

Ixchel Peralta-Martinez

Ixchelperalta4000@gmail.com | (919) 672-5289 | www.linkedin.com/in/ixchel-peralta | <https://github.com/IxchelPM/my-portfolio>

EDUCATION

Master of Science in Data Science

Dec 2024

University of North Carolina Wilmington, GPA: 4.0

Bachelor of Science in Statistics

May 2023

University of North Carolina Wilmington, GPA: 3.6

SKILLS

Languages & Tools: Python, R, SQL, BigQuery, PostgreSQL, Git/GitHub

Visualization: Tableau, R Shiny, ggplot2, Excel charts, Presentation, Documentation

WORK EXPERIENCE

Data Analyst Intern | Pearson

Jun 2024 – Aug 2024

- Conducted in-depth analysis of software data using R and machine learning to improve catalog integrity.
- Applied generative AI, NLP, and machine learning to deduplicate entries, reducing redundancy and enabling \$30K in potential savings.
- Built interactive R Shiny and Excel dashboards for stakeholder use, improving data catalog consistency and transparency.

Graduate Research Assistant | UNCW

Sep 2023 – May 2024

- Collaborated on generative AI/LoRa model to replicate an artist's style using Stable Diffusion and advanced image generative techniques.
- Enhanced the performance of text-to-image AI models by designing domain-specific prompts and fine-tuning outputs.

Statistical Analysis Intern | Biostudy Solutions LLC

Sep 2022 – May 2023

- Cleaned and analyzed clinical trial data for pharmacokinetic and in vitro studies.
- Created and validated tables and listings for FDA-compliant submissions, ensuring 100% data accuracy.
- Supported 10+ studies through effective clinical data management and visualization using SAS and Excel.

PROJECTS

Spotify Song Popularity Prediction

- Employed exploratory data analysis techniques and visualization, including pie charts, histograms, boxplots, scatter plots, bar charts, and heat maps in R to uncover key data features.
- Utilized R to implement and evaluate diverse models including linear discriminant analysis, logistic regression, k-nearest neighbor, decision trees, random forests, bagging, and AdaBoost.
- Created a detailed report summarizing data feature analysis and model performance evaluation, demonstrating the best-performing model.

Global Airline Fleet Data Cleaning & Transformation

- Developed a SQL-based data pipeline to clean and standardize fleet records for 100+ global airlines using raw data. Removed symbols and cast string values to integers to normalize cost and fleet size fields.
- Recalculated total fleet values and aircraft counts using SQL expressions and stored cleaned output in a view for downstream analysis.
- Standardized airline names and restructured fields to support analysis across manufacturers and regions.