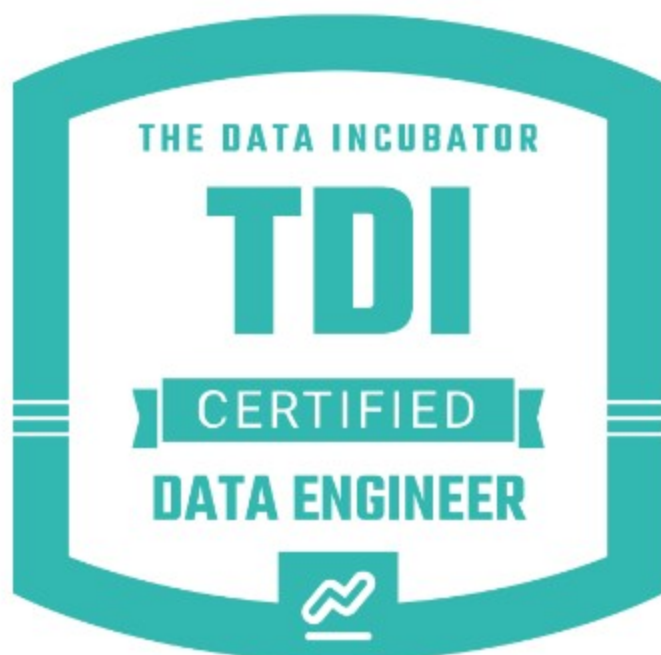




This badge was issued to [Gonsalge Sureka Almeida](#) on July 17, 2023

[Verify](#)

[Celebrate](#)



Data Engineering Fellowship

Issued by [Pragmatic Institute](#)

The earner of this badge has successfully completed mini-projects and capstone project with a score of >90%. They use machine learning tools to derive insights, develop data storage systems and create relational/non-relational data schemas with SQL and python. They curate data pipelines to transform streaming and static data, utilize AWS tools to create, transform, store and analyze data, and utilize Airflow to integrate workflow management systems.

Skills

Airflow

Amazon Web Services

Athena

Communication

Data Storage Architectures

DynamoDB

EMR

Glue

IAM Management

Kafka, Kinesis

Lambda

Machine Learning

MongoDB

Pandas

Presentation

Pricing Schemas

PyTest

Python

QuickSight

Redshift

Relational Databases

S3

Scikit Learn

Spark

Spark Streaming

SQL

Visualization

Earning Criteria

- ☞ The earner completed two mini-projects associated with the data wrangling module, scoring at least a 90%. They used Python libraries and SQL to gather, clean, organize and analyze messy real-world data. They built a social graph of social connections of the population and used that to determine influential people within the group. For the SQL mini-project, they wrote complex SQL queries to extract information from an NYC database of restaurant inspections, revealing common types of violations.
- ☞ The earner completed the machine learning module and successfully completed the ML mini-project, scoring at least a 90%. They used Python and Scikit Learn to develop a machine learning model. They combine existing estimators and transformers via pipelines and feature unions, and they are able to develop custom estimators and transformers to predict venue popularity from these features. Finally, they have built an ensemble model combining several smaller models to achieve better performance.
- ☞ The earner has completed the spark mini-project, scoring at least a 90%. They have parsed, cleaned, and processed a 10 GB set of XML files of user actions on a Q&A website. Using this behavior, they have answered questions about user behavior to predict the long-term behavior of new users. They have trained a word2vec model and a classification model on tags associated with questions. They have worked with RDDs and DataFrames, and they have implemented a machine learning pipeline using Spark ML.
- ☞ The earner has completed two mini-projects associated with the amazon web service module. They have used AWS cloud services to build pipelines that can handle big data, and monitor these deployables through cloud tools. In both cases they have used a new Amazon Web Service tool to construct data pipelines, giving them the ability to research and implement new cloud services.
- ☞ The earner has completed the Airflow mini-project. They have implemented a real-world Airflow workflow manager that handles AWS cloud services and database system tools. They have developed their own architecture to manage the movement of data within the pipeline.
- ☞ The earner has completed a capstone project to the satisfaction of TDI's instructors. The capstone project requires the use of real-world data to help solve a practical issue. They used modern data engineering techniques to produce a pipeline to address a real-world data issue. They have used modern cloud services to derive actionable insights from that data and assembled this work into a deliverable suitable for non-experts.
- 👤 Earner attended courses from The Data Incubator, 21001 N Tatum Blvd Ste 1630 #642, Phoenix, AZ 85050, (480) 515-1411, thedataincubator.com, admissions@thedataincubator.com