

## **Summer Intern Project Report: Manufacturing Technology Transfer Center (MTTC) Benchmark and Training Effectiveness Model**

This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, and was sponsored by (JPL Summer Internship Program) and the National Aeronautics and Space Administration

**Author:** Sara Madjdi-Sorkhabi

**Mentor:** Katina Celio

**Date:** September 31, 2025

**Duration:** 18 weeks

---

### **Executive Summary**

The objective of this project was to enable data driven decisions with metrics that analyze training specific feedback loops to create key performance indicators (KPI) within the “Training Center”: Manufacturing Technology Training Center. This was achieved by conducting a full data analysis on Instruction Build Assemble and Test (IBAT) reports, Inspections Reports (IR), and Course Evaluation to understand workforce capability, identify training deficiencies, and gauge post improvement outcomes. My focus of the project was to run an analysis on one particular JPL required course [Data Sample A].

---

### **1.0 Preliminary Knowledge**

1. What is an IBAT?
  - a. IBAT stands for **Instruction Build Assembly and Test**. It is a controlled document used to record step-by-step instructions for the fabrication, assembly, test, inspection/handling of hardware. There is a record of who created an IBAT and who performed them. The document is a plan created prior to executing an activity. It should capture the instructions and communication amongst all people involved with the IBAT. Cognitive Engineers, Quality Assurance Engineers, and Technicians are involved with the IBAT
2. What is IR?
  - a. IR, Inspection Report, is the action of measuring, examining, and verifying the characteristics of a product and comparing the results against specifications/requirements for compliance. An IR must include observations/findings of an inspection, provide information regarding the state of the hardware, including the context of the inspection, and document the official “stamp of approval”. The purpose of creating an IR is to document hardware conformance against specifications, requirements, configurations, and document hardware

discrepancies and corresponding dispositions. It also provides hardware traceability when investigating inspection history.

3. What is a Cog-E?
  - a. Cog-E is short for Cognizant Engineer. They generate step-by-step instructions about how the hardware is to be built and assembled. These instructions are within the IBAT. Cog-Es are the engineers that construct the necessary IBATs.
4. What is QAE?
  - a. QAE stands for Quality Assurance Engineer, they are responsible for reviewing the instructions within the IBAT and making sure inspection steps are inserted where required. They also are responsible for the Inspection Reports written about the IBATs.
5. What is a technician?
  - a. A Technician is the person that follows the instructions in the IBAT. They create the hardware the Cognitive Engineers instructed for the missions.
6. What is [Data Sample A]?
  - a. JPL's [Data Sample A] is one distinct course intended to teach students (technicians, engineers, and potential instructors) to properly perform workmanship for a specific subset of skills and knowledge based on roles and responsibilities.
7. How are they connected?
  - a. Being certified in [Data Sample A] entails that the employees (all flight technicians, Cog-Es, and QAEs) can accurately execute that specific subset of operations and tasks on processes for JPL instruments and spacecrafts. Some IBATs and IRs require knowledge on [Data Sample A], so taking the course allows for a better understanding of how to construct and successfully perform an IBAT as well as generate more accurate inspections in an efficient manner. There are elements that require the appropriate detail and attention that only JPL flight technicians and engineers provide.

## **1. Project Purpose and Objectives**

### **1.1 Problem Statement**

There are no data analysis performed on IBATs and IRs to determine whether employees such as Cog-Engineers, QA-Engineers, and Technicians had their certification for [Data Sample A] when creating/performing an IBAT or IR that involved knowledge this course provides. In addition, we do not have an analysis on the student evaluation submission for [Data Sample A]. Although there are systems in place that take in information such as the evaluations and reports employees complete, we do not have a metric system that analyzes each of them for performance improvements, correlation detection determining whether what is being taught in the course impacts the missions. If there is an impact, detect the type of impact.

The following are questions that were answered to convert the data to useful information:

- How can we identify if an IBAT or IR requires [Data Sample A]?
- How to identify if the employees assigned to the reports for [Data Sample A] tasks were certified to [Data Sample A] during the duration of the time they worked on the report?
- How can we cross reference the two reports?
- How can I use the course evaluation and IRs to improve the current [Data Sample A] course or monitor the impact from a change to the course?

This project also had a beyond-the-scope assignment if there was time to complete. Currently, the training center does not have a dashboard that documents the current Specifications the curriculum is following. This can cause inconveniences when Specification Documents are revised and published while the course curriculum follows an old version. Not having the most up-to-date Specification Document can hinder the structure of expectations JPL has on their employees.

## **1.2 Project Goals**

To perform an analysis on cleaned IBATs and IRs within a specific timeframe of interest [Population Sample A] that involves [Data Sample A]. In addition, to perform a data analysis on the course evaluations that students complete at the end of their training to determine the students' opinion on the course and whether the course improved. Lastly, create a system that alerts the instructors when a new Document is revised by JPL that affects the course material. This allows the Manufacturing Technology Transfer Center to keep their course material up to date and aligned with JPL and NASA's expectations. To achieve the following I portioned my project into 4 assignments...

1. Come up with metrics that can extract IBATs and IRs that involve [Data Sample A]
2. Determine whether the people were [Data Sample A] certified during the start and completion duration for the instructions in the IBAT or writing the IRs were [Data Sample A] Certified.
3. Provide visual data analysis of the course evaluations from the student. Determine if the course improved their performance, their opinion on the course.
4. The course is structured based on Specification Documents. If a specification document is revised and updated, theoretically the course also should be revised and updated if the document is being used in the course. I am to create a dashboard that highlights all the specification documents used in the [Data Sample A] course. This dashboard will include the current specification document being used, the most recent specification document, and when the document was last revised. When a specification document is updated on the Internal JPL website, the dashboard will alert the instructors using that document. This portion of the project is to help the training center have more accurate courses aligned with the requirements. This result allows for improvements in efficiency,

accuracy, and safety, and promotes cost effective results when working with flight hardware.

### 1.3 Success Criteria

My definition of success for this project depends on the level of difficulty involved. Given that I attempted something that has never been successfully done before, one major measure of success would be developing an accurate and generalized algorithm that is supported by clear, well-organized code documentation. This alone would be a significant achievement, as it lays a strong foundation for future enhancements by other data scientists and analysts.

Another key indicator of success is whether team leaders can easily interpret large volumes of data through the visual outputs I've created. If visualizations enable effective, data-driven decisions, then the system is serving its purpose. Additionally, building a scalable template that evolves with improved datasets will ensure continued accuracy and adaptability over time.

Considering the initial datasets were raw, complex, and potentially inaccurate, simply creating a reliable and efficient system that supports the training center's goals would validate the success of my efforts as both a data scientist and analyst.

### 1.4 Scope and Limitations

The following datasets were given to me: Course Evaluations, Training Certificate, and IRs. These datasets were made fully available, with rarely any limitations on usage. They played an important role in analysis and algorithm development. Although there were some challenges, they were able to be resolved overtime.

In addition to the prior datasets listed, there were two used for this project that required extraction from a low-level internal database using custom SQL queries. Two datasets on the IBATs were utilized:

1. The IBAT dataset containing entries likely related to the course [Data Sample A].
2. The second IBAT dataset lists all accepted instructions tied to the [Population Sample A].

The SQL queries used to extract and filter this data will be included in the process documentation.

While this data source was deemed the most relevant and appropriate metric system during my internship, it introduced several limitations and uncertainties regarding the accuracy and reliability of both the data and the results derived from it.

One major constraint was the **lack of structure and standardization** in the data. The IBATs were often written in freeform text for the convenience of engineers, making it difficult to search for key terms such as "[Data Sample A]". This inconsistency in formatting limited our ability to programmatically classify or analyze the data effectively.

Another significant limitation was the **absence of technician attribution**. There is no standardized column in the database where technicians log their names or sign off on completed work. In some cases, engineers would sign off work as technicians completed on behalf of them. As a result, once data is stored in the low-level system, it becomes difficult, if not impossible, to trace who performed a specific action. This is especially problematic, as technician certification is a critical compliance factor; only certified individuals are authorized to carry out specific IBAT instructions. Without this data, we could not reliably verify whether the proper personnel were involved in the execution of each task.

These limitations influenced both the design and interpretation of the algorithm and analysis, and they represent areas where further data governance and system improvements could yield significantly better outcomes in future iterations.

## 2. Methodology and Process

### 2.1 Overall Approach

The steps taken to complete the first project are listed below in a high-level strategy:

1. Create a to do list and the reasoning behind the task at hand. Construct ideas on what tools and knowledge are required to achieve the purpose of the project; this correlates to figuring out the necessary datasets, how and where to process the data for cleaning and visualization purposes. Create another list of datasets needed along with the necessary features.
2. Research where the datasets are stored, what each dataset signifies, and how to obtain the dataset. Gathered all the required datasets.
3. Upload datasets on to Visual Studio Code for data preprocessing execution. Data preprocessing includes understanding the current data and making alterations to improve the dataset for proper data visualizations. This process includes feature engineering and handling missing data.
4. Once the datasets are cleaned, merge the datasets.
5. Approve final datasets to perform data visualizations techniques that will help answer the questions we had created from step 1 from this high-level approach

The process of this project followed the **DMAIC** acronym, which stands for **Define, Measure, Analyze, Improve, and Control**. At the beginning of any project, it's important to be patient and take the time to fully understand the problem at hand. While this initial reflection may seem unproductive, it lays the foundation for better performance by enabling a deeper grasp of the project's purpose. This understanding not only clarifies the task but also connects it to the broader goals of the organization. I prioritized giving ample, uninterrupted time to the first phase, Define. A well-defined project serves as a guide for the remaining phases and becomes a reliable reference point when challenges arise.

### 2.2 Process Timeline

Duration of my internship: 18 weeks

WEEK 1: Get situated at JPL (badging, computer pick up), and review summer/project expectations. Understand the training center flow, the motivation behind the training center, and how your project plays a key role for training optimization. Request access for required websites.

WEEK 2-3: Understanding the Key terminology needed to start the project (IBAT, QAE, Tech, etc.) and how to use the required systems to retrieve the data needed. Understand how all the systems connect on a conceptual basis, retrieved all data from required documents. Create road maps and list possible challenges that will be faced. Reach out to people that can help make your project possible.

WEEK 4-5: Load and clean the datasets. This requires inserting new necessary columns and removing unnecessary columns

WEEK 6: Apply necessary mergers amongst all datasets, consider the necessary visuals for all datasets needing to be visualized

WEEK 7-9: Once datasets are finalized, load to Power BI to apply data visualizations to support your analysis and conclusions. Reassess metric system used to improve accuracy.

WEEK 10-11: Complete the Report and Presentation work for project 1

WEEK 12-13: Start the thought process for D-documentation Dashboard Alert. Create basic interactive Interface. Contact necessary people to access approval and Integration process

WEEK 14-18: Start the Dashboard for D-Doc Spec Alert, Write down possible/current complications. Record in Report and Presentation

## **2.3 Tools and Resources**

### **Project 1(Training Center Analytics):**

- Resources accessed:
  - Quality Assurance website for IR:
  - JPL internal website-> IBAT website
  - Training Center Database for Training Certification
  - [Data Sample A] Handbook
  - JPL GitHub for storing code on the cloud
  - Badge extraction
  - Student Course Evaluation Dataset
- Frameworks:
  - Excel
  - Visual Studio Code
  - Python

- o JPL GitHub
- o SQL: for data extraction for IBAT
- o Power BI

## **Project 2 (D-doc Dashboard Alert):**

- Resources accessed:
  - o Internal JPL document repository for documents
  - o JPL GitHub for storing code on the cloud
- Frameworks:
  - o Visual Studio Code
  - o SQL
  - o Python
  - o JPL GitHub
- Another resource was reaching out to people for more information about the various systems within JPL

## **2.4 Key Decisions and Rationale**

### **Project 1**

**Data Extraction:** Currently, there is not a feature for IBATs and IRs that can filter based on a course or unique components of a course. This made it a challenge to extract all the data that involved [Data Sample A]. My metrics used to extract the IRs was to load all the [Population Sample A] data and manually parse through the reports. For the IBATs, two different datasets were used: the whole [Population Sample A] data with the status being accepted, the data with filters relating to [Data Sample A]. The second dataset was used as the main dataset and whatever that was not found in that dataset was included in [Data Sample A], was appended from the [Population Sample A] full dataset. There were very few correlations between the IBATs and IRs, which indicates the some crucial IBATs were missing; that was the rationale for having two IBAT datasets. After careful consideration, the issue was with the IBATs and not the IRs because majority of the IRs were related to [Data Sample A] when skimming; we found the IRs to be more reliable than the IBATs. To improve this project, the next person should consider a better method for the initially extracted data. The better the data is, the more accurate the data analysis is to the true values.

**Data Algorithm decisions:** The operation or project determines the requirement of certified employees for IBATs or IRs. In addition, there were debates on whether Cognitive Engineers should be certified in relevant courses. These constraints were taken into consideration when creating the algorithm that calculates the Final Certification Results. The algorithm does not include the requirement for Cognitive Engineers (Cog-E) to be certified in the total count, and instead we compared the average certified Quality Assurance Engineers (QAE) to our threshold (50%). In this case the threshold meant either half or majority of the people must be certified to

consider the report was status as “Certified Workers”. The second key decision relates only to IBATs. The IBAT dataset could not extract Technician information like the Cog-E and QAE. Because of this issue, we could not check their certification status. Technicians are to follow and perform the IBAT instructions, therefore, they are top priority for certification status. This is a constraint worth pursuing further to create a more accurate and reliable certification analysis for IBATs.

---

### 3. Implementation Details

#### 3.1 [Phase 1 /Gathering the Datasets]

SQL Query Ran to retrieve IBAT dataset with [Data Sample A] key words:

```
"select h.7777_number, h.7777_type, h.project, h.77777, s.////_number, s.buy_off_9999999, s.buy_off_9999999_sign_date, s.buy_off_qa,
s.buy_off_88_sign_date , i.*
from 777.edw_7777_header_v h, 777.edw_7777_77777_v s, 777.777_instructions i
where l=1
and h.777_type = '999999'
and h.project = '[Population Sample A]'
and h.777777 = '88888888'
and s.7777_number_id = h.7777_number_id
and h.7777_number_id = i.7777_entity_id
and s.////_number = i.////_number
and i.7777_entity_id = h.7777_number_id
and (i.instructions like '%11111%'
or i.instructions like '%22222%'
or i.instructions like '%33333%'
or i.instructions like '%444%'
or i.instructions like '%5555555%'
or i.instructions like '%66666666666%'
or i.instructions like '%7777%'
or i.instructions like '%88888%'
or i.instructions like '%99999%'
or i.instructions like '%00000%'
or i.instructions like '%####%'
or i.instructions like '%^^^^%'
or i.instructions like '%&&&&%'
or i.instructions like '%(((('%
or i.instructions like '%)))))%'
or i.instructions like '%*****%')"
```

Filters used to retrieve IR dataset:

- Status = ‘Closed’
- Project = “[Population Sample A]”
- Type = “777777”
- Once data retrieved, skimmed through instructions to keep following that were [Data Sample A] related

Filters used to retrieve [Data Sample A] Certified dataset:

- All JPLers with some type of certification of [Data Sample A] during [MM/YYYY]- [MM/YYYY]-

Filters used to retrieve Course Evaluation dataset:

- All students who completed the Course Evaluation Survey that took [Data Sample A]

### 3.2 [Phase 2/Cleaning Datasets and Applying Visualizations]

Detailed description of project tasks:

#### Project 1:

- Removed empty entries in the Cog-E name and QAE name columns for IBATs and IRs.
- Applied feature engineering techniques to clean the dataset and keep columns useful for this project.
  - This process includes reducing unnecessary premade columns and adding columns necessary for merging and visual analysis.
- Build a function that separates the Full Name and Organization of a JPL employee and uses the information to retrieve their badge number from a different website.
- An algorithm that crosses references the [Data Sample A] certification dataset, retrieves the data, calculates the total people certified, calculates the average people certified and compares it to the threshold for final certification indication.
  - To indicate a JPL employee is certified, we crossed referenced the badge numbers and if they had a certification during the entire time they were involved with the IBAT or IR. Not everyone was involved from the start to end of the IBAT. I felt it was fair to only cross reference their certification for the time **they** were involved.
- Using Power BI to create a Dashboard of Visualizations that helps the team understand the mass data
  - A visual of the ratio, certified during IR to non-certified IR AND certified during IBAT to non-certified IBAT
    - Impact of certification at time of operation performed
    - Measure compliance with institutional and project requirements
    - For the future: Can use this visual to understand the certified IR to non-certified IR ratio trends over the course of many projects.
  - A visual of how many IRs do not have an IBAT
    - It is not mandatory for all IRs to have an IBAT attached to them. However, if it refers to an IBAT, it **MUST** have the IBAT number in the report as mentioned in the IR training course. This visual helps filter out non-workmanship related IR's. The program does the search for you and directly indicates which IRs the team should look at to make sure IR training course protocols are executed.

- o Pareto Chart for defect-code used in IRs
  - The main purpose of defect-codes is to help organize the type of discrepancy witnessed during the inspection. Based on the assumption that all the IRs we have gathered from the [Population Sample A] is related to [Data Sample A]. Noticing the most frequent defect-code listed, this can help determine what the focus in material the course instructors should have for [Data Sample A].
- Using Power BI to create an interactive dashboard to capture analysis on Course Evaluations. The purpose is to measure knowledge and engagement gaps to strategically plan impactful and efficient ways we utilize our limited resources in the training center.
  - o Prior to visuals, there are some data cleaning and calculations that needed to be done. The categorical rating criteria was converted to numerical for better analysis trends. For example, that step was crucial to take the overall average rating and create trending visuals.
  - o A filter to distinguish the timeline user wishes to dissect
  - o Information icon for type of data being analyzed
  - o Overall Average Rating of the course based on Quarter and Year? This overall rating includes the student's opinion on the instructor, work environment, course, preparation, materials, etc. Used a Line graph and Gauge chart to show overall rating distribution
  - o Bar graph for student and type of certification taken
  - o Work environment rating distribution
  - o Total number of Submitted Course Evaluations from 0-to-100 (100 being the goal set)
- Refer to Section 5 (Results and Outcomes) to see the illustrated metrics.

**Project 2:**

- The following are needed to create a dynamic feedback loop for Curriculum/ Specification Document tracker
  - o Create a database that stores datasets regarding the courses and specification documents they use along with the date they implemented this specification document into the course curriculum.
  - o Create an interface that allows for updates to this dataset when they update the course curriculum.

- Connect this database to RA.
- This allows for the team members in charge of RA to create an interactive dashboard that helps the training center identify which course curriculums need to be updated.

### **3.3 [Phase 3/Refinement] (last update 7/23/25)**

After project execution, code revisions were made to improve organization and conciseness. Current folders and files were refined, and an OOP (Object Oriented Programming) format was implemented to the code. Reduced code redundancies within the data preprocessing portion by creating necessary classes.

---

## **4. Challenges and Adaptations**

### **4.1.1 Technical Challenges [Project 1]**

The following bullets are the key challenges of project 1:

- 1. Retrieving IBATs and IRs that relate to [Data Sample A] Course accurately.**
  - a. There is no system in place that can help filter out IBATs and IRs relating to [Data Sample A]. Because of that the first month of this project was to learn about [Data Sample A] and the key words in [Data Sample A] that can be used to extract reliable data. This process took excessive time, and the accuracy might not be that great considering, me a data analyst pulled the data based on my month-long knowledge of [Data Sample A]. In addition, since IBATs are written in freeform, engineers vary in how they phrase concepts relating to [Data Sample A]. Due to the freeform structure, the possibilities of expressing a step that relates to [Data Sample A] knowledge are endless.**
- 2. Retrieving the badge number for each JPL employees involved in IBAT and IR**
  - a. The IBAT dataset includes the full name and organization number of the employees that performed certain tasks. Because people's name change and employees can transfer to different organizations, relying on those two factors for checking if they were certified is not probable. Instead, we propose to use the current information to extract their badge number, a constant and reliable form of authentication.**
  - b. Another challenge we had to face was the lack of information on the technicians working on the IBATs. The missing but crucial piece of data is due to a process flow operation. Many of the engineers sign off on behalf of the technicians. This is a major issue when it comes to data analysis because the data we extract does not consider that the engineer signed off on the**

technician's behalf. This means from a data perspective, that valuable information is lost.

3. **There is no common form of assessment to analyze how the [Data Sample A] course is improving IBATs and IRs.**
  - a. **To determine whether the IBATs and IRs are improving by the course provided, there should be common work for assessment that both the IRs and Course curriculum should have. Although both have their own form of assessment, since there isn't a common form amongst the two, from a systems level someone will need to determine a way to seek data correlation to draw conclusions and complete this part of the project.**

#### **4.1.2 Technical Challenges [Project 2]**

**The following bullets are the key challenges of project 2:**

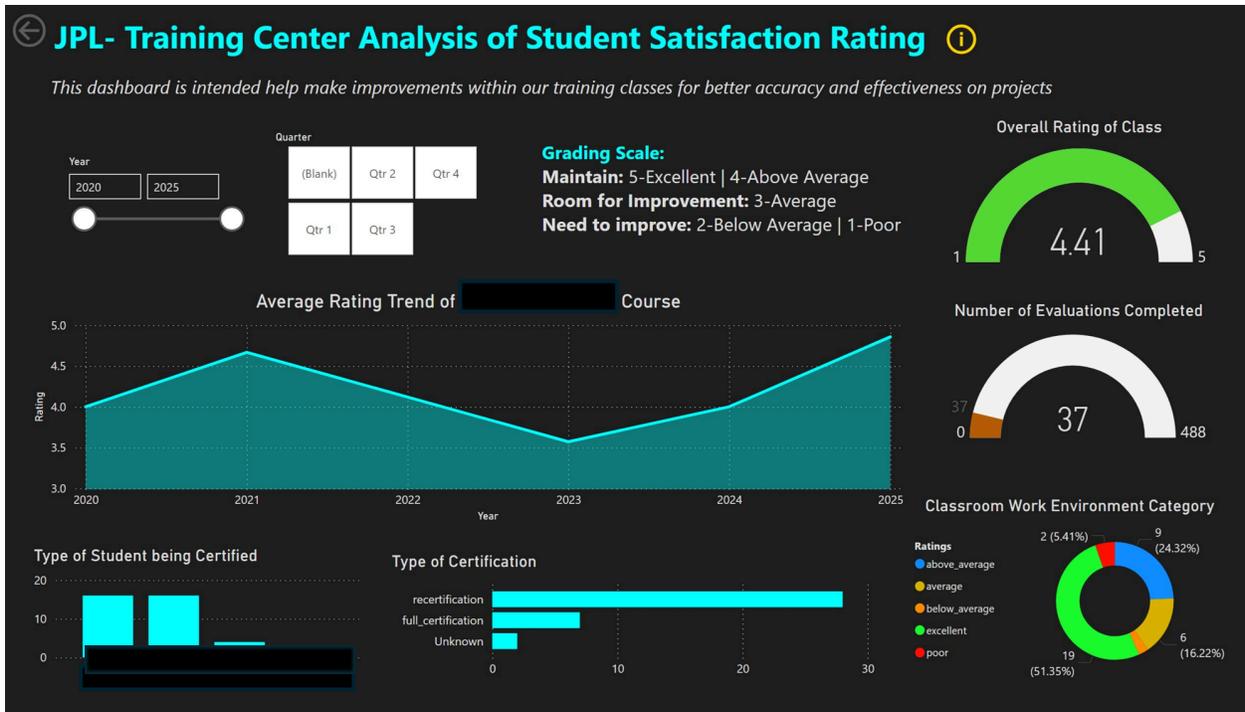
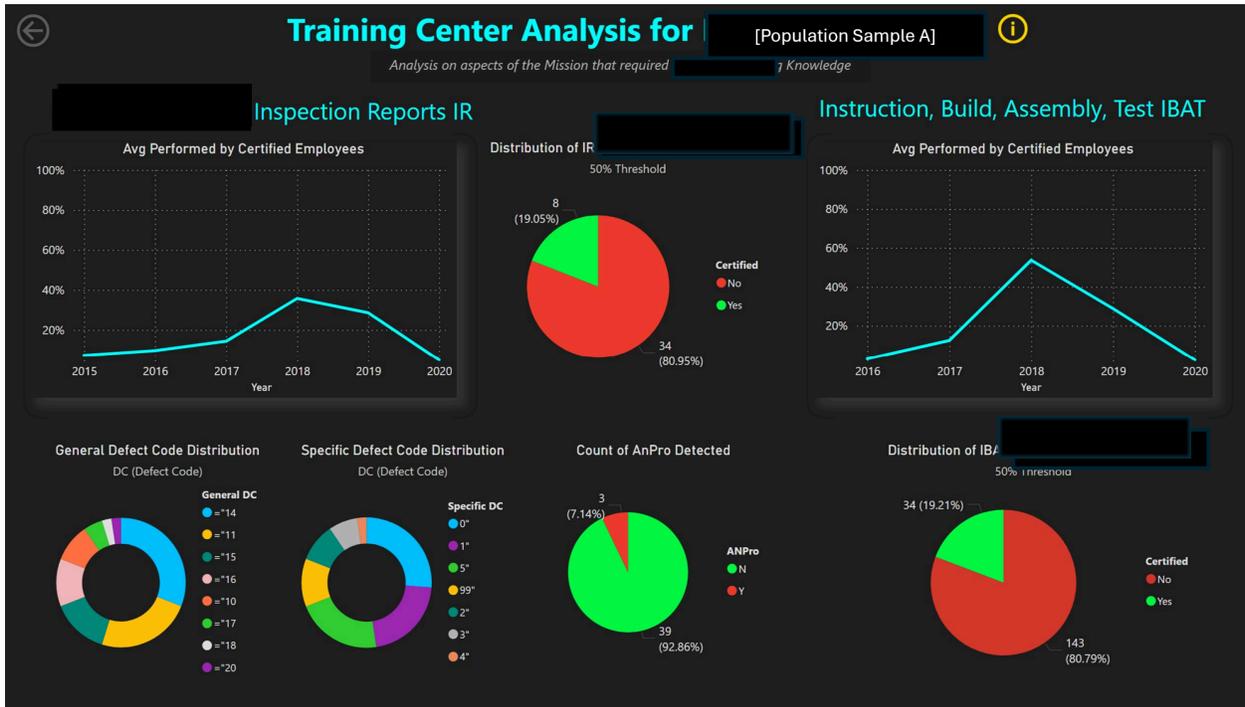
1. **There is no feature that allows for the updates on multiple documents in a dataset on a cadence.**
  - a. **The current INTERNAL JPL WEBSITE and RA system only allows employees to subscribe to a specific Specification Document not multiple. The subscription sends the user an update regarding changes to the revision status on the Spec-Doc. I wanted to access the data in an automated way and upload the data onto a Dashboard for the JPL leads in the 512 division. This data and dashboard are crucial because it will automatically keep track of the updates, and which courses may be affected by the update.**
2. **Coming up with a script that will save the data into a database that can be easily accessed. INTERNAL JPL WEBSITE does not allow for any APIs attached to it. The following are some attempted ideas that have failed:**
  - a. **Use an SFTP, secure file transfer protocol, client to put the necessary files onto a server. This server can be linked to the dashboard for a user-friendly interface. SFTP failed because RA page does not allow data to be sent to a server. Data can only be sent through email.**
  - b. **Web scraping was another method, however, due to the credentials and log in information required to access INTERNAL JPL WEBSITE and RA, that is not an option either.**

**We concluded that there is no way to create this dashboard for the training center on its own server. Since INTERNAL JPL WEBSITE can only connect with RA for consistent, up-to-date information about D-Docs, the best approach for this project is to collaborate with the RA team to create this dashboard for the training center within the RA webpage. RA already has a built in API with INTERNAL JPL WEBSITE, so the data flow is already in place.**

---

## 5. Results and Outcomes

### 5.1.1 What Was Delivered [Project 1]



### 5.1.2 What Was Delivered [Project 2]

For reference, this image was taken on 9/22/2025. This date is relevant to determine the correct calculation for the time lapse.

Course	Specification Document	Version in Use	Recent Version	Recent Revision Date	Time Lapse
Insert Course	Insert a D-doc	v1.0	v1.2	7/4/2025	80 days behind
		v2.0	v2.0	7/8/2025	Up to date

### 5.2 Performance Metrics

I was impressed with the outcome and how easy it was to create an interactive dashboard using Power BI. There are still more enhancements and abilities that can be integrated within this project. However, the fundamentals set during this project went beyond my expectations.

---

## 6. Future Improvements and Recommendations

### 6.1 Immediate Next Steps

**Improvement:**

The Student Evaluation Submission Dashboard is an analysis on Static Data. An immediate fix can be to make it a dynamic dataset with a desired cadence of mentor’s choice.

**Rationale:**

This is important because it will make the dashboard a real time analysis

**Implementation Approach:**

Build a data pipeline that automates the code to extract data on a cadence. To build a pipeline, you must identify the data source, write a program script, choose a scheduling tool, select a Candance to automate the workflow. Currently data is not that big, however, if it weren’t to get larger, use an incremental extraction instead of full extraction for better efficiency

**Expected Impact:**

There is no need for someone to consistently update the dataset to have the most recent analysis. Since analysis would no longer depend on an employee updating the dataset frequently, it is more accurate and efficient.

### 6.2 Medium-term Enhancements

### **Improvement:**

Like the Student Evaluation Submission Dashboard, the IR and IBAT Dashboard Analysis is also Static Data. Unlike the Student Evaluation dataset, for the IBAT and IR Dashboard to use dynamic data, it is most likely not immediate because of the data cleaning and preprocessing implemented.

### **Expected Impact:**

There is no need for someone to consistently update the dataset to have the most recent analysis. Since analysis would no longer depend on an employee updating the dataset frequently, it is more accurate and efficient.

## **6.3 Long-term Vision**

Before this project, several large internal datasets, some containing over 350,000 entries, were not analyzed due to the complexity of accessing and consolidating them. These datasets came from different functional areas and were not originally designed to be cross-referenced, making it difficult to extract broader insights.

The program I implemented solves this by creating a centralized, automated system that retrieves, cleans, and connects data across these sources. In addition, the dashboard I built displays a visual analysis of the cleaned datasets that viewers can interact with. It enables users to identify patterns, track trends over time, and answer strategic questions that had previously gone unexplored.

By surfacing insights from previously untapped data, the dashboard has become a valuable tool for improving visibility, informing long-term planning, and enabling faster, more confident decision-making. Its scalable design also ensures that it can grow with future data needs, supporting ongoing analytical initiatives across the team.

## **6.4 External Approach Worth Exploring**

One of the challenges I encountered during this project was the lack of existing tools or filters to efficiently extract IBAT (Integrated Build and Assembly Tool) and IR (Inspection Report) data related to [Data Sample A] for [Population Sample A]. Because these documents are written in free-text format and use varied language depending on the author, it was difficult to consistently retrieve relevant entries using keyword-based search alone.

To address this, I had to first build my own understanding of [Data Sample A] through internal training materials and then manually construct queries using key terms from that knowledge. While this approach provided some results, the data may still be skewed or incomplete due to inconsistencies in phrasing across different reports.

For future work, I recommend implementing an Artificial Intelligence–based classification system capable of scanning free-text IRs and IBATs and identifying whether they relate to [Data Sample A]. This AI streamline can also suggest reformatting IBATS to allow for easier and cleaner data extraction. This would reduce manual effort and improve accuracy, particularly for analysts who may not have domain expertise in the subject area. Ideally, such a system would allow users to filter data by technical topic, making the retrieval process significantly more efficient and user-friendly.

---

## 7. Knowledge Transfer

### 7.1 Key Files and Documentation

Section 2.3 of this report will give you the material I needed for this project.

Might need to Request Access for the following:

Github:

Power BI:

### 7.2 Dependencies and Setup

What does someone need to know to continue this work?

The following list are sections of this report I believe are **crucial** to read through and understand for someone that plans to continue this work:

- *Executive Summary*
- *Project Purpose and Objectives (1)*
- *Preliminary Knowledge (2.A)*
- *Overall Approach (2.1)*
- *Tools and Resources (2.3)*
- *Key Decisions and Rationale (2.4)*
- *Phase 2/Cleaning Datasets and Applying Visualizations (3.2)*
- *Challenges and Adaptations (4) all the sections listed*

### 7.3 Contacts and Resources

Possible JPL Employees to contact for project understanding and improvement:

- POC#1 (*My Mentor, Training Center Lead Instructor*)
- POC#2 (*Group Supervisor*)
- POC#3 (*Data Engineering, Data Science, Knows the system all around*)
- POC#4 (*Data Science, uses DS to Optimize Training Center*)
- POC#5 (*Hands-on experience with preliminary material, hands-on challenges faced*)

- POC#6 (*IBAT retrieval, handles Database and structure of needed data*)
- POC#7 (*Data Manager, help direct you to people that created the Database*)

## **7.4 Gotchas and Important Notes**

Asking for help, documenting everything, even the things they believe are minimal, and clarifying what is being asked of you. Start by making the project/task/problem as simple as possible, this helps you build a high-level thought process, from there you can expand on the project. This allows you to think strategically, gives you and others something to build off, and the stress/overwhelming feeling decreases. Also give yourself time to process the information. If it is not fully processed, you won't know what to implement and execute. Understanding the purpose of the project allows for questions and uncertainty that have not been considered by you or your mentor to rise. This helps make you become an analytical thinker, and problem solver.

---

## **8. Reflection and Lessons Learned**

### **8.1 What Worked Well**

- Using python to write the programming for data cleaning and integration.
- Using Power BI for visuals.

### **8.2 What Could Be Improved**

- The process is used to extract data involving [Data Sample A]. This was explained more in depth in **6.4 External Approach Worth Exploring**

### **8.3 Skills and Knowledge Gaps**

- Strong foundational understanding of python
  - How to use ChatGPT to your advantage to produce projects at a faster pace
  - Good analytical skills
  - Understanding how to extract data from servers in a cadence
  - Knowledge on how to run an analysis on raw dynamic data
- 

**Document Version:** 2.0

**Last Updated:** 11/5/2025