

# Owen Liu . Jarvis Consulting

I am an aspiring industrial engineering new grad from the University of Toronto. Throughout my undergrad, I specialized in data analytics, operations research, and project management; I had one research assistant experience regarding statistics approaches in the stock market using Python; Besides, several industry-relevant core projects in my undergrad have built up my critical thinking, problem-solving and fast-learning skills; As a current Jarvis employee, I am enhancing my skill sets with Docker, RDBMS/SQL, Python data analytics, Hadoop, Scala/Spark etc. Apart from being a data analytics and data science enthusiast, I am a total people person. I have confident conflict-solving and interpersonal skills. Aiming to be a competent trilingual speaker in 2022, I have proficient verbal and documentation skills in English. I also have intermediate French skills, but I have been hammering it since graduation and tackling a French DELF B2 exam in June 2022. Most importantly, I am highly empathetic. I tend to be 'there' with stakeholders first instead of conducting premature problem-solving regardless of the feelings of the stakeholders.

## Skills

**Proficient:** Python (NumPy, Pandas, Sklearn, Seaborn, Matplotlib, Tensorflow, etc.), RDBMS/SQL, Postgresql, Agile/Scrum, R, Simulation (With tools like Simio), MS Access & Excel, Statistics (Hypothesis testing, A/B Testing, Model Fitting)

**Competent:** Java, NoSQL, Tableau, Hadoop, Optimization (Mixed Integer Programming, Dynamic Programming), Forecasting. Production/Revenue Modelling, Decision Tree Learning (CART, Random Forest), Feature Engineering (Regression, Clustering, Grid Search)

**Familiar:** Linux/Bash, Oracle, Machine Learning (Neural networks, K-nearest Neighbours), Scala, Spark, Azure & AWS & Google Cloud

## Jarvis Projects

Project source code: [https://github.com/jarviscanada/jarvis\\_data\\_eng\\_OwenLiu](https://github.com/jarviscanada/jarvis_data_eng_OwenLiu)

### Cluster Monitor [GitHub]:

- Developed a cluster resource monitoring program that collects data CPU hardware specifications using Bash Shell Scripts.
- Provisioned a PostgreSQL instance using Docker.
- Stored static and dynamic real-time data in the psql instance.
- Automated dynamic real-time data retrieval every one minute using crontab.
- Inserted the specifications into two tables using Data Definition Language.
- Resolved three business-level questions with data analytics tools in SQL.

### Python Data Analytics [GitHub]:

- Delivered data analytics report to the Jarvis LGS team via Jupyter Notebook and Github.
- Provisioned a Jupyter Notebook instance using Docker.
- Used a docker network to connect the Jupyter Notebook Instance and a Postgresql Instance.
- Loaded the .sql file in Jupyter notebook as a data frame with sqlalchemy, Postgresql connection, and Python Numpy.
- Conducted Elementary Data Analysis (EDA) with Numpy operations.
- Answered to eight business questions with Pandasql, Python NumPy, pandas, matplotlib and seaborn.
- Presented insights inspired by the business answers with descriptive, predictive, and prescriptive analysis recommendations.

**Hadoop** [GitHub]: In Progress

**Spark/Scala** [GitHub]: Not Started

## Highlighted Projects

**Capstone Project: the eCA System. An Information Retrieval Digital Platform for EV Stations:** Developed a full-stack Web Application that showcases electric vehicle charging stations(EVCS) data to EVCS Owners and Managers using Python, FastAPI, React, AWS, and Postgresql Database. Conducted research regarding use cases, user stories and personas. Instantiated data ingestion service in migrating raw data from the client's database using the AWS Migration Service. Stored the migrated data in our database and made a Multi-AZ Backup DB. Instantiated data processing service

for applying algorithmic transformations and recorded them in our database using Python supported by AWS EC2. Scheduled the data transformation jobs on AWS batch to keep the processed data up to date. Established PostgreSQL connection to migrated database. Transformed the historical data catered to the interest of end-users inspired by customer demographic research. Featured station metrics, including current power, station availability, average charging hours for each port, etc. Presented the station metrics to end-users with SQL-empowered OLTP Queries. Featured aggregated metrics, including monthly/weekly charging power in total, weekly /monthly visits to each port, etc. Presented the aggregated metrics to end-users with SQL-empowered OLTP Queries. Displayed over ten options of bar plots, line plots, histogram plots of the historical data with Python Seaborn and matplotlib upon user requests. Performed data analysis towards transformed historical data using python-empowered OLAP Queries in the data warehouse and provided insights to EVCS management accordingly. Instantiated the API service, a RESTful backend API written in Python, deployed on AWS Elastic Beanstalk to utilize auto-scaling feature. Deployed the frontend UI in an AWS S3 bucket and served end-users via CloudFront. TLS Encrypted the data transfer between the backend API and the user interface. Used AWS Cognito as the OAuth provider for backend secure access. Conducted industry-level usability test and generated reports spanning the functionality, ease of use, aesthetics, recommendations of the web app. Received critical acclaim from the supervised professor and the industry client when the product was demoed.

**A Developer Job Satisfaction Predictor.:** Predicted job satisfaction and competence from a poll survey conducted by the StackOverFlow community using regression, random forest classifier, and clustering. Performed data cleaning to NA values, amputation to unnecessary attributes, parsing and concatenating to raw data, reordering and mapping data with Jupyter Notebook. Performed data visualization through seaborn, sklearn and matplotlib tools in Python. Drew bar plots, box plots, in-plots of job satisfaction aggregated by different factors. Attained preliminary insights through these graphs. Constructed a baseline model that always returns the mean of the target variables in the train set. Initialized 80% of the data as a training set and 20% for a testing set. Attained the precision score of 0.35 of prediction using pandas in Python. Categorized different job satisfaction levels and competence levels into numerical categorical variables using pandas in Python. Constructed a logistic regression model predicting the performance of categorical variables, achieving a precision score of 0.43 in Python. Constructed a Random Forest Classifier with tuned hyperparameters, achieving a precision score of 0.44 in Python. Constructed a K-Nearest Neighbours Classifier with k=500 for each categorical variable, attaining a precision of 0.37 in Python. Analyzed a confusion matrix of all the models' precisions through Seaborn. Calculated the proportion of data points that were predicted lower than true values for each model using Python. Selected Random Forest Classifier as the best performing model for job satisfaction, Logistic Regression Classifier for competence. Constructed a job satisfaction optimization model based on the score results from the selected results above in Python. Ran the simulation model 100 times, attained the optimized result with different sets of biased constraints in Python. Conducted sensitivity analysis towards different sets of constraints. Delivered insights through predictive analysis to the simulation model result and predictive model results in Python.

**Airport Passenger Movement Simulator.:** Modelled historical data from several CSVs, including passenger arrivals, check-ins, and security screening, bag drops, in R. Analyzed and tested the fitness of presumed statistical distributions over the historical data with the Kirmogorov and Smirnov Test using R. Designed a Simio model to simulate the real-time passenger movement. Identified congestion in different modules of the screening process through Simio model performance. Found 10% of passengers will miss the flights. Provided and instantiated budget-permitting upgrading business decisions in Simio. Decreased the proportion of flight-misses to 4% from performance of the upgraded Simio Model.

**An Online Earphone Encyclopedia Website.:** Developed a fully interactive website including functions like search, sort, review, account creation and edition using Java JDBC, MS Access Database, Git, and Apache Tomcat. Established an RDBMS in Microsoft Access from scratch. Connected MS Access to the Java backend using the Youcanaccess.jar file. Deployed and practiced Agile/Scrum project management methodology. Worked with a 7-person team dynamically through GitHub. Gave a demo/project release speech on behalf of the team.

## Professional Experiences

**Data Engineer Consultant., Jarvis (2021 - present):** Performed various data retrieval, analysis, and report building jobs for the Jarvis LCA team using Shell scripts and docker-provisioned postgresql. Performed data analysis, visualization, and business question proposals for a Jarvis Client with docker-provisioned PostgreSQL and Jupyter Lab.

**Research Associate, Peking University (2019.6 - 2019.8):** Optimized an investor problem through stochastic approximation in Python. Constructed a pricing model of an American call option in pandas Python. Attained the volatility where the predicted price of the option has an optimal fit to the actual price through stochastic approximation in Python. Simulated price samples with a random generator in NumPy Python. Caught the feasible results with steepest descent algorithm, non-linear programming, Monte Carlo simulation, and gradient estimation. Visualized the distributed

result in matplotlib. Retrieved finance data from a dummy database with SQL. Picked up stock market basics in a short time span.

**Store/Warehouse Associate, TJX Canada (2018.9 - 2019.6):** Enhanced interpersonal skills with customers with English and French. Worked as a team in-store to address customer concerns.

**TOEFL/GRE Tutor, Billions Education (2018.6 - 2018.8):** Delivered insights in English to over 20 numbers of ESL mentees with TOEFL writing and speaking tests and GRE Verbal test. Good listener. Listened to mentees' concerns before course plans and collected their feedback after each sector of the course.

## **Education**

**University of Toronto (June 2021),** Bachelor of Applied Science, Industrial Engineering

## **Miscellaneous**

- Graduate Record Examination(GRE): Verbal 166/170, Quant 169/170, Global Top 5%.
- Microsoft Azure DP600 Certificate
- Yale World Youth Economic Forum: 2016 Outstanding Candidate
- Volunteer @ Toronto International Film Festival 2017
- Volunteer @ Toronto Pride 2016 and 2017
- French B1, Achieving full fluency in 2022