
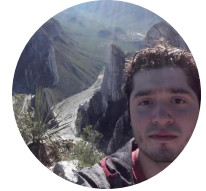


# Marco DE LOS SANTOS H

## Data Scientist | Physicist

in [linkedin.com/in/mdelossantosh](https://www.linkedin.com/in/mdelossantosh)  [github.com/positronn](https://github.com/positronn)  
☎ +52 811 852 7920 @ [marcodelossantos@protonmail.com](mailto:marcodelossantos@protonmail.com)  
📍 Monterrey, N.L., Mexico  
📅 Place and Date of Birth: Mexico, February 1994



Data Scientist with 5+ years of experience in the fields of Data Management and Data Analysis. Open to learn new things and ideas. Ability to analyze systems and data, teach, solve problems and work in team. I'm interested in Physical and Computational Sciences, Statistics, Software Development, Artificial Intelligence (Machine Learning and Deep Learning), Data Management & Integration, Data Governance, Analytics and Business Intelligence and Mathematical Modeling. Commitment with professional ethics, responsible, patient, attentive (helpful) and moral conscious. Communication skills, organized and precise, I also have a strong aptitude towards self taught and research-related work.

Currently I develop Analytics, Data Management and Integration at Tecmilenio, building the foundations for a Analytics Platform that will gather, process, model and serve data from various modules of a LXP (Learning Platform).

## SKILLS AND TOOLS

<b>Skills</b>	<b>Data Science/Analysis, Data Visualization, Process Automation, Data Wrangling REST API, Deep Learning, Machine Learning, ETL-ELT, Database Development</b>
<b>Programming</b>	<b>Python 3, R, SQL, bash</b>
<b>Python Frameworks</b>	pandas, scipy, sk-learn, numpy, keras, matplotlib, seaborn, Flask
<b>R Frameworks</b>	baseR, tidyverse, tidymodels, renv, lubridate
<b>Data Bases</b>	MySQL, PostgreSQL, MongoDB, dbt
<b>Software</b>	Anaconda, Tableau, Power BI, KNIME, MS Excel, MS SharePoint, git, draw.io
<b>Cloud</b>	Azure ML Studio, Azure ML Services, Snowflake, Fivetran
<b>Operating Systems</b>	Mac OS X, Windows Server, Linux

## PROFESSIONAL EXPERIENCE

<b>July 2021 Present</b>	<b>Data Scientist   Digital Transformation Department, TECMILENIO, Mexico</b> <ul style="list-style-type: none"><li>➤ Technical lead in the implementation of a Modern Data Stack: Fivetran (EL/Replicator) + Snowflake (DWH) + dbt (Transformation/Modeling) + Tableau (Viz)</li><li>➤ Sentiment Analysis of Messages from Community Platforms with the VADER model</li><li>➤ User Activity profiling using Feature Engineering and K-Means</li><li>➤ Text Mining and Analysis of Soft and Hard Skills demand for labor vacants in the Information Technology sector</li><li>➤ Web scraping and Analysis of Digital Courses Pricing in relation to their attributes, such as Length, Popularity, Rating, Discipline, etc., from different Learning Platforms such as Platzi, EdX, etc.</li><li>➤ Analysis of user conduct and interaction in chat platforms (Discord)</li><li>➤ Data Modeling and Pipeline Design and Maintenance in Snowflake + dbt</li><li>➤ Design of Metrics and KPIs together with Business Professionals according to Business Objectives and Processes.</li><li>➤ Dashboard design in Tableau</li><li>➤ Data cataloging and documentation</li><li>➤ Support in the design of Data Products for e-learning applications</li><li>➤ I also belong to the Data Governance Team, working with the compliance of Data Regulation, Best Practices and mapping the Data Journey of users in our platform.</li></ul> <div><span>Python 3</span> <span>R</span> <span>SQL</span> <span>Snowflake</span> <span>dbt</span> <span>Tableau</span> <span>Fivetran</span> <span>PostgreSQL</span> <span>Azure ML Studio</span> <span>Open Web Scraper</span></div>
<b>October 2020 June 2021</b>	<b>Data Scientist   Data Science Senior Developer, KALYPSO, Mexico</b> <ul style="list-style-type: none"><li>➤ Data Pipeline Engineering</li><li>➤ Data Analysis</li><li>➤ Development of predictive models using ML</li><li>➤ Linear and Non-Linear optimization of ML Models for prescriptive purposes</li></ul> <div><span>Python 3</span> <span>pandas</span> <span>sk-learn</span> <span>scipy</span> <span>seaborn</span> <span>PostgreSQL</span> <span>Power BI</span></div>

August 2019  
October 2020

#### Data Scientist | HR Reporting & Analytics - Strategic Workforce Planning, PEPSICO, Mexico

- Turnover/Churn Analysis for NFL employees
- Made some python utilities to work with NLP using LDA, Sentiment Analysis and Translators with labor/environment surveys
- Data Analysis for Insight finding with Human Capital Information
- Experiment Design and Hypotheses Testing
- Collaboration with Global teams to implement new technologies and data science best practices
- Implementation and Coordination of a HR Function in a GBS (Shared Services) framework for doing analytics with Human Capital Information provided to countries in LATAM.
- Key Contributor in the stabilization of SLAs (Delivery Time) from 60 to 99.9% in 3 months.
- Automation of 80% of the operation reports with the creation and maintenance of a data pipeline made with Python and Pandas.
- Coordination of creation and maintenance of Executive Dashboards in Tableau showing Human Capital Information such as People Cost, Diversity, Equity, Turnover Analysis for Talent Acquisition, etc.
- Validation of Information Quality in development databases with R scripts.

Python 3 pandas R tidyverse Teradata Tableau Power BI MS SharePoint Azure Web Services

June 2018  
August 2019

#### Data Scientist | Digital Transformation and Innovation Department, AXTEL, Mexico

Design and implementation of prototypes using emergent technologies, specially automation and analytics in order to help and improve various tasks and assignments of the different departments in the company.

- Developed a Back-End project that collected information from several routers of the different carriers of the company via ssh and consolidated this information in a mySQL database to then expose its analytics in a website via an API made with Python Flask.
- Developed a detector and classifier of manufacturing pieces with computer vision (opencv) and deep learning open-source libraries (tensorflow/keras) within a co-creation framework. In this project I used several CNN architectures together with other cv techniques such as Template Matching, Thresholding, Morphological Operations, etc., and embedded systems, Raspberry pi and Arduino.
- Mentored other department teams in their study of python and data analysis.

Python 3 Pandas sk-learn scipy seaborn mySQL OpenCV Keras RaspberryPi Flask SQL

August 2017  
March 2018

#### BI Analyst Jr., AXITY, Mexico

- Support in Oracle and SAP tools such as Modeling in SAP Business Objects (Information Design Tool, WebIntelligence).
- Data flow analysis and quality assurance between Databases (ETL) in Oracle Data Integrator.
- Data Analysis and manipulation in Oracle Essbase (EAS).
- Financial data analysis and validation.

Oracle SQL Oracle Data Integrator SAP Business Objects

## LANGUAGES

Spanish ● ● ● ● ●  
English ● ● ● ● ○  
German ● ● ● ○ ○  
Portuguese ● ○ ○ ○ ○

## APTITUDES

- Passionate about Science
- Self Taught
- Autonomous
- Concise and Precise
- Analytical

## ACADEMIC FORMATION

2017 B.Sc. in **Physics**, Computational Physics Orientation, **Universidad Autonoma de Nuevo Leon**  
Undergraduate Project in Molecular Dynamics | Advisor: Sergio Mejía  
Phase transition study of Mono-Metallic bulks via Molecular Dynamics

## PROJECTS

### SIMPLECONVTRAINER

2019

 [github.com/positronn/SimpleConvTrainer](https://github.com/positronn/SimpleConvTrainer)

A simple CNN trainer optimized for vision purposes: fast for prototyping in a simple way. What I use this program/utills for is to explore the hyperparameter tuning space in a broad way, then choose some prospective model and continue the hyperparameter adjustment in a more detailed manner.

OpenCV Keras scipy matplotlib