

Alonso J. Gutierrez Moran - Data Engineer

+1 954 513 9734 | alonsojgutierrez@outlook.com | alonsojg.github.io

Work Experience

Data Engineer | Capgemini America | New York, U.S. | (Remote) | Mar 2022 - Present

- Designed and developed ETL pipelines to extract, transform, and load data from various sources into a centralized data warehouse.
- Utilized SQL and Python for data manipulation and cleaning tasks.
- Implemented data pipelines using Hadoop and Spark for distributed processing and analysis.
- Collaborated with cross-functional teams to understand data requirements and provide solutions for data-related issues.
- Improved data processing efficiency by optimizing pipelines and implementing data partitioning strategies.

Data Engineer/Business Analyst | SEI Investments Company | Contract | Pennsylvania U.S. (Remote) | Sep, 2022 - Present

- Collected, organized, and analyzed large sets of data to inform business decisions
- Used Excel and SQL to manipulate and extract data
- Presented findings through clear and concise reports and visualizations
- Identified opportunities for process improvement and collaborated with cross-functional teams to implement solutions

Data Analyst | Freelance | Remote | Jun, 2018 - Mar 2018 - Mar 2022

- Collected, organized, and analyzed large sets of data to inform business decisions
- Conducted statistical analysis and data modeling to inform business decisions
- Created dashboards and reports using Tableau to track key performance metrics
- Improved data quality and accuracy through regular audits and maintenance of data pipelines

Project Experience

Processing Enron Debacle Data (ML, NLP, Python, Pandas, Numpy):

- Enhanced search ML classifier algorithms for machine learning, along with the primary component analysis, stratified data shuffling, grid-searching and score-based model-selection.

Gene Expression Visualization (Python, Pandas, Numpy, d3.js/dimple.js, HTML):

- Statistical testing optimization using Python modules to quickly sort through a large set of gene data and find most influential genes used to set apart two very similar types of Leukemia.

Education

B.S. Degree in Mathematics | University of Central Florida | 2018 - 2021

Data Analyst Nanodegree | Udacity

Data Engineering and Machine Learning using SparkData | IBM

Distributed Computing with Spark SQL | University of California, Davis

[Others...](#)

Skills: SQL, Python, Hadoop, Spark, ETL, GCP, Excel, Tableau, Power BI, Data visualization, Statistical analysis, Business intelligence, Machine learning, Strong communication and collaboration skills.