



# Diego Villagrán Salazar

Data Science & AI Analyst (Banking / Risk)

Mexico City, Mexico  
+52 55 7207 1183  
diegovillasal@gmail.com  
<https://www.dvillagrans.dev>  
in dvillagrans  
Hybrid-ready

## Summary

Production-focused Data Science student (ESCOM-IPN, 2026) with experience building AI-powered data pipelines and analytical systems for decision support. Currently leading LLM-based document extraction at EyeNet (200+ docs/day, 92% extraction accuracy) and supporting 10K+ daily transactions. Strong background in statistical modeling, ML pipelines, A/B testing, and dashboarding for non-technical stakeholders.

## Education

2022–2026 **B.Sc. Data Science**, *Escuela Superior de Cómputo (ESCOM-IPN)*, Mexico City  
GPA: 8.5/10.0. Relevant coursework: Statistical Modeling, Machine Learning, Big Data, Probability, Databases.

## Experience

Apr 2025 – **AI & Data Automation Analyst**, *EyeNet*, Remote

- May 2026 ○ Designed and implemented LLM-based document analysis pipelines (OpenAI, Gemini) achieving 92% extraction accuracy across 200+ docs/day.
- Built automated data-quality monitoring with CI/CD and alerting, catching >95% of issues pre-production.
- Developed data infrastructure (PostgreSQL, MongoDB, Redis) supporting 10K+ daily transactions across services.
- Automated ETL workflows with Python and SQL, reducing manual effort by 65%.

2025 – **AI Model Analyst**, *Private GPU Cluster*

- Present ○ Benchmark open-source LLMs (LLaMA, Mistral) vs. proprietary APIs; produce accuracy/latency/cost reports for model selection.
- Build reproducible evaluation frameworks and document methodology for internal stakeholders.

## Selected Projects

2024 **HSBC-relevant: AML / Compliance Analytics Analog**

Developed clustering and anomaly-detection prototypes analogous to transaction monitoring for AML use-cases.

2024 **COVID-19 Risk Profiles**

Applied K-Means and Fuzzy C-Means on 30M+ health records to derive 9 risk profiles; produced interactive dashboard and reproducible pipeline.

2024 **NYC Ride-Hailing Analytics**

Built fare prediction model ( $R^2 > 0.85$ ) and airport classification model (92% accuracy); delivered Streamlit dashboard for operations.

2024–2025 **TimeUp — Business Analytics Platform**

Designed dashboards and ETL pipelines; reduced admin reporting work by 80% for SME clients.

2024 **India Air Quality ETL System**

Engineered scalable ETL on Azure Databricks processing 2M+ daily records from 500+ sensors.

## Technical Skills

Core Python (advanced), SQL (advanced), Pandas, NumPy, Scikit-learn

ML XGBoost, TensorFlow, PyTorch, Clustering, A/B Testing, Statistical Inference

LLM OpenAI API, Gemini API, LLaMA, Mistral, NLP, prompt engineering  
& AI

Data PySpark, ETL/ELT, Azure Databricks, BigQuery, PostgreSQL, MongoDB  
Engineering

[11pt,a4paper,sans]moderncv [][moderncvstyleclassic moderncvcolorblue

[utf8]inputenc [T1]fontenc [english]babel [scale=0.78,top=0.9cm,bottom=1.6cm]geometry

# Diego Villagrán Salazar

*Data Science & AI Analyst (Banking / Risk)*

Mexico City, Mexico

+52 55 7207 1183

+52 55 7207 1183

✉ [diegovillasal@gmail.com](mailto:diegovillasal@gmail.com)

🌐 <https://www.dvillagrans.dev>

**in** [dvillagrans](#)

**in** [dvillagrans](#)

Hybrid-ready

## Summary

Production-focused Data Science student (ESCOM-IPN, 2026) with experience building AI-powered data pipelines and analytical systems for decision support. Currently leading LLM-based document extraction at EyeNet (200+ docs/day, 92% extraction accuracy) and supporting 10K+ daily transactions. Strong background in statistical modeling, ML pipelines, A/B testing, and dashboarding for non-technical stakeholders.

## Education

2022–2026 **B.Sc. Data Science**, *Escuela Superior de Cómputo (ESCOM-IPN)*, Mexico City  
GPA: 8.5/10.0. Relevant coursework: Statistical Modeling, Machine Learning, Big Data, Probability, Databases.

## Experience

Apr 2025 – **AI & Data Automation Analyst**, *EyeNet*, Remote

- May 2026 ○ Designed and implemented LLM-based document analysis pipelines (OpenAI, Gemini) achieving 92% extraction accuracy across 200+ docs/day.
- Built automated data-quality monitoring with CI/CD and alerting, catching >95% of issues pre-production.
- Developed data infrastructure (PostgreSQL, MongoDB, Redis) supporting 10K+ daily transactions across services.
- Automated ETL workflows with Python and SQL, reducing manual effort by 65%.

2025 – **AI Model Analyst**, *Private GPU Cluster*

- Present ○ Benchmark open-source LLMs (LLaMA, Mistral) vs. proprietary APIs; produce accuracy/latency/cost reports for model selection.
- Build reproducible evaluation frameworks and document methodology for internal stakeholders.

## Selected Projects

2024 **HSBC-relevant: AML / Compliance Analytics Analog**

Developed clustering and anomaly-detection prototypes analogous to transaction monitoring for AML use-cases.

2024 **COVID-19 Risk Profiles**

Applied K-Means and Fuzzy C-Means on 30M+ health records to derive 9 risk profiles; produced interactive dashboard and reproducible pipeline.

2024 **NYC Ride-Hailing Analytics**

Built fare prediction model ( $R^2 > 0.85$ ) and airport classification model (92% accuracy); delivered Streamlit dashboard for operations.

2024–2025 **TimeUp — Business Analytics Platform**

Designed dashboards and ETL pipelines; reduced admin reporting work by 80% for SME clients.

## 2024 India Air Quality ETL System

Engineered scalable ETL on Azure Databricks processing 2M+ daily records from 500+ sensors.

### Technical Skills

Core	Python (advanced), SQL (advanced), Pandas, NumPy, Scikit-learn
ML	XGBoost, TensorFlow, PyTorch, Clustering, A/B Testing, Statistical Inference
LLM & AI	OpenAI API, Gemini API, LLaMA, Mistral, NLP, prompt engineering
Data Engineering	PySpark, ETL/ELT, Azure Databricks, BigQuery, PostgreSQL, MongoDB
BI	Power BI, Tableau, Streamlit, Grafana
Cloud & DevOps	AWS, Azure, Docker, Git, CI/CD

### Certifications

Google AI Essentials; Build a Data Warehouse with BigQuery; BigQuery ML for Predictive Modeling; Kubernetes management.

### Languages

Spanish (native); English (B2 — professional working proficiency).