Danlei Qian

Arlington, VA | 202.569.0819 | danleiqdata@gmail.com | [Linkedin](#) | [Portfolio](#)

SKILL HIGHLIGHTS

Programming & Scripting Languages: Python, R, SQL, HTML, CSS, JavaScript (D3.js), Scala
Tools and Applications: Tableau, Power BI, Google Charts, Dash Plotly, Advanced Excel, AWS, Git
SQL & NoSQL Database: MySQL, SQLite, DynamoDB, MongoDB, Neo4j
Big Data Ecosystem: HDFS, YARN, Spark, Sqoop, Kafka, Hive, Impala, KUDU, Hbase

EDUCATION

The George Washington University, Washington, DC   December 2020
***Master of Data Science (*GPA: 4.0/4.0)**
Coursework: Machine Learning, NLP, Time Series Analysis, Data Visualization, Data Warehousing, Cloud Computing, A/B Testing

Nanjing University of Information Science & Technology, Nanjing, China   June 2016
***Bachelor of Accounting (*GPA: 3.6/4.0)**
Awards: Presidential Scholarship (2016); Top3 of 100 in IMA Student Case Competition: Lean Supply Chain Management (2016)

WORKING EXPERIENCE

ShineWing International   Nanjing, China
Associate Consultant   *October 2015 – January 2018*
- Performed data-driven internal and external audit for over ten companies; extracted data from clients' databases using SQL and leveraged analytical and visualization techniques to effectively isolate insights for assessing clients' internal control, operating performance and financial statements using Excel routinely.
- Leveraged data wrangling and profiling using Excel pivot tables and set up visualizations using internal BI platform to evaluate credit risk management for commercial banks; identified non-performing loan of over 100 million dollars.
- Conducted ratio and trend analysis on multiple financial accounts such as inventory, sales, payroll cost and purchasing cost to identify fraud risk of clients' financial statements.
- Built regression models to predict the monthly sales with multiple factors such as market size for an e-commerce company; performed a comparison with actual values to identify unusual fluctuations.
- Developed automated Excel workflow with VBA to evaluate construction cost for over 1000 telecom towers for China Mobile in collaboration with construction consulting teams, resulting an increase of 50% in working efficiency.
- Summarized analysis reports to provide timely advice and collaborated with client personnel to make adjustments.

Suning Commerce Group Co., Ltd.   Nanjing, China
Intern in Financial Data Center   *July 2015—September 2015*
- Implemented ETL (extract, transform, load) process through SAP system; performed data cleaning and integration into required formats for financial analysis using Excel.
- Assisted IT support engineers with the maintenance of SAP system by collecting requests from accountants.
- Examined the integrality and authenticity of invoices; coordinated issues with vendors and business departments.

TECHNICAL PROJECTS

Mechanisms of Action Multi-label Classification (Real Kaggle Competition) | *Python*
- Developed a classification algorithm to classify drugs with multiple labels with tabular data.
- Selected features using T-test and Permutation Importance; conducted feature engineering with PCA and K-means.
- Ensembled three designed DNN models based on transfer learning, TabNet and ResNet, using Pytorch and TensorFlow; optimized hyperparameters with Optuna on AWS.
- Won a Silver Medal with Top 2% rank (64th of 4000+ teams).

[**Global Energy Visualization Web**](#) **|** *Tableau, Google Charts, D3.js, HTML, CSS*
- Developed an interactive web based on HTML & CSS to visualize the global energy data of 50 years.
- Conducted data profiling and transformed data into required format like json by Python.
- Leveraged D3.js to build a force-directed graph to display the network of energy, and build bar-chart race, bubble animations to dynamically display the data change over time.
- Built Tableau Dashboard with an interactive map syncing with bar charts and embed pie charts into tooltips; created analytical charts using Google Charts API for html embedding.

Rossmann Store Sales Timeseries Analysis | *Python*
- Performed Time Series Analysis on sales data including stationarity, seasonal decomposition. autocorrelation.
- Predicted the future sales with Holt-Winters, Multilinear Regression, ARMA models; identify the optimal order (p,q) by self-developed generalized partial autocorrelation (GPAC) table and Auto ARIMA function from Pmdarima library; achieved the accuracy of over 97%.

Home Credit Default Risk Analysis | *Python*
- Predicted customers' repayment ability using KNN, MLP, Decision Tree, Random Forest, XGBoosting and Logistic Regression and used Grid Search for hyperparameter tuning; achieved the highest AUC score of 0.7.
- Preprocessed credit data with one-hot encoding and oversampling.