

GLEICE SOUZA CHAVES

Los Angeles, California | (747) 220-4099 | gchaves@usc.edu | <https://linkedin.com/in/gleiceschaves>

EDUCATION

University of Southern California

Los Angeles, CA

Master of Science Applied Data Science

August 2024-May 2026

- Database Systems, Machine Learning, Data Mining, Research Methods, Applied Natural Language Processing

University of Technology - FIAP

Sao Paulo, Brazil

Bachelors of Database | GPA: 3.7

January 2019-December 2022

- Data Modeling, Data Warehousing, NoSQL/SQL, BI, Cloud Services, Data Governance, Business Development, Python

SKILLS

- Data & BI: Python, SQL, DAX; Pandas, scikit-learn, PyTorch; Power BI (Tabular Editor, semantic models), PostgreSQL, MongoDB
- ML & NLP: LLMs(GPT-4/Claude/Llama); Natural Language Processing, prompt engineering; RAG & embeddings, vectors DBs(Chroma/Pinecode), LangChain/LangGraph, A/B testing, Hypothesis Testing, Linear and Logistic Regression, scikit-learn
- Cloud & Pipelines: ETL/ELT; partitioning; performance tuning (SQL/Spark); Azure (Fabric, Databricks), AWS (S3, Athena, EC2, Lambda, DynamoDB, SageMaker), GCP (BigQuery), C#, Java, Alteryx

EXPERIENCE

Keck Medicine of USC

Los Angeles, CA

Data Engineer Intern

March 2025-Present

- Implement an end-to-end NLP Q&A layer on a Power BI/SSAS tabular model; expose KPI/measure metadata, synonyms, and lineage to Copilot, raising exact-match and faithfulness on a 100-question benchmark
- Own enterprise BI/tabular semantic model (Power BI/SSAS, Tabular Editor, DAX) spanning 300+ tables; author hardened SQL Server views to encode business rules and standardize KPI definitions/documentation
- Diagnose query/model bottlenecks with DAX Studio, VertiPaq Analyzer, and SQL Profiler/Query Store; eliminate CROSSJOIN and bi-directional filter hotspots
- Design, author, and maintain advanced SQL CTEs and parameterized views codify business rules (dimensions, hierarchies, period-to-date, SCD) using reusable patterns to keep calculations consistent across reports and models

Diagnostics of America

Sao Paulo, Brazil

Data Engineer & Analytics

June 2021-December 2022

- Designed and deployed a federated data lake + warehouse across AWS (S3/Athena), Databricks, and BigQuery, standardizing schemas and pipelines for ~350 hospitals and labs; automated a mission-critical billing process from ~20 days to hours
- Developed pipelines with Alterio and SQL, implemented incremental loads, RI/outlier/duplicate checks, and lineage/audit logs to raise reliability and simplify downstream analytics
- Created and delivered C-suite Power BI dashboards for service-line and patient-level costing, optimized DAX/SQL models and queries and established daily refresh SLAs to ensure reliable, decision-ready reporting

Link School of Business

Sao Paulo, Brazil

Business Analyst

December 2019-May 2022

- Planned marketing budget allocation across channels (search, social, display, email); established ROI models and reallocated spend to maximize ROAS under CPA targets; performed monthly forecast & pacing reports
- Developed and executed market research in Excel(Power Query, Pivot Tables) with business faculty to identify high-potential startups, establish an grant scoring model and prioritize pipeline to optimize grant allocation and expected ROI

PROJECTS

RAG assistant for finance KPIs - USC

August 2025

- Built a semantic-aware RAG assistant for finance KPIs by exporting 300+ DAX measures from Tabular Editor, embedding in Chroma (sentence-transformers) with retrieval evaluation and dashboards

Human-AI Trust Calibration - PI: Gale Lucas - USC Institute

May 2025

- Co-authored annotation codebooks; performed thematic coding and qualitative analysis; authored methods and documented inter-rater reliability (Cohen's κ) to inform calibrated human-agent teaming

ChatDB: Natural-Language Database Query System - USC

March 2025

- Built a Streamlit + LangChain text-to-SQL/Mongo app turns English into parameterized PostgreSQL SQL and MongoDB aggregations with schema introspection, LangSmith tracing/evaluation and optional Ollama local LLM