Francis Tao

8688 Hazelbridge Way, Richmond, BC V6X0R6 | P: 778-681-6828 | E: ty199695@gmail.com | LinkedIn | GitHub

Professional Summary

A diligent and ambitious professional with expertise in data analytics, passion in machine learning, exceptional communication skills and multilingual ability. A vast skill set of statistical inferences, time series analysis, predictive analysis and data visualization..

Skills

Database: PostgreSql, MySQL, T-SQL, SSMS

Statistical Programming: Python, SAS, RStudio

Data Visualization: Tableau, Power BI, Matplotlib, Seaborn

Data Analytics and Science: Pandas, Numpy, Sklearn

Statistical Topics: Sampling, Hypothesis Testing, A/B Testing

Others: Excel, Access, Outlook, Slack, Jira, HubSpot, SharePoint, Azure DevOps

Work History

RecycleSmart   Richmond, BC
Data Coordinator   *11/2021 – 12/2022*

- Pull the data from **Microsoft SQL Server**, generate 250+/3300+ customized/standard waste diversion reports monthly in **Excel** and present insights of recycling data through **Power Query** and **Power BI** dashboard.
- Clean and transform over 8300 rows of customers/vendors' addresses with **Pandas**, build **Python** validation tools to automate the process by utilizing CanadaPost/BingMap API, computing textual similarity using **theFuzz/Levenshtein Distance** with 75.1% accuracy, converting kmz file to json and creating region/neighborhood codes based on **Polygon** geospatial analysis.
- Write **SQL** query and utilize **Power Query/Power BI** to transform the data and create visualizations which help to address 17.51% of uneven vendor lines and 106.7k invoice discrepancy for a single agreement type from Jun 2021 to Jun 2022.
- Reduce the monthly waste diversion audit workload by an average of 66.2% over the past 12 months by creating a pipeline including cleansing data in **SQL Server** and leveraging libraries **re**, **Pandas**, **dateutil.parser**, **datetime**, **calendar** and **self def func** in **Jupyter Python**.
- Automate the data migration process of generating over 15k vendor purchase contract lines' unique key, unit price, region code and product id through the entire ETL in **SQL server(CTE, Subquery, Join, Concat, Dense_Rank, Case When)** and **Python(lambda, pandas)** for new ERP(**Dynamics 365**) system.

ehsAI   Vancouver, BC
Data Labelling Analyst   *02/2021 – 10/2021*

- Process EHS documents and conduct language analysis including citations, language labelling, splitting, and classifications to feed the NLP model.
- Develop internal QA survey, evaluate the employees' work performance by tracking workflow in **Jira** and analysing assignments including trend prediction based on visualization through **Power BI/Excel Query**, **Pivot**, Charts, **Python Pandas**, Matplotlib, **NumPy**, **Seaborn** and improve the task accuracy by 10% within 6mo.

Providence Health Care   Vancouver, BC
Data Quality Clerk   *01/2020 – 06/2020*

- Assist end-users with registration-related processes, liaise with various departments and escalate problems.
- Extract, sort and aggregate data using **SQL** and verify them from client registry/clinical information reports to maintain data integrity and to ensure compliance with high accuracy of above 85%.

ECCO Shoes Canada   Richmond, BC
Brand Ambassador   *11/2019 – 10/2021*

- Greet customers, help to find the wanted shoes, fix the displays and resolve the return and exchange issues.
- Generate more than $1500 in weekly sales with successful marketing, sales and customer relations approaches.

Personal Projects

Bike Share Demand Prediction (Link) – 0.4075 RMSLE(301/3243 on Kaggle, top 9%); Data manipulation with **Pandas**, **NumPy** and datetime; Histogram, scatterplot and bars visualization with **matplotlib**; Modelling and training with **Train_Test_Split**, **LinearRegression**, **RandomForestRegressor** and **XgBoost**.

Heart Disease Prediction (Link) – **GridSearchCV** Logistic Regressoin with 88.5% accuracy; 303 rows of data; Barplot, scatterplot and heatmap by **matplotlib** and **seaborn**; Modelling and training with **LogisticRegression**, **KNeighborsClassifier**. **RandomForestClassifier** and **DecisionTreeClassifier**; **Hypertuning** using **RandomizeSearchCV** and **GridSearchCV**

Voluntary Experience

Quest Food Exchange   Vancouver, BC
Office Assistant Volunteer   05/2019 – 07/2019
- Excel data entry and update referral clients' information into spreadsheet over 2800 rows for 2000+ organizations.

Greater Vancouver Food Bank   Vancouver, BC
Cleaner   11/2021
- Grocery shelf cleaning

Education

The University of British Columbia   Vancouver, BC
Bachelor of Science, Statistics   05/2019
Dean's List 2016 – 2017; Member of Golden Key Society: Top 15% in faculty; Graduated with 82%

Certification

(1). Data Analyst with Python
(2). MCSA: SQL 2016 Database Development
(3). Tableau Desktop Specialist
(4). SAS Certified Specialist: Base Programming
(5). Data Science Professional Certificate