

Ireri Avila

San Francisco, CA | +1-650-398-8571 | [Email](#) | [LinkedIn](#) | [Github](#) | [Website](#)

Summary

Experienced Data Scientist with a physics background and strong analytics engineering skills. Adept at transforming raw data into actionable insights using SQL, Python, and visualization tools to drive business growth. Proven ability in automating workflows and building scalable dashboards, reducing manual effort while enhancing operational efficiency. Recognized for translating complex datasets into clear narratives for technical and non-technical audiences.

Skills

- **Programming & Analytics:** Python, SQL, NoSQL, A/B Testing, Spark, PyTorch, ACL Analytics, LaTeX, Transformers, HuggingFace, NLP, Spacy, Statistical Analysis
- **Data Engineering & Visualization:** MongoDB, DataBricks, Airflow, Plotly, Tableau, PowerBI, Data Modeling, dbt, Looker, LookML
- **Cloud & Databases:** Google Cloud Platform, AWS, Oracle, BigQuery
- **Tools & Collaboration:** Git, GitHub, Salesforce, English, Spanish, French, Italian

Professional Experience

Give Us The Floor | *Data Scientist*

Oct 2023 - Present

- Developed and deployed NLP-driven ML models to classify rare, high-sensitivity messages in unstructured group chat data—achieving under 1% Type II error in highly imbalanced datasets similar to fraud detection use cases—demonstrating advanced analytical and SQL proficiency.
- Reduced human intervention by 90%, resulting in annual savings of \$90K and impacting more than 2,000 users.
- Enhanced message retrieval by 80% for 400K messages through a Python script that improved API request efficiency and date filtering.
- Engineered and maintained an Apache Airflow DAG to automate end-to-end ML pipelines, from raw data ingestion to model inference, reducing manual oversight by 60% and increasing system reliability.
- Created comprehensive onboarding materials and a structured curriculum for incoming data science interns, cutting ramp-up time from one month to two weeks and bolstering team productivity.

Grupo Bimbo | *Data Analyst*

Mar 2021 - May 2022

- Developed 15 analytical scripts using Oracle Database connections to generate audit reports, highlighting strong SQL skills and data integrity focus.
- Implemented Python scripts to automate monthly tasks, reducing manual efforts by over 3 hours per month and ensuring scalable operations.
- Collaborated with cross-functional teams to design and deploy a dynamic risk assessment dashboard for fixed assets using Power BI, reducing reporting time by 70% across five organizations and delivering clear, actionable insights.
- Validated data consistency for over 2 million records across three Azure data lakes through advanced SQL queries, ensuring reliable data governance.
- Optimized and refined data processing scripts for large datasets, reducing computing time by 40% while demonstrating robust analytical problem-solving abilities.

National Institute of Astrophysics, Optics, and Electronics (INAOE) | *Data Scientist*

May 2019 - May 2020

- Data mining of around 250K records of data from 28 different sources using Beautiful Soup.
- Implemented a Python script for data normalization, analysis, and time series modeling, reducing data preprocessing time by 30%.
- Created data visualizations using Matplotlib and Seaborn, transforming complex datasets into clear, interpretable insights, leading to a 25% reduction in analysis time.

Education

University of San Francisco

Jul 2023 - Jun 2024

M.S in Data Science

- **Coursework:** Data Acquisition, Machine Learning, Regression, Time Series, A/B Testing, Relational Databases (SQL), NoSQL

Universidad de las Américas Puebla

Aug 2016 - May 2020

B.S in Physics

- **Coursework:** Object-Oriented Programming, Matrix Theory, Probability and Statistics, Computer Simulation

Projects

[AI Generated Podcasts](#) | *Google AI Hackathon*

May 2024

- Podcast generation of research papers by interactions between two LLMs.
- Built Python scripts to integrate with Google Gemini's API, retrieving and processing Q&A data from documents, and generating conversational transcripts using advanced prompt engineering techniques.
- Designed code to connect with Google's Text-to-Speech API. Generated podcast audio for both LLM speakers using Google's Text-to-Speech API.

[Group Listening Dynamics with Spotify API](#)

- Executed project's pipeline in Airflow through Google Cloud Composer as well as deployment in Google Cloud Platform for data retrieval from Spotify's API for a group of users, keeping over 5K data points in a MongoDB cluster for later processing through SparkSQL.
- Co-authored an insightful Medium post showcasing the end-to-end data pipeline and visualization results, facilitating knowledge sharing within the analytics community and garnering over 100 views.