**Joye Lee (Data analyst/data engineer)**

**PROFILE SUMMARY**

* Accomplished Data Analyst/ Marketing Analyst with 5+ year experience offering solid technical background in relational and NoSQL databases, Business Intelligence, Python Data Analytics, Big Data Analytics, and Cloud Computing.
* Partner with stakeholders to identified business requirements from internal & external sources.
* Acquire and merge legacy data which resides in MS Access, MS Excel, relational and NoSQL databases, and Cloud Platform databases.
* Understand the legacy data, target data model, data cleansing strategies & approaches and create data mapping.
* Define and create database objects, tables, stored procedures, views, triggers and SSIS packages.
* Experience in maintaining and improving database integrity and performance.
* Design and develop ETL (Extract, Transform, and Load) processes that transform a variety of raw data, flat files into SQL database.
* Analyze problems by executing large and complex SQL query, applying analytical methods (segmentation, forecasting etc.) as needed.
* ETL data into data warehouses such as Microsoft Azure, Amazon S3, Teradata and Oracle
* Develop data exploratory analysis (EDA) using Python (NumPy, Pandas, SciPy)
* Generate data-driven routine report and ad hoc report with business insights by SSAS & SSRS in MS BIDS, and also using other data visualization tools such as Tableau, Power BI, Google Data Studio and Python (Matplotlib, Seaborn).
* Provide analytical support and monitor key performance indicators, highlighting trends and analyzing causes of changes.
* Define business metrics and perform statistical analytics such as A/B Testing, Hypothesis test, t-test.
* Extract website data using web crawling with API in Python (Requests, Beautiful Soup, Selenium).
* Experience in Cloud Platforms such as AWS (Lambda, EC2, RDS, DynamoDB), Microsoft Azure, Google Cloud.
* Experience in big data tools such as Hadoop (HDFS, Hive, Pig, MapReduce) and Spark (PySpark, MLlib).
* Interact cross-functionally with a wide set of stakeholders and teams at varied levels of the organization to enact data maintenance process.
* Self-motivated individual with strong communication skills, great attention to details and the ability to multitask, set priorities and work efficiently in a high-paced environment with competing demands.
* Excellent team player with demonstrated experience of working diplomatically and collaboratively with all levels of staff to ensure successful implementation of assignments.

**Technical Skills**

**Languages:** Python, R, SAS

**Relational Database:** MySQL, MS SQL Server, Teradata, HiveQL, Spark

**NoSQL Database:** MongoDB, DynamoDB, Amazon Redshift

**Business Intelligence:** Tableau, Power BI, Google Data Studio

**Big Data:** Hadoop, Hive, Spark

**Version Control System:** GitHub

**Models:** Linear Regression, Logistic Regression, Classification, SVM, KNN, Clustering, Time Series, NLP, A/B Testing, Hypothesis Testing, ANOVA

**Project Experience**

**Data Analyst / Khol’s, Milpitas, CA** Jan 2020 – Till now

Project Description: Created user behavior analysis for E-commerce transactions

* Applied **Python (Pandas, NumPy and Sci-kit learn)** to load, explore, modify, cleaning, and extract the valuable data (one hot transaction)
* Merged and loaded daily fetched data to **AWS S3** bucket using **AWS CLI**
* Imported data from AWS S3 to **AWS RDS MySQL** automatically with **AWS Lambda**
* Connected **Tableau** and AWS RDS MySQL and plotted data in each segment to understand trends and relationships (line plot, bar chart, pie chart)
* Extracted Key Performance Indicators from user behavior features in multiple dimensions by using **SQL/PL SQL** and Python
* Visualized statistical result of KPIs in **Python (Matplotlib, Seaborn, Plotly)** to present trends and analyze changes
* Classified user according to the **RFM** (Recency, Frequency, Monetary) model, and designed marketing strategy for each category
* Developed **Sentiment Analysis** for reviews to evaluate the performance of each product using **Python (requests, Beautiful soup)**
* Built collaborate filtering recommendation model using **Spark (PySpark, MLlib)** with ALS (Alternative least square) method
* Connect Tableau with Python (TabPy) and MS SQL Server to build interactive dashboards to view online sales performance in real-time

**Data Analyst / Citi Bank, San Francisco, CA** Aug 2017 - Dec 2019

Project Description: Built a prediction model for the Credit Cards department of the bank to effectively detect credit card client default payment.

* Queried credit card user data from **MongoDB** using **R**
* Transformed **Jason** format data to data frame, merged and integrated the data set
* Used **R** to explore and clean the data, worked with Credit Cards department to deal with inconsistent format, mislabeled data, missing values, outliers
* Graphed views in **Power BI** to better understand data in multiple dimensions by creating bar chart, line graphs, cross table graphs and correlation matrix heatmap
* Utilized **Feature engineering** and **Feature selection** approaches using **R (DataExplorer, SuperML),** to create features and choose most important ones from them
* Performed train and test sets splitting and used to train **Machine Learning Logistic, Decision Tree and Random Forest Regression** models
* Optimized parameters by applying **Grid Searching Hyperparameter tuning** method
* Evaluated model performances through a variety of metricsto compare **ROC-AUC, Precision and Recall** of each regression classifiers with **R (Rpart, caret, mlr)**
* Generated analytical report of the prediction performance, monitoring key performance indicators, highlighting trends and analyzing causes of changes

**Marketing Analysis Manager** / **Renhe Scientific Instrument, Shanghai, China** Jul 2014 - Jun 2017

* Managed a 3-person marketing team for a leading chemical instrument agent company
* Cooperated successfully with over 20 world-famous foreign instrument brands in the Chinese market
* Successfully designed **CRM database** for the company, **migrated** and **integrated** the data collected from various sales channels, to improve database management efficiency
* UtilizedCRM to continuously update, monitor, and analyze inventory, orders and customers with the consolidated data from all the sales channels
* Provided creative solutions to analyze marketing campaigns and business performance
* Evaluated sales performance of each channel in multiple dimensions using **SSAS**
* Created customized report using **SSRS** to present overall assessment
* Partnered with sales and product departments to define and generate a new **pricing system** for each distribution channel based on analytical reports and data-driven insights on pricing
* Performed monthly, quarterly and yearly analysis report with the data from CRM system, to present the sales performance, trending and time series to CEO, sales and product departments
* Observed perspectives on the market, achieved 10% average revenue growth per year

**Education BackGROUND**

**California State University, East Bay, USA**

Master of Science in Business Analytics

**Central China Normal University, China**

Bachelor of Science in Chemistry Education