

AARON IVAN ORTIZ URZUA

Data Scientist | Business Intelligence & Analytics

Queretaro, Mexico | Remote (US / LATAM) or hybrid in Queretaro and Mexico City

aaron@aaronurzua.com | linkedin.com/in/aaron-urzua- | aaronurzua.com

PROFESSIONAL SUMMARY

Data Scientist and Statistical Industrial Engineer with 5+ years building the reporting and prediction layer for operations in manufacturing, retail technology and public statistics. Replaced recurring manual reporting with automated data pipelines and self-serve dashboards in Power BI, R Shiny and Python; built an NLP pipeline that coded 83% of 148,310 free-text survey records without human review; and owns the predictive maintenance program that has cut reactive incidents by up to 9%. Works end to end: SQL extraction, ETL, modeling, deployment and delivery to the teams that use it.

CORE COMPETENCIES

Data Analytics · Business Intelligence · Data Visualization · Dashboard Development · Reporting Automation · ETL and Data Preparation · Data Cleaning · Predictive Analytics · Time Series Analysis · Forecasting · Anomaly Detection · Predictive Maintenance · Root Cause Analysis · KPI Reporting · Natural Language Processing · Statistical Analysis · Stakeholder Reporting · Data Storytelling

TECHNICAL SKILLS

Languages & Querying: Python, R, SQL (MySQL, SQL Server), DBeaver, JavaScript

Analysis & Machine Learning: pandas, NumPy, scikit-learn, XGBoost, Gradient Boosting, Random Forest, Decision Trees, ARIMA, time-series decomposition and forecasting, TensorFlow, Hugging Face Transformers

Visualization & BI: Power BI, Plotly, Highcharts, Leaflet, ggplot2, Matplotlib, Seaborn, DataTables, data storytelling, Microsoft Excel (advanced reporting)

Analytical Apps & Delivery: R Shiny, flexdashboard, Streamlit, Flask, FastAPI, Docker, REST APIs, Git, GitHub

Data Engineering & Platforms: ETL and data pipeline design, data quality validation, SQL optimization, ServiceNow incident data, Microsoft Azure (AZ-900 certified), Databricks

Methods: Predictive maintenance, anomaly detection, feature engineering, Pareto analysis, root cause analysis, statistical process analysis

PROFESSIONAL EXPERIENCE

Toshiba Global Commerce Solutions | Senior Data Scientist

Mexico City, Mexico | May 2023 - Present

- Own the predictive maintenance program for a national retail chain's in-store hardware (POS terminals, self-checkout units, receipt printers, back-office servers), **cutting reactive incidents by up to 9% against a baseline of about 14,000 a month.**
- Train and tune the XGBoost failure-prediction models as the program's only data scientist, over the PAS monitoring event history (platform, failure type, serial number), so maintenance is scheduled preventively instead of after a breakdown.
- Forecast failure rates for newly deployed equipment with no service history, using ARIMA and time-series decomposition to set an initial baseline and project expected monthly replacements.
- Deployed a Hugging Face transformer that classified free-text service-request notes as failure vs. non-failure, while the program was run out of the US organisation, **automating 60% of a review that until then was fully manual.**
- Design and optimize SQL extractions over device event and alert tables, and engineer per-equipment features from a stream of isolated alerts, producing the datasets the predictive models run on.
- Build Python ETL pipelines over the client's ServiceNow incident history with automated data quality checks (nulls, duplicates, type and range validation), removing recurring manual preprocessing from the monthly analysis cycle.
- Run Pareto and root cause analysis on failures by store and lane to isolate dominant failure modes, and drive the resulting reliability actions with cross-functional operations and field-service teams.
- Deliver the Power BI and R Shiny dashboards and executive reporting on top of the program, tracking failure trends by platform and equipment for the operations team.

Optimus Steel USA | Junior Data Scientist

Houston, Texas, USA | Mar 2022 - May 2023

- Built a tap-temperature prediction model for the ladle refining furnace (decision trees, random forest, gradient boosting) that **cut required samples per heat from 6 to 3 (-50%)**, saving about \$30 per heat.
- Served the model through an interactive web app where operators select steel grade and enter process variables, so predictions run on the floor without an analyst in the loop.
- Replaced the monthly bundle-scrap report, **3 days of analyst work every month**, with an automated MySQL-to-HTML pipeline in R, delivering a self-service dashboard with Pareto analysis of non-conformance codes, heatmaps and stacked breakdowns.
- Implemented an automated oil-leak detection and alerting system on hydraulic equipment that pushes alerts to a Telegram group and to email the moment a threshold is crossed, shortening the gap between anomaly and maintenance response.
- Containerized analytical applications with Docker and deployed model-facing web tools (Streamlit, Flask) so plant stakeholders could run predictions without analyst involvement.

INEGI - National Institute of Statistics and Geography | Data Analyst

Aguascalientes, Mexico | Mar 2021 - Mar 2022

- Developed the automated SINCO/SCIAN coding system for the National Survey of Occupation and Employment (ENOE), reaching **82.84% automatic coding for occupation and 84.02% for economic activity across 148,310 free-text records**.
- Designed the five-stage NLP data pipeline behind it: text normalization, synonym-dictionary substitution, special-table matching and access-table lookup with INCLUDE/EXCLUDE rules, each stage assigning a confidence indicator to the resulting code.
- Delivered the system as a navigable R Shiny application documenting methodology, execution and results, with word-frequency analysis, word clouds, sentiment analysis and interactive result tables for the survey team.
- Produced national and state-level reporting on EEN data-capture modality (internet, digital media, paper) for all 32 states as interactive HTML reports with Highcharts visualizations, Leaflet choropleth maps and searchable data tables.
- Cleaned and prepared ECIS 2021 and ENAFIN 2021 survey datasets and processed occupation and economic-activity variables used in institutional analysis and publications.

EDUCATION

Universidad Autonoma de Aguascalientes (UAA) | B.Eng. Statistical Industrial Engineering

Aguascalientes, Mexico | 2021

- EGEL-CENEVAL Outstanding Performance Award (Testimonio de Desempeno Sobresaliente), national graduate assessment.
- Peer Tutor (Tutor Par) and Institutional Recognition (CCB), Universidad Autonoma de Aguascalientes.

CERTIFICATIONS

Cloud: Microsoft Certified: Azure Fundamentals (AZ-900)

Python (IBM): Python for Data Science; Machine Learning with Python; Deep Learning with TensorFlow; Data Visualization using Python

R (IBM): Data Visualization with R; Machine Learning with R; Using R with Databases

Professional Standard (A2 Capacitacion, 2021): Python (ECPY), R (ECSR), MySQL (ECSQL)

Version Control: Professional Git Course; In-depth GitHub Course

LANGUAGES

Spanish - Native | English - Professional working proficiency