SUMMATIVE PORTFOLIO

Matthew Paver - Projecting Success

PROJECTING SUCCESS

CONTENTS

PROJECT

PROJECT 1

01

AUTOMATED APPRENTICESHIP ONBOARDING

Breakdown

Situation:

- To improve the efficiency when onboarding Apprentices using Robotic Process Automation (RPA)
- Initially, the process was done by hand and is an exercise of duplicating information and performing calculations.
- This process was created for our HR team to use where the project was complemented with a user interface to ensure the team is able to understand the tool at its highest level.

Task:

- To Automate the process of capturing Apprentice data, store into an SQL database and retrieve for manipulation into the users' contract and learner plan
- This was to be used for each learner that had joined the Apprenticeship, in which the manager selects a button that issues the documents to be created as it triggers a process.

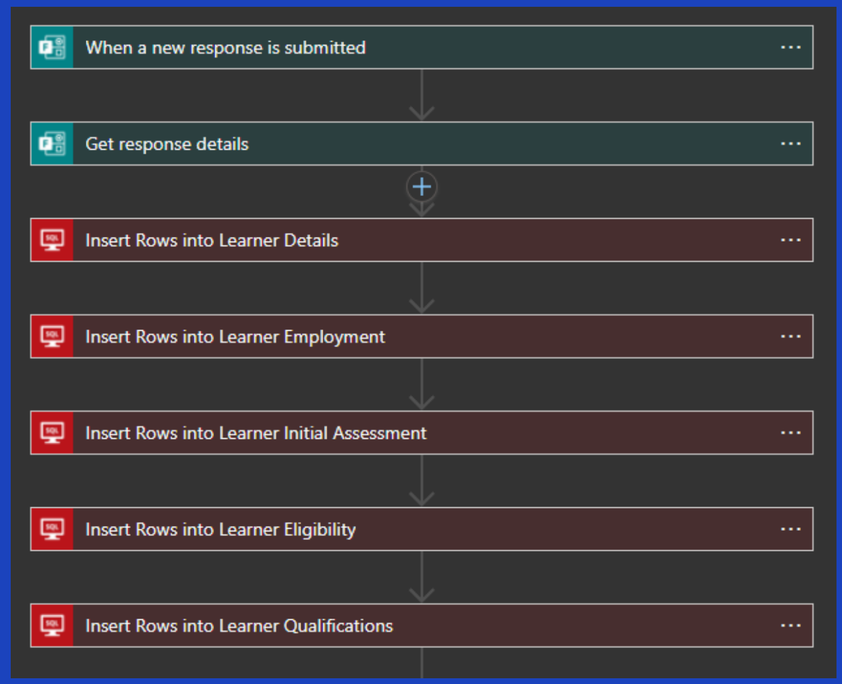
Action:

- Retrieve data from Microsoft Form
- Store data into a SQL database
- Retrieve data when required to create learner documents
- Link this with a pre-established application so the process can be used

Result:

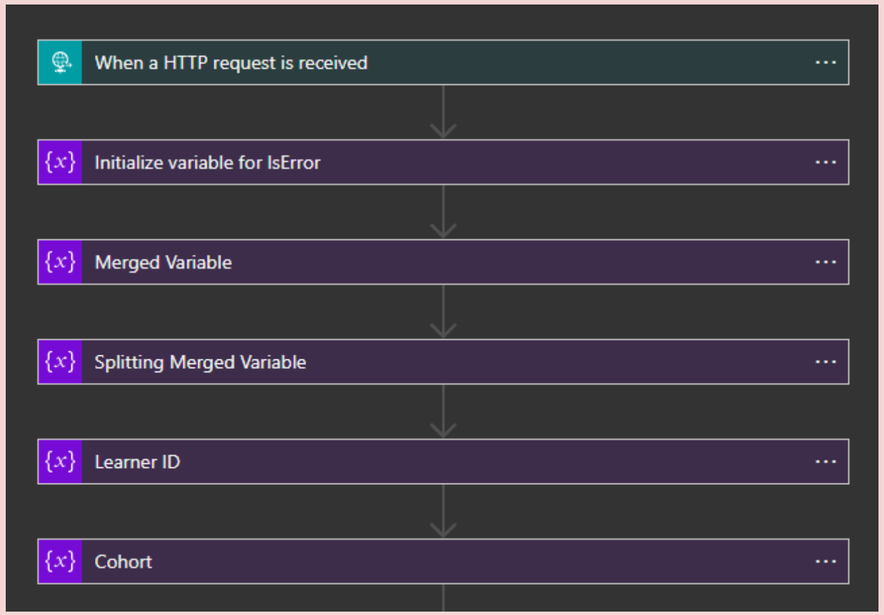
- Automated learner documents
- A secure location to store learner information
- Reduce manual labour of the Apprenticeship team to allow the time to be used in developing the course.

PHASE 1



OVERVIEW

PHASE 2



THE PROJECT

Onboarding Learners is a constant process within Projecting Success. Every 2 months, a new group of learners are brought into our Project Data Academy Apprenticeship. Therefore, the apprenticeship team has to create a personalised plan for the learner over the 15-month scheme; e.g including personal details, off the job hours and a calculation of how much of the course they can be taught due to a pre-initial screening.

Why did we decide to automate this?

I took the initiative to have a meeting with the Apprenticeship Manager to work out what parts of their job is mundane and that can be automated. This was known as the breakdown of effort.

Apprenticeship scaling										
	Per course	Internal work	Per cohort	Per Learner per cohort	Total per intervention	Number of interventions	Automation potential	Person	Comments	Matt P thoughts
Onboarding										
Part 1				0.03	0.03	1		S.Sullivan		
Part 2				0.03	0.03	1		S.Sullivan		
Check eligibility				0.03	0.03	1		S.Sullivan		
Commitment Statement				0.06	0.06	1		S.Sullivan		
L&D plan				0.06	0.06	1		B.Morris		
Contract				0.06	0.06	1		S.Sullivan		

Below, you can view the tasks the manager has and how many times they have to intervene with the process to ensure it is successful. I then asked as a RAG (Red, Amber Green) status what they thought could be automated.

I did this to gain an understanding of what they thought was the capability of Automation. Following this, it allowed me to provide them greater support in how anything that can be clicked, dragged, dropped, or typed can be automated. This opened up the possibility to automate the onboarding section.

THE PROJECT

The problem statements set were:

- As an Apprentice Manager, I want to be able to issue a contract once the learner has been confirmed onto the Apprenticeship
- As an Apprentice Manager, I want learner documents to be added to the correlated company and learner.

Why was this set?

The Apprentices Manager would like to issue a contract as it saves them manually copying and pasting all of the learner details into a word and excel document. It also reduces the chance of human error. This then allows them to use their time on something more intellectually demanding and can develop the course further.

Having the learner documents correlated to the company and learner means that the Apprentice manager does not need to drag and drop files as it is automatically done. This should be also ready within 3 minutes of pressing the issue the contract button which would be much quicker than doing this process manually.

THE PROJECT - BREAKDOWN

This Project split into 2 phases:

PHASE 1 - Collection of data

- When a form has been filled in by a learner
- Retrieve the details from the form
- Insert the data into the correlated tables within our Project_Data_Academy database
- Once all the information has been stored, an empty learner plan and contract is added to their learner folder ready for phase 2 where the apprentice team is also notified

PHASE 2 - Creating the learner documents

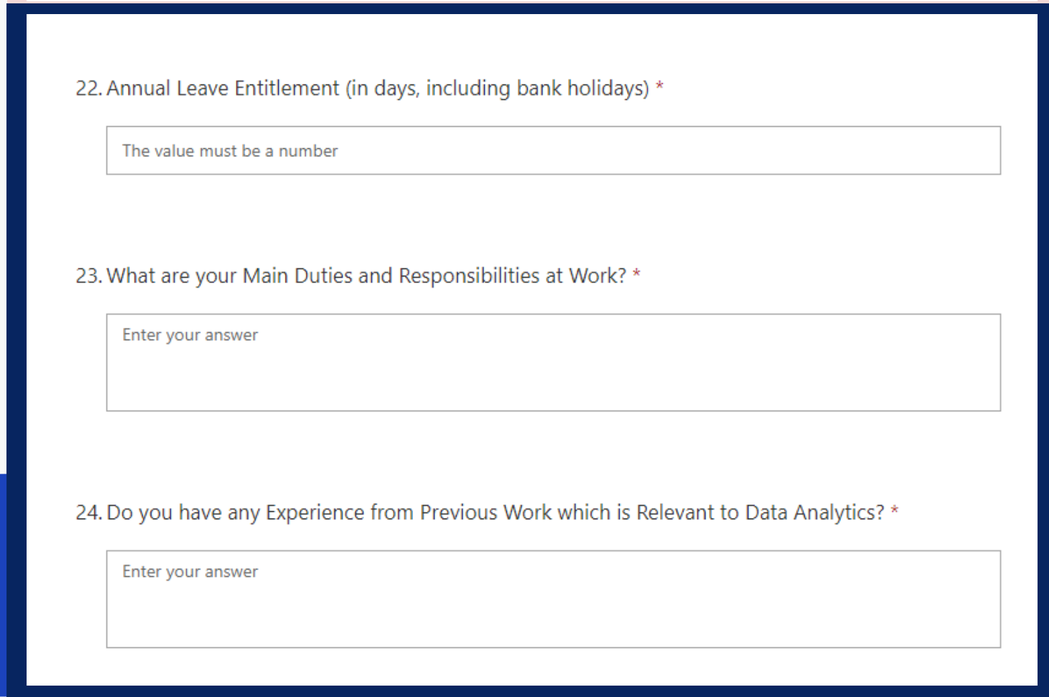
- The apprenticeship manager issues a contract
- Retrieve the details from the SQL database
- Manipulate the data such as working out the age of the learner by their birth date.
- Once all the calculations have been made, the learner plan and contract are populated

THE PROJECT PHASE 1

Learner Perspective

The Learner is sent the Eligibility Screening form which consists of 106 questions.

Once the form is submitted by the learner, it triggers the process to take the details and store them in the SQL database whilst notifying the team that a new learner has completed their eligibility form.



THE PROJECT PHASE 1

End User Perspective

This process is built using a Logic App which is a service from Azure. The reason we use this is due to the seamless integration with the SQL Server Management Studio 18 to connect to our Project Data Academy database.

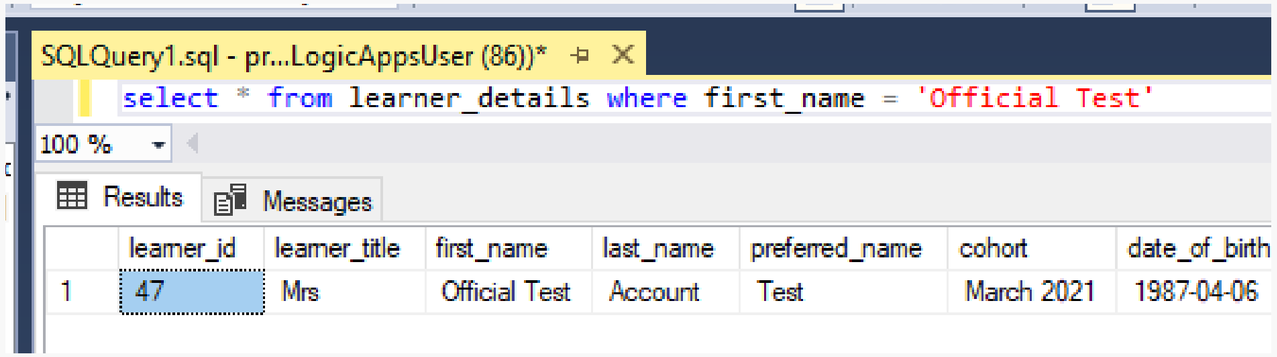
Here is a successful run of Phase 1. You can identify through this what data is added to the SQL database and where the files are created.

The next page highlights what type of information is gathered and how its added to the database.

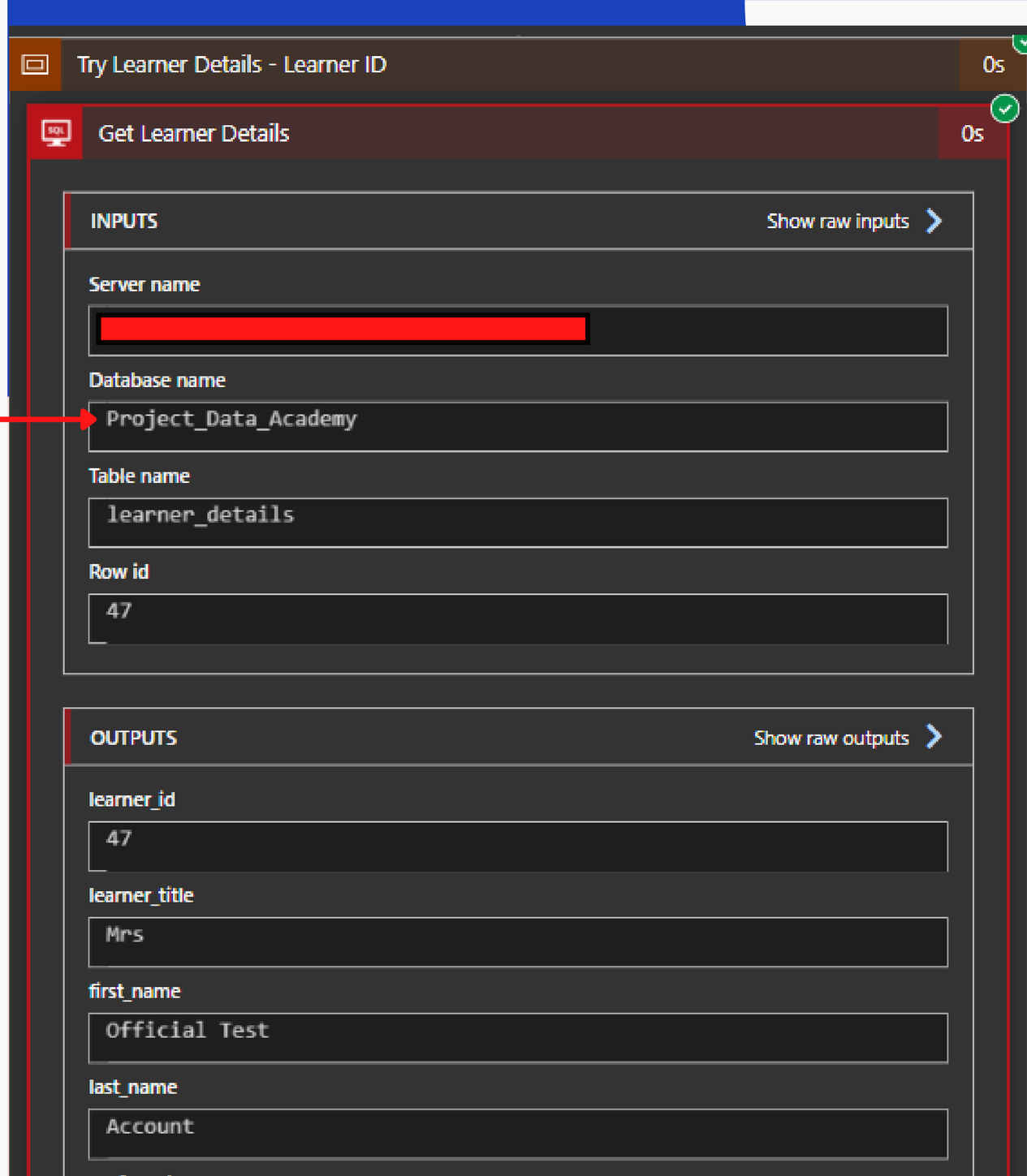
THE PROJECT PHASE 1

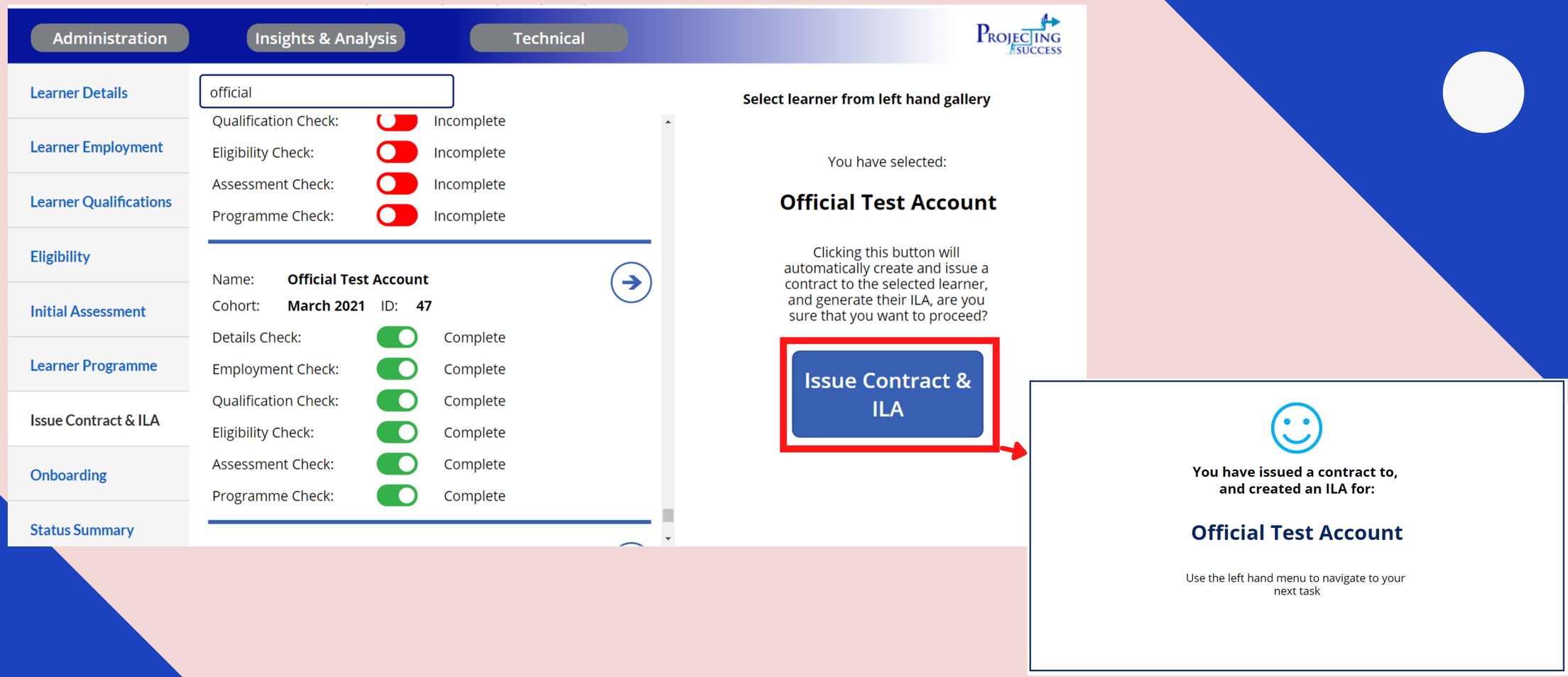
This highlights that a row has been added to the learner_details table in the database.

Below is an example of how the data is structured in the database using SQL Server Management Studio 18



	learner_id	learner_title	first_name	last_name	preferred_name	cohort	date_of_birth
1	47	Mrs	Official Test	Account	Test	March 2021	1987-04-06





Application created using Microsoft PowerApps

THE PROJECT PHASE 2

For this process to start, the Apprenticeship Manager selects the Issue contract & ILA button which triggers the next flow

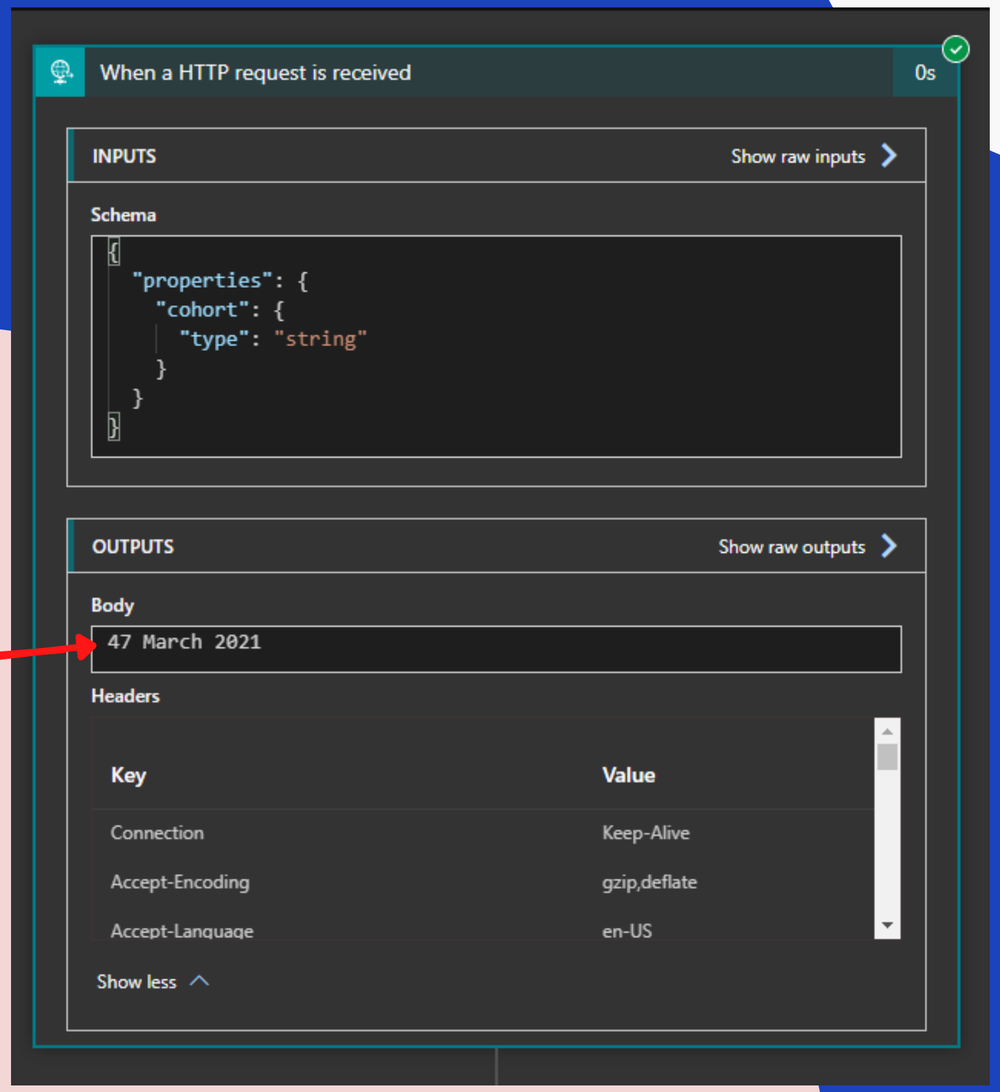
This can only happen once all the checkboxes have been selected by the Apprenticeship Manager as they have to manually check to ensure that they applicable for the Apprenticeship

THE PROJECT PHASE 2

The process accepts the data enclosed in the body of the request (the button) and takes specific details from the learner which are the ID and cohort.

Name: **Official Test Account**

Cohort: **March 2021** ID: **47**



The ID identifies the unique learner and the cohort is added from the PowerApp and is then used to retrieve the rest of the row for the learner.

THE PROJECT PHASE 2

Data Manipulation

To split the string, I used a function to split after the ID. Otherwise, there would be an error due to the ID being processed as '47 March 2021' rather than '47'

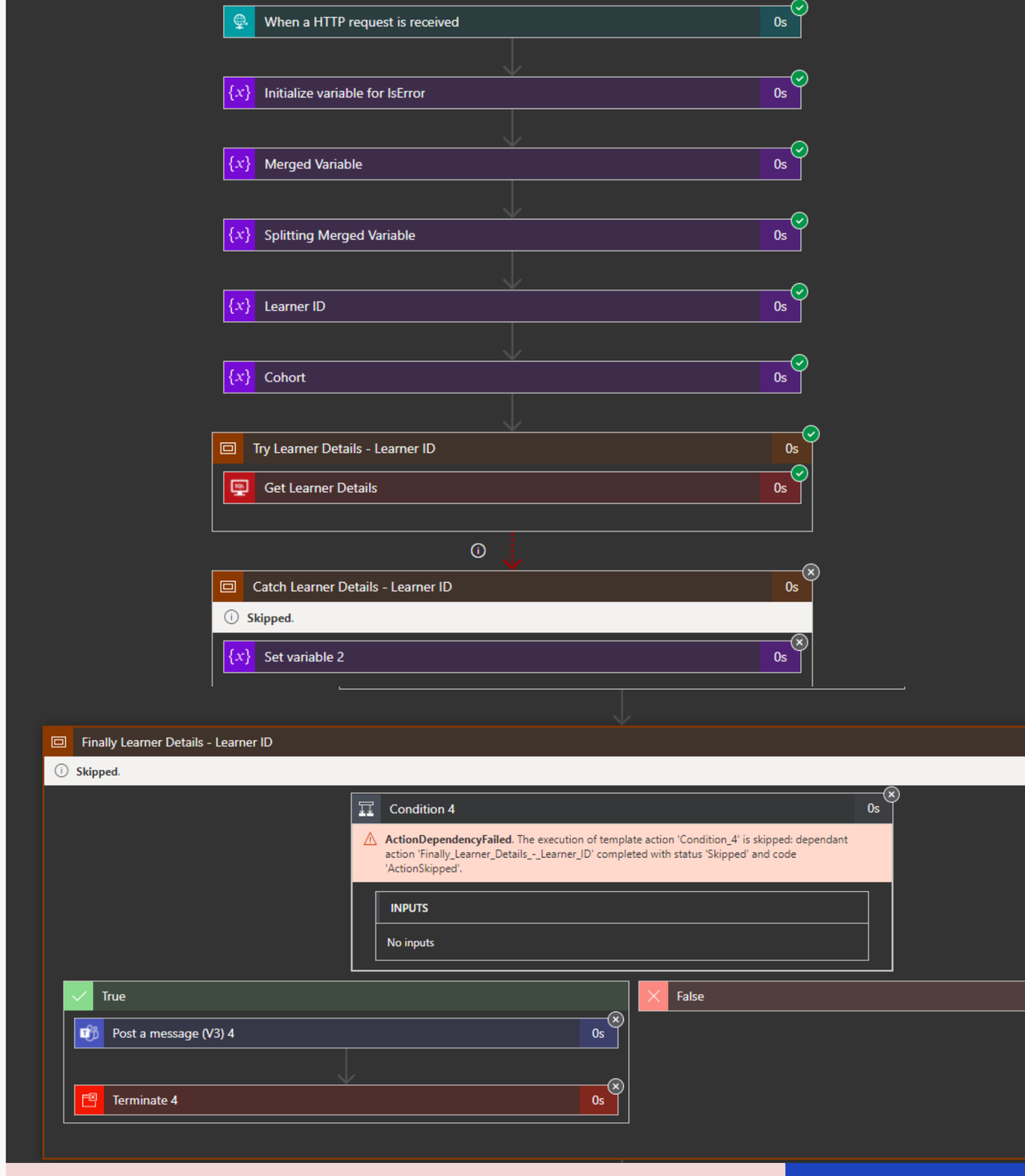
I then passed the Learner ID and cohort into the 'Get Learner Details' which is retrieving the data from the SQL database.

Try/Catch Statement

This is within a Try catch block as through my user acceptance testing, if the ID is not recognised it fails.

Now, if the ID is not found, it will set the IsError variable I created earlier to True in the catch block.

Afterwards, it will try to post a message in Teams to notify the automation support team that the flow has stopped working due to the ID being incorrect. The flow will then terminate.

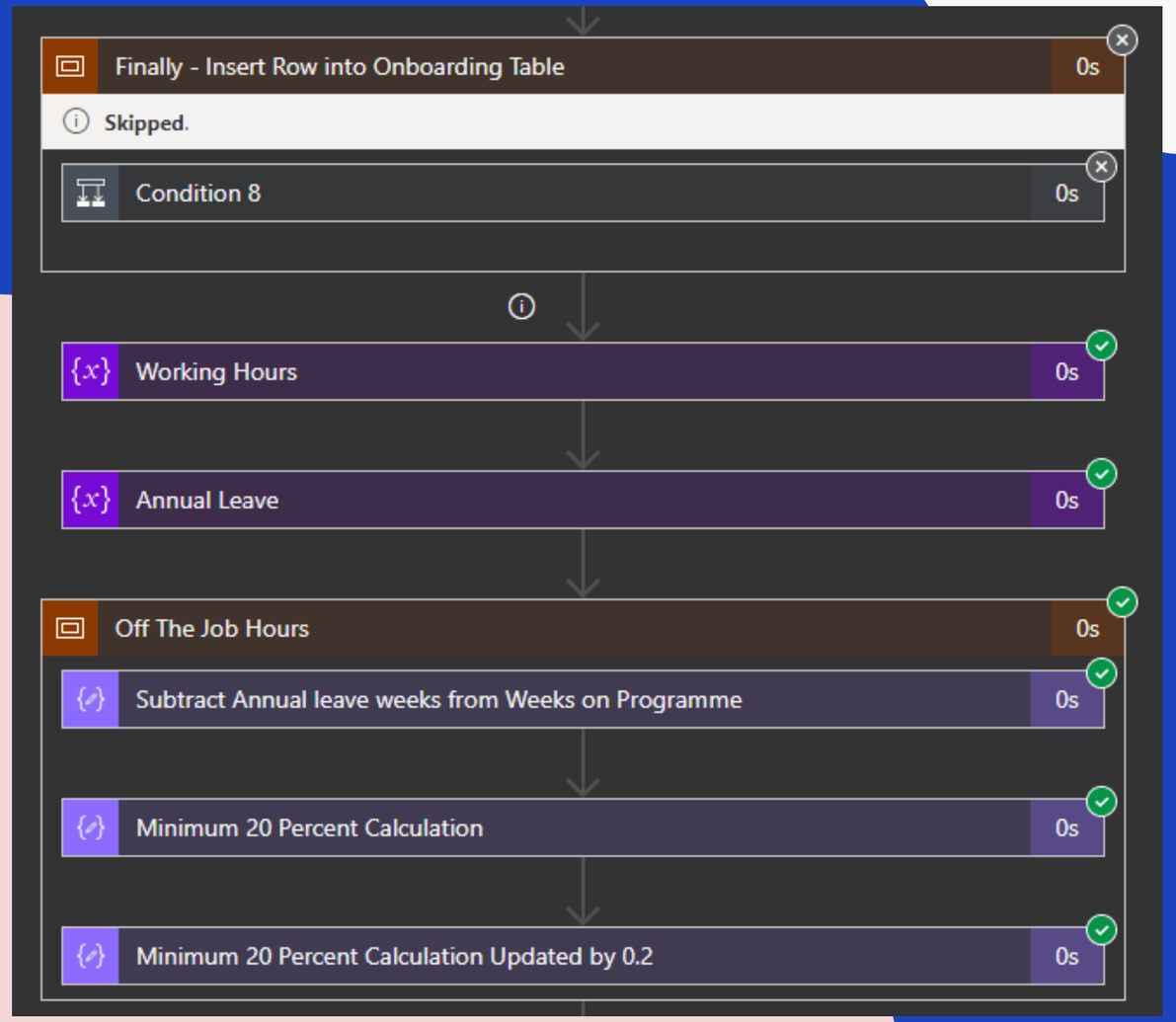


THE PROJECT PHASE 2

The process continues to have the try/catch blocks in order to populate the learner documents.

It is then followed by a calculation to determine the amount of off-the-job hours they must do in order to complete the apprenticeship. This is based on the learner committing 20% of their working hours for the course.

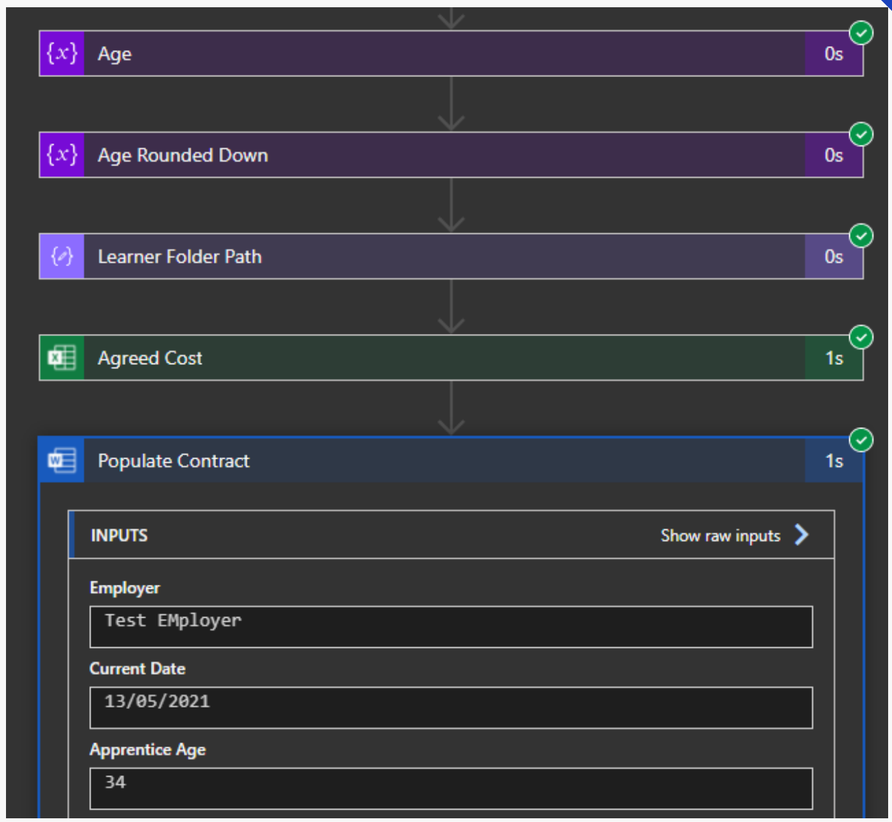
This requires multiple functions in order to calculate the amount of off-the-job hours, which displays my understanding of data manipulation in this instance.



THE PROJECT PHASE 2

Updating the Learner Plan

After retrieving all the information from the SQL database, I can add the unique learner details to the learner plan.



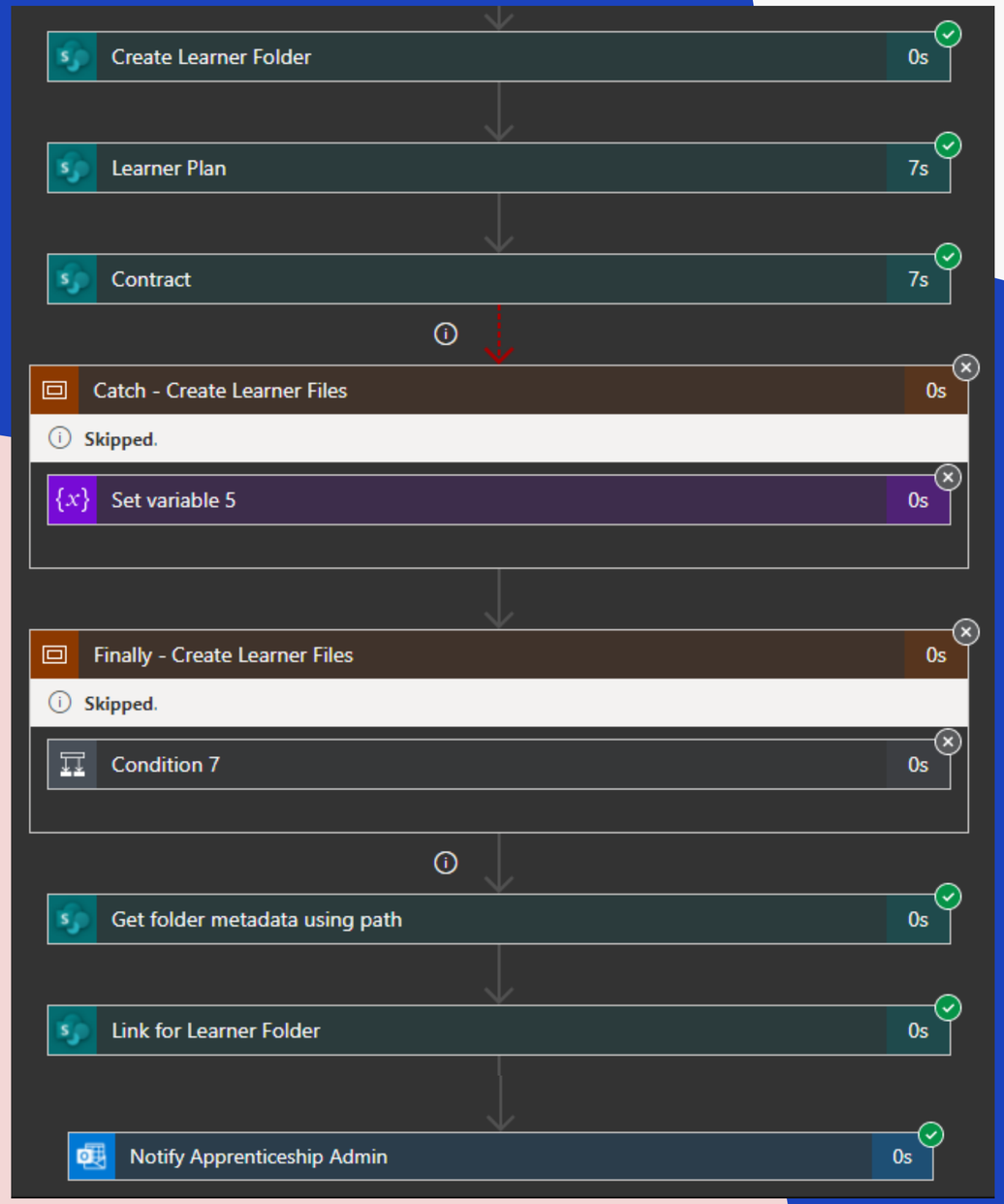
Populating the contract

From the learners' birthdate, I have created a function to determine the learners' age and add this into the contract. This enables dynamic data manipulation as the birthdate will change dependent on the learner.

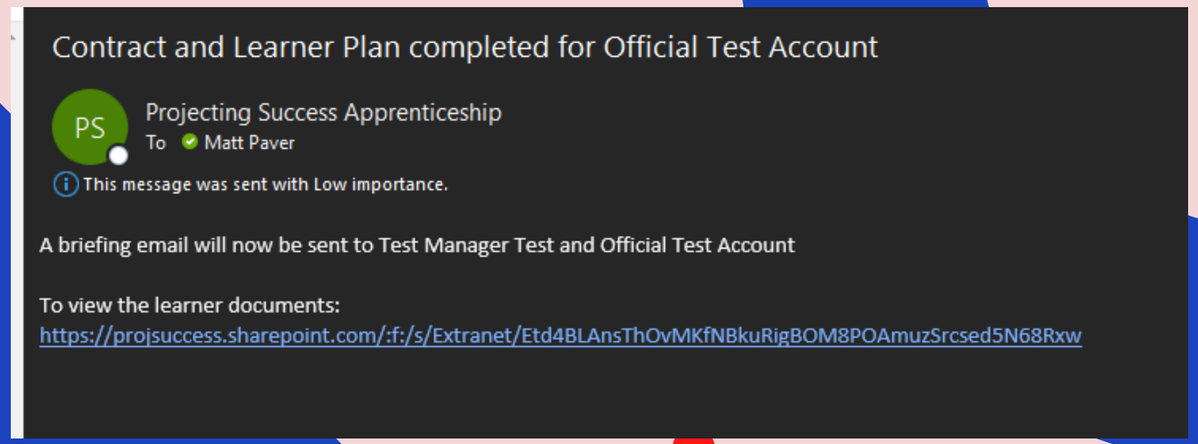
THE PROJECT PHASE 2

This section ensures that the learner documents are collated and added to the learner folder. If it is unable to create a learner folder, as there may be an update to the contract, it will go through the catch and finally block which will simply add the contract and learner plan into the previously created learner folder.

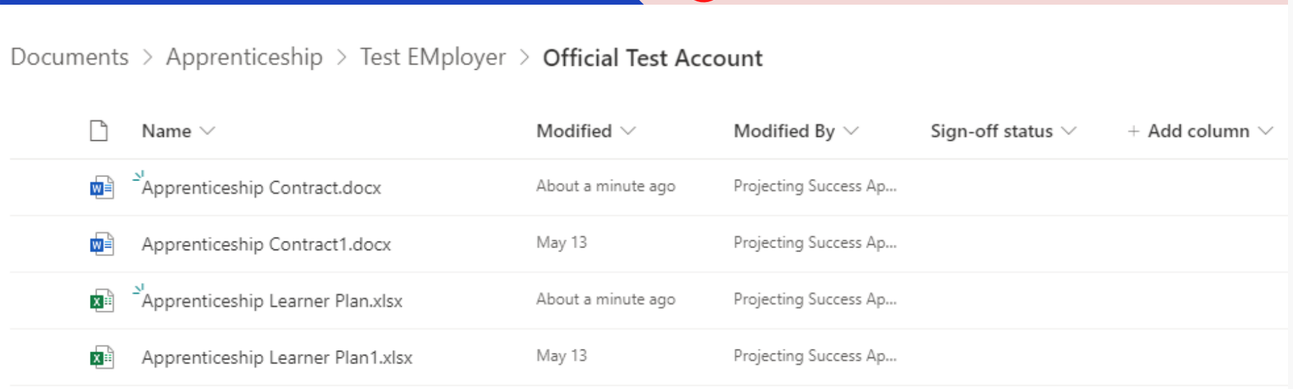
Finally, this process ends by notifying the Apprenticeship admin team, line manager, and the Apprentice. that the learner documents have been created and can now be signed off by the learner and line manager.



THE PROJECT PHASE 2 OUTPUT



Here is the email that is sent to the users. It includes a link to view the documents.



Rather than overwriting the documents, it has a counter on the end of the file name to ensure that personal data is not lost. This was a consideration in the rigirous testing process.

THE PROJECT PHASE 2

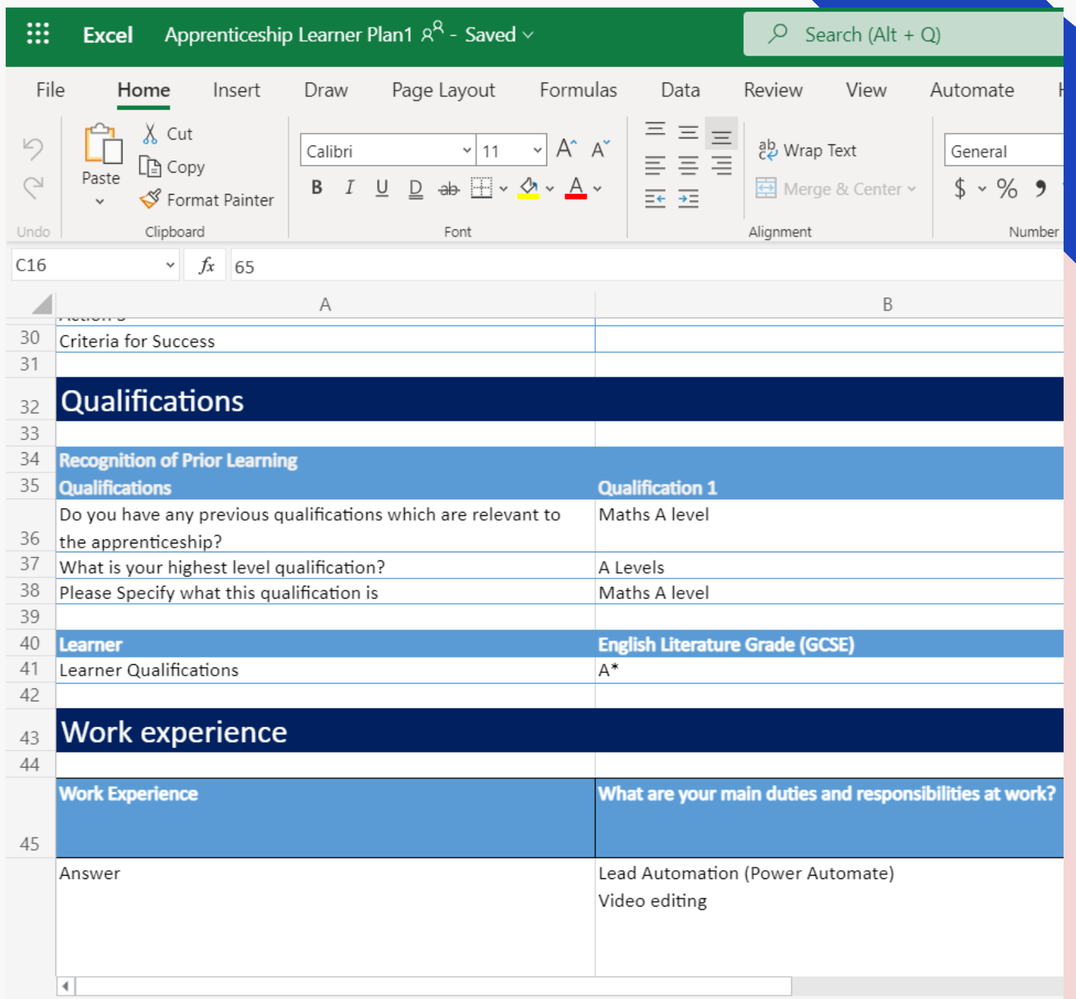
Specific Terms	
Apprentice Name:	Official Test Account

Apprenticeship Training Services Agreement 2

Apprentice Date of Birth:	1987-04-06
Age at Start of Apprenticeship:	34
Apprentice Address:	Test Cottage,Test Street,Test TownTest County,Test Post Code
Type of Apprenticeship:	Level 4 (Project) Data Analyst
Date of commencement of apprenticeship:	2021-03-04
Duration:	15 months
Forecast Completion Date (final day)	2022-08-01
Off-the-job Training (hours)	448

Populated Contract

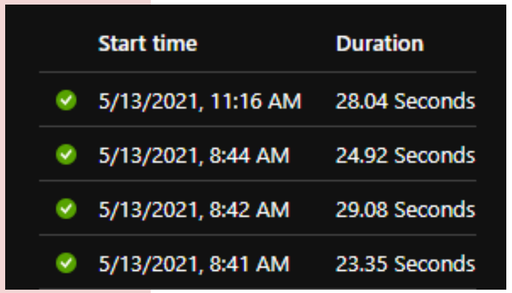
The unique data for the learner has been successfully extracted from SQL



Populated Learner Plan

The Learner Plan has been populated with previous qualifications and Off the job hours

How will this help?



Start time	Duration
5/13/2021, 11:16 AM	28.04 Seconds
5/13/2021, 8:44 AM	24.92 Seconds
5/13/2021, 8:42 AM	29.08 Seconds
5/13/2021, 8:41 AM	23.35 Seconds

- The Apprenticeship Manager will be able to save a vast amount of time on copying and pasting files and information as the process takes approximately 26 seconds. The evidence for this is highlighted through the screenshot of the duration of runs. Before automating, this took around 15-30 minutes per learner.

- The Apprenticeship Manager will be able to spend this time developing the course which will improve the experience for the learners as it is more seamless.
- There is little to no chance of Human error as all calculations are put through the pre-established functions. It also means that there is no spelling errors unless the learner has done this through the eligibility form. Therefore, manual intervention is still required by all parties to ensure they are happy with the information inputted.

Were the problem statements solved?

- As an Apprentice Manager, I want to be able to issue a contract once the learner has been confirmed onto the Apprenticeship

The Apprenticeship Manager is able to issue a contract through the PowerApp. I have created a HTTP trigger to add a front-end so it is easier to use for the manager.

- As an Apprentice Manager, I want learner documents to be added to the correlated company and learner.

The documents are placed in the company name/Learner Name/Files so this has also been completed

How can this be improved?

- I would like to improve this process by calculating the discount a learner has. The discount is currently worked out by how a learner rating themself in how competent they are in each area of the course. This section is slightly more complex but it would be a welcome addition as we would only need to check that the discount is correct.

- The Apprenticeship Manager will be able to spend this time developing the course which will improve the experience for the learners as it is more seamless.

- This could also be improved with a factor of a learner who is coming back from a break in learning as the contract will need to be updated. This could either be added through the PowerApp or Logic App.

Project 2

POWER BI DASHBOARD

Building a Power BI Report using a government API

Overview

Establishing insights from government health data

Situation:

- Through Government data, there is a clear need on the analysis of projects. Data is collected, visualisations are made but there is no lessons learned in why projects over run.

Task:

- 'As a portfolio manager, I want to look across all of my projects for a leading indicator to identify projects that are going to be over budget so I can implement an intervention and reduce the cost over run.

Action:

- Connect to live, government health data via an API.
- Highlight the ratio of projects that are under/over budget.
- Display the budget variance by the initiative name and if the project has been completed using visualisations.

Result:

- To understand the costings of a project and why a project could be over budget
- To understand the reason why projects are over and to start delving deeper into why this was the case.

THE PROJECT

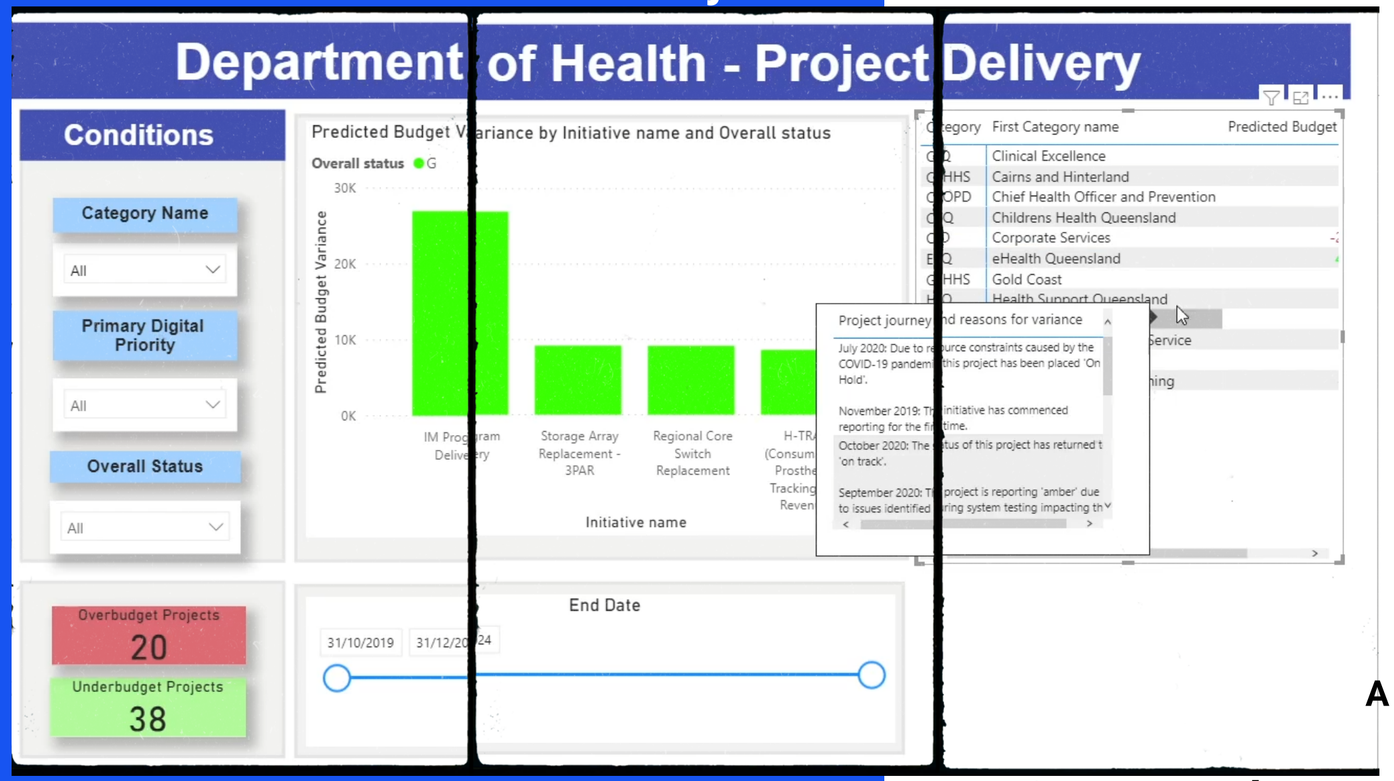
Why did I use Government health data?

1 **This project was created to highlight to our community at a meetup the current issue that companies are facing of collecting data but not evaluating the results and creating actions from it.**

2 **It demonstrates the capability of data analysis using public data where the same tools and methods can be used for community members to apply this to their own datasets.**

3 **We have found through our events there is a gap in knowledge in why data visualisations are used. We have identified that user stories are missing from a scenario so a user can never truly complete their dashboard as they don't know who they are aiming the solution at.**

Budget Variance of Projects



Conditions of Project

Reasons why individual projects were overbudget

THE PROJECT

For this project, I wanted to explore the capability of creating a dashboard that changed with the dynamic data.

This allows for the data to be changed without any disruption to the dashboard; providing it follows a similar data structure.

A video walkthrough of my solution to the project is in the zipped gateway folder labelled 'Project 3 Walkthrough'

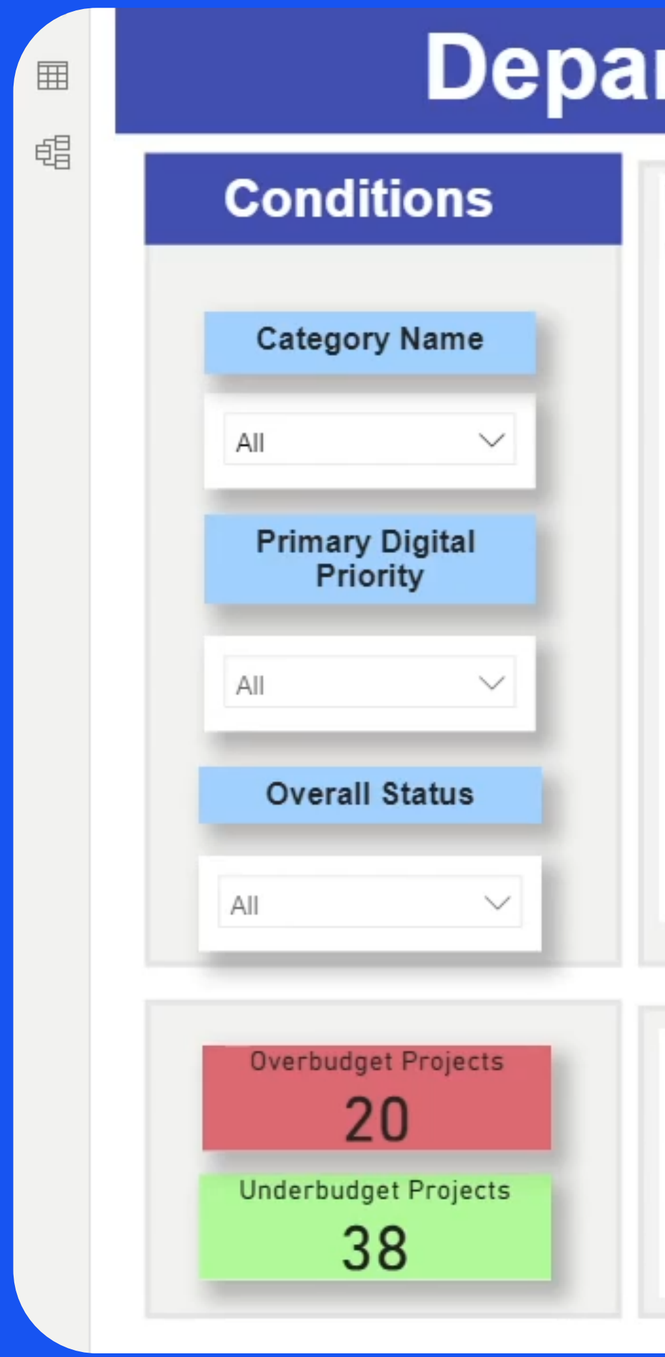
What did I do with the data?

To identify whether a project was going to be under or over budget I had to use DAX (Data Analytical Expressions) to:

- Calculate the duration of each project (the difference between the start and end date of the project)
- Calculate the budget burn rate per day by dividing the approved expenditure by the duration of each project
- Calculate the days a project has been going for it it is not completed
- Calculate the actual burn rate of the project per day.
- Calculate the predicted budget variance by subtracting the actual burn rate from the predicted. If this is negative, then the project is over budget

This links back to the user story proposed of 'As a portfolio manager, I want to look across all of my projects for a leading indicator to identify projects that are going to be over budget so I can implement an intervention and reduce the cost over run.' as you are able to identify the amount of projects that are over budget.

To further this, there are filters for the projects so the project manager can identify if projects are over budget as the manager can filter the categories of the projects to determine whether there is a correlation that certain categories of the projects are generally more over budget.

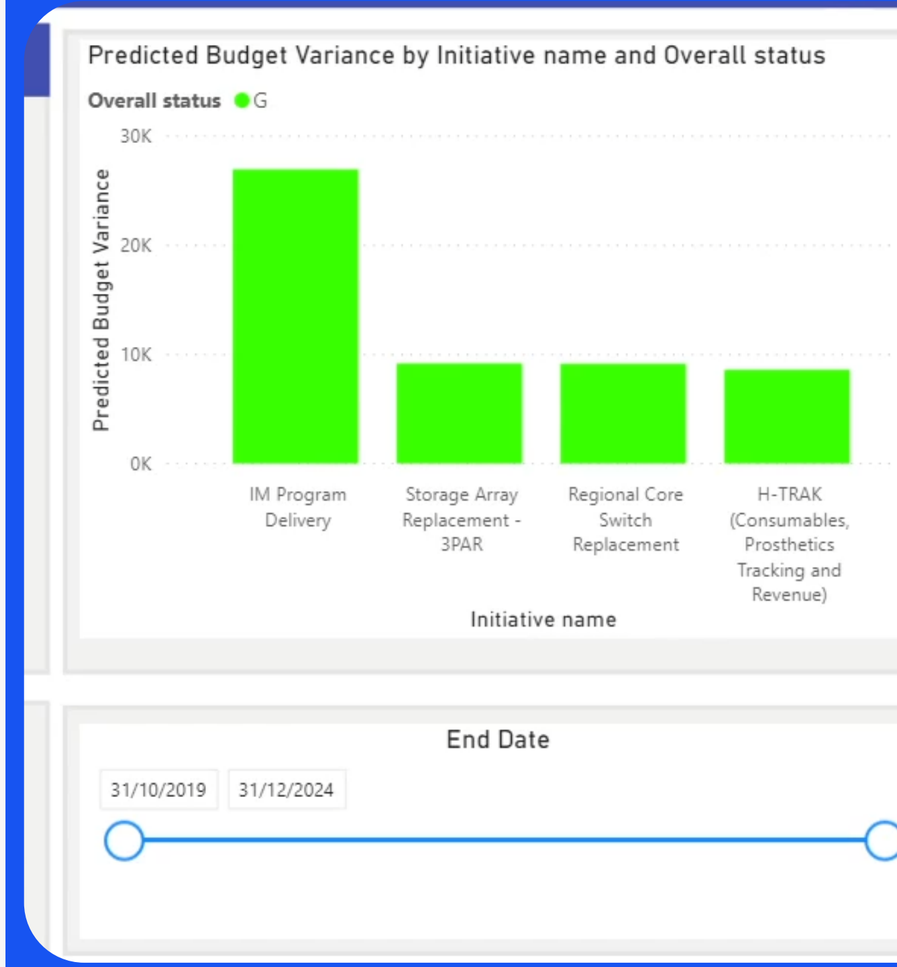


Linking back to the problem statement

Following the previous calculations I wanted to evaluate what insights we can gain from the data:

- I used a bar chart to highlight which projects are completed and had the most variance in its budget in comparison to the approved expenditure. The portfolio manager would want to view this as it highlights which categories of projects vary the furthest from the initial thought cost.
- This can be drilled down further using the End Date timeline as this can identify in what year did this approved expenditure vary the most.

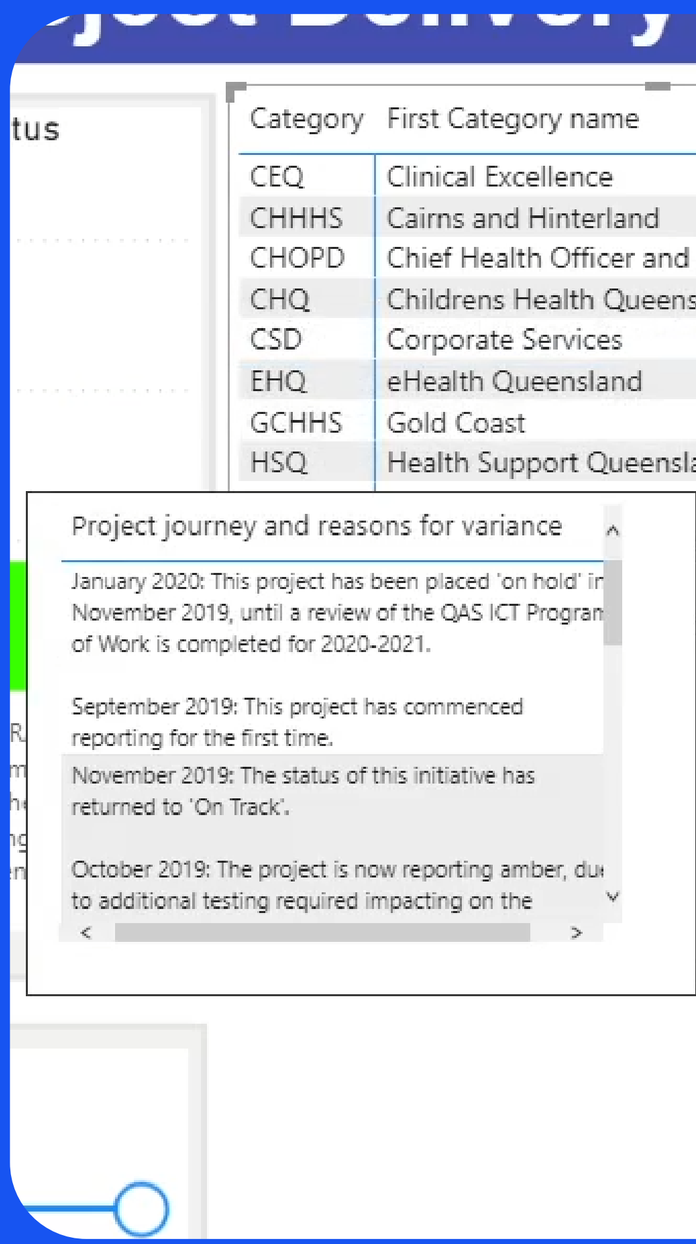
However, this is not explaining why the projects varied so much in their budget. We will now look at the project journey and the reasons in its variance.



Linking back to the problem statement

This particular visualisation is highlighting a timestamped project journey and the reasons for its variance. A hover-over tooltip was added to provide additional information in why certain project categories may have been delayed.

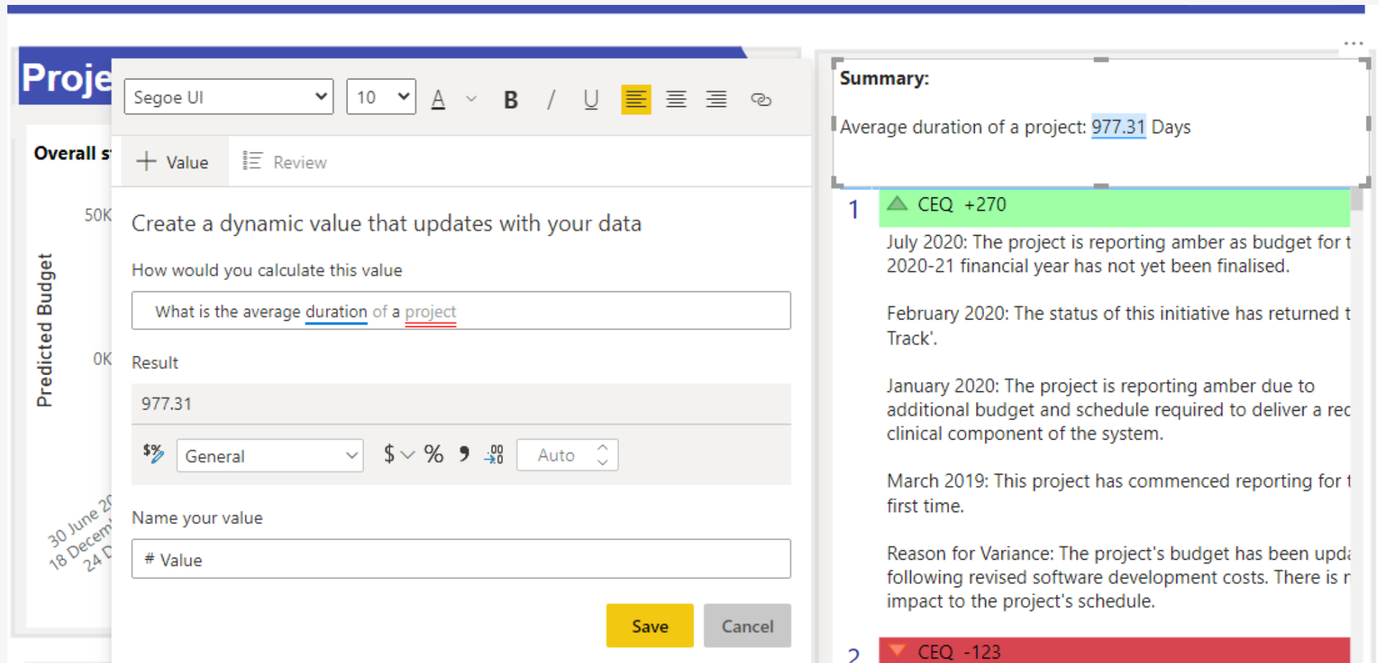
The portfolio manager would want to view this as they may want to understand why some project budgets varied. It allows the manager to drill down per project and follow up in why a project may have been delayed.

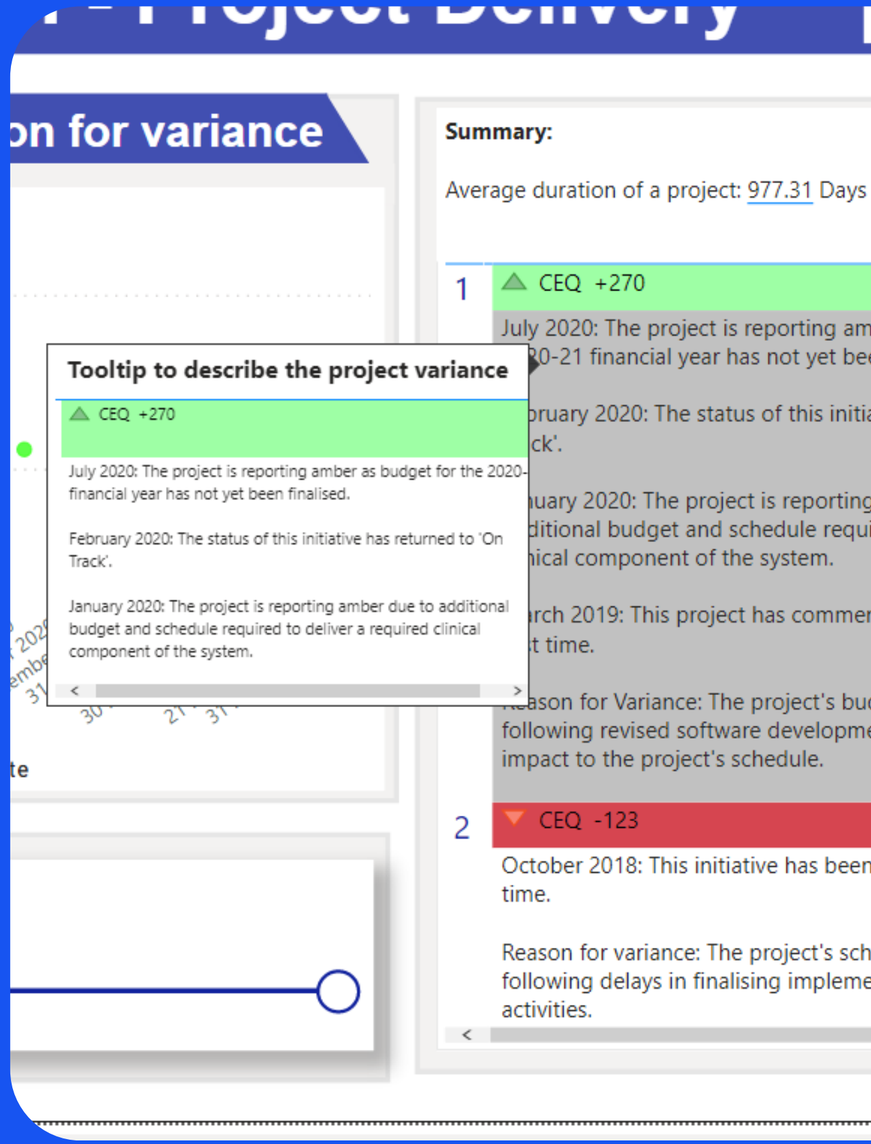


Linking back to the problem statement

I have recently revised this project and improved the hover-over tooltip. Here, you can now identify which projects are Over/Under budget and by how much. Here you can see CEQ overall saved £270 per day.

In addition, you can also see below machine learning has been utilised to answer a proposed question. I asked 'What is the average duration of a project' which provided an answer of 977 days. This is a visualisation known as a smart narrative that adds context to what the data visualisations are displaying and summarises the findings from the overall solution.





How can this be improved?

Through the reasons projects varied lessons learned could be established

Lessons learned would be vital for the portfolio manager as it would ensure that the same reasons a project may have been delayed doesn't happen again.
There can be a documented solution with the best practices in how to deal with problems that may arise within the project delivery.

Forecasting the budgets using a machine learning model

With more time interval data of a project, you can evaluate in what part of the year is the most optimum to start a project. This will create a pipeline of projects as the portfolio manager can understand the most effective time to run a project.
This can also be improved by using feature importance where the factors in why a project is late are all listed and then using machine learning you can extract keywords from the description and identify the likeliness the project will be delayed due to its predisposition to certain factors

Was the problem statements solved?

'As a portfolio manager, I want to look across all of my projects for a leading indicator to identify projects that are going to be over budget so I can implement an intervention and reduce the cost over run.'

The problem statement has been answered to an extent. Through the dashboard, you are able to identify the descriptive statement in why a project is over budget. However, I would further this by extracting the key problem statements so it's easier for the portfolio manager to view the problems at a higher level.

In addition, the 'implementing an intervention' section can be resolved with what I suggested previously by creating a list of lessons learned. This would reduce the cost over run as the portfolio manager is aware of why a certain project may be delayed/over budget so can try to reduce the risk by looking at the best practices. This could be taken from a different site as it could be an environmental/site factor in why the project was delayed.

How will this help?

The Portfolio Manager now has a high level dashboard of why Projects are over budget

The Project Manager can view the dashboard and instead of looking at raw data can evaluate the visualisations from the dashboard which saves the manager time as they do not need to collate or clean the data as it has already been done.

The Portfolio Manager can share this with the site manager/workers

Providing access to other colleagues creates transparency of the project. The site manager and workers understand what factors projects maybe predisposed to and try to avoid them as they understand that this problem has occurred in the past

Predictive "Hit/Release" Modelling

PROJECT 3

Situation

It is important for us to be able to understand when hits or releases are likely to occur within a project so that our team is able to either mitigate potential hits or maximise releases. We want to be able to see the likelihood of a hit happening at least two months in advance based on historical project data.

Task

As a business, we want to be able to understand the likelihood of a hit happening at least two months in advance of occurrence. Can we develop a machine learning model that can predict the probability of a hit using historical precedence of hits occurring.

Action

To build a model based on this historical data to predict the likelihood of either a hit or release occurring for a project.
Train the model on a portion of the data saving the remaining data to test your model. We are also interested to understand which attributes are the most predictive of a hit or release occurring.

Result

A machine learning model that can be run across the entire active project portfolio to predict project hits and releases using metrics that are tracked and stored on a monthly basis.

SITUATION – User Story

Client: I've called this meeting to try and understand why so many of our projects in this portfolio are taking Hits each month and I'm not getting any warnings from my team leads.

Our Users demand consistency and predictability and will go elsewhere for services if we can't improve.

Consultant: I find that by the time we get sight of emerging issues that it's too late to raise early warnings or discuss mitigations with the team without there being some sort of financial impact.

Our processes are just too slow to react and we need more time.

I can produce productivity curves and provide you with forecasts, but they aren't very reliable. The forecast are only based on a basic interpretation of our productivity to date, they're too linear and inflexible.

Client: I wish I had a script to predict when our projects are going to take a hit months in advance.

TASK

Problem Statement Summary

- To Identify and Predict when Gross Margin Variance becomes a hit or a release
- Need 2 months to identify and implement corrective actions

Key Questions

- How many projects are going to experience a hit?
- Who is accountable for the project(s)?
- What is the monetary impact of a hit/release?
- What are the key influencers?

KPIs

- A league table of Projects heading towards a hit or release
- Ranking projects dependant on value as well as probability of pain.
- Maximise financial impact of avoiding hit/release
- Heat map of key Influencers

Data Sources

- Financial data for 360 projects
- User input performance scoring for a variety of disciplines and functions with in the project

ACTION

Applications used

Python – A programming language and software to clean and perform the operations to solve the project

Power Bi – A data visualisation tool where the cleaned data from python will be imported and visualised.

Packages used

Pandas– A data analysis and data manipulation tool to be imported into Python. I have used it for data cleaning.

Scikit Learn– To import the Naive Bayes supervised learning method to determine how well the testing data fits against the model.

Numpy – Imports mathematical functions to coincide with the Gaussian Bayes Model

ACTION

79%

Forecasting Accuracy

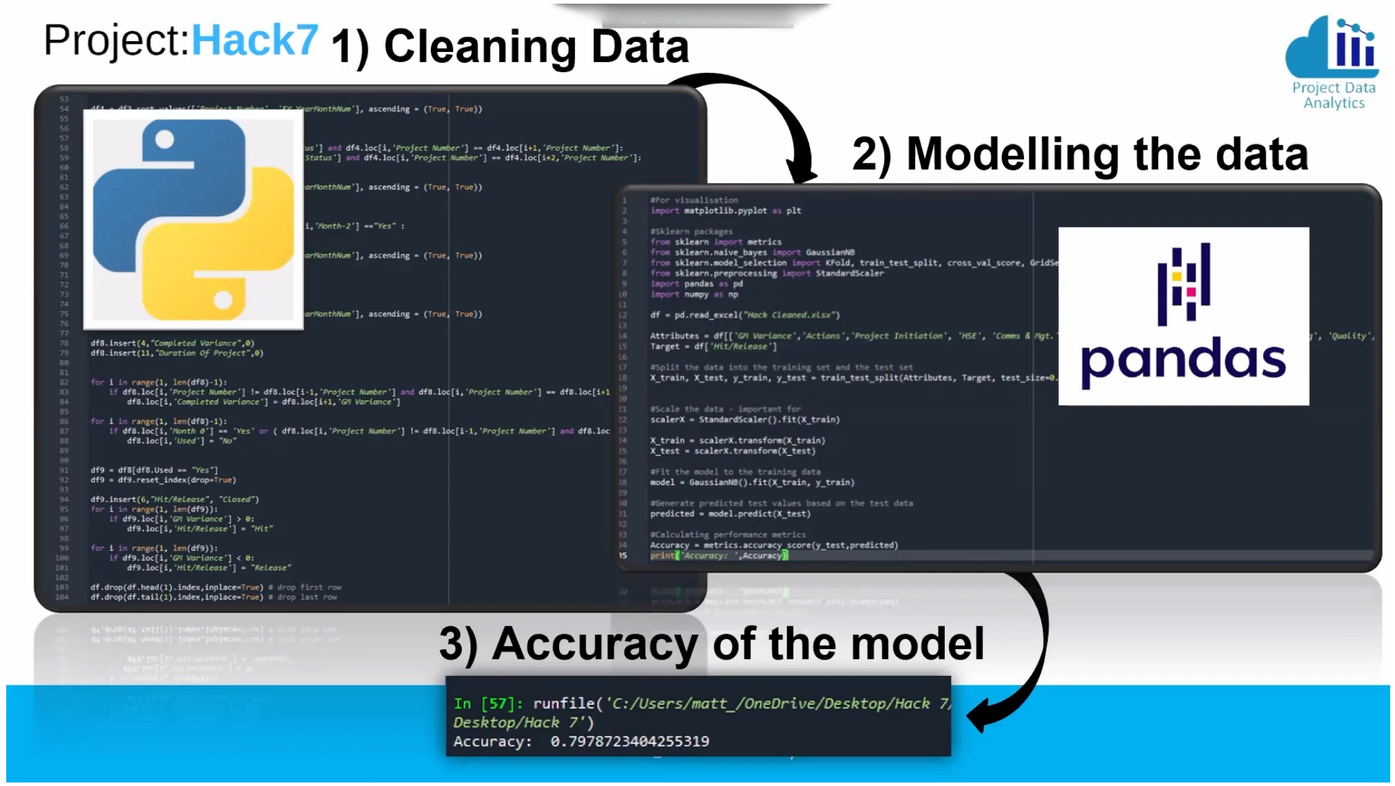
2000

Projects cleaned

The 2 processes we created:

1)Cleaning Data – reduced the redundant rows to ensure that the next stage of modeling the data does not contain extraneous variables
2)Modelling Data – I used various tools as previously mentioned to create a powerful prediction of whether the project will be a hit or release within a Gaussian Naive Bayes model.

From this, I identified that the accuracy of the model is 79% which is a great building block to improving the model in the future.



ACTION

Additionally, we used the Feature Importance technique that assigns a score to input features based on how useful they are in predicting a target variable.

In our case, we used this to look at the financial features of each project and how they influence the target variable GM Variance

FEATURE IMPORTANCE CODE

```
    # fill null values
    numeric_data = numeric_data.fillna(0)


    ########## FEATURE CREATIONS ##########

    # create variable of columns with RAG status
    rag_data = data[['Actions', 'Project Initiation', 'HSE', 'Comms & Mgt.',
                     'Interfaces & GID', 'Change Control', 'Tech & Engg', 'Quality',
                     'Staff & Resources', 'Field & Site', 'Schedule', 'Financial',
                     'Risk & Opp']]

    rag_data = data[['Actions']]#, 'Project Initiation', 'HSE', 'Comms & Mgt.']]

    # fill null values in RAG columns
    rag_data = rag_data.fillna('UNKNOWN')

    # create dummy variables
    #rag_dummies = pd.get_dummies(rag_data, columns=['Actions', 'Project Initiation', 'HSE', 'Comms & Mgt.',
    #                 'Interfaces & GID', 'Change Control', 'Tech & Engg', 'Quality',
    #                 'Staff & Resources', 'Field & Site', 'Schedule', 'Financial',
    #                 'Risk & Opp'])
    rag_dummies = rag_data.astype('category')

    # convert dates to datetime
    #data['date'] = pd.to_datetime(data['

    ########### SCALE DATA ############

    # create scalers
    y_scaler = MinMaxScaler(feature_range
    x_scaler = MinMaxScaler(feature_range

    # scale y data
    y_data = y_scaler.fit_transform(data[
    #y_data = data[['GM Variance']]

    # concatenate x data
    #x_data = pd.concat([numeric_data,rag_
    x_data=numeric_data

    # scale x variables
    x_data = pd.DataFrame(x_scaler.fit_tr
```

OUTPUT

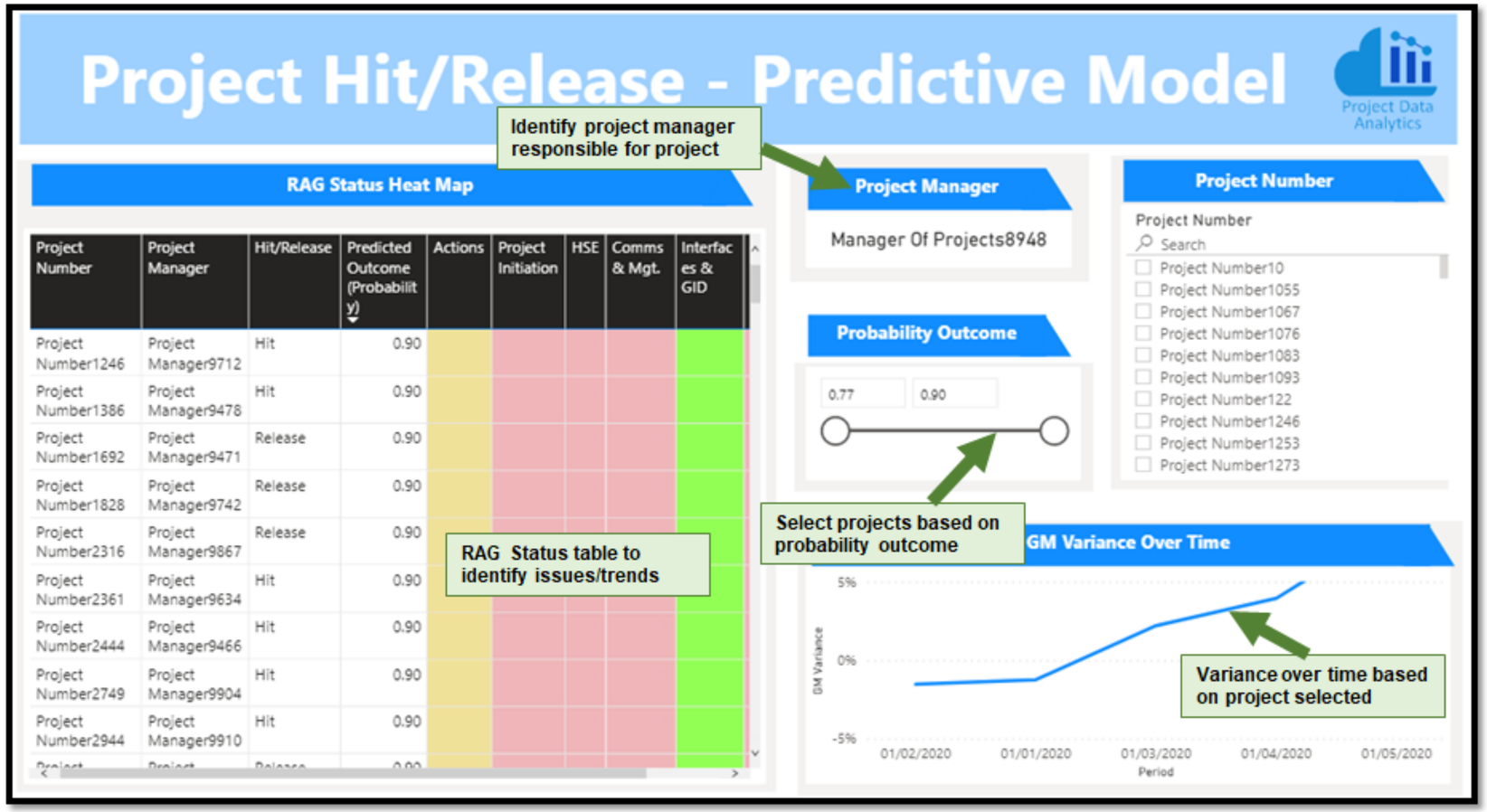
	Parameter	Std. Err.	T-stat	P-value	Lower CI	Upper CI
POC %	0.0004	0.0013	0.3031	0.7620	-0.0021	0.0029
USD Total TP	1.4627	0.3409	4.2908	0.0000	0.7918	2.1336
USD Base TP	0.2900	0.0324	8.9474	0.0000	0.2262	0.3538
USD VC	0.0013	0.0057	0.2234	0.8234	-0.0099	0.0124
USD PTD Cost	0.7686	0.3285	2.3395	0.0200	0.1220	1.4151
USD PTD Rev	-1.1821	0.5415	-2.1830	0.0298	-2.2478	-0.1163
USD PTD GM	0.1544	0.0648	2.3826	0.0178	0.0269	0.2820
USD PTD GM/h	0.0115	0.0063	1.8148	0.0706	-0.0010	0.0239
USD EAC Rev	-0.4077	0.0835	-4.8827	0.0000	-0.5721	-0.2434
USD EAC GM/h	-0.0196	0.0070	-2.8181	0.0052	-0.0333	-0.0059
USD ETC Rev	-0.8085	0.0950	-8.5129	0.0000	-0.9954	-0.6216
USD ETC Cost	0.3559	0.0841	4.2335	0.0000	0.1904	0.5213
USD ETC GM	1.3232	0.0999	13.246	0.0000	1.1266	1.5198
USD ETC GM/h	-0.0009	0.0031	-0.2821	0.7781	-0.0071	0.0053
USD Budget	-0.6589	0.0317	-20.756	0.0000	-0.7214	-0.5964
USD CM POC	0.0226	0.0087	2.5919	0.0100	0.0054	0.0398
USD Poc Rev Adjust	-0.0112	0.0097	-1.1564	0.2485	-0.0302	0.0079
EAC GM%	-0.0051	0.0037	-1.3729	0.1709	-0.0125	0.0022
ETC GM%	0.0040	0.0052	0.7741	0.4395	-0.0062	0.0142
PTD GM%	0.0463	0.0094	4.9303	0.0000	0.0278	0.0648
Unbilled 31-60	0.0151	0.0064	2.3464	0.0196	0.0024	0.0278
Unbilled 60+	-0.0227	0.0043	-5.2635	0.0000	-0.0312	-0.0142
Debt 60+	-0.0177	0.0042	-4.2216	0.0000	-0.0259	-0.0094

The smaller the P-value, the more significant the attribute is.

RESULT

To collate the findings, I decided to create a dashboard. This provides an overview of the different projects and the likeliness of a project being on time.

Through this dashboard, Project Managers can put in place timely targeted interventions in the projects most at risk.



How will this help?

Within the project, there are various commercial applications and benefits

01

Aligning with KPIs

Project Hits will often align with missing Client KPI’s on consistency and predictability.

02

Client Perception

This tool will shape client’s perception of the company and their value as a dependable & predictable partner.

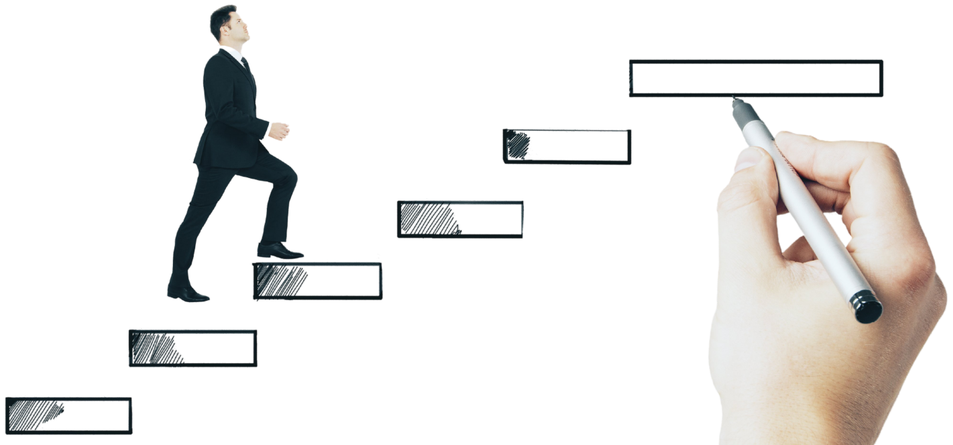
03

Reduced Project Costs

Project costs will be reduced by avoiding expensive interventions required by the time project difficulties become deeply entrenched. Earlier action attracts less project debt, benefiting the company and their Clients.

Future Development

This project provides a lot of growth. I found that in the time I had, it was an exercise of data cleaning and the fundamentals of time-series forecasting. Through richer data, the model will learn on the contract data earlier in the project lifecycle and grow into a powerful and accurate predictive tool.



01

Develop a working front page League Table identifying high probability projects

I was only able to produce a proof of concept in the time provided.

02

Time Series Element

I was unable to do this due to time constraints. However, this provides the foundations of a model that could be added to so that we can predict hits or releases over time. We just need more time phased data. We could go back to the data holder to target the needed data

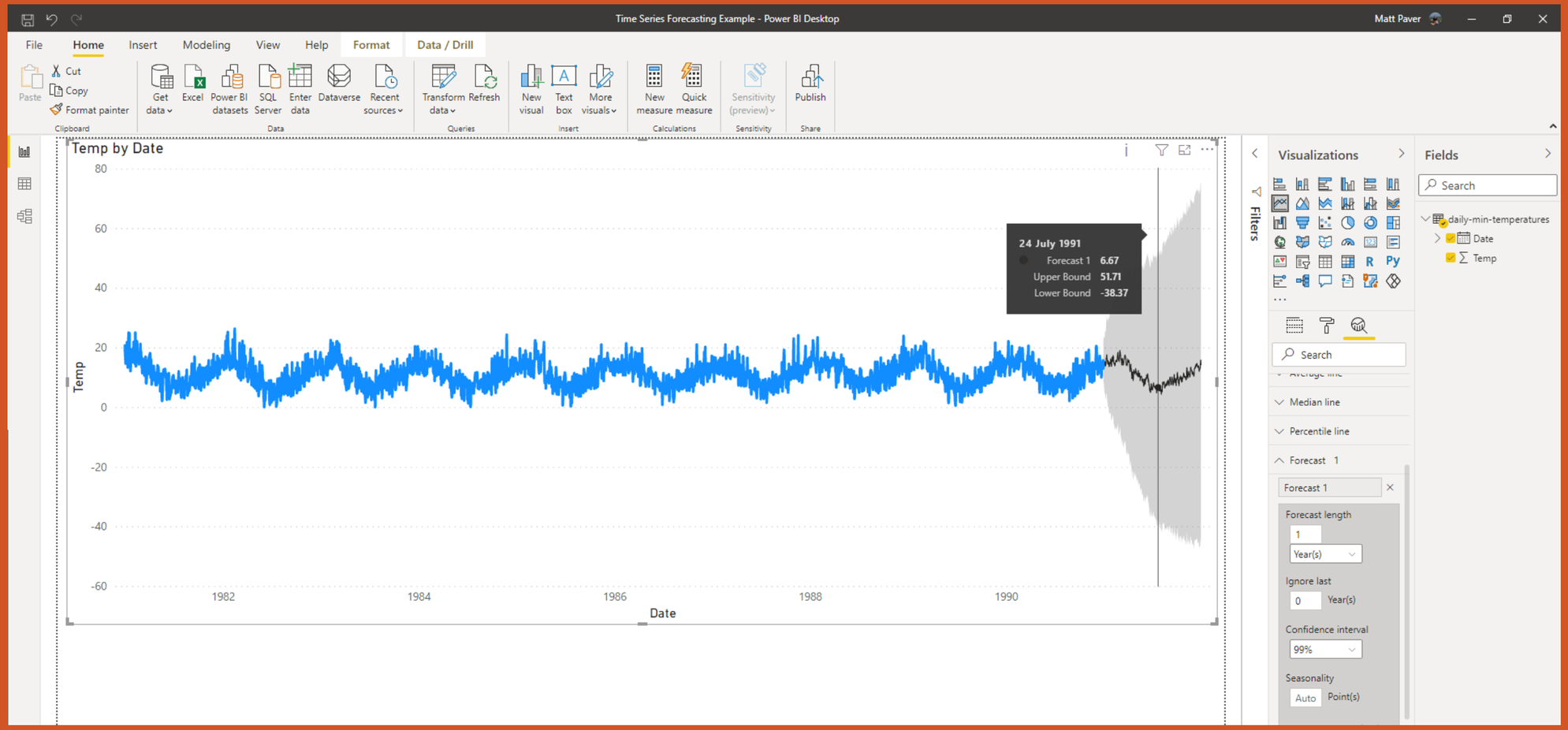
03

Heat Map

Develop heat map in Power BI to identify reasons for high probability projects

Time Series Forecasting

Project 4



Here is an example of how I have used a Time Forecasting method to predict the temperature for the next year from historic data.