# Jaime Andrés Castañeda

*Data Scientist*

jacasta2@gmail.com | +57 (300) 222-4658 | Colombia | linkedin.com/in/jaime-data-science


## Summary of Qualifications

- Researcher with ~15 years of experience, leading and collaborating in projects that leverage experiments and data science to gain insights into decision-making
- Engineering professional that applies data science and machine learning in research and industry projects
- Rigorous, organized and ready to give a hand to my teammates and learn from them

## Skills

- **Programming languages:** Python, SQL
- **Analytics:** data wrangling, data visualization, descriptive and inferential statistics, econometrics, machine learning, NLP, simulation, optimization
- **Tools:** Pandas, Numpy, Matplotlib, Seaborn, Scikit-learn, Keras, FastAPI, Streamlit, SageMaker, Postgres, Git, Power BI, Stata, SPSS, Excel, Qualtrics, LimeSurvey

## Relevant Experience

**Data Scientist,** *Colombian ICT Ministry* Mar 2023 – Dec 2023 | Remote

- Built a composite index for prioritizing internet connectivity initiatives for schools using multiple techniques such as principal components analysis and factor analysis in Python
- Prepared 40,000+ rows of school data using Python, performing data integration, imputation, transformation and feature engineering to improve its quality
- Compared indexes using multiple tools such as Bland-Altman plots, accuracy metrics and customized functions in Python, providing insights into the indexes' impact on the target population

**Data Science Fellow,** *Correlation One* Mar 2022 – Jul 2022 | Remote

- Collaborated with five teammates to analyze motorcycle accidents in Bogotá for prioritizing road safety operations, pulling accident data from a GIS web service into a local Postgres database using Python and SQL
- Analyzed 200,000+ rows of accident data, creating heatmap visualizations based on crosstabs that relate accident type, accident severity and vehicles involved using SQL, Python and Seaborn
- Developed an unsupervised machine learning model by engineering continuous and categorical features using the K-Prototype algorithm in Python, defining prioritization levels for road safety operations

**Researcher,** *Universidad del Rosario* Apr 2016 – Jan 2022 | Bogotá, Colombia

- Co-led a funded research project that evaluated the role of expediting prices in inventory replenishment, implementing an in-person and online behavioral experiment with 400+ participants
- Examined 25,000+ rows of replenishment decisions, creating customized box plot visualizations of the decisions grouped by expediting prices and averaged over subjects using Python and Seaborn
- Fit several supervised machine learning models using random-effects panel data models in Python, providing insights into how expediting prices impact replenishment decisions

## Additional Experience

- Built an NLP app that recommends papers based on the similarity between the papers' abstracts and the user's interests using text processing, vector embeddings, Pinecone and Streamlit 2024
- Deployed a ML app that compares the influence of a pair of NBA teammates' stats on their team's winning chances using logistic regression, SHAP values, Hopsworks and Streamlit 2023
- Implemented an API endpoint for my data science fellowship project using FastAPI to allow users to retrieve predictions from the clustering model 2023
- Built a data pipeline module for my data science fellowship project to update the local database with monthly accident data from the GIS web service 2023

## Education

**Ph.D., Economics,** *Università della Svizzera italiana* 2013 | Lugano, Switzerland