ERICA(XUAN) ZHOU

Data Analyst/Engineer

MD / DC / North VA • (240)308-3976 • Ericazhou900506@gmail.com

KEY CREDENTIALS

Skills: Data Analysis, SQL/T-SQL, Data Visualization, Data Mining, Machine Learning, NLP, Front-end Development
Technical Tools: MS SQL Server, Tableau, Python, HTML, CSS, JavaScript, VS code, Microsoft Office Suite

WORK EXPERIENCES

SQL Data Analyst, Sita Health & Life Sciences May 2020 – Current

- Wrote queries in relational databases, tables with joins, indexes, user-defined functions (UDFs), views, stored procedures and triggers using T-SQL.
- Worked with integration services (ETL) for transferring and reviewing data from heterogeneous sources (Excel, CSV, flat files, JSON); performed data normalization and integrity validation.
- Performed data analysis and calculations using SQL queries for ad-hoc analysis and reporting.
- Extensive experience in Dimensional Modelling, ER Modelling, Star Schema/Snowflake Schema, Fact and Dimension tables, and Operational Data Store (ODS).

Cost Data Analyst, China Resources Land Co., Ltd Apr 2011 – May 2016

- Involved in building and maintaining relational database (supplier, customer, inventory etc.) using SQL with migrating information from all historical data in various file formats and sources.
- Performed cost data analysis to produce total cost projections based on historical data and market trends using regression models; compiled analytical and predictive reports to support project life cycle.
- Performed data analysis and built predictive models to produce annual reports for projects budget planning.

PROJECTS

Relational Database Design for E-ZPass® Toll Management System Dec 2018

- Designed an E-ZPass® toll management system database with a set of features for user account management, auto toll billing, payment and statistical analysis.
- Created data models with Lucidchart; built SQL procedures and functions to enable user account management, vehicle and transponder registration, auto toll billing, auto payment, toll computing, trips and bills look up.

Relational Database Design for Car Rental Operations May 2018

- Independently designed a database system for car rental business using SQL server; acquired business logic from online resources; identified and defined main functions and created data models using Lucidchart to enable future operations.
- Created tables and integrated data using SQL for order management processes; performed analysis using temp tables, CTE, and built-in functions to identify most profitable car models.

NLP Project for Hate Speech and Offensive Language Detection on Tweets Dec 2019

- Produced a classifier using Python (Jupyter Notebook) that classifies tweets as hate speech, offensive, or neither by comparing a plethora of models and algorithms.
- Conducted exploratory analysis and optimized data scale through visualization of cleaned data and elimination of redundant features.
- Models used include K-Nearest Neighbours (KNN), Random Forest, Support Vector Machine (SVM), Stochastic Gradient Descent Classifier with Linear SVM, and Recurrent Long Short-Term Memory Neural Network.
- The best performing algorithm was Stochastic Gradient Descent Linear SVC with best F1 score (92%) and micro average (93%), scalability, and speed.

EDUCATION

University of Maryland, Baltimore County Baltimore, Maryland May 2020
Master of Science in Information Systems GPA 3.8/4.0

REFERENCES UPON REQUEST