Juan Felipe Toro Salgado

Bogota D.C., Colombia torosalgado.jt@gmail.com +573016049641 LinkedIn Github www.juanfts.com

PROFILE

I am engineer with expertise in cloud computing and data integration. My experience includes designing and maintaining efficient data pipelines using tools like Apache Airflow (Google Cloud Composer) to streamline ETL processes and ensure seamless data flow into centralized warehouses like Google BigQuery. Skilled in Python, SQL, and cloud platforms such as Azure and GCP, I excel in handling large-scale data processing, statistical analysis, and data modeling with tools like DBT Cloud, transforming raw data into valuable insights.
Currently pursuing a Master's in Applied Artificial Intelligence, I stay at the forefront of technological advancements, leveraging my skills to optimize performance and ensure data quality. I thrive on collaborating with cross-functional teams to integrate analytics solutions and deliver actionable insights that drive innovation and strategic decision-making. My passion lies in empowering organizations to embrace data-driven strategies and capitalize on cutting-edge technologies for impactful outcomes.

EXPERIENCE

Data scientist

Amapola tech , Bogota **September 2024 – now**

- Designed and implemented robust (DAGs) using Apache Airflow to automate and streamline complex workflows, including ETL processes, ensuring high availability and fault tolerance. Developed scalable, maintainable, and resilient pipelines for the ingestion, transformation, and loading of structured and unstructured data into the BigQuery data warehouse. Leveraged Google Cloud Composer to manage Airflow environments at scale, ensuring seamless integration with existing cloud infrastructure. Conducted thorough data validation, implemented error-handling mechanisms, and optimized task scheduling and retry strategies to maintain workflow efficiency and data integrity. Continuously monitored DAG and pipeline performance, promptly resolving failures to minimize downtime and enhance reliability.
  Achievements: Successfully automated a debt management modeling pipeline using Apache Airflow,reducing processing time by 50%, significantly enhancing decision-making capabilities and operational efficiency in financial risk assessments.
- Partnered with teams to deploy machine learning models into production and embed analytics solutions into core business processes. Provided technical guidance on the integration of data pipelines with advanced analytics tools. Conducted rigorous testing and optimization to ensure models performed reliably and efficiently in production environments. Facilitated knowledge sharing to enable better utilization of analytics tools across teams.
  Achievements: Conducted fine-tuning of pre-trained models for domain-specific applications, leading to better performance metrics and business impact.
- Architected and implemented advanced data transformation pipelines using DBT Cloud to model complex business logic and produce analytics-ready datasets. Employed rigorous version control practices to manage changes and ensure reproducibility. Collaborated with stakeholders to gather requirements and design models that aligned with organizational needs. Continuously monitored and refined data models to optimize performance and adapt to evolving business requirements.
  Achievements: Successfully designed and deployed scalable data transformation pipelines that streamlined complex business logic, reducing data processing time by 40%.

Junior Engineer

Online Building plans, Bogota **February 2023 – March 2024**

- Coordinated closely with project managers and the engineering team to deliver essential information across all engineering disciplines involved in a project, leveraging software tools like Revit and CAD. This involved swiftly meeting their needs and furnishing comprehensive construction documents including Mechanical Design and Mechanical Calculations, within tight timelines and following design protocols established by the office and the current Florida Building Code
  Achievements: Successfully delivered documentation for 10 projects per month ensuring client satisfaction and timely delivery of solutions.
- Create a Power BI dashboard with the Miami-Dade database efficiently extracts real estate business leads, offering insights into market trends, property listings, and investment opportunities.
  Achievements: Demonstrated exceptional analytical skills in translating complex real estate data into actionable insights, empowering stakeholders to make informed decisions.
- Create data extraction algorithm for Florida contractors uses web scraping to gather business details from online sources like directories and websites. It compiles information such as company names, contact details, and services offered, allowing contractors to quickly access and analyze market data to inform strategic decisions and foster business growth in Florida's construction industry.
  Achievements: This innovative solution significantly reduced manual effort and time required for data gathering, resulting in a 40% increase in data accuracy and completeness

Engineer

Maternal and Child Clinic Casa del Niño **October 2021 - February 2023,**

- Responsible for managing the procurement process for fixed assets and biomedical equipment in our institution. Working closely with various stakeholders to identify needs, research options, negotiate contracts, and ensure the timely acquisition and installation of equipment to support our operations and deliver high-quality services.
  Achievements: Cost Optimization of 9% for the acquisition of assets and biomedical equipment
- Responsible for maintaining accurate and up-to-date records and documentation to ensure compliance with audit requirements and regulatory standards. With attention to detail, excellent organizational skills, and the ability to work efficiently both independently and as part of a team.
  Achievements: Successfully increased the pass rate of audits by 30% through the implementation of comprehensive quality control measures
- Create and maintain inventory levels of assets, biomedical equipment, and Hardware. This includes setting par levels, conducting regular counts, and tracking usage patterns.
  Achievements: Creating and maintaining EXEL database specifically dedicated to managing hospital assets and creating visualization dashboards.

PROJECTS

MLOps Classification – Insurance Company

- Implemented an end-to-end MLOps workflow to classify customer risk levels for an insurance company. Automated data ingestion, feature pipelines, training, model registry, CI/CD and deployment using MLflow, Docker and GitHub Actions.
- Models are promoted to production only if they improve KPIs, and served via serverless API for underwriting and collections decisioning.

Real-Time Streaming Pipeline – Call Center Analytics

- Developed a real-time streaming pipeline to process live call center events (calls, agents, durations, outcomes) using Kafka + Flink + PostgreSQL/BigQuery.
- KPIs and operational metrics update instantly in Grafana, enabling real-time monitoring, SLA tracking and proactive decision-making.

Batch ETL Orchestration – **Airflow + dbt + GCP + Looker**

- Designed and orchestrated batch data pipelines using Apache Airflow and dbt on Google Cloud.
- Automated ingestion, transformations and data quality checks feeding BigQuery models used for business analytics dashboards in Looker, supporting daily financial and operational reporting.

Cloud Resume, Res Azure with Blob Storage, Functions, CosmosDB, and GitHub Actions, Webpage

- Webpage with HTML,JavaScript and CSS
- Developing a CI/CD workflow with GitHub
- Deploying a static web app with a serverless database using gcp services

Data scraping for yellow pages to extract business details Python, pandas, Scrapy, Github

- A data scraping and data analysis tool to fetch data from yellow pages (no API).
- The program extracts specific business information from Yellow Pages, such as business names, addresses, contact details, types of services or products offered, ratings, and reviews.
- Gathered and analyzed data from over 1000 business to identify patterns of business in certain areas.

Assessing Underground Water Quality in Mexico using Data Science, pandas, scikit-learn, NumPy, matplotlib.pyplot seaborn, GitHub

- Identify the quality of underground water sources in Mexico using data-driven techniques and provide valuable insights for environmental monitoring and resource management.
- Algorithms used: clustering (KMeans).

EDUCATION

Master of Science in Applied Artificial intelligence. Majored in Intelligent software.

Interest in Statistics, Deep learning, Machine learning, Data analysis, cloud computing, AI • Tecnologico De Monterrey • International campus

Bachelor of Science in Mechanical Engineer

Interest in Mathematics, Algebra, Programming, Statistics • Escuela Colombiana de Ingeniería Julio Garavito Bogotá

SKILLS

- Programming Languages: Python, SQL, HTML,CSS,SQL,noSQL,java script
- Technology: Jupyter, Power BI, tableau, GCP, AZURE, Visual Studio Code,GitHub CI-CD Libraries: pandas, scikit-learn, TensorFlow, Keras, Sea born, Scrapy, beutifullsoup, selenium

CERTIFICATIONS

- Azure Fundamentals AZ- 900

COURSES

- Python Basics for Data Science (IBM) Data Analysis with Python (IBM)
- The Structured Query Language SQL (University of Colorado Boulder) AI for Everyone: Master the Basics (IBM)
- Google Cloud Computing Foundations: Cloud Computing Fundamentals

SOFT SKILLS

- CriticalThinkin
- Data Interpretation
- Communication Skills Teamwork
- Attention to Detail

LANGUAGES

- English: B2
- Spanish: Native