



**DOUGLAS** COLLEGE

COMMERCE & BUSINESS ADMINISTRATION  
COMPUTING STUDIES & INFORMATION SYSTEMS  
COMPUTER AND INFORMATION SYSTEMS (PBD)

CSIS 4495-050: APPLIED RESEARCH PROJECT

Progress Report 3:  
**End-to-End Data Engineering Solution for HR Analytics**

**Student Name:** Matheus Filipe Figueiredo | 300389657

**Instructor:** Dr. Bambang Sarif

NEW WESTMINSTER/BC  
FALL/2025

## 1.0 Work Hours

---

**Student Name: Matheus Filipe Figueiredo**

Date	Number of Hours	Description of Work Done
10/12/2025	2.5	Initial Study on the Databricks logic and website.
10/18/2025	1.5	Created the Report file, adding the sections based on the template on BlackBoard
10/18/2025	2.5	Created the Notebook to create the dataset for the gold layer
10/19/2025	3.5	Updated the Notebook to create the dataset for the gold layer
10/19/2025	2.5	Created a New Power BI using the new dataset
10/12/2025	2.5	Initial Study on the Databricks logic and website.
10/18/2025	1.5	Created the Report file, adding the sections based on the template on BlackBoard
10/18/2025	2.5	Created the Notebook to create the dataset for the gold layer
10/19/2025	3.5	Updated the Notebook to create the dataset for the gold layer
10/19/2025	2.5	Created a New Power BI using the new dataset
10/21/2025	1	Edited the POWER BI section in the Midterm Report
10/22/2025	1	Add and update Data Engineering part for Gold Layer in Midterm Report
10/23/2025	1	Created the Progress Report 3

## 2.0 Work Done

---

During the period between the second report and this one, I focused on expanding our data engineering and visualization components while improving the documentation and organization of our project.

At first, I dedicated time to studying the Databricks platform to better understand its logic and how to use it effectively within our data pipeline. Since Databricks was a new tool for me, I spent some time learning about its environment.

Afterwards, with the help of Bruno, I began working on the Gold Layer dataset. Together, we designed and refined the notebook responsible for generating the final dataset, ensuring it aligned with the Medallion Architecture principles.

Once the dataset was finalized, I built a new Power BI dashboard using the updated data. The main goal was to make sure the visualizations accurately represented the structure and logic of our pipeline. This updated dashboard provided clearer insights of the new dataset.

In parallel, I also created the initial version of the Midterm Report document, using the template provided on Blackboard. I organized the sections based on our progress so far, adding the proper information from the initial Report to it.

---

## 2.0 Repo Update

---

All the work completed during this period was added and updated in our project's GitHub repository. Specifically, I uploaded the notebook for the Gold Layer generation, the new version of the Power BI dashboard (Version 2), and the first draft of the Midterm Report.

Since the upload, I've made some updates to the Power BI file, improving the visuals, adding the slicers, adjusting formatting, and ensuring all data fields display correctly. These updates helped make the dashboard more dynamic and suitable for presentation.