

Daniel Pereira da Costa

daniel.pereiracosta@hotmail.com | linkedin.com/in/daniel-pereira-da-costa | danieldacosta.vercel.app | github.com/DanielDaCosta
California, United States | (213) 732-6265

EDUCATION

University of Southern California

Master of Science in Applied Data Science

Los Angeles, United States

August 2022-May 2024

École Centrale Paris

Master of Science in Engineering - Dual Degree with PUC-Rio

Paris, France

September 2016-December 2019

Pontifical Catholic University of Rio de Janeiro (PUC-Rio)

Bachelor in Control and Automation Engineering, Minor in Mathematics

Rio de Janeiro, Brazil

January 2014-December 2019

TECHNICAL SKILLS

Programming Python (Pandas, PySpark), R, SQL (PostgreSQL, MySQL, SparkSQL), NoSQL (MongoDB), JavaScript, TypeScript, C/C++

Cloud Platform Amazon Web Services (AWS), Google Cloud Platform (GCP)

ETL Tools Amazon Elastic MapReduce (EMR), AWS Step Functions, MLflow, AWS Glue, Airflow

Other Tools Amazon QuickSight, Power BI, React, Tableau, Redshift, Apache Spark, Databricks, PyTorch, Terraform, Docker, Git

WORK EXPERIENCE

Rose AI

Machine Learning Engineer

Jun 2024-Present

California, United States

- Fine-tune **LLM agent** to enhance chart generation capabilities on Rose AI, offering financial analysts a robust data solution experience. Developed a cost analysis platform to estimate the costs of 5 AI agents based on **GPT-4o**.

Intel Corporation

Data Analyst

May 2023-May 2024

California, United States

- Developed an all-encompassing **PowerBI** dashboard utilizing data analytics and data visualization to proactively identify performance deviations in 50+ AI workloads running on Ponte Vecchio GPUs across single and distributed system configurations.
- Designed **RAG** application using **LLMs** and **Llama-Index** to ingest and associate unstructured customer engagement updates from disparate sources (e.g., PowerPoint, JIRA, Excel) into a single PowerBI based reporting for several high priority engagements.

Akad Seguros (Argo Group-Insurance Company)

Data Engineer

January 2022-February 2023

Rio de Janeiro, Brazil

- Orchestrated design of 600GB Data Lake on **AWS** through **S3**, **Glue**, **Lambda**, and **Step Functions**; published to [AWS Blog](#).
- Collected and ingested over 100GB of data from 8 different on-premises and cloud-based servers including Azure SQL, SQL Server and **PostgreSQL** into **Data Lake** via **Data Migration Service** and ETL jobs; utilized **Databricks Delta Lake** to ensure data integrity.
- Defined and deployed **Data Warehouse (50GB)** components and architecture on AWS applying tools such as S3 for storage, **EMR Spark** for big data processing, and **Athena** for data analysis; conducted data modeling for 15+ tables using a snowflake schema.

Cyberlabs (PSafe – AI Powered Cybersecurity)

Data Scientist

Jul 2019-December 2021

Rio de Janeiro, Brazil

- Engineered Machine Learning model with 71% f1-score to reduce churn and increase retention from 6 million users' cybersecurity app; deployed model in **AWS** utilizing **Spark Clusters EMR** for big data processing and **EC2** instance for ML inference with **MLFlow**.
- Guided a team of 4 in creating a Churn Predictive Model (LSTM network) with 72% precision and 83% recall for business analysis within a telecom company; leveraged quantitative analysis techniques to identify customer churn patterns across 10 key metrics.
- Spearheaded a team of 3 in constructing a system aiding 20+ companies in controlling COVID-19 spread for 3,000 employees; implemented event-driven **serverless architecture** on **AWS** employing Infrastructure as Code (**Terraform**) and **Lambda**, **RDS** and **API Gateway**. Utilized **Amazon QuickSight** to present data to over 30 different clients.
- Built microservice application that generates up to 10 statistical metrics and insights per minute from Real-Time Object Detection Model; increased 100 establishments' decision-making productivity by 20 hours/month through crowd counting data collection.

Softbank Robotics Europe

Embedded Software Engineer

June 2018-December 2018

Paris, France

- Created firmware module in **C/C++** embedded in a microcontroller for a robot's inductive sensor (LDC1312/4).
- Presented product analysis to firmware team to validate sensor's effectiveness against 2 KPIs.

PROJECTS/PUBLICATIONS

Driving Insurance Industry Innovation with a Data Lake on AWS with Databricks Delta Lake

[AWS Blog](#)

Amazon Web Services, Elastic MapReduce, Glue, Step Functions, Database Migration Services

- Architected a Data Lake to store over 600GB of historical and incremental data, implementing change data capture (CDC) from relational databases across three data layers—Raw, Staging, and Curated—to ensure robust governance and observability.

Machine Learning pipeline with Airflow and Elastic Container Service

[Towards Data Science](#)

Amazon Web Services, Airflow, Docker, Elastic Container Service (Fargate), Elastic Container Registry

- Created an Airflow pipeline and deployed containerized application to a serverless compute engine (Fargate) for machine learning inference; published tutorial on Towards Data Science.