

Carlos Miguel De Leon

(650) 438 3716 - cmdeleon1@gmail.com - www.linkedin.com/in/cmmeddata - cmmeddata.github.io - South San Francisco, CA

Education

San Francisco State University | B.S. Computer Science

12/2025

- Co - founded and Vice President of the Data Science Society at SFSU
- Genentech - PINC Scholar
- Certificate in Data Science and Machine Learning in Biotech

Experience

Software Engineer for Data Platforms Intern

5/2024 - 08/2024

MethaneSAT & Environmental Defense Fund | Remote

- Developed a Python-based script that streamlined satellite data acquisition, resulting in significant improvements in operational efficiency and accuracy.
- Automated satellite information detailing, eliminating manual processes and ensuring real-time data availability.
- Predicted downlink KSAT filenames and locations, enhancing satellite data management and decision-making processes.

Data Analyst Intern

02/2024 - 08/2024

GLIDE San Francisco | San Francisco, CA

- Conducted data cleaning and preprocessing on hundreds of survey datasets to ensure accuracy and consistency.
- Performed exploratory data analysis (EDA) on categorical data from residents in the Tenderloin district of San Francisco.
- Utilized insights from EDA to identify trends and inform data-driven decision-making.

Data Science Mentor

08/2023 - 05/2024

TheDataJam | Pittsburgh, PA

- Guided a group of high school and community college students through a comprehensive semester-long curriculum covering data analysis, Programming in R and Python, advanced excel techniques and introductory machine learning concepts.
- Provide hands-on experience by assigning real-world data analysis projects, enabling students to apply their skills in practical scenarios.

Projects

Measuring the Economic Empowerment of Women

- Examined global gender inequality using indicators such as a country's Gender Inequality Index (GII), Education, Age, Labor force Statistics and Parliament status.
- Employed Python pandas for data cleaning, addressed missing values with sci-kit learn, K-Nearest Neighbors (KNN) and employed multiple linear regression using statsmodel for data driven insights.
- Presented clear visual insights, facilitating a comprehensive understanding of women's economic empowerment challenges globally and translated complex statistical findings into impactful visualizations using Tableau.

Data Exploration: Car Break-ins in San Francisco

- Gathered, cleaned, and analyzed a large dataset of over 700k reported crimes in San Francisco, ensuring data accuracy with Python(pandas)
- Created informative data visualizations using Matplotlib and Seaborn, simplifying complex data for easy comprehension
- Analyzed crime data trends, identified actionable insights and effectively communicated findings to support data-driven decision making

Assessing COVID-19 in San Francisco: A closer look at Neighborhoods and its impact on COVID-19 Cases

- Conducted comprehensive analysis of San Francisco neighborhood's impact on COVID-19 using excel.
- Analyzed independent variables such as poverty rate, geographic data, population and ethnic demographics
- Utilized data visualization and regression analysis to gain insight on local pandemic dynamics, contributing to COVID-19 understanding and showcasing excel data analysis skills.

Skills

- **Programming:** Python (NumPy, Pandas, Sci-kit Learn, Statsmodel), R, SQL (MySQL)
- **Visualization:** Python (Matplotlib, Seaborn), R, Tableau