

# Qi Tang

Address: Saint Louis, MO 63130 || Cell: (314)203-5663 || [qi.t@wustl.edu](mailto:qi.t@wustl.edu)

LinkedIn: <https://www.linkedin.com/in/allen-tang-798641224>

## SUMMARY

Master of Finance (STEM), aiming at finding data-related positions, with sufficient project experience and working experience in **machine learning** (recommendation system/series forecast), **big data analyzing** (Databricks/Azure), **business intelligence** (Tableau/Power BI), and solid programming skills in **Python and SQL**.

## EDUCATION

<b>Washington University in St. Louis</b> <i>MO, USA</i>	Aug 2021 - Dec 2022
Master of Finance in Wealth and Asset Management(STEM)	<i>GPA: 3.7/4.0</i>
Coursework: Python and Data Science, A/B testing, SQL, Text mining	
<b>Nanjing University of Finance &amp; Economics</b> <i>Nanjing, China</i>	Sep 2017 - Jun 2021
Bachelor of Economics in Finance	<i>GPA: 3.7/4.0</i>
Coursework: Microsoft Office, Statistics, Data Analysis Technology Based on Spark	

## SKILLS

### Machine Learning Techniques:

- Supervised Algorithms: Neural Networks (CNN, RNN), SVM, Naive Bayes, KNN, Text mining (NPL)
- Unsupervised Algorithms: Clustering (K-means, DBSCAN), Principal component analysis (PCA)
- Big Data Techniques: Azure Databricks

**Business software:** Tableau, Power BI, Microsoft Office

## PROJECT EXPERIENCE

**San Francisco Crime Analysis in Apache Spark** Oct - Nov 2021

[https://github.com/Allen9809/Big-Data-Project-1\\_SFC-crime-data-analysis](https://github.com/Allen9809/Big-Data-Project-1_SFC-crime-data-analysis)

- Built data processing pipeline based on Spark RDD, Spark Dataframe and Spark SQL for big data OLAP.
- Got business insights of the analysis of data, such as the top-3 dangerous districts, percentage of resolution for different category of crime, then post hints toward police policy.
- Explored the spatial distribution of incidents by K-means clustering algorithm, plotted and found the k value with the best silhouette performance, which is 3, then trained the optimal model.

**Movie Recommendation System based on Spark ALS** Dec 2021 - Jan 2022

[https://github.com/Allen9809/Machine-Learning-Project\\_Movie-Recommendation-System-based-on-ALS](https://github.com/Allen9809/Machine-Learning-Project_Movie-Recommendation-System-based-on-ALS)

- Performed data ETL on origin movie data from GroupLens and find all movie genres by Spark SQL and OLAP.
- Fitted the ALS model on the training data(80%), then select the hyper-parameters through grid search and 3-fold cross validation to tune the model, and the RMSE on the testing data is 0.88.
- Used the model to make k movie recommendations to users with given userId by Spark SQL.
- Found top k similar movies for a given movieId by KNN algorithm, and distance evaluation are based on both the Euclidean distance and Cosine distance.

## WORK EXPERIENCE

### Nestle, Purina

*Saint Louis, USA*

*Data science internship.*

May – Aug 2020

- Built a rice consumption amount forecast model for the Purina supply chain department, with an accuracy of 95%, which is acceptable to apply to the business and even 10% higher than the previous model, and it can be generalized to almost all commodities besides rice.
- Developed an ingredient allocation optimization tool for the Purina supply chain department, which can help them rate different series of products based on their current business focus (e.g. profit, supply), thus guiding factories prioritize the production of all products.

### Nestle, Purina

*Saint Louis, USA*

*Associate data scientist.*

Oct 2020 - now

- Assist to develop a truck load optimization engine, which allows the procurement specialists to consolidate the order into full trucks. The average truck utilization is around 98%. It is currently being tested in the warehouse.