

# Phil Salazar

+1 (707) 853-3757 • sphilazar@gmail.com • linkedin.com/in/phil-salazar • phil-salazar.github.io

I am a full stack data scientist skilled at developing ML pipelines, implementing ML models, and incorporating generative AI and LLMs in lending data analysis. I have 5 years of experience in technology and recently graduated from Harvard, where I focused on statistics and financial services. Proficient in AWS, Python, R, React, and SQL.

## WORK EXPERIENCE

### Grupago

09/2024 - Present

#### Data Scientist • Contractor

- Leverage deep learning techniques to extract predictive features from unstructured data using AWS SageMaker, Keras, Python, and SQL.
- Created machine learning pipeline, including a credit risk model using machine learning to analyze loan economics in international markets.

### Bankaya

05/2023 - 08/2023

#### Credit Risk Analyst • Internship

- Tech Stack: AWS, Airflow, dbt, Snowflake, SQL, Python, MLFlow
- Implemented new credit risk strategy for returning customers using machine learning in AWS, Python, and SQL, improving overall credit margin of international loan business by more than 3 percentage points.

### Indigov

01/2022 - 08/2022

#### Data Engineer

- Created efficient Python scripts, bash scripts, ETL pipelines, and data ingestion from external vendor API, ensuring high data quality and seamless data flow, accelerating client onboarding by 25%.
- Managed efficient execution of data pipelines and provisioning cloud resources using Prefect and Terraform, reducing overhead resources by 18%.

### GoodLeap

11/2019 - 11/2021

#### Business Intelligence Data Analyst

- Tech stack: AWS, Python, Airflow, Delta Lake, Databricks, PySpark, SQL, Cube.js, React
- Served as lead for Agile business intelligence team, delivering more than 95% of portfolio reporting and interactive functionality on time to diverse business stakeholders and C-suite executives.
- Gathered technical requirements from business stakeholders to develop over 40 complex SQL queries, over 30 data models, over 20 Python ETL & ELT pipelines, and over 75 React dashboards.

### Amazon

06/2017 - 09/2018

#### Retail Vendor Manager

- Grew \$20M vendor portfolio by 23% year over year using SQL to analyze sales trends.
- Served as vendor lead time subject-matter expert, collaborating with supply chain analysts to reduce inventory lead time by 12%, resulting in over \$7M additional revenue.

EDUCATION

Master in Public Administration in International Economics

Harvard University

08/2022 - 05/2024

Advanced Statistics, Econometrics, Advanced Macroeconomics, Consumer Financial Services, Managerial Finance, Corporate Finance, MLOps, Deep Learning

**Course Assistant:** Advanced Microeconomic Analysis (Prof. Luis Armona, Prof. Jie Bai)

**Research Assistant:** Randomized Control Trial (RCT): Economic impact of trade fairs in Chile (Prof. Jie Bai)

Bachelor of Science in Engineering

Stanford University

08/2012 - 06/2016

PROJECTS

Prescribe.ai

Developed generative AI web application for under-resourced medical professionals in Latin America to predict accurate medication prescriptions for observed patient symptoms. Built end-to-end: created virtual environments in Docker, deployed large language model (LLM) in Google Cloud Platform (GCP), and built frontend application in React.

Predicting Disease Burden in Brazil

Developed 2-stage machine learning model in R to predict municipalities with highest disease burden to prioritize investment in health infrastructure in collaboration with Brazil's Ministry of Health.

SKILLS

Cloud Platforms & Infrastructure

- AWS
- Databricks
- Docker
- GCP
- Terraform

Machine Learning

- Deep Learning
- Keras
- Tensorflow

Programming

- Github
- NumPy
- Pandas
- PySpark
- Python
- R
- React
- sklearn
- SQL
- Unix