ANTONNY PICALUA

Ingles, Frances, Portugués, Español

Antonny is a passionate and self-driven Data Analyst with strong foundations in data manipulation, statistical modeling, and data visualization. With hands-on experience in Python, SQL, and data preprocessing techniques, Antonny has developed solid analytical thinking and a data-driven mindset. His recent projects focus on handling imbalanced datasets, conducting hypothesis testing, detecting fraud patterns, and generating dynamic reports for decision-making. He combines technical skill with curiosity and a deep interest in machine learning, marketing analytics, and artificial intelligence trends.

CONTACTO

(57) 3504241139 - Bogota, D.C , Colombia - josepicalua64@gmail.com - josepicalua63@gmail.com

PROFESSIONAL EXPERIENCE

Remote | **Dec 2024 - Present**

Hands-on analytical projects and advanced data processing experience

- Conducted end-to-end data analysis using Python, including data wrangling with Pandas, NumPy, and the creation of reusable preprocessing pipelines.
- Managed and analyzed structured datasets from simulated healthcare, marketing, and operational domains, focusing on real-world business questions.
- Performed statistical testing (A/B testing, p-values, confidence intervals, binomial distributions) to drive insights and support data-driven decisions.
- Addressed imbalanced datasets using resampling techniques such as SMOTE and undersampling; evaluated results using precision, recall, F1-score, and confusion matrices.
- Developed exploratory data analysis (EDA) reports to detect patterns, anomalies, outliers, and trends, using visual libraries like Matplotlib and Seaborn.
- Executed SQL queries on PostgreSQL and MySQL databases to extract, filter, and join tables, optimizing queries for performance and scalability.
- Designed data cleaning workflows: removing duplicates, handling missing values via imputation (mean/median/mode), and performing data normalization and encoding. 2012 - 2014
- Applied feature engineering techniques to improve model input quality, including feature scaling, binning, and interaction terms.
- Practiced machine learning models for classification tasks (e.g., fraud detection), focusing on model evaluation and proper data splitting (train/test/validation).
- Created mini case studies combining business understanding with technical implementation, especially in healthcare analytics and consumer behavior prediction.

SKILLS

Data Analysis & Statistics

- Exploratory Data Analysis (EDA)
- Statistical Testing (A/B Testing, Hypothesis Testing, p-values)
- Handling Imbalanced Datasets (SMOTE, undersampling)
- Outlier Detection & Missing Data Treatment
- Data Imputation (mean, median, mode)
- Feature Engineering & Selection
- Data Normalization & Encoding Techniques

Programming & Tools

- Python (Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn)
- SQL (PostgreSQL, MySQL – joins, aggregations, filters, subqueries)
- Jupyter Notebooks for workflow development
- Microsoft Excel (Formulas, Pivot Tables, Conditional Formatting)
- Git (basic version control)

SKILLS

Machine Learning Foundations

- Classification Models (Logistic Regression, Decision Trees)
- Model Evaluation (Accuracy, Precision, Recall, F1-Score, Confusion Matrix)
- Train/Test/Validation Splits
- Basic understanding of anomaly detection and fraud analytics

Soft Skills

- Analytical Thinking & Problem Solving
- Self-learning and Independent Research
- Documentation & Clear Reporting
- Adaptability to New Tools and Workflows

HOBBIES & INTERESTS

- Data Storytelling: Enjoys transforming raw data into compelling narratives using visualizations and analytical insights.
- Learning Emerging Tech: Passionate about exploring topics like artificial intelligence, machine learning, and automation tools.
- Problem Solving: Enjoys tackling logic puzzles, probability games, and brainteasers that enhance analytical thinking.
- Reading & Research: Regularly reads articles and case studies related to data science, healthcare analytics, and fraud detection.
- Fitness & Wellness: Committed to personal health through daily physical activity, mindfulness, and balanced routines.
- Volunteering in Education: Interested in helping others learn data tools and analytics through peer support or online forums.

KEY PROJECTS

📊 **Imbalanced Dataset Analysis for Fraud Detection**
Python, Pandas, Scikit-learn

- Built a pipeline to detect fraudulent transactions in an imbalanced dataset using oversampling techniques (e.g., SMOTE).
- Performed EDA to uncover hidden patterns and trained classification models to distinguish legitimate from fraudulent records.
- Evaluated model performance using precision, recall, and F1-score to handle class imbalance effectively.

🧪 Statistical Testing on E-commerce Campaigns
Python, NumPy, SciPy, Seaborn

- Simulated A/B testing scenarios to analyze the effectiveness of digital marketing campaigns.
- Applied binomial distribution and hypothesis testing to validate business assumptions.
- Visualized the results to communicate findings clearly to non-technical stakeholders.

🗃️ **SQL-Based Patient Data Analysis**
PostgreSQL, MySQL
Designed and executed advanced SQL queries to extract, clean, and analyze patient data from relational databases.
Applied aggregation, filtering, and CASE logic to generate customized reports and summaries.
Used SQL joins and subqueries to correlate clinical observations across multiple tables.

📉 **Data Cleaning & Preprocessing Framework**
Python, Pandas
Created reusable functions to automate data cleaning, including null detection, duplicate removal, and feature scaling.
Built a flexible script that detects outliers and imputes missing values based on column distribution.
Prepared data for use in machine learning workflows with proper encoding and normalization.

📈 **Visual Exploratory Data Analysis (EDA)**
Matplotlib, Seaborn, Excel
Designed EDA dashboards to analyze trends and detect anomalies in datasets from healthcare and consumer domains.
Used histograms, boxplots, heatmaps, and pair plots to reveal key variables affecting outcomes.
Provided actionable recommendations based on visual insights.