








# TENNISON YU

## CONTACT

-  248-385-8969
-  tennisonyu@gmail.com
-  linkedin.com/in/tennisonyu
-  tyu0912.github.io/
-  github.com/tyu0912/
-  72 Townsend St #306,  
San Francisco CA
-  Canadian Citizen/American  
Permanent Resident

## SKILLS

### Python:

scikit-learn, keras, pytorch,  
pyspark, flask, nltk, numpy,  
pandas, matplotlib

### Data Querying:

MySQL, Snowflake, HIVE,  
Elasticsearch

### Development & Cloud Services:

Docker, Jenkins, Logstash, AWS  
[EMR, ECS, Lambda, Sagemaker]

### Machine Learning & Analysis:

Logistic & Linear Regression,  
K-NN, K-means, Time Series,  
Neural Networks, Hypothesis  
Testing

## EDUCATION

### Master of Information and Data Science (MIDS)

UC Berkeley // APR 2020

### Master of Engineering Science (M.E.Sc): Chemical/Biochemical

Western University // AUG 2014

### Bachelor of Science (B.Sc): Biochemistry, Minor: Math/Stats

McMaster University // APR 2012

## PROFILE SUMMARY

Data science professional with 5+ years of analytical and machine learning experience. Particularly interested in productionizing ML models, developing pipelines, and ensuring fast and efficient infrastructure. Currently, my biggest hobby is creating new cooking recipes.

## EXPERIENCE

### DATA SCIENTIST

BEAM SOLUTIONS // SAN FRANCISCO // MAY 2019 – Present

A financial compliance software company aiming to simplify and enhance transaction monitoring and the detection of money laundering, fraud and other criminal activities.

- Build RESTful-APIs using Docker, Flask, and Pyspark for data ingestion, aggregation and general access by clients.
- Addressed big-data computing needs by spearheading deployment and ongoing maintenance of AWS EMR clusters and related frameworks.
- Worked with banking clients on data integrity, anomaly model development, and overall integration of transaction data.

### DATA ANALYST

DOXIMITY // SAN FRANCISCO // SEPT 2018 – MAY 2019

A social network platform that aims to connect physicians and provide them with latest content in terms of news, academic research, and institutional ratings.

- Simplified hypothesis testing of marketing email content by building a pipeline to stage different formats and the calculations afterwards.
- Created new more targeted categories for classifying medical news using TF-IDF method to vectorize articles and k-means for grouping.

### CLINICAL ANALYST // BI DEVELOPER

CHEN MEDICAL CENTERS // MIAMI // MAY 2015 – DEC 2018

A medical center enterprise, ChenMed's model revolves around preventative care and has its own IT department to build and manage a robust medical record system.

- Helped find medical form inconsistencies by building a web dashboard powered by a K-NN classifier that grouped filled vs unfilled forms.
- Prepared scripts for data ingestion/scraping to MySQL, KPI calculations, and dashboard refreshes.
- Data consultant and advisor for all East Virginia markets as well as corporate medical, management, and operation teams.

## Research and Personal Projects

- [Using a temporal-shift-model on MobileNetV2 backbone to detect falls](#)
- [Predicting water likelihood at Serengeti using satellite imagery](#)
- [A comparison of GPT-2 and XLNet for sentence generation](#)
- [Multiplicative effects of novel therapeutics on fungal pathogens](#)
- [A model to predict spectral properties of nanotubes in surfactant solutions](#)