

DIEGO SÁNCHEZ ROJAS

DATA SCIENTIST

 **diegosreco@gmail.com**
 **5561299166**
 **<https://www.linkedin.com/in/diegorojas12>**
 **Aguascalientes, Aguascalientes**

EDUCATION

M.A. Methods for the Analysis of Public Policies (METPOL)

Center for Research and Teaching in Economics (CIDE)

may 2019- september 2021
Thesis: "Short-run Effects of COVID-19 Pandemic on the Labor Market Dynamics in Mexico".

E-poster in annual conference Society for Medical Decision Making (SMDM) 2021.

B.A. in Economics

Metropolitan Autonomous University (UAM)
august 2014- december 2018

Thesis: "Wage gaps between formal and informal workers"

MAIN SKILLS

- **Data Extraction, Analysis, and Visualization**
- **Applied Statistics and Econometrics (forecasting, causal inference, and impact evaluation)**
- **Decision Analysis (deterministic and probabilistic cost-effective and cost-benefit analysis, microsimulation models)**
- **Machine learning (forecasting, regression classification, and NLP)**
- **Design, Implementation, and Analysis of Socioeconomic Experiments**

SOFTWARE AND PROGRAMMING

- **R (tidyverse-tidymodels packages, plotly, leaflet , sf, NLP , bst, survival, tseries,ranger, forecast, msm, lm4, svy, inference)**
- **Python (numpy, pandas, polars, statsmodels, scikit-learn, tensorflow, beautifulsoup, nltk, leaflet, seaborn, pyodbc, pyspark)**
- **SQL (SQL Server, PostgreSQL)**
- **Snowflake**
- **SAS**
- **Git**
- **Tableau**
- **Office**
- **Stata**

WORK EXPERIENCE

DATA ANALYTICS COORDINATOR

Confie Insurance - Freeway Insurance | since march 2023

Coordination of data analytics and data science projects. Develop, automate, optimize, and deploy ETL processes, inference frameworks, and machine learning models. **Achievements:** Development of the ETL process and classification of deposits for the different carriers with high accuracy. Forecast of the commissions using econometric models of time series and machine learning models. Automation of the reconciliation between deposits and statements for all Carriers, in addition to the BI tools for visualization.

Tools: SQL Server, Python, R, Tableau, and Snowflake

TAX RESEARCH AND INFORMATION ANALYSIS ASSISTANT MANAGER

Tax Administration Service (SAT) | may 2022 - march 2023 | CDMX

Develop of ETL process and analysis of big data. Creation, implementation, and evaluation of time series and machine learning models for forecasting. Evaluation of tax policies through microsimulation models and impact evaluation techniques.

Achievements: Using a microsimulation analysis, I evaluated the effect of the modification to the tax on the production of alcoholic beverages and beer. I built and implement a nowcasting model using machine learning and econometrics methods. **Tools:** SAS (Enterprise Guide), Python, R, Teableau, and Azure Databricks (PySpark and SQL).

ANALYTICS EXECUTIVE

Porter Novelli Microsoft Latam account | september 2021- may 2022 | Remote

Data extraction and collection using web scraping techniques and ETL process. Application of natural language processing (NLP) techniques and machine learning models for text classification with the aim of identifying insights, and trends, and creating KPIs to measure campaign impact. **Achievements:** I systematized the ETL process for data. I built and implement a machine learning model for automatic text classification. **Tools:** Python, R, and Power BI.

JR. CONSULTANT (PROJECT)

Digital Government Coordination of the Aguascalientes City Hall | june 2021- august 2021 | Aguascalientes

Data analysis and evaluation of the program that collects citizen reports (Program 072). **Achievements:** Classification of reports, georeferencing, demand forecasting for capacity and service time, and construction of benchmarks. **Tools:** Python and R.

JR. RESEARCHER (PROJECT)

Anti-corruption System of Aguascalientes State | march 2020-may 2020 | Aguascalientes

Design and analysis of an experimental survey where, through a treatment vs. a control, we measure the performance of social communication campaigns that encourage citizen participation. **Tools:** SurveyMonkey and

RESEARCH ASSISTANT

Metropolitan Autonomous University (UAM)-Azcapotzalco | january 2018- april 2019 | CDMX

Data analysis, implementation of microeconomic models (casual inference), impact evaluation models, time series models and teaching of practical labs. **Tools:** Stata and R.

COURSES AND CERTIFICATIONS

- **Data Scientist Associate Certificate | DataCamp 2024-2026:**
<https://www.datacamp.com/certificate/DSA0017593368246>
- **Machine Learning Specialization | Stanford University and DeepLearning.AI in Coursera | 2023:**
<https://www.coursera.org/account/accomplishments/specialization/SKL37VU99VJ9>
- **Career Track Data Scientist with Python | Datacamp | 2021**
- **Multivariate Time Series in R | National Laboratory of Public Policies (LNPP) | 2021**
- **Bayesian Statistics: From Concept to Data Analysis | the University of Santa Cruz California (remote) 2021**
- **Career Track Data Scientist with Python | Datacamp | 2021**
- **Career Track Data Scientist with R | Datacamp | 2022**
- **Microeconomic Modeling with Cross-sectional and Panel Data in Stata | UAM | 2018**

LANGUAGES

- **Spanish: Native**
- **English: Intermediate**
- **Italian: Basic**