Ireri Avila

Email | LinkedIn | GitHub | (+1) 650-398-8571 | San Francisco, CA

Data Scientist with strong problem-solving skills and 5+ years of experience, driving impactful process improvements in industry and academia.

Education

M.S in Data Science *Jul 2023 – Jun 2024*
University of San Francisco *San Francisco, CA*
   Relevant Coursework: Data Acquisition, Machine Learning, Regression, Time Series, A/B Testing, Relational Databases (SQL), NoSQL

B.S in Physics *Aug 2016 – May 2020*
Universidad de las Américas Puebla *Puebla, MEXICO*
   Relevant Coursework: Object-Oriented Programming, Matrix Theory, Probability and Statistics, Computer Simulation

Professional Experience

Data Scientist *October 2023 - Present*
Give Us The Floor *Remote*
- Designed and implemented machine learning models, leveraging NLP techniques along with transformers for binary classification.
- **Reduced human intervention by 90% saving $90K per year** and impacting more than 2,000 users.
- Improved message retrieval through API requests by 80% with a Python script for effective date filtering for 400K messages.
- Created an Apache Airflow DAG for scheduled runs of full pipeline from data fetching to final classification.
- In charge of training new data science interns.

Data Analyst, Internal Global Audit *Mar 2021 – May 2022*
Grupo Bimbo *Mexico City, MEXICO*
- Creation of **15 analytics scripts** for audit reports from an Oracle Database connection
- Implementation of Python scripts for the automatization of monthly programmed tasks, **saving over 3 hours per month.**
- Worked closely with different teams to create a Dynamic Risk Assessment dashboard for fixed assets using Power BI.
- Validation of information and data consistency of Azure data lakes through SQL queries.
- Optimized and redesigned scripts that previously crashed for large inputs of data while reducing **computing time by 40%.**

Data Scientist, Research *May 2019 – May 2020*
National Institute of Astrophysics, Optics, and Electronics (INAOE) *Puebla, MEXICO*
- Data mining of around 250K records of data from 28 different sources using Beautiful Soup.
- Led the creation of a Python script to normalize, analyze, and model data using time series analysis.
- Data visualizations using Matplotlib and Seaborn for the aforementioned data, as an intern.

Projects

AI Generated Podcasts - Google AI Hackathon
May 2024

- Podcast generation of research papers by interactions between two LLMs.
- Developed 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

- Designed and implemented 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 that data in a MongoDB cluster for later processing through SparkSQL.
- Medium post to show pipeline and visualization results.
-

Color Analysis with AI

- Obtain tonality and seasonality based on pictures, using neural networks for classification.
- Mined and curated dataset for all 12 categories for color analysis.
- Designed neural networks for multiclass classification.

Skills

- Python, SQL, NoSQL, MongoDB, DataBricks, Transformers, HuggingFace, Google Cloud Platform, AWS, A/B Testing, Airflow, Spark, PyTorch, NLP, Oracle, Git, GitHub, Spacy, Plotly, Tableau, PowerBI, ACL Analytics, LaTeX, English, Spanish, French, Italian