Roger Qiu

Dallas, TX | rogerqiu7@gmail.com | 832-349-0490

LinkedIn | Portfolio

EDUCATION

UNIVERSITY OF SAN DIEGO | M.S. APPLIED DATA SCIENCE   09/2022 - 08/2024

Coursework: Statistics, Time Series Analysis, Applied Machine Learning, Natural Language Processing

SAINT LOUIS UNIVERSITY | B.S. HEALTHCARE ADMINISTRATION, INFORMATION TECHNOLOGY MANAGEMENT   08/2015 - 05/2019

Coursework: Health Operations Management, Enterprise Applications Development, Systems Analysis and Design

EXPERIENCE

DATA ENGINEER   Boston, MA

Cohere Health   02/2023 - Present

- Developed and optimized dozens of Python and Spark ETL pipeline jobs, reducing job runtime in Airflow by up to 90% by redesigning and creating more efficient code
- Automated in-depth exploratory data analysis (EDA) and preprocessing of millions of records daily within minutes by creating dozens of packaged functions, stored in our GitHub

FINANCIAL ANALYST   Houston, TX

National Oilwell Varco   10/2022 - 02/2023

- Leveraged Alteryx and SQL to create dozens of automated data pipelines and workflows, saving hundreds of hours of manual ETL processes
- Increased management decision making capabilities by developing and presenting Power BI dashboards to report financial data efficiently using visualizations and dashboard summaries

FINANCIAL ANALYST   Itasca, IL

WernerCo   10/2021 - 10/2022

- Modeled and presented monthly financial data from company income statements using IBM PA (Cognos TM1), Excel, and Power BI, guiding departments to millions in savings of quarterly budget

FINANCIAL ADVISOR   Chesterfield, MO

Merrill Lynch   08/2019 - 10/2021

- Advised clients on numerous financial products including equities, fixed income, alternative investments, life insurance, annuities. Performed analytics on investment options, managing portfolios of millions of dollars.

PROJECTS

Car Price and Trim Prediction

- Utilized preprocessing methods such as normalization, one-hot encoding, TF-IDF vectorization (NLP), and Principal Component Analysis (PCA) to handle high-dimensional data. The best models, Logistic Regression for trim and Linear (Ridge) Regression for price, were optimized using GridSearchCV.

Patient Management System

- Patient database system created using Java for data storage and Java Swing for the graphical user interface.

SKILLS

Python, Apache Spark, Apache Airflow, AWS Athena (SQL), Github, AWS Sagemaker, Power BI, AWS EMR, Java, API, Supervised Machine Learning, SKlearn, Natural Language Processing (NLTK), Dimensionality Reduction (PCA)