Anuar Hage

Email: anuar.hage@gmail.com

LinkedIn: www.linkedin.com/in/anuar-hage

Education

University of Oklahoma Norman,OK

Bachelor of Science - Engineering Physics **August 2014 - June 2018**

Skills

Programming Languages: Python, PostgreSQL, MySQL

Frameworks: Pandas, Numpy, Scikit-Learn, PyTorch, LangChain

Tools: Excel, Power Bi, Tableau, GitHub, MySQL, Superset, Redshift

Platforms: PyCharm, Jupyter Notebook, Visual Studio Code, Sublime, Big Query

Languages: English, French, Spanish

Work Experience

Senior Technical Project Analyst | Dimagi, Inc. **January 2024 - Present**

- Partnered directly with non-technical business leaders to scope analytical questions, translating complex data into actionable insights that accelerated payment cycles for field workers from months to days.
- Leveraged teamwork principles of Scrum to resolve impediments and achieve sprint goals.
- Managed the full analytical lifecycle within a Scrum framework, collaborating with technical teams to deploy data products that accelerated deployment of vaccinators in at-risk areas by 40%.
- Engineered end-to-end data pipelines by extracting and modeling data from SQL databases and APIs to build high-usage BI dashboards for regional deployment teams.
- Implemented robust data cleansing and validation protocols, improving reporting accuracy by 30% and establishing reliable data foundations for operational metrics.

Technical Analyst | Dimagi, Inc. **December 2022 - December 2023**

- Acted as a dedicated analytical counterpart to stakeholders, leading requirements gathering to design metrics that directly supported business and GTM objectives.
- Designed and deployed interactive dashboards using Apache Superset, providing cross-functional teams with continuous visibility into time spent by users amongst different application forms.
- Streamlined internal data request management and communication standards during sprint planning, reducing turnaround time for ad-hoc business analysis by 20%.

Projects

Large-Scale Data Pipeline & Feature Engineering [Python, Redshift, Pandas, Riot Games API] **February 2022 - October 2022**

- Engineered an automated data ingestion pipeline using Python to extract, parse, and clean over 2.5 million raw match records from the Riot Games API, handling rate limits and data validation.
- Designed and implemented a relational database schema in Redshift to store and model the extracted data, optimizing for efficient querying and downstream analytical use.
- Conducted extensive Exploratory Data Analysis (EDA) and feature engineering using Pandas and NumPy to identify key win-condition metrics and player behavioral trends, transforming unstructured data into a structured format ready for predictive modeling.

Legal RAG System & Semantic Search [Python, LangChain, Pinecone, Embeddings] **November 2025**

- Engineered a Retrieval-Augmented Generation (RAG) system to perform semantic search and automated question-answering across the entirety of the Mexican Penal Code.
- Extracted, cleaned, and chunked dense, unstructured legal texts, tokenizing using Qwen3-235b-A22B-Instruct-2507 and transforming them into vector embeddings using llama-text-embed-v2 to capture semantic meaning.
- Designed the retrieval architecture using LangChain and Pinecone, enabling highly accurate, context-aware information retrieval from complex documentation.