COOPERATIVE EDUCATION REPORT

A Practical Exploration of Healthcare Data Analytics at Deerhold Ltd.: Challenges Faced, Solutions Applied, and Key Learnings

WRITTEN BY:

Devansh Lath

Student ID: 6508040052

This Report Is Submitted In Partial Fulfillment of the requirements for Cooperative Education

Faculty of Business Administration,

Academic Semester 2/2025

Siam University

Title: A Practical Exploration of Healthcare Data Analytics at Deerhold Ltd.: Challenges Faced, Solutions Applied, and Key Learnings

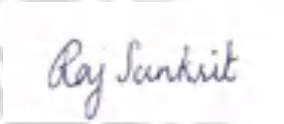
Written By: Mr. Devansh Lath

Faculty: Bachelor of Business Administration (Finance and Investment)

Academic Advisor: Mr. Raj Sankrit

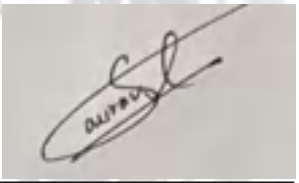
We have approved this cooperative education report as partial fulfillment of the cooperative education program semester 2/2025.

Oral Presentation Committee



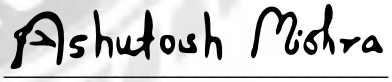
(Mr. Raj Sankrit)

Academic Advisor



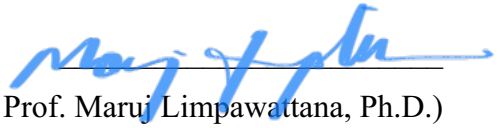
(Mr. Saurav Adhikari)

Job Supervisor



(Mr. Ashutosh Mishra)

Cooperative Education Committee



(Asst. Prof. Maruj Limpawattana, Ph.D.)

Vice President and Director of the Cooperative Education Department

Project Title: A Practical Exploration of Healthcare Data Analytics at Deerhold Ltd.: Challenges Faced, Solutions Applied, and Key Learnings

Credits: 5

Prepared By: Mr. Devansh Lath

Academic Advisor: Mr. Raj Sankrit

Degree: Bachelor of Business Administration

Major: Finance and Investment

Faculty: Business Administration Semester/ Academic Year: 2/2025

Abstract

This report titled ***"A Practical Exploration of Healthcare Data Analytics at Deerhold: Challenges Faced, Solutions Applied, and Key Learnings"*** presents a comprehensive overview of the Co-operative Education experience undertaken at Deerhold as a Data Analyst Intern. The primary objective of the internship was to bridge the gap between theoretical knowledge and practical application in data analytics, healthcare data processing, and ETL pipeline validation. The study highlights key activities, including working with Machine-Readable Files (MRFs), processing JSON and Parquet datasets, performing exploratory data analysis (EDA), and validating large-scale datasets using Python and SQL (Amazon Athena).

The report explores the challenges encountered during the internship, including complex data structures, inconsistent dataset mappings, ETL join issues, and difficulties in tracing data lineage across multiple layers. Various practical approaches were applied to address these issues, such as cross-validation between parsed and ETL datasets, SQL-based verification, and structured root cause analysis. Overall, the internship contributed significantly to the development of technical skills in Python, SQL, and data analysis, while providing valuable insights into real-world healthcare data systems and analytics workflows.

Keywords: Healthcare Data Analytics, Machine-Readable Files (MRFs), ETL Validation, Python, SQL (Athena), Data Quality, Pricing Transparency, Healthcare Claims Data.

ACKNOWLEDGEMENT

I would like to extend my sincere gratitude to Kathmandu College of Management and Siam University for providing me with this cooperative education period. A sincere thanks to my Academic Advisor, Mr. Raj Sankrit, and the faculty members for allowing me to use this internship as part of my learning experience. Their guidance and support greatly helped me in completing this internship report.

I would also like to thank Deerhold Ltd for providing me with the opportunity to undertake this internship and gain valuable practical insights in the field of Data Analytics. This exposure has allowed me to bridge the gap between theoretical knowledge and practical application in a professional setting. My sincere appreciation to my supervisor, Mr. Saurav Adhikari, and the team members of the Data Department at Deerhold for their continuous support during this internship period. Their willingness to share knowledge and involve me in various tasks has significantly contributed to enhancing my practical understanding of data analytics.

Devansh Lath
Student ID: 6508040052

Table of Contents

List of Tables

List of Figures

List of Abbreviations

SQL: Structured Query Language

API: Application Programming Interface

ETL: Extract, Transform, Load

MRF: Machine-Readable Files

HIPPS: Healthcare Insurance Prospective Payment System

BCBS: Blue Cross Blue Shield

UHC: UnitedHealthcare

IPF: Inpatient Psychiatric Facility

IPPS: Inpatient Prospective Payment System

IRF: Inpatient Rehabilitation Facility

MSA: Metropolitan Statistical Area

NPI: National Provider Identifier

CHAPTER 1: INTRODUCTION

Deerhold Ltd. is a premier global software development and data network company headquartered in Lexington, Massachusetts, with its primary offshore development center located in Kathmandu, Nepal. Founded by seasoned technology entrepreneurs, Deerhold specializes in high-capacity data engineering, custom software solutions, and specialized SaaS products for the healthcare industry. The company was established to bridge the gap between complex Western enterprise needs and high-level engineering talent in South Asia and Japan. By leveraging a follow-the-sun operational model, Deerhold provides continuous development cycles that allow international businesses to scale their digital infrastructure rapidly and efficiently.

Figure 1: Company Logo

1.1 Company Profile

Deerhold Ltd. commenced its operations in 2018, built upon a foundation of decades of experience in the U.S. healthcare and technology sectors. The company operates an integrated service ecosystem that encompasses Data Analytics, Cloud/DevOps, AI/Machine Learning, and Custom Application Development. Unlike traditional outsourcing firms, Deerhold focuses on "Deep Tech" challenges, such as parsing massive, unstructured machine-readable files (MRFs) and ensuring HIPAA-compliant data environments. This technical rigor allows the company to serve as a strategic partner for large-scale enterprises that require sophisticated data warehousing and business intelligence tools to remain competitive in a regulated global market.

The company utilizes a diversified business model that pairs high-touch client services with proprietary product innovation. A flagship example is the PRIZM platform, a SaaS solution designed to provide transparency in healthcare pricing. This dual approach—building bespoke solutions for clients while developing scalable internal products—ensures that Deerhold remains at the cutting edge of software architecture. With a workforce of over 500 professionals and

support centers spanning the United States, Nepal, Japan, and Singapore, Deerhold maintains a truly global footprint. The organizational structure is designed around cross-functional pods of data scientists, software engineers, and domain experts who collaborate to solve high-impact problems for the healthcare, fintech, and logistics industries.

1.1.1 Company's Mission

To deliver high-quality, cost-effective digital technology and software development services by leveraging a global network of expert engineers, while providing innovative SaaS-based healthcare solutions that drive transparency and informed decision-making in the U.S. healthcare market.

1.1.2 Company's Vision

To be a global leader in data engineering and software innovation, recognized for creating a seamless bridge between technology and business strategy through a commitment to excellence, integrity, and the continuous growth of our global engineering community.

1.1.3 Product Portfolio

Deerhold Ltd. offers a sophisticated and data-centric portfolio of services and proprietary SaaS products, specifically engineered to support large-scale enterprises in navigating complex digital landscapes. The portfolio is anchored by three primary service pillars—**Software Development, Data Engineering, and Cloud/DevOps**—all of which are complemented by the company's flagship **PRIZM** healthcare analytics suite. This integrated approach allows Deerhold to function as a strategic technology partner, managing the entire data lifecycle from ingestion and cleaning to advanced visualization and compliance-ready reporting.

Product/Service	Description

Custom Software Development	Development of scalable web and mobile applications tailored for enterprise requirements with focus on performance, security, and usability.
Web Application Development	Building interactive and data-driven web platforms with strong backend processing and real-time functionality.
API Development & Integration	Designing and integrating APIs to enable seamless communication between healthcare systems, databases, and third-party services.
Enterprise Application Solutions	Developing enterprise-level systems for large-scale data processing, workflow automation, and reporting needs.

Table 1: Software Development Services

Product/Service	**Description**
Data Engineering & ETL Pipelines	Designing and maintaining ETL pipelines for processing large healthcare datasets (e.g., MRF files) into structured outputs.
Data Processing & Transformation	Handling JSON, Parquet, and other formats and converting them into analysis-ready datasets.
Data Analytics & Validation	Performing data validation, reconciliation, and quality checks across multiple datasets and carriers.
Large-scale Data Processing	Managing and analyzing high-volume healthcare datasets to generate insights and ensure data consistency.

Table 2: Data Engineering & Analytics Services

Product/Module	Description
PRIZM Rates	Benchmarks negotiated healthcare rates across payors and providers for pricing comparison and analysis.
Cost Transparency Tool	Estimates patient-level out-of-pocket costs for medical procedures using real-world pricing data.
Network & Provider Analytics	Compares providers across procedures, codes, and regions to evaluate cost and efficiency.
MRF Analytics Engine	Processes Machine-Readable Files (MRFs) and converts raw healthcare pricing data into structured insights.

Table 3: PRIZM Healthcare Analytics Platform

1.1.4 Company's Core Values

Deerhold operates on a set of fundamental principles that guide its internal culture and client relationships:

- *Value-Driven Results:* Delivering high-fidelity solutions at a fraction of the cost through a robust global support network.
- *Deep Domain Expertise:* Utilizing over 25 years of experience specifically in the healthcare and data analytics space to ensure accuracy and compliance.
- *Global Collaboration:* Fostering a "follow-the-sun" model that integrates talent from the U.S., Nepal, and Japan to ensure continuous productivity.
- *Integrity & Quality:* Maintaining rigorous standards in software quality assurance (QA) and data security (HIPAA compliance).
- *Empowerment:* A commitment to upskilling engineers in developing regions and supporting social innovation projects.

1.1.5 Strategies of the Company

- *Global Delivery and "Follow-the-Sun" Strategy:* Deerhold leverages its geographical diversity by maintaining a 24/7 operational cycle. By coordinating between its U.S. headquarters and its engineering hubs in Nepal and Japan, the company ensures continuous development and support, significantly reducing time-to-market for complex enterprise projects.
- *Domain-Specific Expertise Strategy:* Rather than being a generalist firm, Deerhold focuses heavily on high-barrier industries like U.S. healthcare. By specializing in HIPAA compliance, insurance data structures, and Transparency in Coverage regulations, the company establishes itself as an indispensable strategic partner for fiduciaries and healthcare providers.
- *Data-First Engineering Strategy:* The company prioritizes the structural integrity and scalability of data. Their strategy centers on the "Deep Tech" ability to ingest, parse, and clean massive, unstructured datasets (such as billions of rows in Machine-Readable Files), transforming raw information into refined, actionable intelligence for business decision-making.
- *Hybrid Service-Product Model:* Deerhold balances high-touch custom consulting with the development of proprietary SaaS products like the PRIZM platform. This dual strategy creates diversified revenue streams while allowing the company to use its internal products as a "proof of concept" for its technical engineering capabilities.
- *Rigorous Quality and Security Strategy:* Operating in highly regulated sectors, Deerhold integrates automated Quality Assurance (QA) and strict security protocols into every stage of the lifecycle. Their strategy ensures that every solution is not only functional but also meets international standards for data privacy and cybersecurity.
- *Talent Empowerment and Social Responsibility:* The company follows a "Karma-based" philosophy, investing heavily in the training and upskilling of its engineers in Nepal and Japan. By fostering a high-performance culture and participating in social innovation projects, Deerhold ensures long-term talent retention and a reputation for excellence in the global tech community.

1.2 Organizational Structure

Deerhold Ltd. operates with a global matrix organizational structure designed to bridge its dual identity as a U.S.-based strategic consultancy and a high-capacity offshore engineering firm. This structure is built for scalability and 24/7 productivity, allowing the company to manage complex data engineering and SaaS product development across multiple time zones. Unlike traditional top-down hierarchies, Deerhold utilizes a "pod-based" approach within its departments, ensuring that specialized talent is deeply integrated into every client project.

The leadership is centralized at the executive level in Lexington, Massachusetts, where strategic vision, U.S. healthcare compliance oversight, and business development are managed. This top-tier leadership coordinates directly with the operational and engineering hubs in Kathmandu, Tokyo, and Singapore. This cross-border alignment ensures that high-level business requirements are seamlessly translated into technical execution by specialized teams.

1.2.1 Hierarchy of Management

A. *Top Management (Executive Leadership)*: The executive team is responsible for the company's global strategy, financial health, and technological vision. They operate primarily from the headquarters in Lexington, Massachusetts, and oversee the international subsidiaries.

Founder & Executive Chairman – Dr. Rudra Pandey: A pioneer in the IT industry, he provides the overarching vision for Deerhold, leveraging his experience from previous successful ventures like D2Hawkeye and Deerwalk to guide the company's long-term growth and global partnerships.

Chief Executive Officer (CEO) – Jeff Gasser: Leads the company's day-to-day global operations, steers innovation strategies, and ensures the delivery of high-value solutions to clients in the healthcare and technology sectors.

Chief Administrative Officer (CAO) – Muna Joshi: Oversees the critical pillars of finance, administration, and quality assurance, ensuring the organizational infrastructure supports the company's rapid scaling.

Chief Product Officer (CPO) – Ed Hausman: Drives the product vision and strategy, particularly for the **PRIZM platform**, ensuring that Deerhold's SaaS offerings meet the evolving regulatory and market needs of the U.S. healthcare system.

SVP of Sales & Marketing – Pete Titas: Directs global sales initiatives and market positioning, focusing on building long-term relationships with healthcare payers, TPAs, and enterprise clients.

Figure 2: Executive Leadership

B. *Mid-Level Management (Functional Leadership):*

This level is responsible for the tactical execution of projects and managing the large talent pool across the company's various offshore centers.

- **Director of Nepal Operations – Sachin Karanjit:** Manages the primary engineering hub in Kathmandu, focusing on technology-driven services, local business operations, and maintaining high standards of client satisfaction.
- **Departmental Leads (Data, DevOps, QA):** These individuals oversee specific functional

"Centers of Excellence." They manage teams of software engineers and data analysts, ensuring that agile methodologies and HIPAA-compliant practices are followed across all sprint cycles.

- **Product Owners & Project Managers:** Act as the bridge between the client and the engineering pods, organizing product backlogs and ensuring that delivery timelines align with the executive strategic roadmap.

1.2.2 My Job Position

During my internship, I am working as a **Data Analyst Intern** and supporting the data engineering and analytics teams in processing and analyzing large-scale **healthcare** datasets. My responsibilities include assisting in the ingestion and parsing of Machine-Readable Files (MRF) from major U.S. insurance carriers like Cigna, BCBS, and Aetna. I am utilizing Python and Pandas to perform data cleaning and insight generation, ensuring that complex JSON and CSV files were accurately transformed for downstream analysis. I also gained experience in writing and optimizing **SQL queries (AWS Athena)** to validate data integrity and extract specific provider pricing information.

I am placed under the **Data Analytics/Engineering Department**, working closely with senior data analysts and engineers under direct supervision. The department is responsible for maintaining the architectural backbone of the **PRIZM** platform and ensuring high-fidelity data transparency for the U.S. healthcare market. My role involves supporting routine data validation tasks, checking records for specific identifiers, and assisting in the analysis of Medicare pricing logic and HIPPS/IPPS/OPPS billing codes. This experience helped me understand practical big data workflows, the importance of **HIPAA** compliance, and how massive datasets are managed and visualized in real-world business operations.

The key objectives of my internship were:

- *Understand Healthcare Data Pipelines*: Gain in-depth knowledge of Machine-Readable Files (MRFs) and their role in healthcare pricing and transparency.
- *Perform Data Analysis*: Analyze large-scale datasets using Python and SQL to extract insights and identify patterns.

- *Validate ETL Processes*: Examine ETL pipelines to ensure accurate transformation, mapping, and completeness of data across multiple carriers.
- *Investigate Discrepancies*: Identify and resolve inconsistencies in healthcare datasets.
- *Develop Technical and Analytical Skills*: Strengthen skills in Python, SQL (Athena), and data handling best practices.
- *Generate Actionable Insights*: Produce structured reports and observations that contribute to improving data quality and operational decision-making.

1.3 My Intention and Motivation to Join Deerhold as My Co-Op Studies Workplace

Choosing Deerhold as my Co-op Studies Workplace was driven by my academic interest in data analysis and my growing inclination toward working with real-world datasets. As a student aiming to build a career in the data analytics field, I was particularly interested in gaining practical exposure to large-scale data processing, analytics, and problem-solving in a professional environment. Deerhold's focus on healthcare data analytics and pricing transparency aligned well with my goal of applying theoretical knowledge to real-life data challenges.

Additionally, I was motivated by the opportunity to work with complex datasets such as Machine-Readable Files (MRFs), and to develop hands-on experience with tools like Python and SQL. Being part of an organization that handles large volumes of structured and unstructured data allowed me to understand how data pipelines, ETL processes, and validation techniques function in real-world scenarios. The dynamic and analytical work environment at Deerhold provides an excellent platform to enhance both technical and problem-solving skills. It allows me to improve my ability to analyze data, identify inconsistencies, and derive meaningful insights. Furthermore, collaborating with experienced professionals has helped me gain a deeper understanding of industry practices and standards.

Overall, this internship plays a significant role in strengthening my foundation in data analytics and preparing me for future career opportunities. By being part of Deerhold, I am confident that I will grow both technically and professionally while contributing effectively to the organization's objectives.

1.4 Strategic Analysis of Deerhold (SWOT Analysis)

A SWOT analysis was conducted based on available information and internship observations to understand Deerhold's position in the healthcare data analytics domain and assess its operational and competitive strengths.

Strengths	Weaknesses
Specialization in healthcare data analytics, particularly Machine-Readable Files (MRFs), enabling deep and structured healthcare pricing analysis.	High complexity of healthcare datasets and ETL systems makes debugging, validation, and onboarding challenging for new analysts.
Strong data engineering and analytical capabilities using Python, SQL (Athena), and ETL pipelines for handling large-scale datasets.	Dependence on structured and complete parsed/ETL datasets may lead to discrepancies when data quality issues arise.
Robust ETL and data-processing infrastructure integrating parsed files, master tables, and carrier-level datasets for comprehensive analysis.	Layered joins and transformations make data lineage tracing and interpretation technically demanding.
Domain-driven healthcare transparency products such as Shopping and PRIZM solutions supporting actionable business insights.	Limited simplicity in interpreting highly transformed healthcare datasets and pricing outputs.
Opportunities	**Threats**
Increasing demand for healthcare pricing transparency and analytics solutions in global healthcare markets.	Regulatory changes in healthcare data privacy and transparency policies may impact data accessibility and compliance requirements.
Potential integration of AI and machine learning to improve automation, predictive analytics, and data intelligence.	Growing competition from healthcare analytics companies and large technology firms entering the healthcare data industry.

Expansion of healthcare data products into new payor systems, compliance tools, and healthcare markets.	Data quality and integration issues from external sources may affect analytical reliability and reporting accuracy.
Rising adoption of cloud-based healthcare analytics platforms and large-scale data engineering solutions.	Rapid technological changes may require continuous infrastructure upgrades and workforce upskilling.

Table 4: SWOT Analysis

1.5 Objectives of the Co-Operative Studies

This report serves as a reflection of my learning through Co-operative Education Studies as a Data Analyst Intern at Deerhold. The significance of my report and the objectives of this cooperative education study are:

- To apply theoretical knowledge of data analysis, databases, and programming in real-world healthcare data environments, particularly in working with Machine-Readable Files (MRFs) and large-scale datasets.
- To understand how data pipelines and ETL processes are designed and maintained in a professional setting, including data extraction, transformation, validation, and loading across multiple systems.
- To gain practical experience in handling structured and semi-structured data formats such as JSON and Parquet, and converting them into usable datasets for analysis.
- To develop the ability to use analytical tools such as Python (Pandas) and SQL (Amazon Athena) for data extraction, cleaning, exploration, and validation.
- To improve skills in identifying data inconsistencies, investigating discrepancies across datasets, and performing root cause analysis in complex healthcare pricing data.
- To understand healthcare pricing structures and concepts such as IPPS, OPPS, IRF, and IPF, and how they are reflected and processed in real-world datasets.
- To enhance professional skills such as attention to detail, analytical thinking, problem-solving, and documentation while working in a structured data-driven environment.

CHAPTER 2: CO-OP STUDY ACTIVITIES

2.1 My Job Description

During my internship in the Data Analytics team at Deerhold, my primary role was to assist in analyzing healthcare datasets and supporting data validation processes across multiple systems. This included working with Machine-Readable Files (MRFs), processing large datasets using Python, and supporting ETL validation activities. I was also involved in handling structured and semi-structured data formats such as JSON and Parquet, converting them into usable DataFrames, and performing data cleaning and transformation for analysis. Additionally, I supported exploratory data analysis (EDA) to identify patterns, inconsistencies, and pricing trends within healthcare datasets. Furthermore, I contributed to SQL-based data validation using Amazon Athena by querying large-scale production datasets and cross-checking results with Python-based outputs. I also assisted in investigating discrepancies across different healthcare carriers such as Cigna, BCBS, Aetna, and UHC by comparing parsed, ETL, and master-level datasets.

Overall, my role focused on supporting the data analytics team in ensuring data quality, validating ETL pipelines, and generating meaningful insights from complex healthcare datasets while gaining practical exposure to real-world data engineering and analytics practices.

2.2 My Job Duties and Responsibilities

During my internship in the Data Analytics team at Deerhold, my duties and responsibilities mainly involved supporting healthcare data analysis, validating ETL processes, and assisting in maintaining data accuracy across multiple datasets. I worked with large-scale Machine-Readable Files (MRFs), where I was responsible for processing structured and semi-structured data, performing analysis using Python and SQL, and supporting cross-dataset validation tasks. In addition to data processing, I was also involved in identifying inconsistencies in healthcare pricing data and assisting in their investigation through coordination with senior analysts.

Major Duties and Responsibilities

- *Data Processing and DataFrame Preparation*: Converted large JSON and Parquet-based MRF datasets into structured Pandas DataFrames. Performed data cleaning, transformation, and normalization to prepare datasets for analysis and ensure consistency across different data sources.
- *Exploratory Data Analysis (EDA)*: Conducted exploratory data analysis to identify patterns, pricing distributions, and structural inconsistencies in healthcare datasets. Generated insights to support further validation and investigation tasks.
- *SQL-Based Data Validation (Amazon Athena)*: Used SQL in Amazon Athena to query large production datasets. Performed filtering, joins, and aggregations to validate Python-based outputs and ensure consistency between ETL and parsed datasets.
- *ETL and Data Pipeline Validation*: Assisted in validating ETL workflows by comparing parsed data, ETL outputs, master tables, and carrier-level datasets. Identified discrepancies caused by transformation logic and join operations.
- *Carrier-Level Data Analysis*: Analyzed healthcare pricing data across multiple carriers such as Cigna, BCBS, Aetna, and UHC. Compared IPPS, OPPS, IRF, and IPF datasets to identify inconsistencies in record availability and classification.
- *Root Cause Investigation of Data Issues*: Participated in deep-dive analysis to investigate pricing and classification discrepancies. Identified issues such as one-to-many joins, incorrect code-type mappings, and missing records in parsed datasets.
- *Data Quality and Consistency Checks*: Performed validation checks to ensure accuracy, completeness, and consistency of datasets across multiple layers, including parsed, ETL, and master-level data.

2.3 Activities in Coordinating with Co-Workers

During my internship at Deerhold, I worked closely with my supervisor, senior data analysts, and other team members in the Data Analytics department. Most of my tasks required continuous communication and clarification, especially when dealing with complex datasets, inconsistent records, and unclear mappings in healthcare pricing data. I regularly coordinated with team members to understand data structures, validate ETL logic, and clarify how specific fields and

classifications should be interpreted in different datasets.

Coordination also played an important role when working on shared analytical tasks such as carrier-level comparisons, ETL validation, and deep-dive investigations into IRF and IPF datasets. I frequently discussed discrepancies in data outputs, sought clarification on pricing logic, and updated analysis based on feedback from supervisors. In several cases, I also collaborated with team members while cross-checking results between Python outputs and SQL (Athena) queries to ensure consistency and accuracy.

This collaborative working environment helped me understand the importance of teamwork, clear communication, and systematic verification in maintaining reliable data pipelines and ensuring accurate analytical outcomes.

2.4 Contribution as a Co-op Student in the Company

During my Co-operative Education period at Deerhold, I contributed to the Data Analytics team by supporting data processing, validation, and analysis tasks. I assisted in converting raw Machine-Readable Files (MRFs) into structured datasets using Python, which improved data usability for further analysis. I also supported exploratory data analysis (EDA) and SQL-based validation using Amazon Athena to ensure data consistency across multiple datasets. Additionally, I contributed to ETL validation by identifying discrepancies in carrier-level data and assisting in cross-checking parsed, ETL, and master datasets. Furthermore, I helped in investigating data quality issues such as missing records, incorrect mappings, and pricing inconsistencies across healthcare categories. Overall, my contributions supported data accuracy, validation processes, and improved reliability of analytical outputs.

CHAPTER 3: LEARNING PROCESS

3.1 Challenges/ Problems Encountered

During my internship at Deerhold, I encountered several practical challenges while working with large-scale healthcare datasets and ETL validation processes. These challenges mainly involved data quality issues, system complexity, and inconsistencies across multiple data sources. The major challenges are summarized below:

- *Complexity of Healthcare Data and ETL Pipelines*: Healthcare datasets such as Machine-Readable Files (MRFs) are highly complex, involving multiple nested structures, transformations, and joins. According to IBM (2023), complex data pipelines often increase the risk of transformation errors and make debugging difficult due to multiple processing layers. Similarly, while working with ETL workflows, understanding how raw data was transformed into final analytical outputs required deep analysis of multiple stages including parsed, ETL, and master datasets.
- *Data Inconsistencies and Missing Records:* One of the major challenges was dealing with inconsistencies across datasets from different healthcare carriers such as Cigna, BCBS, Aetna, and UHC. In many cases, records were present in one dataset but missing in another, making validation difficult. As highlighted by Researchgate (2024), data inconsistency across distributed systems is a common challenge in large-scale analytics environments and often requires cross-system validation techniques.
- *ETL Join and Mapping Issues:* While analyzing IRF and IPF datasets, issues were observed in data mapping and join logic. The same billing codes were mapped to multiple categories during ETL processing, leading to duplication and incorrect pricing calculations. According to Microsoft Learn (2023), incorrect join logic in ETL pipelines can lead to data duplication and misleading analytical outputs if master tables contain overlapping mappings.
- *Difficulty in Tracing Data Lineage:* Another challenge was understanding how data moved from raw ingestion to final analytical tables. Due to multiple transformation layers, tracing the origin of specific values was complex. Google Cloud (2023) states that lack of proper

data lineage tracking in analytics systems increases difficulty in debugging and root cause analysis.

- *Outliers and Skewed Data Impacting Analysis:* During IRF analysis, a large number of outliers significantly impacted calculated ratios and metrics. According to Towards Data Science (2022), outliers in healthcare datasets can heavily skew analytical results and must be handled carefully to avoid misleading interpretations.

Overall, these challenges helped me understand the real-world complexities of healthcare data systems, particularly in ETL validation and large-scale data analysis. They also improved my skills in problem-solving, debugging, and systematic data validation.

3.2 Approaches Used to Overcome the Challenges

To address the challenges encountered during my internship, I adopted several structured analytical approaches and followed industry-standard data validation practices. I also identified potential improvements for enhancing data quality and pipeline efficiency.

- *Structured Data Cleaning and Validation*: To manage inconsistencies in healthcare datasets, I followed a systematic approach of cleaning and organizing data before analysis. This included handling missing values, correcting data formats, and standardizing fields across JSON and Parquet datasets. I also cross-verified outputs between Python (Pandas) and SQL (Amazon Athena) to ensure consistency across systems.

  This approach aligns with best practices in data quality management, where structured preprocessing and validation are essential for reliable analytics (IBM, n.d.).
- *Cross-Verification Across Multiple Data Sources:* To address discrepancies between parsed, ETL, and master datasets, I used a cross-validation approach. This involved comparing outputs from different layers of the data pipeline and identifying mismatches in pricing, classifications, and record availability across healthcare carriers.

  Such multi-layer validation is a widely recommended practice in ETL systems to ensure accuracy and consistency in data transformation processes (AWS, n.d.).
- *Root Cause Analysis for ETL Issues:* To resolve issues such as incorrect mappings and one-to-many joins, I performed structured root cause analysis by tracing data flow from raw ingestion to final output tables. This helped identify transformation errors, duplication

issues, and incorrect code-type mappings within ETL pipelines.

ETL debugging and data lineage tracking are critical components of modern data engineering workflows for identifying transformation errors (Microsoft Learn, n.d.).

- *Handling Outliers and Data Quality Issues:* To manage outliers affecting analytical results (especially in IRF datasets), I separated and analyzed data with and without outliers to understand their impact on overall metrics. This helped in identifying skewed distributions and incorrect pricing relationships.

  According to Towards Data Science (2022), outliers in large datasets can significantly distort analytical outcomes and should be carefully analyzed before inclusion in final reporting.
- *SQL-Based Validation Using Amazon Athena:* I used SQL queries in Amazon Athena to validate Python-generated outputs. This included performing joins, aggregations, and filtering to ensure consistency between ETL and analytical datasets.

  Cloud-based query engines such as Athena are widely used for scalable validation of large datasets in distributed environments (AWS, n.d.).

3.3 Recommendations to the Company

Based on the challenges identified during my internship and the approaches used to address them, the following recommendations are suggested to improve data quality, ETL processes, and overall analytical efficiency within the organization:

- *Standardization of Data Formats*: The company should implement standardized formats for datasets across all stages of the pipeline, including parsed, ETL, and analytical layers. Consistent formatting of fields such as billing codes, carrier data, and pricing structures will reduce inconsistencies and improve processing efficiency.
- *Strengthening Data Quality and Documentation Practices:* All datasets should include clear metadata definitions, proper field descriptions, and standardized labeling. Improving documentation within datasets and ETL workflows will reduce ambiguity and enhance transparency during analysis and debugging.
- *Improved ETL Validation and Monitoring:* A stronger validation framework should be implemented within ETL pipelines to detect issues such as incorrect joins, duplicate

records, and misclassified billing codes. Automated validation checks can help ensure data accuracy before it reaches analytical layers.

3.4 Learning Outcome from the Co-Op Studies

During my internship, I gained significant practical exposure to real-world healthcare data analytics and ETL processes, which helped bridge the gap between theoretical knowledge and its application. I developed a strong understanding of how large-scale datasets such as Machine-Readable Files (MRFs) are processed, structured, and analyzed using tools like Python and SQL (Amazon Athena). I learned how to handle different data formats such as JSON and Parquet, perform data cleaning and transformation, and generate analysis-ready datasets for further evaluation.

Additionally, I became familiar with key healthcare data concepts such as IPPS, OPPS, IRF, and IPF, along with their role in pricing and reimbursement systems. Working on ETL validation tasks helped me understand how data flows through different stages, including parsed, ETL, and master tables, and how inconsistencies can arise during transformation processes. Beyond technical knowledge, this internship also improved my analytical and problem-solving skills. I learned how to identify data discrepancies, investigate root causes of errors, and validate results using multiple tools and methods. Handling complex datasets improved my attention to detail, logical thinking, and adaptability to different analytical scenarios. Furthermore, interacting with supervisors and collaborating with team members enhanced my communication skills and understanding of professional teamwork in a data-driven environment.

Overall, the experience has strengthened both my technical competence in data analytics and my professional confidence in working with real-world healthcare data systems.

3.5 Application of Coursework to Real Working Situations

The knowledge gained from my Data Analysis and Database Management coursework was directly applicable during my internship, especially while working with large-scale healthcare datasets and SQL-based validation. I was involved in querying datasets using Amazon Athena,

performing joins, filtering, and aggregation, which helped me understand how theoretical database concepts are applied in real-world data systems. Concepts from Data Structures and Programming (Python) were fundamental to my daily tasks. I applied my understanding of Pandas DataFrames, data manipulation, and file handling while working with JSON and Parquet datasets. Tasks such as cleaning data, transforming nested structures, and preparing datasets for analysis strengthened my practical understanding of programming concepts in data analytics.

My coursework in Data Warehousing and ETL Processes was highly relevant while working with parsed, ETL, and master datasets. I was able to understand how data moves through different pipeline stages and how transformations affect final outputs. This helped me analyze issues such as incorrect mappings, duplicate records, and data inconsistencies across healthcare carriers. The Machine Learning and Statistics foundation helped me while performing exploratory data analysis (EDA), identifying outliers, and understanding data distribution patterns. This was particularly useful during IRF and IPF analysis, where outliers significantly impacted results and required careful interpretation. Knowledge from Cloud Computing was useful while working with Amazon Athena and understanding cloud-based data processing systems. I gained insight into how large datasets are stored, queried, and managed in scalable cloud environments.

Overall, my coursework provided a strong technical foundation that supported my ability to perform data analysis, validate ETL processes, and adapt to real-world healthcare data challenges effectively.

3.6 Skills and Knowledge Gained During this Process

- *Data Processing and Python Proficiency*: Gained hands-on experience in working with large-scale healthcare datasets using Python. Improved skills in handling JSON and Parquet files, transforming data into structured formats, and performing exploratory data analysis (EDA) using Pandas.
- *SQL and Data Validation Skills:* Developed practical experience in using SQL (Amazon Athena) for querying, filtering, joining, and validating large datasets. Improved ability to cross-check data across multiple tables and identify inconsistencies.
- *ETL and Data Pipeline Understanding:* Learned how ETL pipelines function in real-world systems, including data ingestion, transformation, and loading. Gained understanding of

how errors in transformation and joins can impact final outputs.

- *Healthcare Data Domain Knowledge:* Developed familiarity with healthcare data structures and concepts such as MRF files, IPPS, OPPS, IRF, and IPF. Understood how these categories are used in pricing and analytical systems.
- *Attention to Detail:* Strengthened ability to detect data inconsistencies such as missing records, incorrect mappings, duplicate entries, and outliers affecting analytical results.
- *Analytical and Problem-Solving Skills:* Improved ability to investigate data issues, perform root cause analysis, and logically interpret discrepancies across different data sources and pipeline stages.
- *Communication and Team Coordination:* Enhanced professional communication skills through regular interaction with supervisors and team members to clarify data logic, resolve discrepancies, and validate findings.
- *Data Handling and Organization Skills:* Developed structured approach to organizing datasets, maintaining clean data pipelines, and ensuring consistency across multiple analytical stages.
- *Understanding of Real-World Data Systems:* Gained practical insight into how large-scale healthcare data systems operate, including how raw data is transformed into structured insights used for decision-making and reporting.

CHAPTER 4: CONCLUSION

4.1 Summary of Highlights of My Co-Op Studies at Deerhold

During my 16-week Co-operative Education placement at Deerhold Ltd, I was involved in a wide range of healthcare data analytics and ETL validation tasks. My work primarily focused on processing and analyzing large-scale healthcare datasets, particularly Machine-Readable Files (MRFs), using Python, SQL (Amazon Athena), and data analysis techniques. Key areas of involvement included:

- Working with JSON and Parquet healthcare datasets
- Converting raw datasets into structured Pandas DataFrames
- Performing exploratory data analysis (EDA) and dataset validation
- Writing SQL queries in Amazon Athena for filtering, joins, and cross-validation
- Supporting ETL validation and healthcare data quality checks
- Investigating discrepancies across parsed, ETL, and master datasets

One of the most impactful aspects of my internship was working on detailed analytical tasks involving healthcare pricing and ETL transformation issues. These included:

- Investigating missing and inconsistent records across multiple healthcare carriers such as Cigna, BCBS, Aetna, and UHC
- Performing IRF and IPF deep-dive analysis to identify outliers and pricing inconsistencies
- Identifying ETL mapping and one-to-many join issues affecting healthcare pricing calculations
- Analyzing billing code mismatches between parsed and ETL datasets
- Validating healthcare categories such as IPPS, OPPS, IRF, and IPF across multiple datasets

These experiences significantly improved my analytical thinking, debugging ability, and understanding of real-world healthcare data systems.

Overall, the internship met and exceeded my expectations by providing practical exposure far beyond classroom learning. I not only developed technical skills in Python, SQL, data processing,

and ETL validation, but also gained deeper insight into healthcare analytics workflows and large-scale data engineering processes. Working closely with supervisors and collaborating with team members also enhanced my communication, teamwork, and professional problem-solving abilities. This experience has strengthened my confidence and prepared me for future roles in data analytics and data engineering by equipping me with both technical competence and practical industry exposure.

4.2 My Evaluation of the Work Experience

My Co-Op experience at Deerhold provided me with valuable hands-on exposure to real-world healthcare data analytics and ETL validation processes. The internship allowed me to work with large-scale datasets, analytical tools, and cloud-based technologies, which significantly enhanced my understanding beyond theoretical concepts. The opportunity to work on data processing, SQL validation, and healthcare pricing analysis made the experience both challenging and highly rewarding.

This internship has contributed directly to my career development goals in data analytics and data engineering by strengthening my practical knowledge and technical skills. I developed proficiency in Python, SQL (Amazon Athena), data transformation, exploratory data analysis (EDA), and ETL validation processes. Working with multiple healthcare datasets and carriers helped me understand complex data structures and improved my ability to analyze, validate, and interpret large-scale data effectively.

Throughout the internship, I demonstrated attention to detail, adaptability, and a willingness to learn. I was able to identify inconsistencies in datasets, investigate root causes of ETL issues, and contribute to data validation tasks, which reflects my analytical and problem-solving abilities. However, I also recognize that there is still room for improvement in areas such as advanced data engineering concepts and large-scale system optimization.

Compared to when I started, I have become more confident in handling complex datasets, communicating technical findings, and working in a professional team environment. Overall, this experience has significantly improved both my technical competence and professional confidence for future roles in data analytics and data engineering.

4.3 Limitations of My Co-Op Studies

During my Co-operative Education period at Deerhold, several limitations influenced the scope and depth of my learning experience. These limitations were mainly related to the complexity of healthcare datasets, internship duration, system access, and large-scale data processing environments.

- The limited duration of the internship (16 weeks) restricted exposure to the complete lifecycle of large-scale data engineering projects, including long-term pipeline optimization, deployment processes, and production-level monitoring systems. As a result, most learning was focused on ongoing analytical and ETL validation tasks.
- Working with multiple healthcare carriers and datasets created challenges due to differences in data structures, pricing formats, and carrier-specific implementations. Considerable time was required for data cleaning, validation, and reconciliation across parsed datasets.
- Initially, understanding healthcare-specific concepts such as MRFs, IPPS, OPPS, IRF, and IPF required additional learning and adaptation, as these areas were highly domain-specific and more complex than traditional datasets.
- The large size and complexity of healthcare datasets also increased processing and validation time, especially while tracing discrepancies across multiple stages of ETL pipelines. In many cases, identifying the root cause of issues required extensive cross-verification between datasets and transformation layers.
- Dependency on existing ETL structures and restricted access to certain internal systems limited deeper exploration of pipeline architecture and backend implementation details. This reduced opportunities to fully understand some production-level engineering processes.
- Additionally, a significant portion of the internship involved data validation and debugging tasks, which limited exposure to advanced areas such as machine learning implementation, predictive analytics, and large-scale system design.
- Despite these limitations, the internship provided valuable practical exposure to real-world healthcare data analytics and significantly enhanced my technical and analytical skills.

REFERENCES

Amazon Web Services. (n.d.). *Amazon Athena documentation*. https://docs.aws.amazon.com/athena/

Amazon Web Services. (n.d.). *What is data quality?* https://aws.amazon.com/what-is/data-quality/

IBM. (n.d.). *Data pipeline*. IBM. https://www.ibm.com/think/topics/data-pipeline

IBM. (n.d.). *Data pipeline overview*. IBM. https://www.ibm.com/think/topics/data-pipeline

IBM. (n.d.). *ETL (extract, transform, load)*. IBM. https://www.ibm.com/topics/etl

IBM Research. (n.d.). *Refactoring ETL flows in the wild*. https://research.ibm.com/publications/refactoring-etl-flows-in-the-wild

Microsoft Learn. (n.d.). *ETL process in data architecture*. Microsoft. https://learn.microsoft.com/en-us/azure/architecture/data-guide/relational-data/etl

ResearchGate. (2023). *ETL (extract, transform, load) best practices*. https://www.researchgate.net/publication/368300449_ETL_Extract_Transform_Load_Best_Practices

ScienceDirect. (2023). *Data pipeline quality and integration issues*. https://www.sciencedirect.com/science/article/pii/S0164121223002509

APPENDICES

Appendix A. Weekly log

Name of Student: Devansh Lath

Name of Organization: Deerhold Ltd.

Name of Job Supervisor: Mr. Saurav Adhikari

Internship Period: 19 January, 2026 - 19 May, 2026

Week 1:

Date	Day	Task Assigned
19-01-2026	Monday	Joined Deerhold Ltd., completed HR formalities, and received introduction to the team, healthcare pricing systems, and MRF datasets.
20-01-2026	Tuesday	Brief introduction to healthcare transparency products including PRIZM and healthcare shopping tools. Studied MRF structures and healthcare pricing concepts.
21-01-2026	Wednesday	Worked with MRF JSON datasets in Python and converted raw JSON files into Pandas DataFrames for analysis.
22-01-2026	Thursday	Continued processing MRF datasets and performed exploratory data analysis (EDA) for structural validation and pricing insights.
23-01-2026	Friday	Worked with Parquet datasets, split datasets into separate DataFrames, performed joins, and prepared consolidated datasets for EDA.

Week 2:

Date	Day	Task Assigned
26-01-2026	Monday	Started ticket investigation involving carrier-level validation and NPI/TIN-based record analysis across healthcare datasets.
27-01-2026	Tuesday	Performed SQL validation using Amazon Athena to identify records across parsed and ETL datasets.
28-01-2026	Wednesday	Investigated IPPS and OPPS record availability across multiple carriers including Cigna, BCBS, Aetna, UHC, and Medcost.
29-01-2026	Thursday	Cross-validated parsed and ETL datasets to identify missing institutional billing records and pricing inconsistencies.
30-01-2026	Friday	Investigated hospital MRF datasets to analyze carrier-specific inpatient and outpatient pricing records.

Week 3:

Date	Day	Task Assigned
02-02-2026	Monday	Performed similar carrier-level investigation and validation tasks for BCBS datasets.
03-02-2026	Tuesday	Started IRF deep-dive analysis by comparing Cigna and BUCA weight ratios with and without outliers.
04-02-2026	Wednesday	Analyzed HIPPS-related outliers and investigated Medicare percentage inconsistencies in IRF datasets.
05-02-2026	Thursday	Investigated billing code discrepancies between parsed and ETL datasets involving HCPCS and HIPPS mappings.

06-02-2026	Friday	Performed root cause analysis on ETL joins and identified one-to-many mapping issues in master procedure tables.

Week 4:

Date	Day	Task Assigned
09-02-2026	Monday	Started IPF deep-dive analysis involving MSA-level comparisons and carrier-level validation.
10-02-2026	Tuesday	Analyzed IPF NPIs across parsed and ETL datasets and validated billing-code-level record availability.
11-02-2026	Wednesday	Investigated discrepancies in MSA-level IPF data across Cigna and BUCA datasets.
12-02-2026	Thursday	Cross-validated billing codes and NPIs in parsed datasets and investigated missing carrier records.
13-02-2026	Friday	Completed IPF validation and summarized findings regarding parsed data availability and ETL inconsistencies.

Week 5:

Date	Day	Task Assigned
16-02-2026	Monday	Worked on advanced SQL validations and Athena-based data filtering for healthcare datasets.
17-02-2026	Tuesday	Performed exploratory data analysis (EDA) on structured MRF datasets and generated analytical observations.
18-02-2026	Wednesday	Validated pricing consistency across multiple healthcare carriers and datasets.
19-02-2026	Thursday	Completed ticket documentation and summarized findings from IRF and IPF investigations.

20-02-2026	Friday	Leave

Week 6:

Date	Day	Task Assigned
23-02-2026	Monday	Worked on parsed and ETL dataset reconciliation using SQL joins and validation queries.
24-02-2026	Tuesday	Investigated carrier-level discrepancies and validated healthcare pricing records.
25-02-2026	Wednesday	Continued healthcare data validation and dataset consistency checks across multiple tables.
26-02-2026	Thursday	Leave
27-02-2026	Friday	Worked on healthcare pricing analysis and ETL-level validation tasks.

Week 7:

Date	Day	Task Assigned
02-03-2026	Monday	Holiday (Holi)
03-03-2026	Tuesday	Work From Home (WFH) – Performed SQL validation and Python-based healthcare dataset analysis remotely.
04-03-2026	Wednesday	Work From Home (WFH) – Continued ETL validation and exploratory analysis on MRF datasets.
05-03-2026	Thursday	Holiday (Election)
06-03-2026	Friday	Work From Home (WFH) – Investigated pricing inconsistencies and validated ETL transformation outputs.

Week 8:

Date	Day	Task Assigned
09-03-2026	Monday	Performed validation checks on institutional billing and pricing datasets.
10-03-2026	Tuesday	Worked on Athena queries for healthcare dataset filtering and pricing validation.
11-03-2026	Wednesday	Continued parsed and ETL dataset reconciliation and identified missing records.
12-03-2026	Thursday	Worked on healthcare pricing comparisons across multiple carriers.
13-03-2026	Friday	Leave

Week 9:

Date	Day	Task Assigned
16-03-2026	Monday	Investigated discrepancies between parsed and ETL datasets and validated transformation logic.
17-03-2026	Tuesday	Performed advanced SQL joins and dataset filtering using Amazon Athena.
18-03-2026	Wednesday	Worked on carrier-level pricing analysis and healthcare data validation.
19-03-2026	Thursday	Cross-validated MRF pricing datasets and generated analytical findings.
20-03-2026	Friday	Investigated record mismatches and ETL-level inconsistencies.

Week 10:

Date	Day	Task Assigned
23-03-2026	Monday	Worked on parsed dataset validation and healthcare billing-code analysis.
24-03-2026	Tuesday	Continued ETL validation and healthcare pricing investigations.
25-03-2026	Wednesday	Performed SQL-based cross-validation between parsed and master datasets.
26-03-2026	Thursday	Investigated outliers affecting healthcare pricing ratios and analytical outputs.
27-03-2026	Friday	Generated analytical observations and summarized ETL findings.

Week 11:

Date	Day	Task Assigned
30-03-2026	Monday	Worked on healthcare dataset validation and pricing consistency checks.
31-03-2026	Tuesday	Performed exploratory data analysis (EDA) on carrier-level healthcare data.
01-04-2026	Wednesday	Continued SQL-based validation and ETL discrepancy investigation.
02-04-2026	Thursday	Investigated billing-code-level mismatches and pricing irregularities.
03-04-2026	Friday	Documented findings from healthcare data analysis and validation tasks.

Week 12:

Date	Day	Task Assigned
06-04-2026	Monday	Worked on Athena queries and healthcare dataset filtering tasks.
07-04-2026	Tuesday	Continued ETL validation and healthcare pricing analysis.
08-04-2026	Wednesday	Performed parsed versus ETL dataset reconciliation and identified inconsistencies.
09-04-2026	Thursday	Investigated institutional billing-class datasets and pricing records.
10-04-2026	Friday	Generated analytical observations from healthcare pricing datasets.

Week 13:

Date	Day	Task Assigned
13-04-2026	Monday	Leave
14-04-2026	Tuesday	Holiday (Nepali New Year)
15-04-2026	Wednesday	Leave
16-04-2026	Thursday	Resumed ETL validation and healthcare pricing analysis tasks.
17-04-2026	Friday	Worked on advanced dataset reconciliation and SQL-based filtering.

Week 14:

Date	Day	Task Assigned
20-04-2026	Monday	Started ticket: Analysis of network_name field and investigated

		carrier-level network inconsistencies.
21-04-2026	Tuesday	Worked on capitated data analysis and investigated pricing structures in healthcare datasets.
22-04-2026	Wednesday	Investigated filtering issues involving Medcost and Quartz datasets by Windsor.
23-04-2026	Thursday	Worked on Impact Network Review and validated network-level healthcare pricing records.
24-04-2026	Friday	Started UHC – Christiana Care Delaware – Inpatient Data analysis and investigated inpatient pricing datasets.

Week 15:

Date	**Day**	**Task Assigned**
27-04-2026	Monday	Continued network_name field analysis and carrier-level validation tasks.
28-04-2026	Tuesday	Continued capitated data investigation and pricing consistency analysis.
29-04-2026	Wednesday	Continued Medcost and Quartz filtering investigation and generated analytical findings.
30-04-2026	Thursday	Worked on Impact Network Review and ETL-level validation tasks.
01-05-2026	Friday	Continued UHC – Christiana Care Delaware – Inpatient Data analysis and pricing validation.

Week 16:

Date	Day	Task Assigned
04-05-2026	Monday	Leave
05-05-2026	Tuesday	Continued healthcare pricing validation and SQL-based reconciliation tasks.
06-05-2026	Wednesday	Worked on parsed and ETL dataset cross-validation and discrepancy investigation.
07-05-2026	Thursday	Continued ETL validation and generated analytical summaries for healthcare datasets.
08-05-2026	Friday	Performed final validation checks and documented analytical observations.

Week 17:

Date	Day	Task Assigned
11-05-2026	Monday	Leave
12-05-2026	Tuesday	Continued healthcare data analysis, SQL validation, and ongoing ticket investigation tasks.
13-05-2026	Wednesday	Worked on ETL validation and healthcare pricing discrepancy analysis across multiple carriers.
14-05-2026	Thursday	Continued parsed versus ETL dataset validation and generated analytical observations.
15-05-2026	Friday	Worked on healthcare data quality checks and ticket documentation.

Week 18:

Date	Day	Task Assigned
18-05-2026	Monday	Continued SQL-based healthcare dataset analysis and validation tasks.
19-05-2026	Tuesday	Summarized findings from ongoing tickets and updated analytical documentation.

Appendix B. Photographs of Internship

Figure 3: Office Entrance

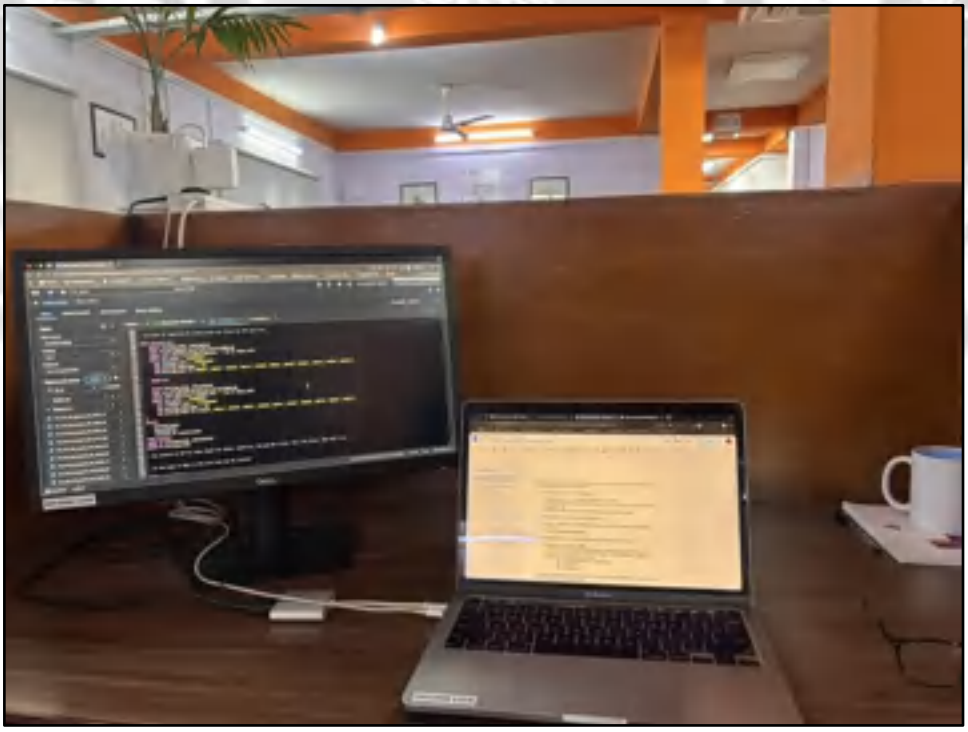


Figure 4: My Workstation

Figure 5: With My Job Supervisor

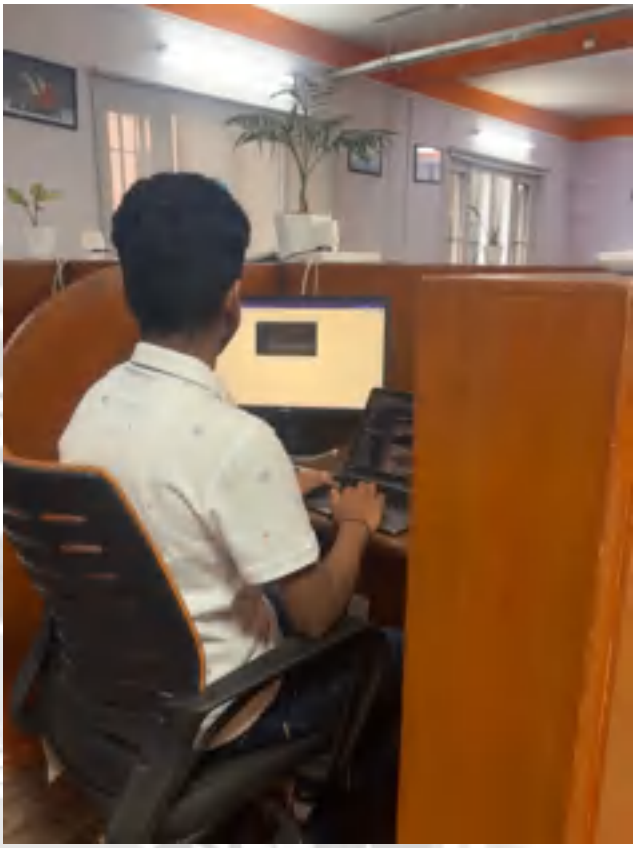


Figure 6: Working on my tasks

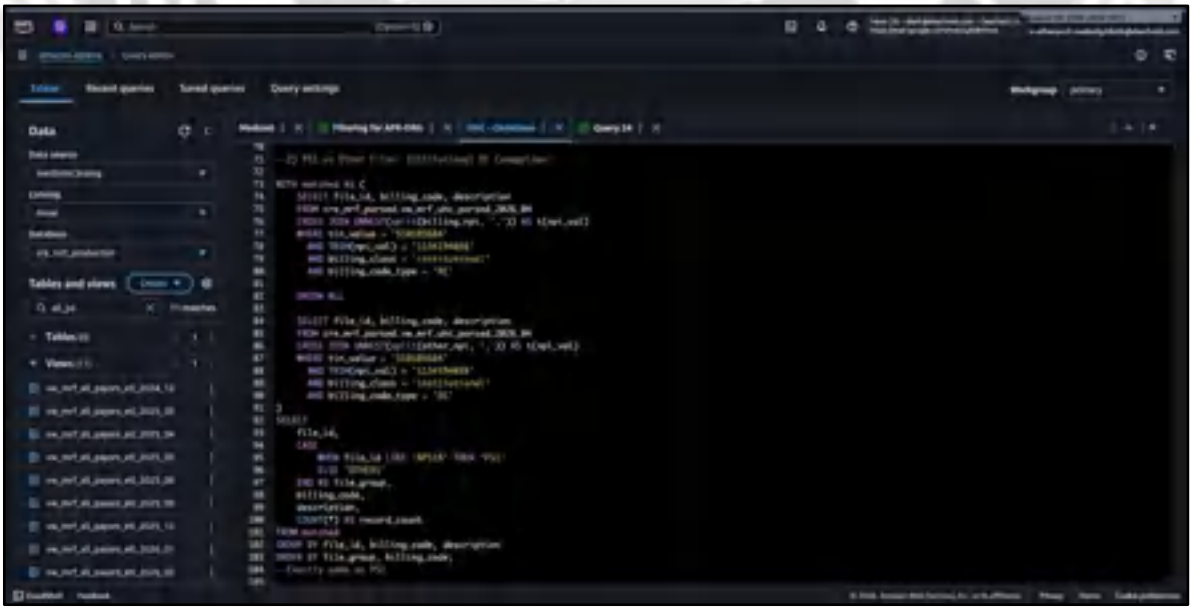


Figure 7: Writing SQL Queries in AWS Athena

Figure 8: Data Cleaning in Python

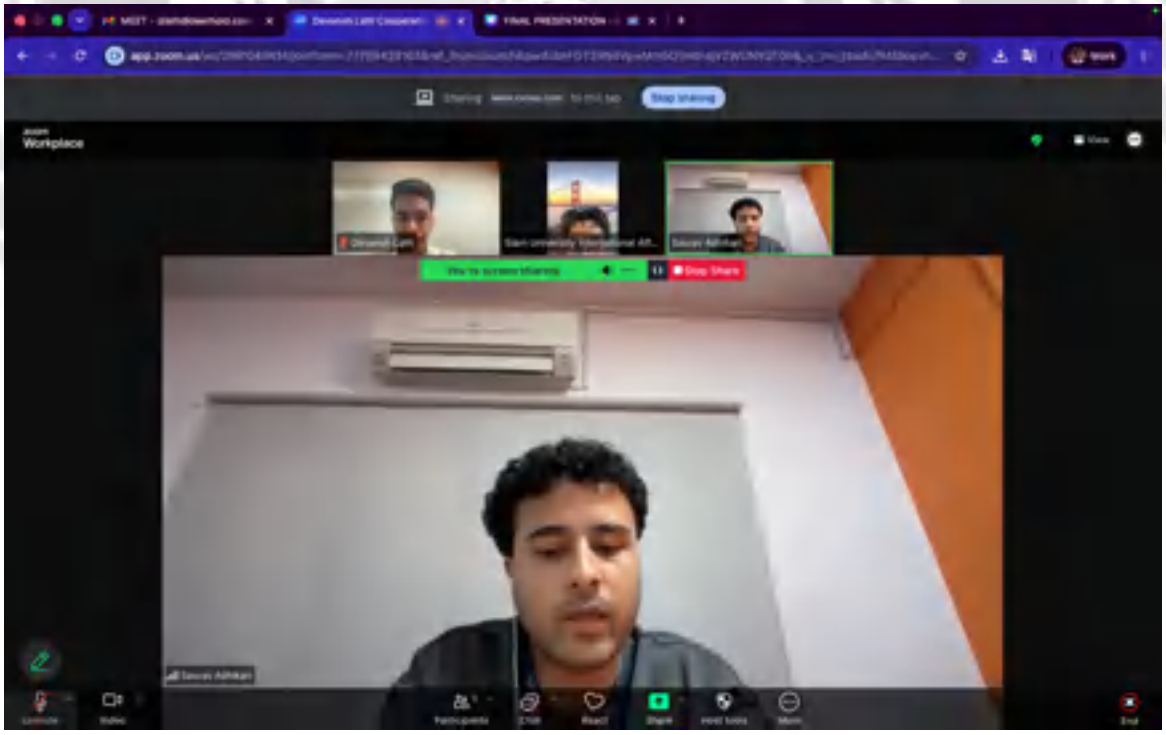


Figure 9: Presentation