SoniaMendizábal
data scientist

about

Name Sonia Georgina Mendizábal Claustro
Email sonia.gmcl@gmail.com

Github Songeo
LinkedIn soniamendizabal
Twitter @SonGeo
Web elated-augusta

interests

statistical models • machine learning • design of experiments • sampling • longitudinal analysis • spatial analyisis • social good research • constant learning • interdisciplinary teams • knowledge transfer

skills

analysis

ML • EDA • DOE • sampling • bootstrap • multilevel modeling • bayesian models • longitudinal data • HTS optimization • multivariate analysis • LSH • implicit recommendation

programming

R • SQL • Python • Bash • C++ Git • Docker • awk Shiny • JAGS • Stan Big Query • Athena • Snowflake GCP • AWS markdown • Tableau • Power BI

communication

Liaison between employers and clients for technical aspects of projects • Presentation of results to technical and non-technical audiences • Guest speaker • Collaboration in multidisciplinary teams

languages

Spanish. *Mothertongue*
English. *Advanced*

experience

Oct/21 - Today

Data Scientist Lead
| RAPPI |Remote Mexico City
Developed and maintained data analysis pipelines to evaluate the impact of COVID-19 via mobility data across Latin American cities. Advised and supported the division's specialists in the development of research. Authored technical documentation for non-technical audiences

Jul/20 - Feb/22

Data Science Consultant
| INTER-AMERICAN DEVELOPMENT BANK | Remote Washington, DC
Developed and maintained data analysis pipelines to evaluate the impact of COVID-19 via mobility data across Latin American cities. Advised and supported the division's specialists in the development of research. Authored technical documentation for non-technical audiences.

Jun/21 - Aug/21

R Programming Tutor
| ECLAC UNITED NATIONS | Remote Mexico City
Conducted a R programming course focused on manipulation and communication analysis skills.

Sep/18 - Apr/21

Data Scientist Senior
| GEEK END STATISTICAL CONSULTING | Mexico City
Designed, launched and operated ad-hoc solutions based on data. Implemented statistical models, machine learning algorithms, experimentation frameworks and database integration. Mentored junior data scientists and manage the delegation of data products. Cooperated and assisted multidisciplinary teams inside organizations. Authored technical documentation and deliverables for non-technical audiences.

Sep/18 - Dic/19

Data Science Consultant
| PHILIP MORRIS INTERNATIONAL MEXICO | Mexico City
Boosted data science solutions within the organization. Promoted data good practices in the daily way of work of the team. Guided technical requirements for the acquisition of data science tools and talent.

Jun/18 - Aug/18

Data Science Fellow
| DSSG AT THE UNIVERSITY OF CHICAGO | Chicago
Collaborated with the Department of Labor in Chile to improve workplace safety. Implemented a proactive approach through data science and predictive tools.

Apr/16 - May/18

Data Scientist
| HEALTH DATA ANALYSIS CENTER - CAD Salud | Mexico City
Implemented and maintained data analysis projects focused on the prevention and improvement of health and social issues in Mexico.

Mar/15 - Jun/18

Data Analytics Consultant
| GEEK END STATISTICAL CONSULTING | Mexico City
Partnered with multidisciplinary teams in the delivery of data solutions.

Jan/11 - Feb/15

Data Analyst Senior
| L15L BUREAU OF STATISTICAL ANALYSIS | Mexico City
Performed data analysis in projects applying statistical and ML models.

Jun/09 - Dic/09

Adhoc Analysis Intern
| AC NIELSEN MEXICO | Mexico City
Aided and enabled the Ad-Hoc team in the development of projects.

education

training

collaborations

Jul 2020   **Estimating COVID-19 cases and reproduction number in Mexico**
M. Anzarut, L.F. González, S. Mendizábal, M.T. Ortiz
Published PDF

lectures

2021 Aug
R software trainer   **R software training course**   UN ECLAC MEXICO
Two month basic and intermediate course on the use of R software for the processing and analysis of statistical data for officials of ECLAC Subregional Headquarters in Mexico, adapted to the needs of the Social Development Unit.

2018
Guest Speaker   **A `bigrquery` example**   ITAM and RLadies CDMX
Real case example using `bigrquery` to compute the births from 2010 to 2016 per mother age.

2018
Guest Speaker   **¿What is Happening with Journalists in Mexico and the World?**
Open Data Day 2018 CDMX
App presentation of analysis and census of journalists killed from January 1992 to may 2017. `Link to App.`

2017
Guest Teacher   **R Introductory Workshop**   Department of Statistics ITAM
Nine hour R introductory workshop addressed to undergraduate students. Material in Spanish published at `Link to Bookdown.`

2016
Guest Teacher   **R Introductory Course**   Department of Statistics ITAM
Nine hour R introductory course addressed to undergraduate students with specific interest in using R for data analysis.

2013
Guest Speaker   **Data Visualization Tools**   Student R User Meeting ITAM
Real case example using functions from R package `ggplot2` for more natural, agile and simpler code to create graphics and enhance data visualization.

2012
Guest Speaker   **Data Analysis using New Tools in R**   First R User Meeting ITAM
Several examples of simpler and easy-to-read code for data manipulation employing functions from R packages `plyr` and `reshape2`.

industry projects

The following list shows a sample of my most-liked projects from 2011 to 2022.

Experimental Design

Audience ratings | Data Scientist Sr
Design, implementation and analysis of an experiment created to assess ratings in open mexican television. **Project report.**

LSH Similarity Join

Movie database | Data Scientist Sr
System build to join movie databases using titles and production year. Several tables including client tables, open movie databases and social networks. **Project report, presentation and interactive tool.**

Recommendation Sys.

Insurance recommendations | Data Scientist Sr
System build to recommend insurance products using different consumption patterns from clients. **Project report, presentation and interactive tool.**

Regression Models

Hierarchical Time Series of TV Viewers | Statistical Analyst
System build to estimate monthly television viewers using daily time series forecast with aggregation constraints adjustments. **Project report, presentation and interactive tool.**

Multivariate Prediction of Hectares Planted | Statistical Analyst
Multivariate dynamic model to predict the number of hectares planted for two substitute seeds using environmental features since 2000. **Project report and presentation.**

Drug Clinical Trial | Analytics Consultant
Bayesian model to assess drug effect significance in a clinical trial for epilepsy. **Project report.**

Approval Prediction using Finger Sensors | Statistical Analyst
Mixed effects model to estimate brand mascot proposals approval using information from finger sensors from children between 3 to 5 years. **Project report.**

Classification Models

Consumer Clusters Prediction | Statistical Analyst
Support vector machine, multinomial regression and decision trees to predict groups form several segmentations of consumers. **Project report and Excel simulator.**

Acquisition Probability Estimation Statistical Analyst
Acquisition probabilities estimation for several proposals using logistic regression. **Project report and presentation.**

Clustering Methods

Segmentation of SMEs | Analytics Consultant
Principal components and hierarchical clustering to create groups of SMEs considering technological advance, technical training and financing. **Project report.**

Behavioral and Demographic Segmentation Statistical Analyst
Behavioral and demographic segmentation of consumers using principal components analysis and hierarchical clustering. **Project report and presentation.**

Bayesian Networks

Consumption Relational Analysis | Statistical Analyst
Bayesian Network to assess the relationship between demographic features, behavioral variables and consumption habits of snacks consumers. **Project report and Hugin simulator.**

research projects

Improving Workplace Safety through Proactive Inspections| Data Science Fellow

Machine learning product using binary classification models to identify facilities at high risk of labor law violations. The problem required several data science tools and field trials.

2018. Private investigation worked jointly between U. Chicago and the Department of Labor in Chile. **Project report submitted.** `Project link.`

Small Steps to Transform the Health System in Mexico | Research Analyst

Analysis of variance of Mexican private health market prices using insurance companies' data and public open data. Presentation of recommendations addressed to government and insurance companies to lessen social impact of catastrophic diseases.

2017. Private investigation worked jointly between CAD Salud and IMCO. *Sponsored by CEEG.* **Project report published.** `Publications link.`

Sexual Violence and Teen Pregnancy in Mexico | Research Analyst

Meta-analysis of sexual violence indicators in teenage girls. Identification and assessment of risk factors in this group.

2017. Private investigation worked jointly between CAD Salud and IPAS Mexico. **Investigation report published.**

Diabetes | Research Consultant

Methodology design and project supervision to attract and retain diabetic patients in treatment using random forests and survival analysis. Identify paths to promote new clinics openings and increase treatment capacity to achieve sickness stabilization in Mexican communities.

2016. Private investigation worked jointly between CAD Salud and CIMAT. *Sponsored by Conacyt and Clínicas del Azúcar.* **Project report submitted.**

New Jobs | Research Analyst

Comparison of the job market in Mexico before and after digital services were introduced using Mexican Social Security Institute (IMSS) data. Estimate economic gains and promote investment for digital platforms.

2016. Private investigation by CAD Salud. *Sponsored by IMSS.* **Article approved for publication.**

Marijuana | Research Analyst

Analysis of public surveys on the perception of Mexicans' about marijuana usage and its impact on society. Comparison of marijuana usage between Mexico and other countries considering violence and mortality. The results will be used to stimulate informed debates about the legalization of marijuana for medical and recreational use.

2016. Public investigation by CAD Salud. *Sponsored by CAD Salud.* **Infographics published.** `Infography 1 link. Infography 2 link.`

Food Diet Patterns | Research Analyst

Model to identify food consumption patterns and association patterns using daily ingest diaries from a probabilistic sample of the Mexican urban population. Identify habits and behavior to aid campaigns that inform and educate people to reduce obesity prevalence in Mexico.

2014. Private investigation worked jointly between L15L and Mexican Health Institute. *Sponsored by Nestlé.* **Book and videos published inside corporation.**

extras

RStudio::Conf	**International Diversity Scholarship Recipient**	2020
Escuela de Datos AME	**Volunteer**	2019
RLadies CDMX	**Volunteer**	2019