

# Zhangtuoming Zhao

(408)-335-7141 | chriszhao9399@gmail.com

## SUMMARY

- Over **7-years'** experience in **data engineering, data analysis, data processing, data modeling, data visualization** and web scraping; applied knowledge to identify, determine, and solve business problems
- **Extracted, cleaned, and transformed** from various datasets to **GCP and AWS**
- Leveraging **Apache Airflow**, and writing **Airflow DAGs** in
- Working experience in **Data Warehousing** concepts and **Extract Transform and Load (ETL)** processes on **AWS, Azure** and **Google Cloud Platform**
- Developed automated data pipelines with **Apache Spark and Airflow** on **GCP and AWS**
- Developed and maintained **Airflow DAGs** to automated ETL pipelines for various purposes
- Strongly experienced in writing advanced **SQL** queries in (**PostgreSQL, Hive, MySql**) with procedures
- Experienced in managing large-scale database systems **NoSQL (MongoDB)** systems
- Expertise in multiple programming languages, and data analytics & visualization tools including **Python, SQL**
- Experience in **Python**, used **NumPy** and **pandas** for statistical analysis and **machine learning methods**
- Use **statistical** and **machine learning techniques** to create scalable analytics solutions
- Strong understanding of **business process flows**, Business Process Reengineering (BPR)
- Develop a deep understanding of a complex technical environment through data and reporting using various data sources and tools
- Experienced in enormous **data cleansing** to dealing with missing values & duplicate rows and performed **Exploratory Data Analysis (EDA)** in the integrated data.
- Performed **sentiment analysis** with NLP to enhance KYC features to improve provided services.
- Performed customer segmentation analysis for categorizing target customer to provide marketing insights
- Created, developed and ensure high-quality easy-access appearance and accuracy of monthly, quarterly reports templates for decision making teams; delivered data-driven presentation
- Created **interactive dashboards** to deliver easily interpreted insights for business performance evaluation helped to improve the business performance
- Conducted **A/B testing** on data and defined metrics for analyzing incremental impact of product features; evaluated whether to launch new features to guide decision making
- Working experience in **R**, managed and analyzed both structured and unstructured data with scientific computing and analysis packages
- Experience in database schemas designed, data transformation from raw data to designed database schemas and data cleansing procedures to extract data
- Generated dashboards of weekly and monthly reports; Also, Generated **story-telling** quarterly business performance interactive dashboards for quarterly summary meeting
- Executed advanced **excel** skills (**VLOOKUP & pivot** table) to visualized and categorized customized reports to make in key performance indicators and insights for cross-functional business evaluations
- Extensively experienced with **advanced Tableau visualization** using Actions, Calculations, Parameters, Maps, Trend Lines, Statistics, Groups to create **interactive dashboards**
- Worked in agile methodology environment to track user stories
- Full lifecycle experience in Business analysis of the Agile Scrum, object-oriented development systems analysis, data wrangling, Core documentation and Process analysis design, development and testing phases of Software Development Life Cycle

## SKILLS & KNOWLEDGE

<b>Language:</b>	Python (numpy, pandas, seaborn, matplotlib, NLTK, TextBlob), Spark, Apache Airflow
<b>Project methodology:</b>	Agile, Scrum, Waterfall
<b>Business Visualization Tools:</b>	Tableau, Power BI, MS Visio
<b>Database management tools:</b>	MySQL, SQL Server, NoSQL, Hadoop, NoSQL with MangoDB

<b>Cloud Platform:</b>	<b>AWS, Azure, GCP, Snowflake</b>
<b>Operating System:</b>	Windows 10, macOS, Linux
<b>UML Diagram:</b>	Use Case Diagrams, Activity Diagrams, Process flow Diagrams, Data Flow Diagrams
<b>Machine Learning:</b>	Classic & penalized Regression Method (Lasso, Ridge), Decision Tree, Clustering, SVM

## MOST RECENT EXPERIENCE

---



Dallas, TX

### Data Engineer

Aug 2021-Current

#### *Description:*

Detail-oriented and experienced data engineer seeking a challenging role at a ride-sharing company, focusing on customer support- related data to drive insights and improve user experiences and customer satisfaction. Bringing expertise in data pipeline development, ETL processes, big data technologies, and cloud platforms to help the company make data-driven decisions.

#### *Responsibilities:*

- Developed and maintained efficient, reliable data pipelines using **Python, SQL, Airflow, Flyte**, and **Spark** for data integration, transformation, and processing of large-scale datasets on cloud platforms such as **AWS S3** and **Google Cloud Storage**.
- Implemented data monitoring, validation, and cleansing processes to maintain data quality and integrity, while ensuring consistency and compliance with industry standards and business requirements.
- Optimized data models and storage systems through advanced techniques like denormalization, partitioning, indexing, and data compression, resulting in improved query performance and efficient resource utilization.
- Designed and implemented **APIs** for data extraction and real-time analytics using **Python**, allowing seamless access to data for various internal and external stakeholders, enhancing data-driven decision-making capabilities within the organization.
- Proactively monitored **data infrastructure performance** and **health**, addressing potential bottlenecks and collaborating with cross- functional teams to resolve data-related issues.
- Collaborated with peers to promote a learning environment and shared best practices, actively contributing to the continuous improvement and innovation.
- Collaborated with data scientists and analysts to deliver customer support data for advanced analytics, utilizing optimized data models
- and storage systems to enable the identification of trends, patterns, and areas for improvement in customer satisfaction.
- Refined and maintained the unified customer support data model for the native customer help system, leveraging data monitoring, validation, and cleansing processes to provide consistent, accurate, and timely data to support teams across the customer-support organization.
- Contributed to a **22%** improvement in customer effort scores within the first quarter of implementation by leveraging data-driven
- insights to enhance targeted support measures, while proactively monitoring data infrastructure performance and addressing potential bottlenecks.
- Seamlessly integrated data from internal agent action monitors and Zendesk, leveraging **real-time ETL** pipelines to calculate centralized KPIs for each agent, supporting informed performance evaluation and decision-making.
- Enhanced data quality tracking for customer service content, supplementing qualitative agent feedback with enriched data from diverse support sources, culminating in the creation of accurate and comprehensive reporting dashboards.
- Realized an **8%** improvement in agent efficiency through the implementation of a data-driven approach while adhering to data security and access control protocols in compliance with industry best practices and regulatory requirements. Collaborated with the Ride-Safety team to automate pre-ride safety predictive Machine Learning

Models, leveraging Airflow for training data backfilling and efficient data aggregation using batched data pipelines.

- Developed real-time pipelines to stimulate the production environment to test the various levels of tolerance tests for the Model.
- These pipelines were built to handle streaming data, leveraging **NoSQL** solutions using **Amazon DynamoDB** for ingesting and processing **real-time** event data efficiently.
- Utilized Google Cloud **Dataflow** and integrated **GCP** services, including **BigQuery**, **Cloud Storage**, and **Pub/Sub**, to design, manage, and optimize data workflows, ensuring efficient data processing and timely delivery of aggregated metrics for diverse stakeholders across the organization.
- Collaborated with cross-functional teams to develop data-driven customer support dashboards and visualizations using BigQuery and Data Studio, empowering stakeholders with actionable insights into key performance indicators and facilitating informed decision-making.
- Developed versatile APIs using **Python**, **AWS Lambda**, and **Google Cloud Functions** to provide seamless access to cleaned and refined data, while ensuring data security and compliance through robust authentication, authorization, and access control mechanisms, accommodating diverse use cases and stakeholder requirements.

## Nokia Corporation

San Jose, CA

### Data Engineer

Mar 2020-Aug 2021

#### *Description:*

Our project purpose was to optimize services provided by leveraging in depth marketing and customer analysis. My responsibility was to collect, track and analyze 5G network performance data from clients to optimize services, and create KPIs for business insights dashboards.

#### *Responsibilities:*

- Designed information management tool to handle 5G network data, helping Nokia solution teams to **track, log and fix** clients' network glitches quickly by reducing service turnaround time by **70%**
- Wrote and maintained end-to-end Data pipelines to process transaction level authorization data stores in the target databases (**Redshift and Big Query**)
- Integrated and operationalized Analytical Models, automated Data pipelines and created automation tools that made data analysis workflows efficient
- **Extracted, cleaned, transformed** and merged customer-level hundreds gigabyte network data from **Redshift** and **BigQuery** using advanced optimized SQL queries
- **ETL** from various data sources like RDBMS, flat files and API into HDFS
- Developed **Apache Airflow** pipelines to automate vendor reports generation on **GCP** and **AWS**
- Implemented QC functions in the **Airflow DAGs** to automatically QC input, intermediate and output tables
- Leveraged Python to build the DAGs files and evaluation functions for QC functions
- Utilized PySpark for big data projects on **GCP** and **AWS** to optimize data pipelines performances by 30%
- Developed and implemented network key performance indicators, helping Nokia in performance **testing and contract negotiations** with clients by highlighting the hidden problems in their networks
- Automated Daily network reports (Radio and Network Switching) to measure network health, improving the operational efficiency
- Prepared weekly, monthly and quarterly network performance testing reports for **all-level audience, from Senior Manager to VP level**, evaluating 5G network expansion and optimization
- Gathered **business and functional requirements** for Retail Marketing department
- Identified distinct groups of customers which informed content and marketing strategy
- Use statistical and machine learning techniques to create scalable analytics solutions
- Analyzed **business trends and performance** regionally, including point of sales analysis of product categories, revenue, and customers segmentation to identify the new business opportunities
- Worked with supervisor to discern significant value drivers and challenges, and determine the key metrics and insights for improving business performance
- Collaborated with cross-functional teams (sales, strategy, IT) to enhance channel program offering

- Conducted **data preparation and wrangling** (EDA, impute, scaling and formatting), **figure engineering** for prediction
- Developed **predictive models** based on logistic regression and data mining technique by Python to optimize email send-time for marketing department, and day-of-week, resulting in a 13% uplift in open rates with email subscriber list
- Performed data migration with incremental data and full data, ensuring high integrity and quality for full 9-years economic cycle data
- Develop and Maintained **Data pipelines** for real-time interactive Dashboards in **Tableau** for unique insights
- Identified opportunities, and helped program management stakeholders to make decisions for improving business performance by implementing changes in the existing data model, and provide ad-hoc reporting dashboard using Tableau
- Developed customer affinity analysis based on prior network activities and geographical locations; Retargeted campaigns led a 20% increase in new/recurring revenue

**Technology/Environment:** MS Office, Python (JSON, YAML, Numpy, Pandas, Sklearn, **PySpark, Apache Airflow**), Tableau, SQL, **AWS, Azure**

## **Blue Cross Blue Shield**

McLean, VA

Data Analyst

June 2019-Feb 2020

### **Description:**

The project was aimed at improving and automating claim management system to improve customer experience by leveraging A/B testing, sentiment analysis, visualizing key matrix to support corresponding team. Another part of the work was to prepare ready-to-use dataset by building data pipeline to assist customer pre-assessment for pricing.

### **Responsibilities:**

- Maintained patient database by using advanced **SQL queries**; supported adjudicators with medical, information and regulatory reviews
- Worked under **Agile Methodology** to promote iterations collaboration, and process adaptability; participated in fast-paced scrum meetings throughout the life cycle of project
- Partnered with business analyst to analyze end-to-end business processes, develop business requirements, **automate full processes by leveraging data pipelines on AWS**
- Extracted million rows of data from primary or secondary sources; **wrangled, formatted, scaled and excluded outliers** from the data
- Applied market analytics by conducting **A/B testing** on website content, timing and channels, resulting in YoY growth in impression/clicks/apps
- Updated workflow, process flow diagram and GAP analysis by identifying existing billing, documenting the enhancements to meet the end state requirements
- Presented the results of non-functional/technical improvements on sprint review meeting by leveraging **Tableau**
- Analyze business requirements and document them into vision document and segregated them into high-level and low-level Use Cases, **UML** diagrams (activity diagrams / State Chart Diagrams)
- Performed **sentiment analysis** with NLP techniques on online surveys, primary doctors reviews by customers to drive multiple processes improvements
- Generated **Tableau interactive dashboard** to visualize key metrics to explore potential customer behavior indicators

**Technology/Environment:** MS Office, Python (Numpy, Pandas, Sklearn), Tableau, SQL, Scrum framework, **AWS**

## Hubei Century Network Technology Inc.

Wuhan, China

Data Analyst

Feb 2017-May 2018

### Description:

The project was to enhance the online payment features. Improved and simplify customers' experience by providing eye capturing snapshot design and customers' experience assistance.

### Responsibilities:

- Gathered information from various stakeholders and documented high level business requirements
- Assist users in account management by providing visualized summaries on weekly spending
- Designed, created and defined **key metrics** applying to data to analyze the past business performance, and worked on generating scripts for churn rate prediction by considering factors such as monetary size, monthly payment, rating, and terms
- Utilized **tree model, random forest model, logistic model and linear regression** model to predict the business performance
- After tuned the parameters of mode, Improved the model performance by around 10%
- Conducted workflow, process diagram and GAP analysis by identifying existing technologies, documenting the enhancements to meet the end state requirements
- Used advanced Microsoft excel to categorize customers; Improved the counseling effectiveness by 20%
- Ran sample tests on different categories and visualized the business performance increasing rates and trend in each category
- Built Tableau visualizations for **ad-hoc** reporting against various Data Sources for stakeholders' meetings
- Analyzed provided dashboards and requirements to improve the performance for faster access of dashboards to the partners

**Technology/Environment:** MS office, Python (Numpy, Pandas, Sklearn), Tableau, SQL

## HireVue Inc.

South Jordan, UT

Data Analyst

June 2015-Dec 2016

### Description:

Helped to generate daily and monthly financial reports by using automation scripts and financial analysis models in R. Categorized customer information by Excel.

### Responsibilities:

- Performed **full cycle** of accounting duties and assisted with budgeting and forecasting on a monthly basis
- Major accomplishment included reducing company's marketing cost and made routine management tasks into automation procedure to improve office efficiency
- Ran the statistical data trends analysis and created financial visualization models using **Excel** and **R**, and monitored the risk appetite and exposure threshold rebalance the portfolio asset allocation
- Built automatic customized analysis and data visualization dashboard to automate the portfolio reporting on daily PnL to portfolio manager

**Technology/Environment:** MS Office, R, Power BI

## Zion Bancorporation

Salt Lake City, UT

Business Analyst

June 2014-Dec 2014

### Description:

Leveraged R programming and advanced Excel (pivot table) to predict churn rate, categories customers, and targeted customers to give one-time discount.

### Responsibilities:

- Worked on generating scripts for churn rate prediction by using **R**, considering factors such as monetary size, monthly payment, rating, and terms

- Utilized tree model, random forest model, logistic model and linear regression model to predict the churn out of customers; tuned the parameters of models
- Used advanced **Excel** skills (VLOOKUP & pivot table) to categorize customers according to **RFM** (Recency, Frequency and Monetary) model; ran sample tests on the models
- Ran sample tests on different categories and visualized the respond rates and trend in each category by Tableau
- Calculate CLV (customer lifetime value) by considering discount rate, category costs, marketing costs, budget and profits
- Construct financial risk investigation models and timely rejected 5% of credit letter requests based on credit default rate

**Technology/Environment:** MS Office, Excel (VLOOKUP & pivot table), R, SQL, Tableau

## EDUCATION

<b>Johns Hopkins University</b>	Washington, DC
<i>Master of Science in Applied Economic</i>	
<b>University of Utah</b>	Salt Lake City, UT
<i>Bachelor of Science in Applied Mathematics</i>	

## CERTIFICATION

<b>HackerRank</b>	<b>SQL(Basic), SQL(Intermediate), Python(Basic)</b>
<b>DataBrick</b>	<b>Apache Spark SQL for Data Analysts</b>
<b>Google</b>	<b>Google Cloud Platform Big Data and Machine Learning Fundamentals</b>