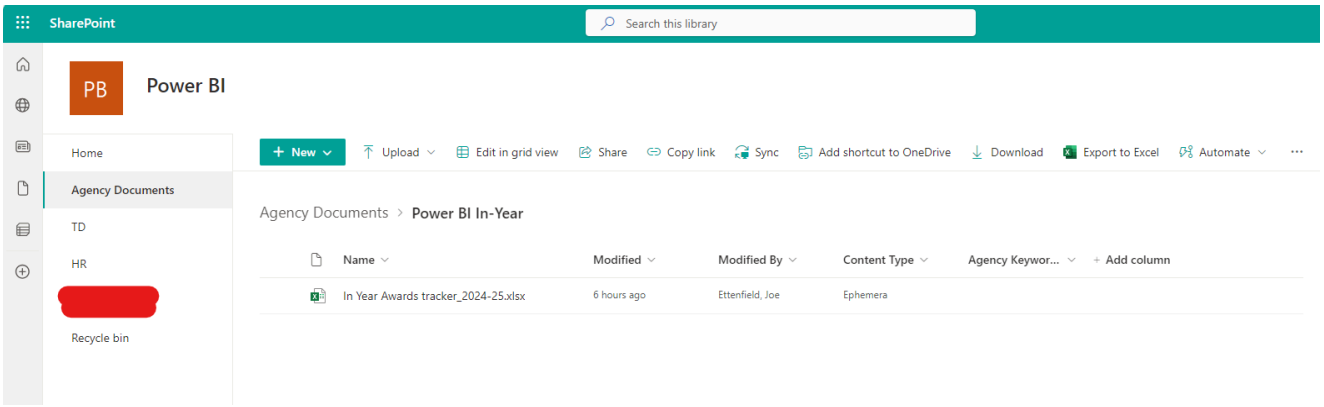


Figure 7 - Actual SharePoint source for dashboard

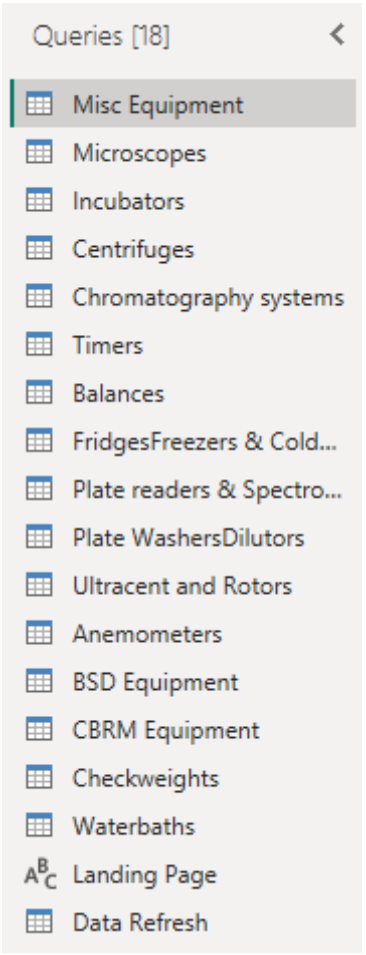


Figure 8 - List of queries for the Equipment service management portion of LRT Dashboard

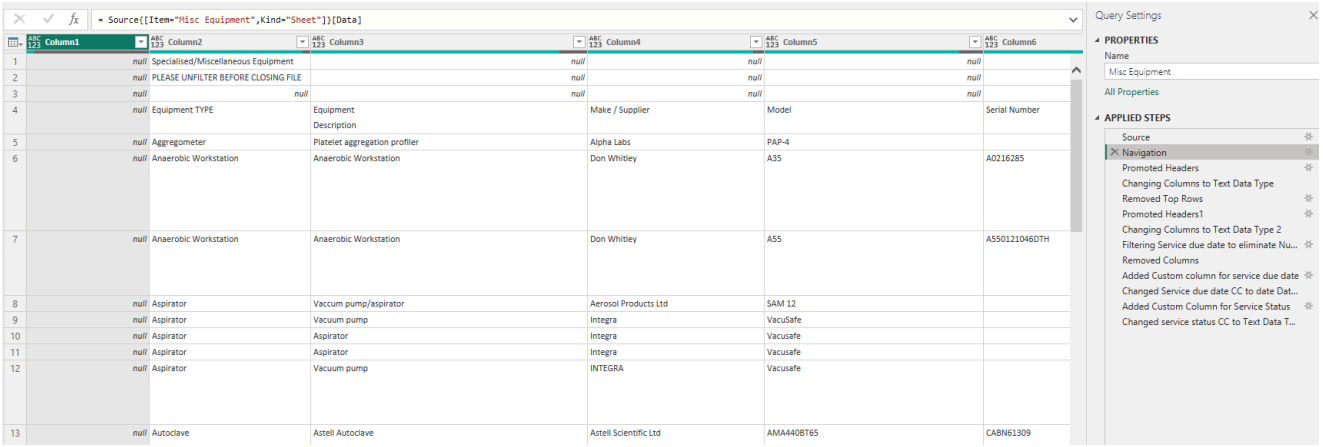
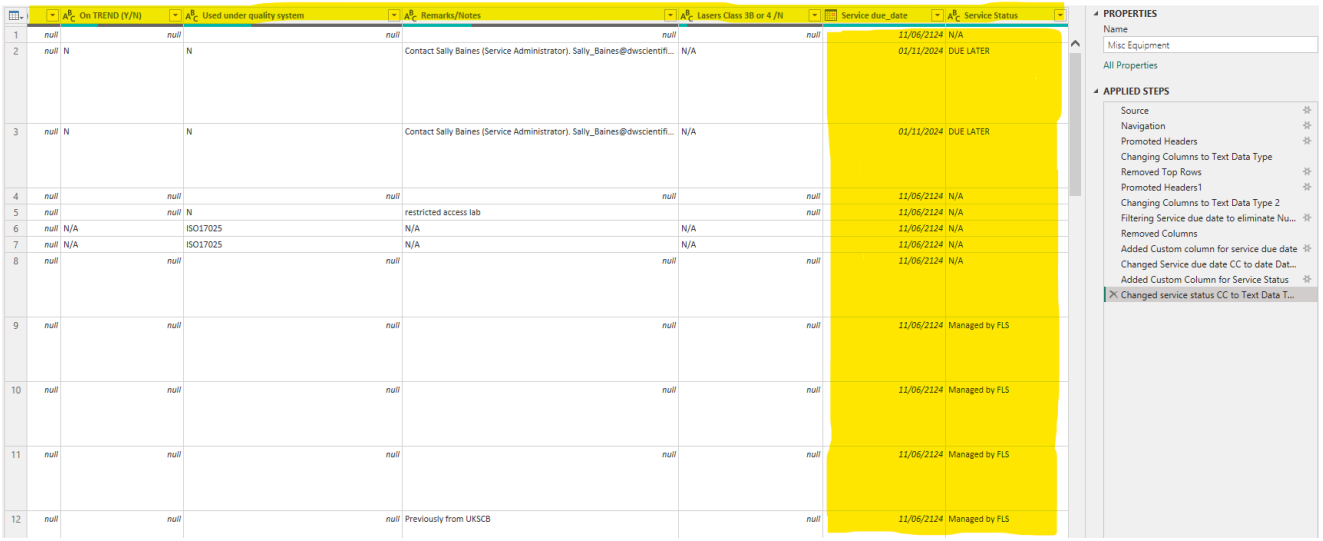


Figure 9 - Data before cleansing and transformation - You can see that there are null values and the columns are not named right, this is because of the way the excel sheet is formatted.



	On TREND (Y/N)	Used under quality system	Remarks/Notes	Lasers Class 3B or 4 /N	Service due_date	Service Status
1	null	null	null	null	11/06/2124	N/A
2	null	N	N	Contact Sally Baines (Service Administrator). Sally_Baines@dwscientifi...	N/A	01/11/2024 DUE LATER
3	null	N	N	Contact Sally Baines (Service Administrator). Sally_Baines@dwscientifi...	N/A	01/11/2024 DUE LATER
4	null	null	null	null	null	11/06/2124 N/A
5	null	null	N	restricted access lab	null	11/06/2124 N/A
6	null	N/A	ISO17025	N/A	N/A	11/06/2124 N/A
7	null	N/A	ISO17025	N/A	N/A	11/06/2124 N/A
8	null	null	null	null	null	11/06/2124 N/A
9	null	null	null	null	null	11/06/2124 Managed by FLS
10	null	null	null	null	null	11/06/2124 Managed by FLS
11	null	null	null	null	null	11/06/2124 Managed by FLS
12	null	null	null	Previously from UKSCB	null	11/06/2124 Managed by FLS

PROPERTIES
Name
Misc Equipment
All Properties

APPLIED STEPS
Source
Navigation
Promoted Headers
Changing Columns to Text Data Type
Removed Top Rows
Promoted Headers1
Changing Columns to Text Data Type 2
Filtering Service due date to eliminate Nu...
Removed Columns
Added Custom column for service due date
Changed Service due date CC to date Dat...
Added Custom Column for Service Status
Changed service status CC to Text Data T...

Figure 10 - Data after cleansing and transformation - You can see the addition of custom columns as well as the column names changing in the highlighted area, there has also been the removal of columns and changing of data types. This sort of data processing/cleaning/transformation has been carried out for the rest of the queries in Figure 9

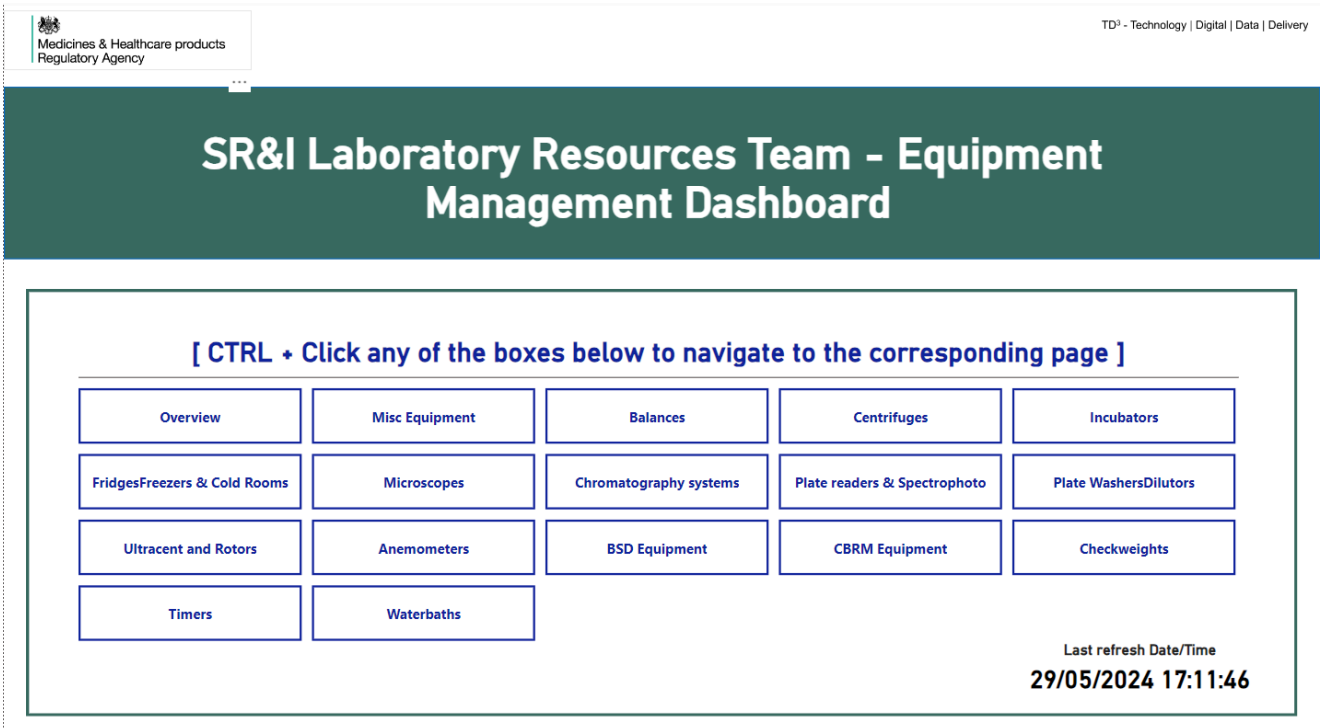
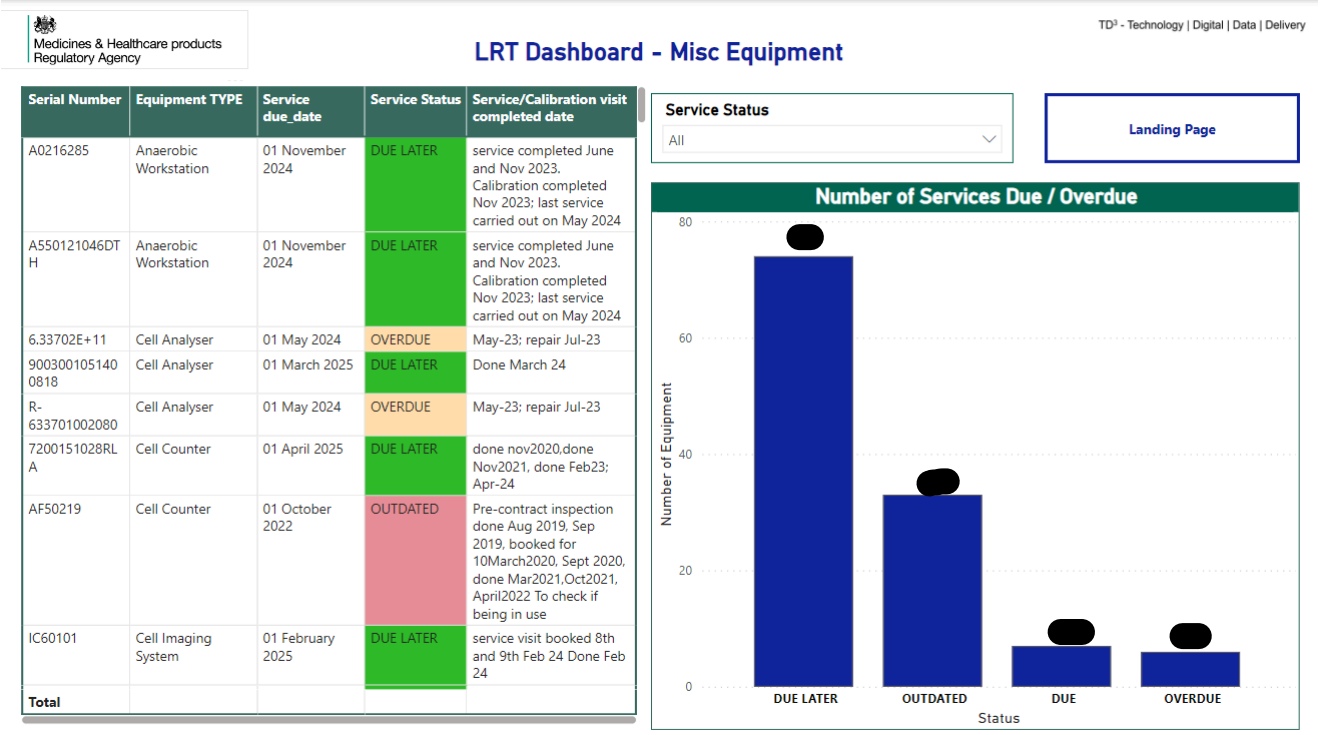


Figure 11 - LRT Dashboard Landing Page



Serial Number	Equipment TYPE	Service due_date	Service Status	Service/Calibration visit completed date
A0216285	Anaerobic Workstation	01 November 2024	DUE LATER	service completed June and Nov 2023. Calibration completed Nov 2023; last service carried out on May 2024
A550121046DTH	Anaerobic Workstation	01 November 2024	DUE LATER	service completed June and Nov 2023. Calibration completed Nov 2023; last service carried out on May 2024
6.33702E+11	Cell Analyser	01 May 2024	OVERDUE	May-23; repair Jul-23
900300105140 0818	Cell Analyser	01 March 2025	DUE LATER	Done March 24
R-633701002080	Cell Analyser	01 May 2024	OVERDUE	May-23; repair Jul-23
7200151028RLA	Cell Counter	01 April 2025	DUE LATER	done nov2020,done Nov2021, done Feb23; Apr-24
AF50219	Cell Counter	01 October 2022	OUTDATED	Pre-contract inspection done Aug 2019, Sep 2019, booked for 10March2020, Sept 2020, done Mar2021,Oct2021, April2022 To check if being in use
IC60101	Cell Imaging System	01 February 2025	DUE LATER	service visit booked 8th and 9th Feb 24 Done Feb 24
Total				

Figure 12 - Equipment service management page for the Miscellaneous Laboratory equipment

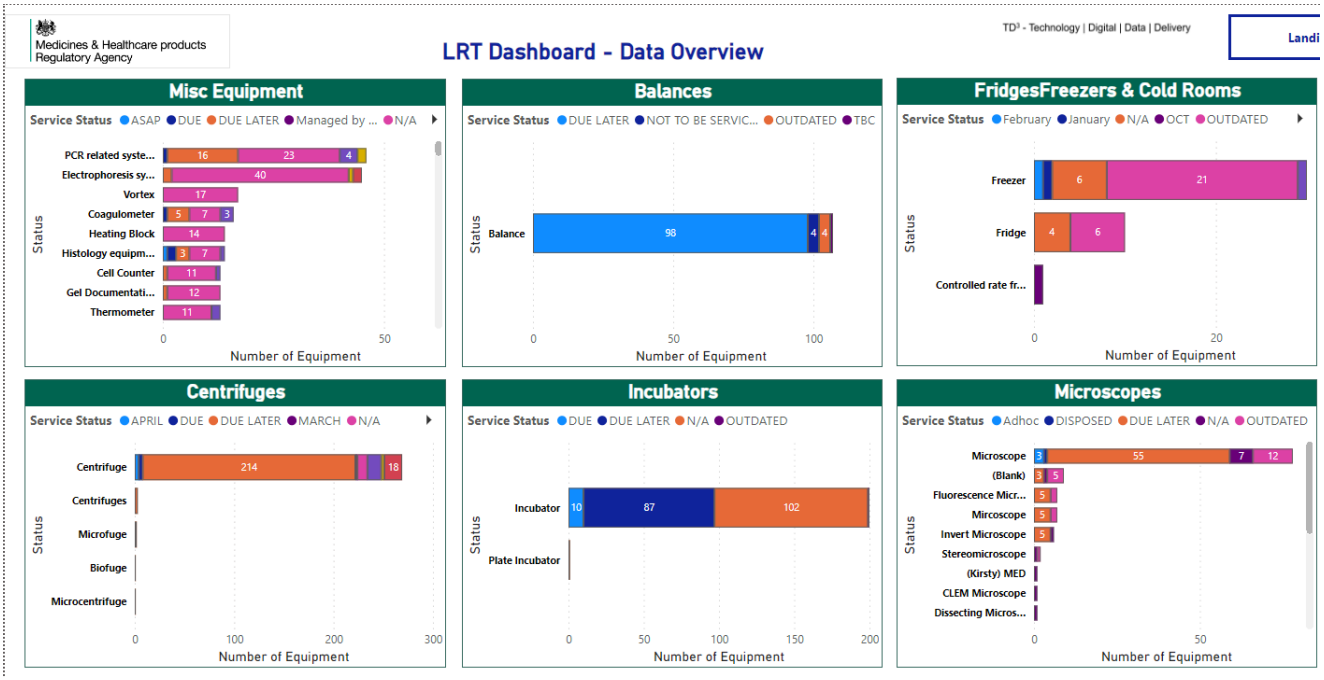


Figure 13 - Data overview page which had been requested later on in the project by the stakeholders

Figure 14 - The other side of the data overview page related to figure 13

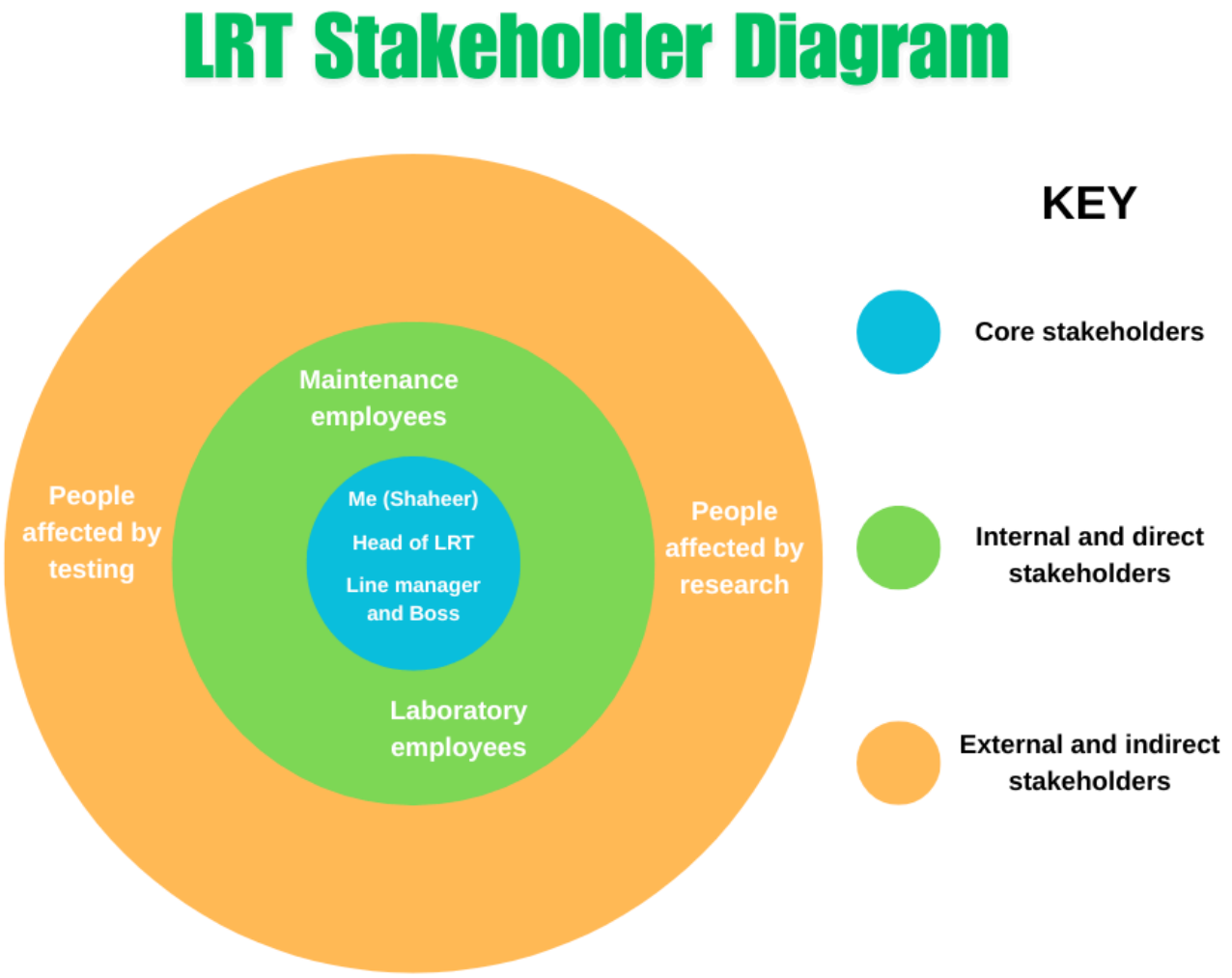


Figure 15 - Stakeholder diagram used to show who this dashboard and project involves and effects.

LRT Equipment maintenance dashboard stakeholder requirements

1. Display equipment and whether their maintenance status

- Create a custom column for the service status
- Potentially some sort of conditional formatting to reduce cognitive load
- Include service calibration date as requested

2. Show an overview of the amount of equipment in each status

- I will have to create a Bar chart to use the Service status custom column
- Potentially use a count of serial number
- Format visualisation

3. Create a way to filter the data by the status

- Make use of a slicer visual but this will have to be done after the Service status custom column is created
- Dropdown menu

3. Create a way to filter the data by the status

- Make use of a slicer visual but this will have to be done after the Service status custom column is created
- Dropdown menu

4. Create an overview of all of the equipment categories

- Once Service status custom column is created, I can generate a graph for each equipment type and make an overview page
- Will need navigation buttons to make the flow of the dashboard more smooth

5. Automated refresh of dashboard

- Redirect the dashboard to point to a live source of data after development with a local source of data
- Upload dashboard to PowerBI Service
- Create a refresh schedule so that the dashboard is up to date
- Monitor the dashboard for any refresh schedule errors

5. Automated refresh of dashboard

- Redirect the dashboard to point to a live source of data after development with a local source of data
- Upload dashboard to PowerBI Service
- Create a refresh schedule so that the dashboard is up to date
- Monitor the dashboard for any refresh schedule errors

Requirements from me:

- **Smooth navigation from page to page using custom buttons**
- **Refresh visual to show how up to date the dashboard is**
- **Clear and consistent format throughout dashboard**
- **Consistent monitoring**
- **Consistent and clear communication with stakeholders**

Figure 16 - Stakeholder requirement analysis document - Listed out main criteria and bullet pointed what I would have to do in order to achieve these requirements

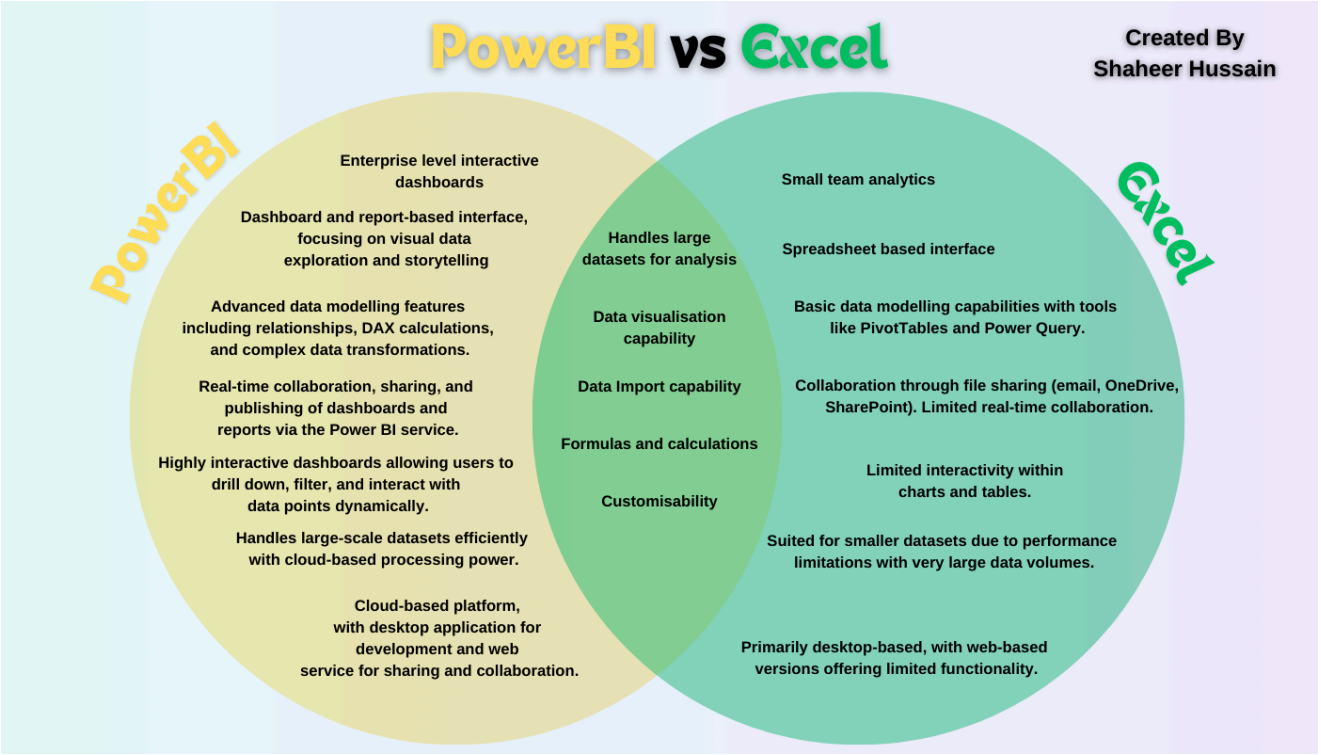


Figure 17 - Tool comparison for LRT project - Powerbi vs Excel **(www.datacamp.com, n.d.)**

Risk / Challenge	Description
Data Inconsistency	Inconsistent data formats, missing values, or mismatched columns between Excel sheets can lead to errors.
Duplicate Data	Combining sheets might result in duplicate entries, skewing analysis results.
Performance Issues	Large data sets from multiple sheets can slow down Power BI performance.
Complex Data Transformations	Complex merging and transformation processes can be difficult to manage and maintain.
Data Refresh Problems	Keeping data updated across all combined sources can be challenging, especially if the Excel sheets are updated frequently.
Schema Changes	Changes in the structure of any Excel sheet (e.g., added or removed columns) can break the data model.
Error Propagation	Errors in the original Excel sheets can propagate through the combined data, leading to incorrect analysis.
Access and Permissions	Ensuring that all users have the necessary permissions to access and refresh the data can be problematic.
Data Security	Sensitive information might be exposed if proper security measures are not implemented during data integration.
Integration Complexity	Integrating data from Excel with other data sources in Power BI can be complex and error-prone.
Version Control Issues	Managing different versions of Excel files can lead to confusion and data discrepancies.
User Training and Expertise	Users may require additional training to handle complex data integration tasks effectively.
Dependency on Excel	Relying heavily on Excel for data sources can limit scalability and flexibility of the Power BI solutions.
Data Validation	Ensuring the combined data's accuracy and reliability requires rigorous validation and testing.
Maintenance Overhead	Regular maintenance is needed to keep the combined data sources up-to-date and error-free.

Figure 18 - Data source combination risks and challenges for data analysis activity.**(Ziad, 2017)**

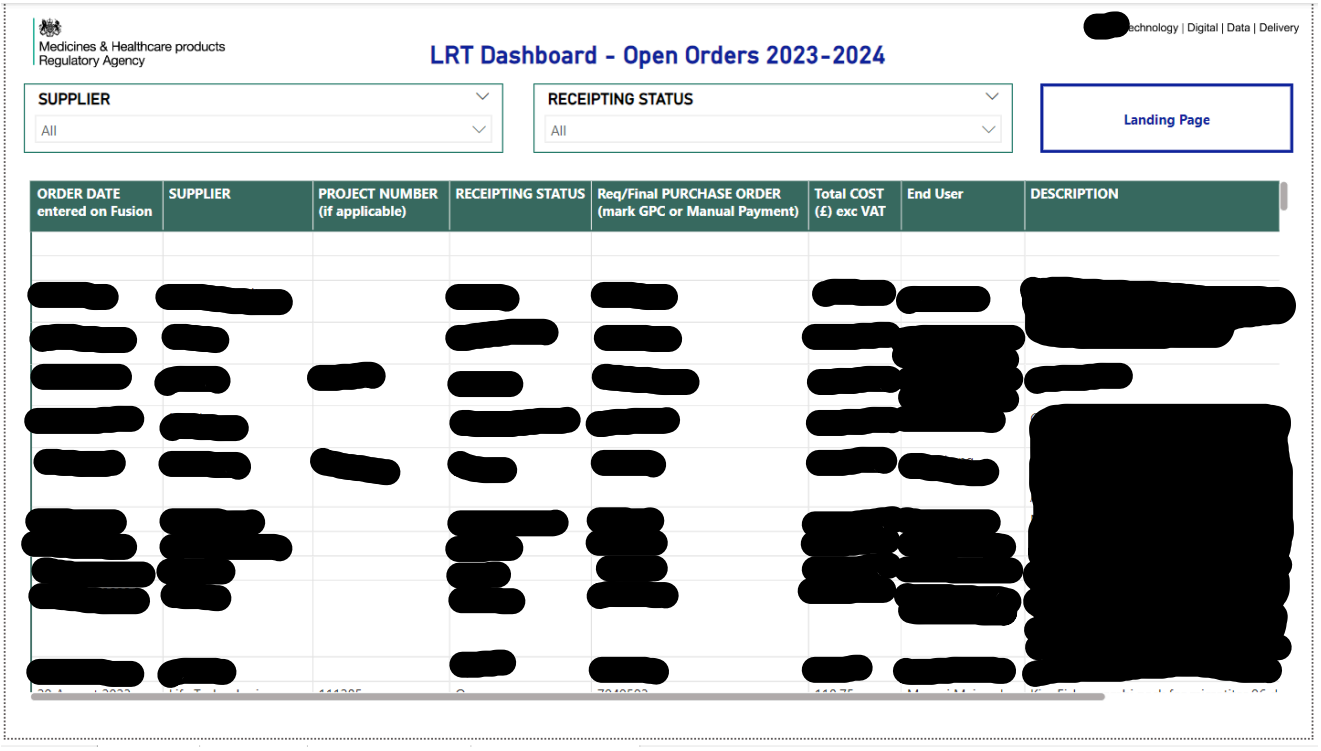


Figure 19 - Open Orders for the years 2023 to 2024 page

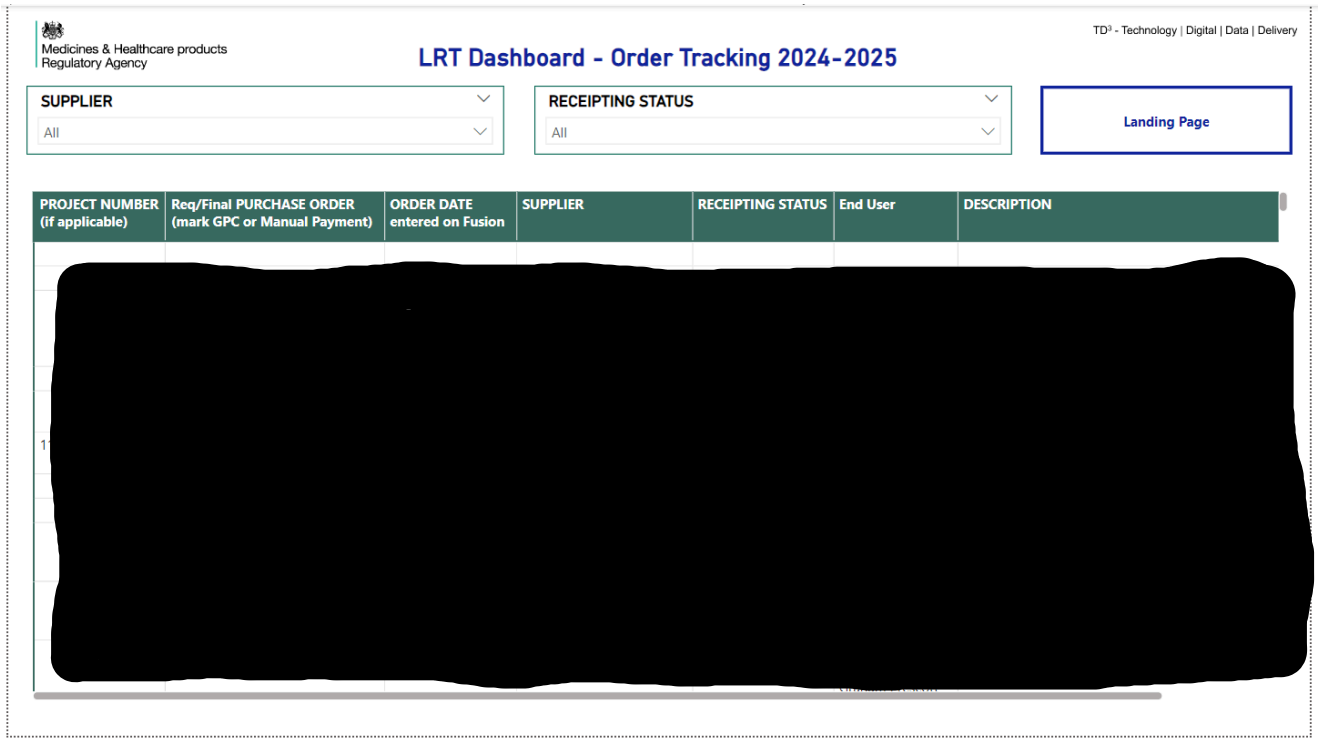


Figure 20 - Order tracking for orders within 2024 to 2025

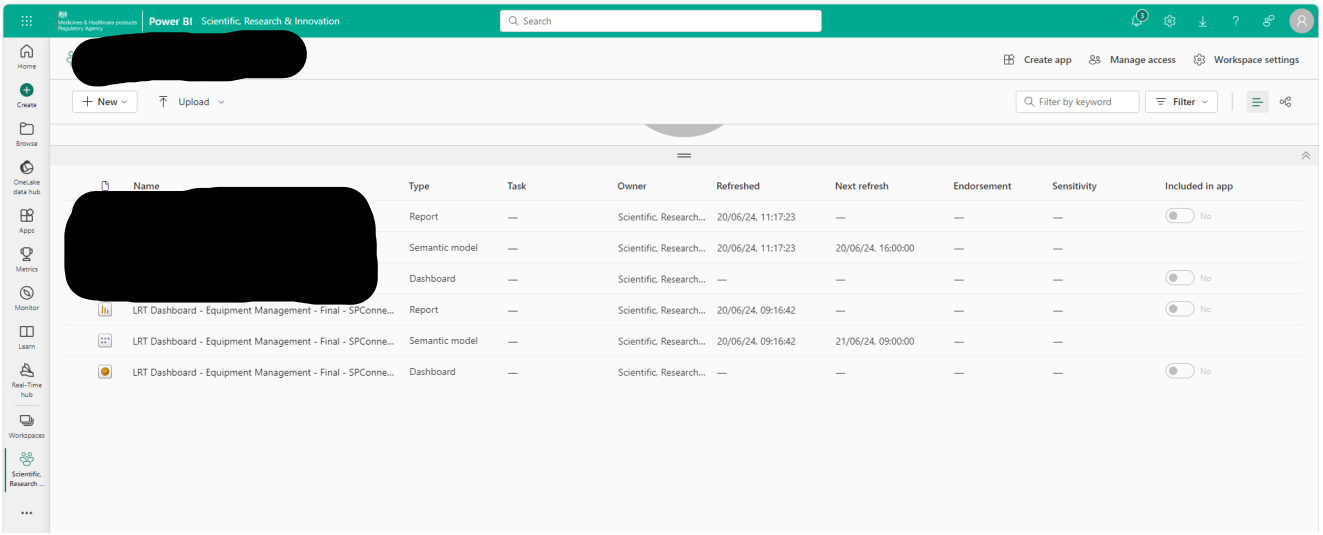


Figure 21 - Deployment to PowerBI Service / Online and accessible to users.

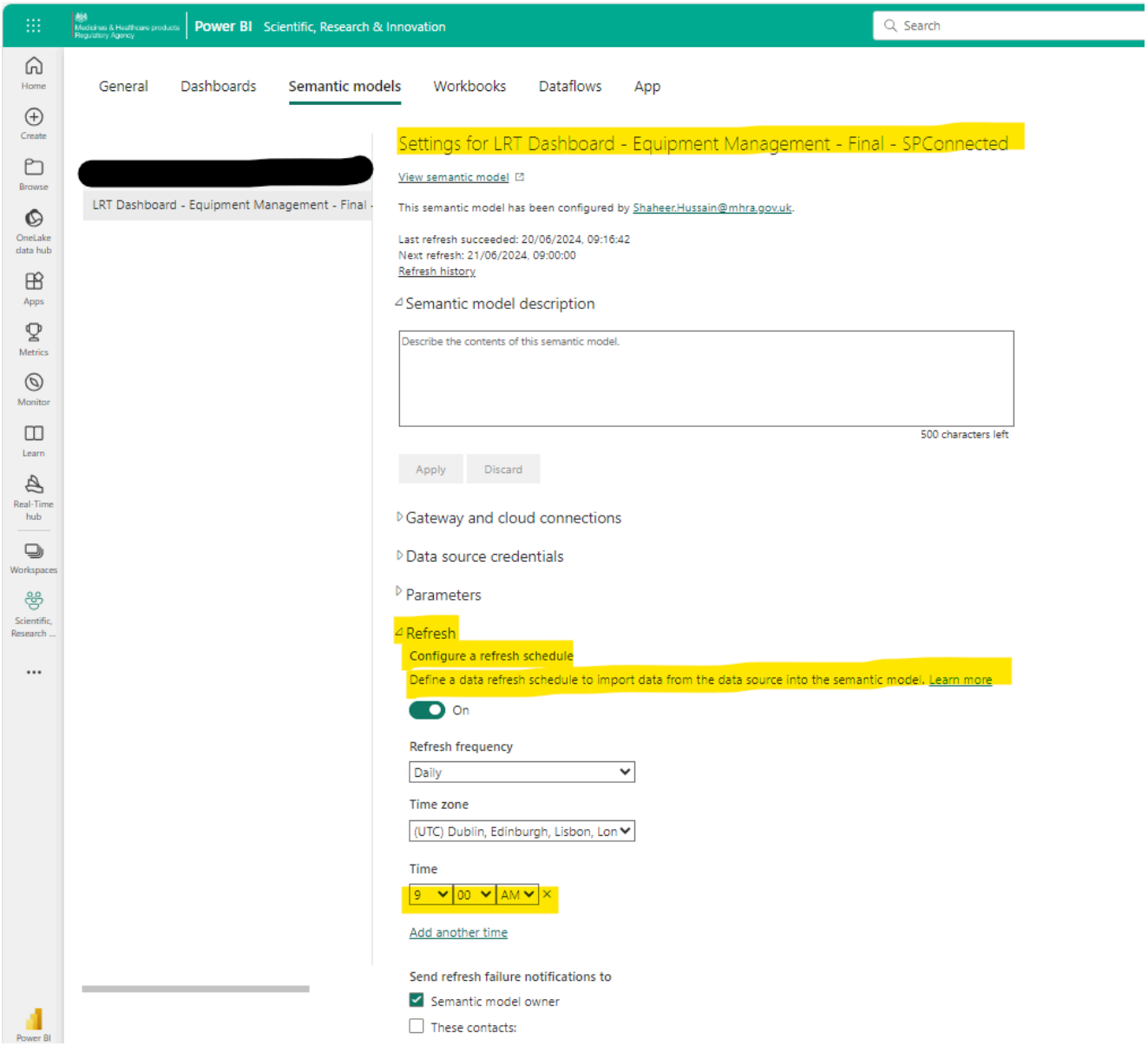


Figure 22 - Refresh setting suggesting live connection and up to date data.

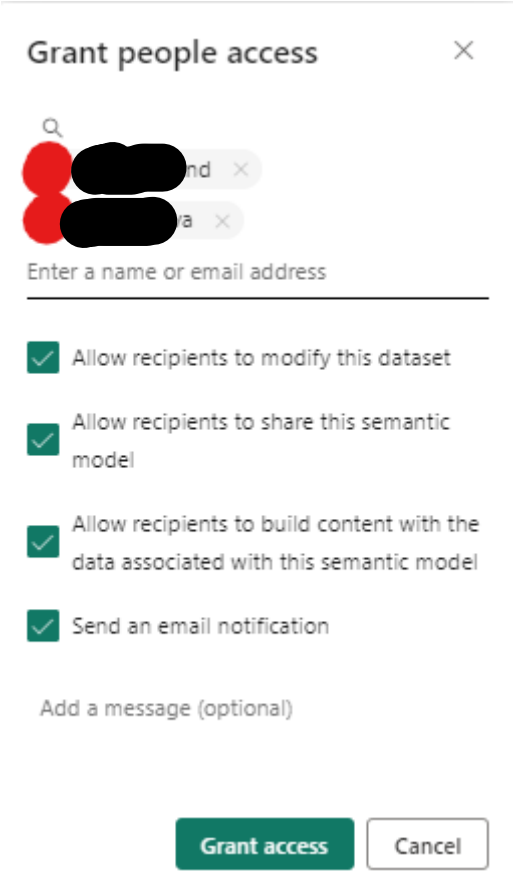


Figure 23 - Granting access to only the stakeholders to prevent any unauthorised access to the dashboard.

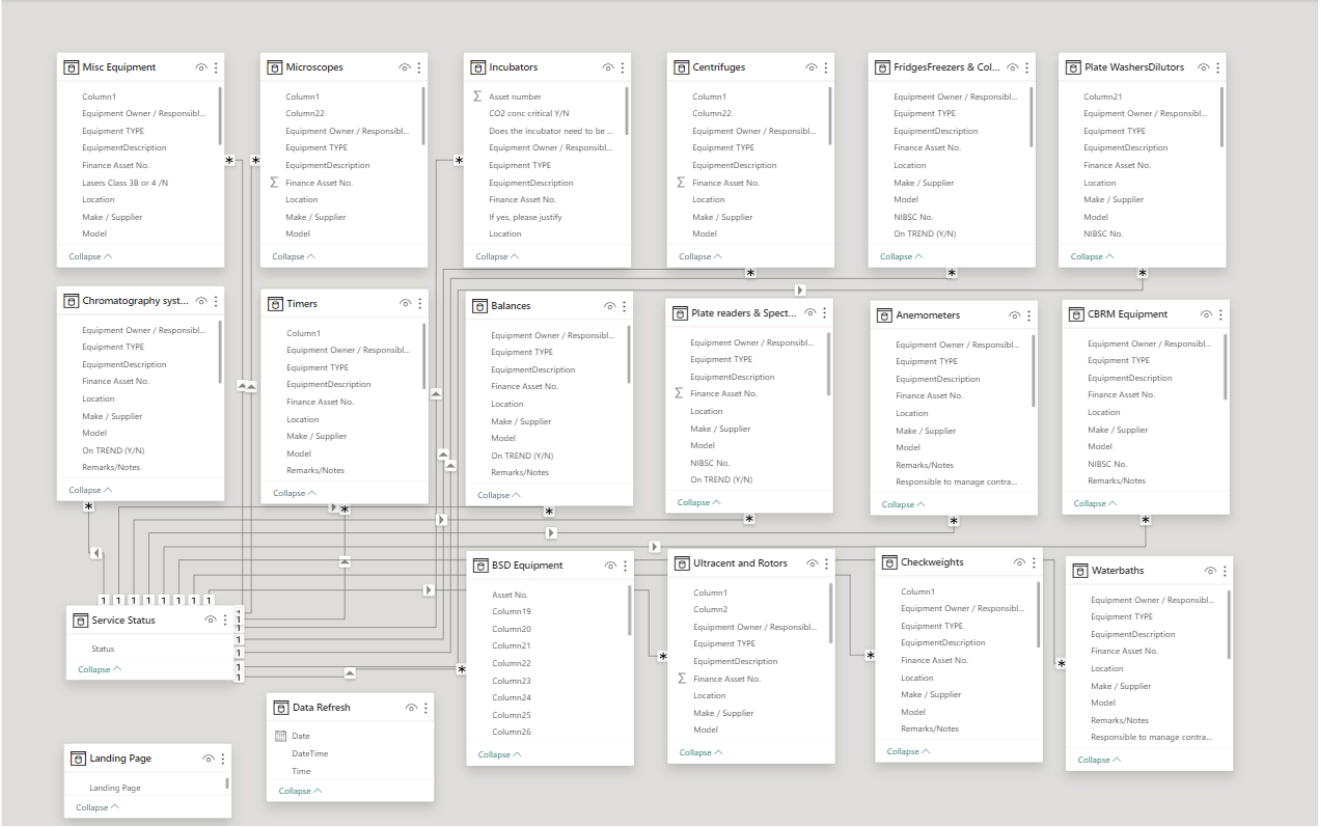


Figure 24 - ERD / Entity relation diagram to show the relationship between the main queries which originate from an excel file in sharepoint to a status table which i manually created in power query

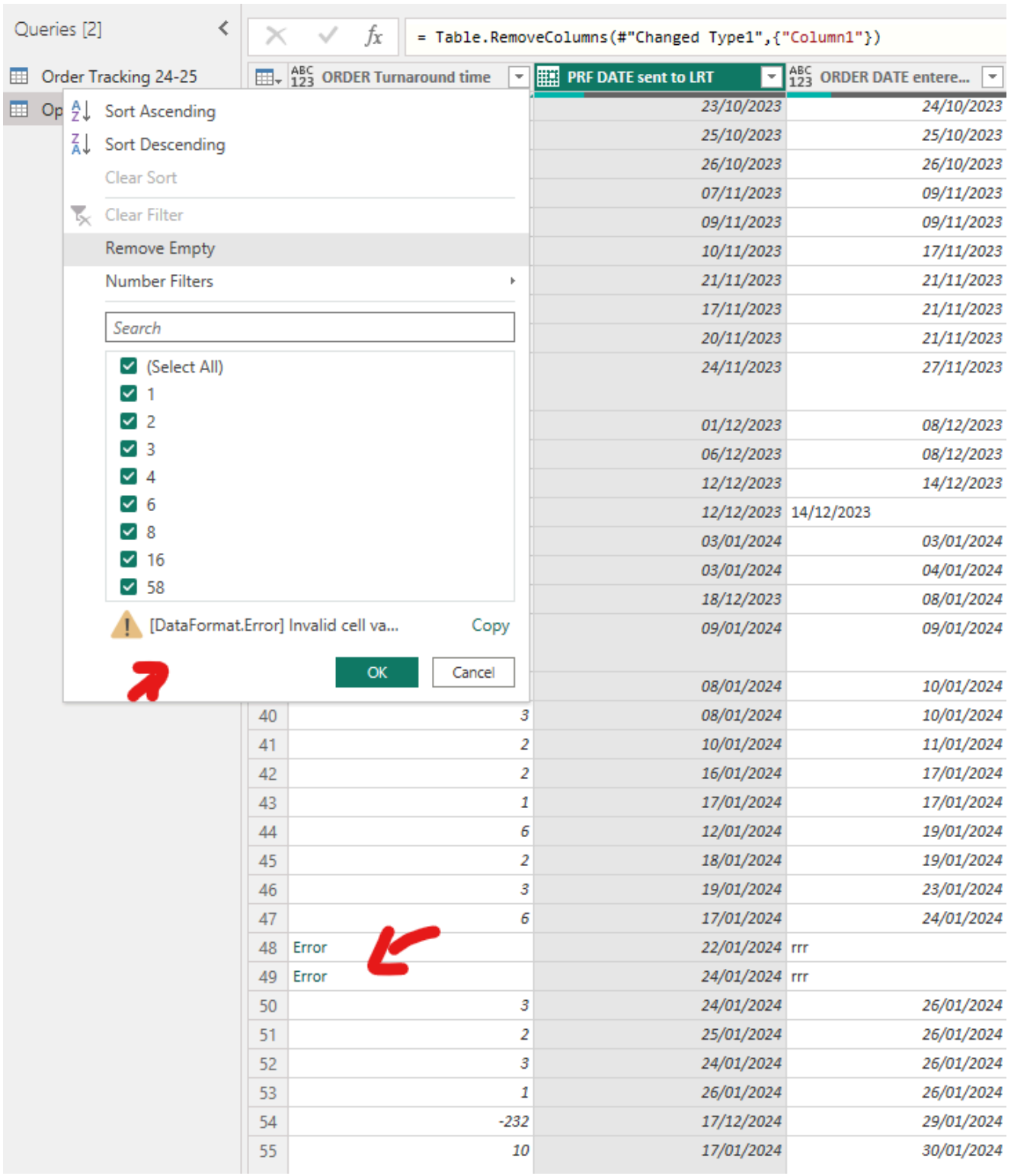


Figure 25 - example of low/bad data integrity and quality which has to be investigated and resolved in order to continue development

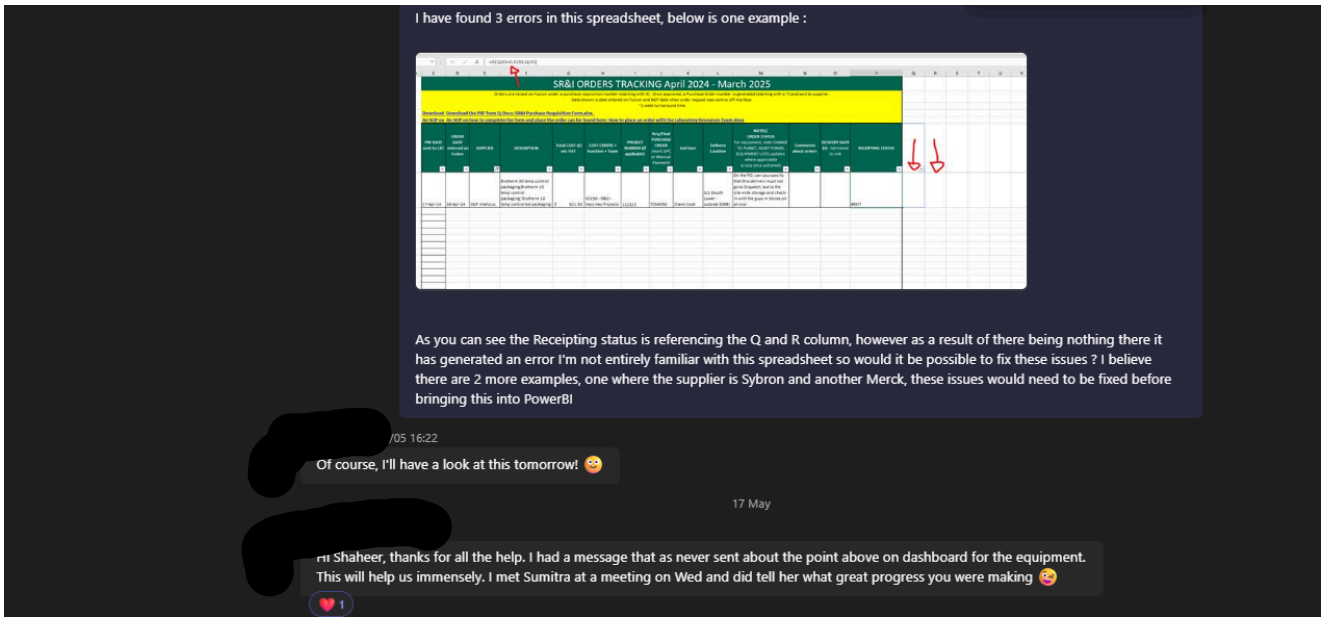


Figure 26 - Identifying and escalating potential data quality risks

Figure 27 - Tool comparison between PowerBI, Microsoft Excel and MicroStrategy for LRT Dashboard.**(www.datacamp.com, n.d.) (Anon, 2020)**

Visualisation comparison for LRT Dashboards (Oetting, 2023) (Mansurova, 2024) (careerfoundry.com, n.d.)

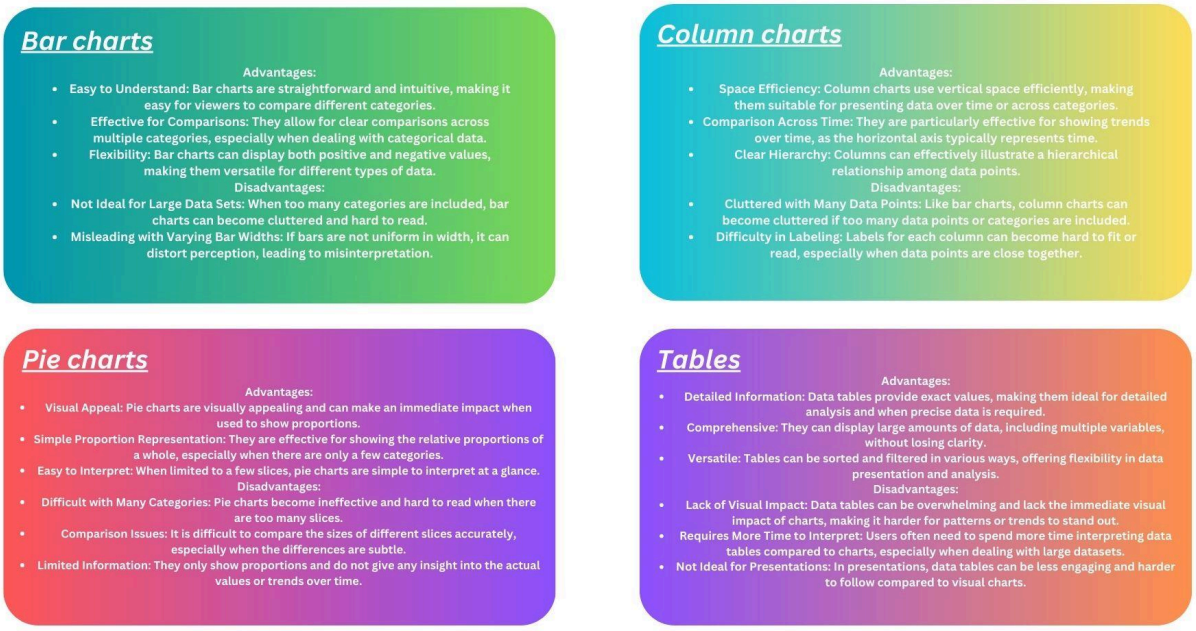


Figure 28 - Visualisation comparison for the LRT Dashboards

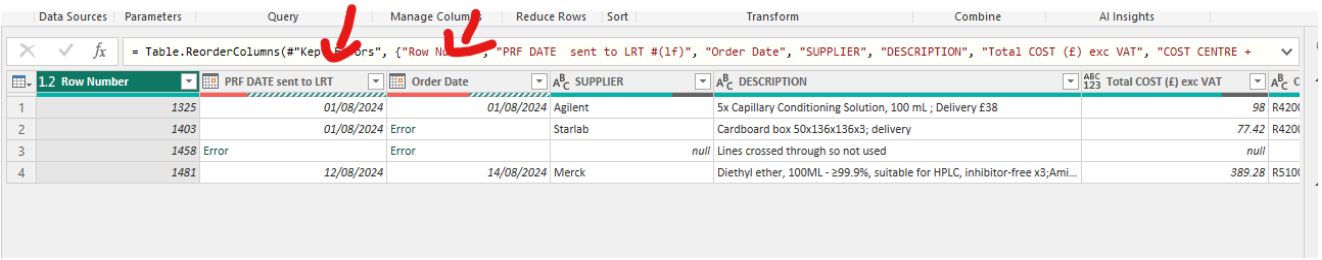


Figure 29 - Data error identification in Power query

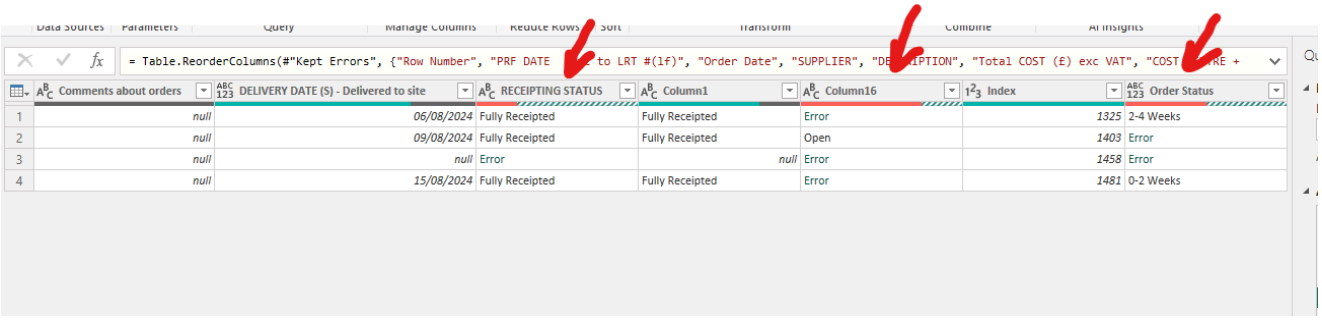


Figure 30 - Data error identification in Power query - scrolled through

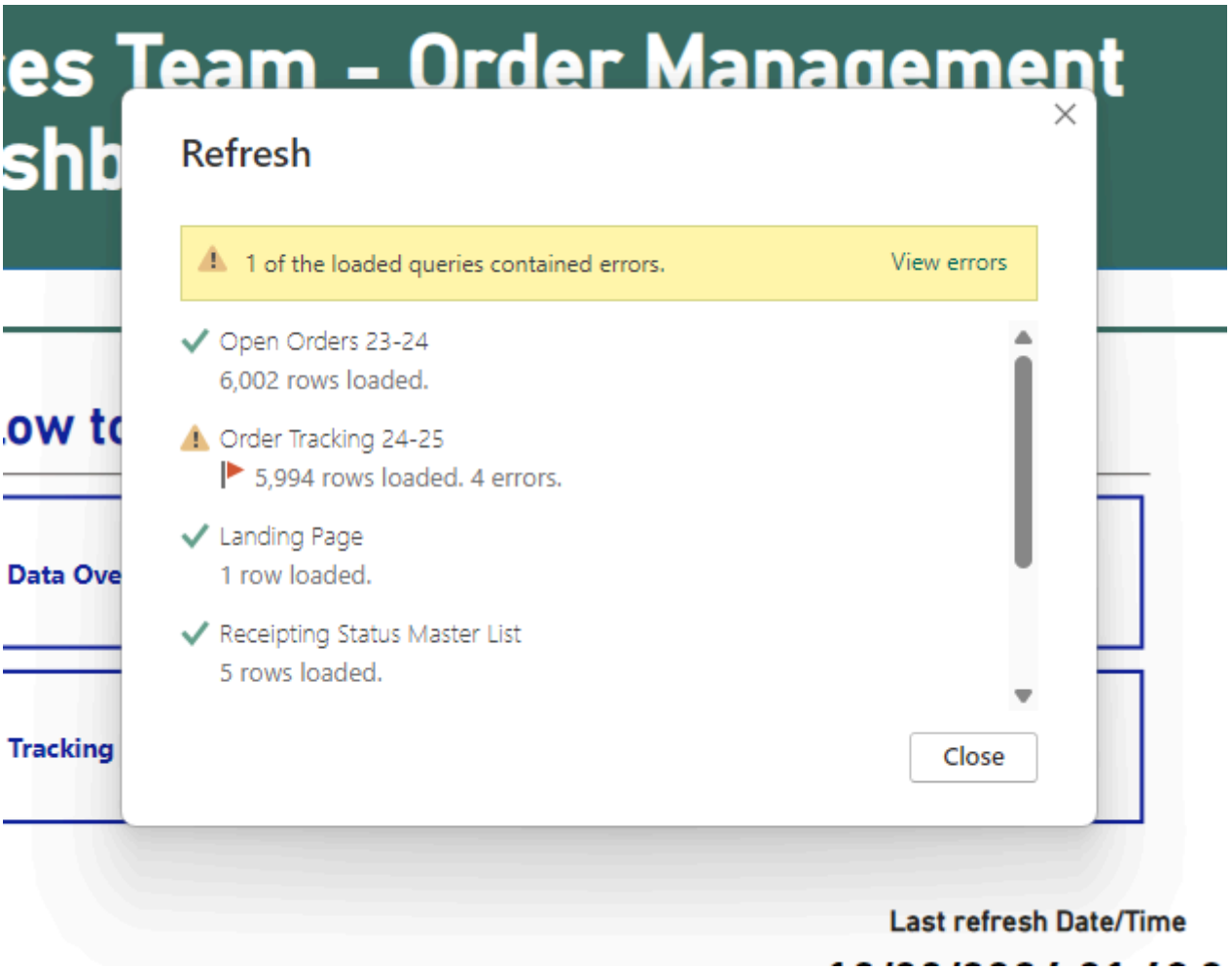


Figure 31 - PowerBI Refresh failure caused by data errors

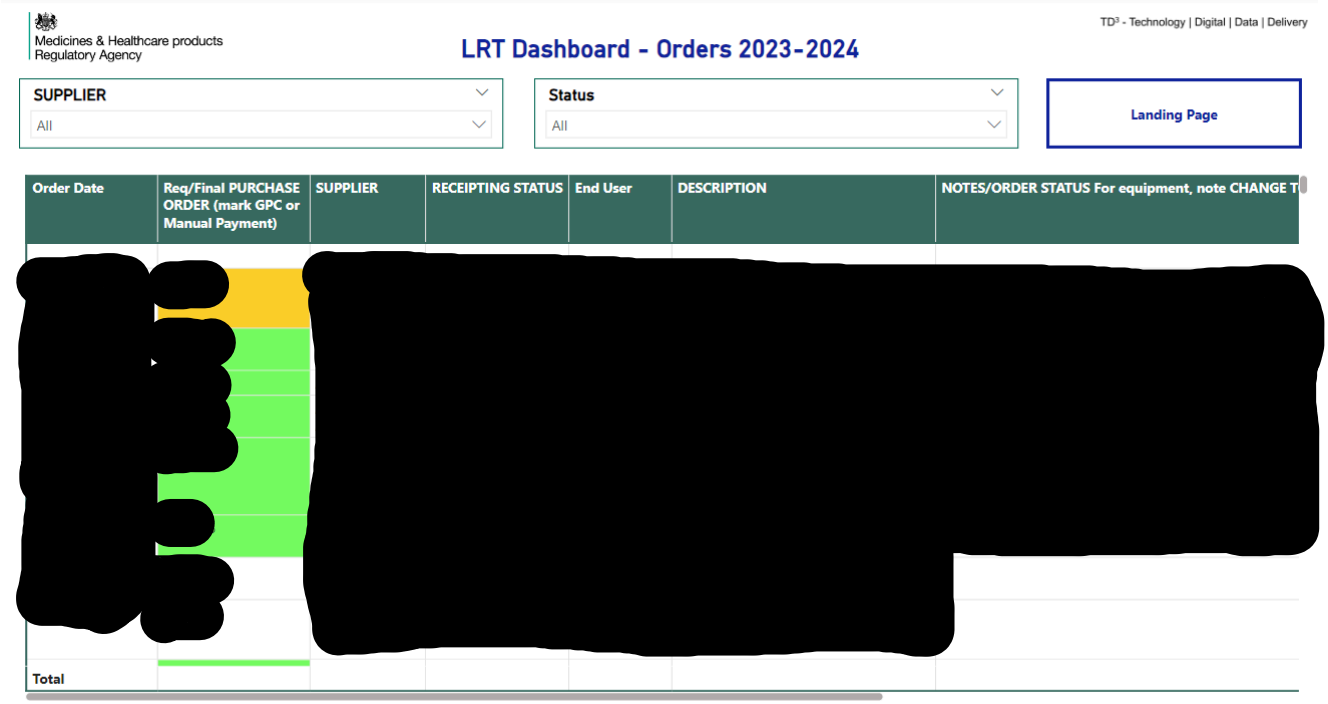


Figure 32 - Final designs for LRT order table page

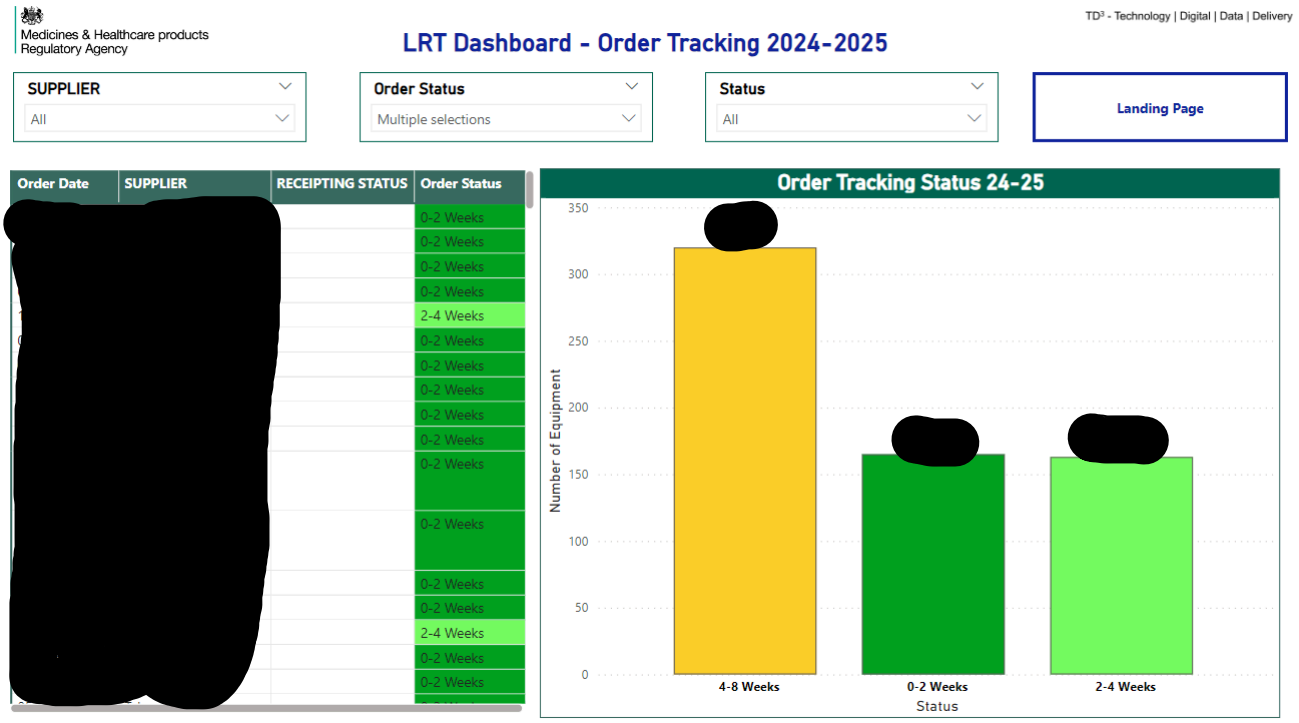


Figure 33 - Final designs for LRT order summary page

Figure 34 - Final design for LRT equipment overview with conditional formatting

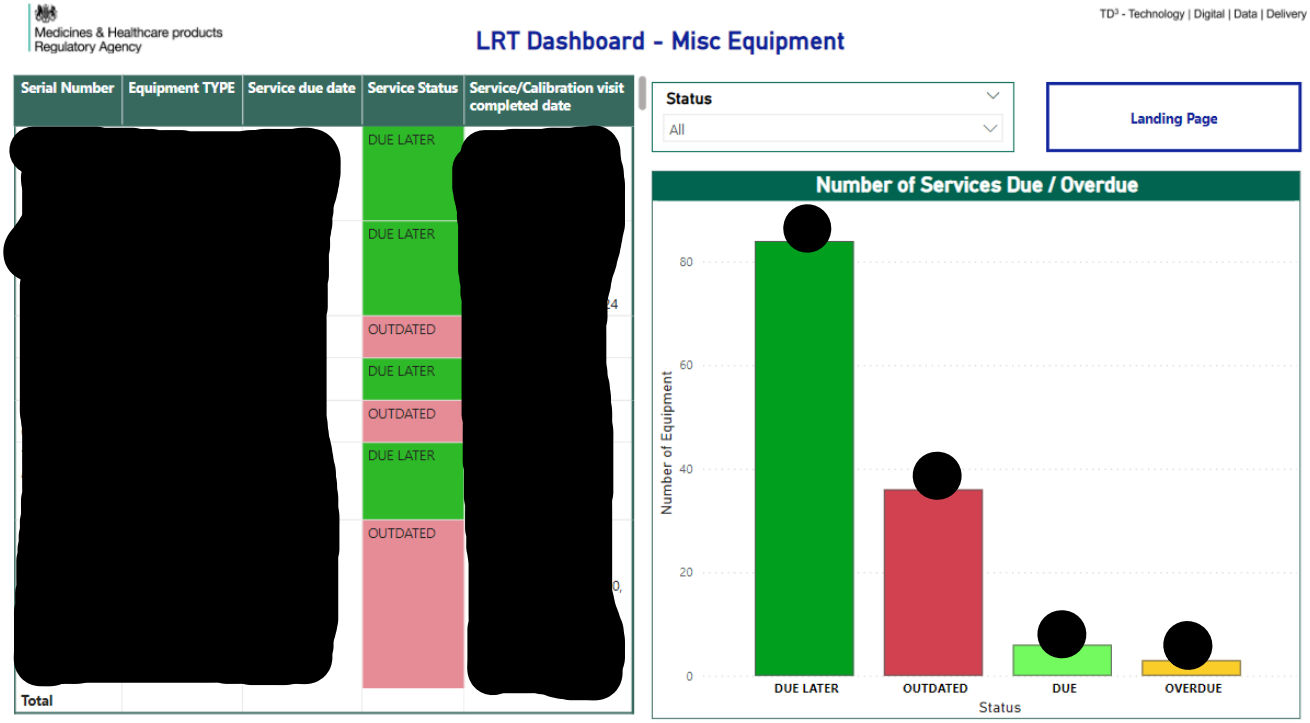


Figure 35 - Final design for LRT equipment pages with proper conditional formatting (Page 1 of 17)

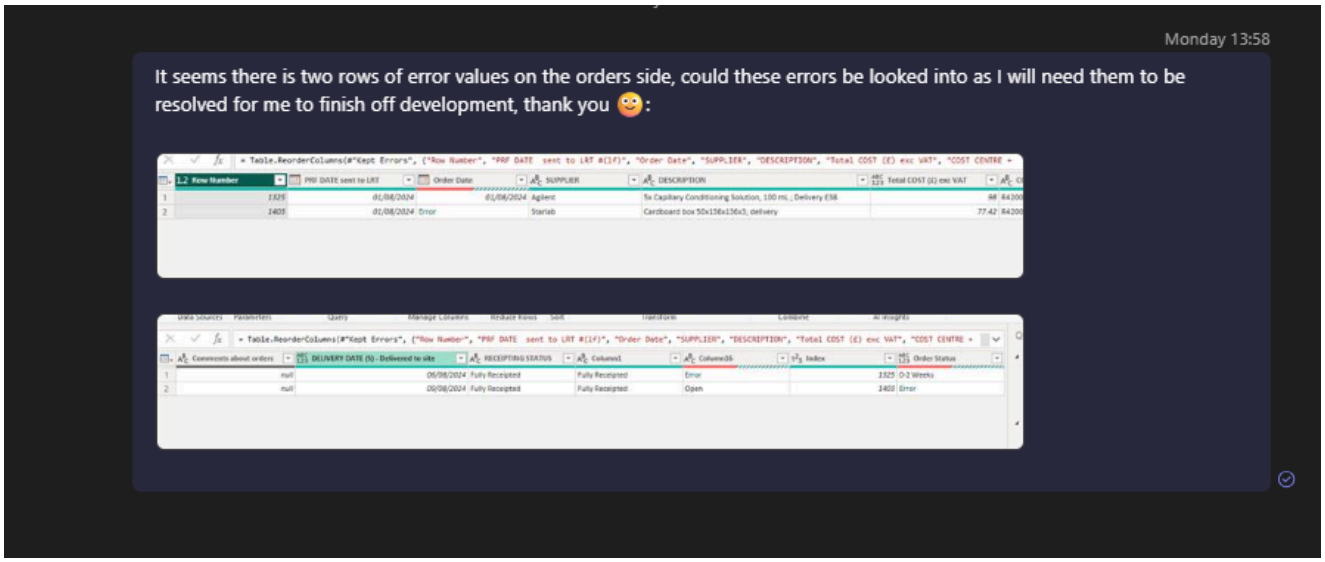


Figure 36 - Communication with stakeholders about data errors

References

Faisal, S. (2024). *17 UX Design Principles to Follow for a Great User Experience*. [online] Thoughts about Product Adoption, User Onboarding and Good UX | Userpilot Blog. Available at: https://userpilot.com/blog/ux-design-principles/.

www.datacamp.com. (n.d.). *Power BI vs Excel: Which Should You Use?* [online] Available at: https://www.datacamp.com/blog/power-bi-vs-excel-which-should-you-use.

Anon, (2020). *Power BI vs MicroStrategy | Which BI Tool Wins In 2022?* [online] Available at: https://www.selecthub.com/business-intelligence-tools/power-bi-vs-microstrategy/.

Ziad, Z. (2017). *Merging Data from Multiple Sources - Challenges and Solutions*. [online] Data Ladder. Available at: https://dataladder.com/merging-data-multiple-sources/.

onlinemanipaleditorialteam (2022). *What is the data science lifecycle?* [online] Online Manipal. Available at: https://www.onlinemanipal.com/blogs/data-science-lifecycle-explained.

Oetting, J. (2023). *Data visualization 101: How to choose the right chart or graph for your data*. [online] HubSpot. Available at: https://blog.hubspot.com/marketing/types-of-graphs-for-data-visualization.

Mansurova, M. (2024). *Visualisation 101: Choosing the Best Visualisation Type*. [online] Medium. Available at:

https://towardsdatascience.com/visualisation-101-choosing-the-best-visualisation-type-3a10838b150d.

careerfoundry.com. (n.d.). *13 Types of Data Visualization [And When To Use Them]*. [online] Available at: https://careerfoundry.com/en/blog/data-analytics/data-visualization-types/.