

## Contact

[www.linkedin.com/in/lanyuealisonmeng](https://www.linkedin.com/in/lanyuealisonmeng) (LinkedIn)

## Top Skills

Matlab  
Nuclear Medicine  
C

## Languages

English (Full Professional)  
German (Elementary)  
Japanese (Professional Working)  
Italian (Elementary)  
Chinese (Native or Bilingual)

## Certifications

Algorithms: Design and Analysis, Part 2  
BerkeleyX: CS110x Big Data Analysis with Apache Spark  
An Introduction to Interactive Programming in Python  
Algorithms: Design and Analysis, Part 1  
BerkeleyX: CS120x Distributed Machine Learning with Apache Spark

## Honors-Awards

National scholarship  
National Scholarship  
National Scholarship  
National Scholarship

# Lanyue Alison Meng

Senior Data Scientist at Google  
San Francisco, California, United States

## Summary

Passionate statistical data scientist with demonstrated experience utilizing machine learning to create predictive models and actionable insights.

My Email: [lanyue.meng@gmail.com](mailto:lanyue.meng@gmail.com)

### My Technical Skills:

- Analytics: understanding business problems and building data-driven solutions
- Machine Learning: Regression, Classification, Clustering, Feature Selection, Model Selection/ Assessment, and Recommendation
- ETL& Data Mining
- Natural Language Processing
- Python Data Analysis: Numpy, Scipy, Pandas, Scikit-learn, NLTK
- R Data Analysis: Dplyr, Lubridate, Glimmer, Caret, TidyR
- Data Visualization: Tableau, ddplot, matplotlib
- Big Data related techniques: Spark, Hadoop, Oozie, Sqoop, Hive, Impala
- Programming Languages: Scala, Python, R, Java, SQL, C, JavaScript, Matlab
- Database: MySQL, SQL Server, TeraData, Vertica, MongoDB

---

## Experience

### Google

5 years 6 months

#### Senior Data Scientist

November 2022 - Present (9 months)

San Francisco, California, United States

YouTube Responsibility

#### Data Scientist

November 2021 - November 2022 (1 year 1 month)

San Francisco, California, United States

YouTube Responsibility

Data Scientist

February 2018 - November 2021 (3 years 10 months)

Zurich, Switzerland

Google Trust & Safety

MedeAnalytics

Data Engineer

April 2016 - August 2017 (1 year 5 months)

San Francisco Bay Area

1. Developed a clinical quality engine in Spark (Scala) to quantify the quality and cost of patient care with over 79 measures

2. Implemented ETL jobs for products in Spark (Scala) and deployed in production Hadoop ecosystem

2.1 Developed in full software development lifecycle

2.2 Supported weekly reporting and analytics for over 150 medical facilities on customized UI platforms

2.3 Built over 70 metrics to measure the quality and the effectiveness of hospital utilizations, to detect their areas of risk and to provide solutions to improve their financial health

2.4 Investigated and debugged data issues such as outdated data and timestamps converted to incorrect time zones

3. Designed and built scaled data pipelines to transfer and process data (1TB daily) using Apache Oozie, Spark, Sqoop and Hive

3.1 Improved instrumentation to ensure the accuracy, concurrency and timeliness of these pipelines and optimized logging

4. Performed ad hoc analysis, data processing and data visualization using Hive and Impala and other SQL scripts

2nd.MD

## Data Analyst

February 2015 - January 2016 (1 year)

Houston, Texas Area

1. Performed data cleaning, data integration, data mining and machine learning on health care claims data (ICD10& CPT Codes)
  - 1.1 Proactively stratified high-risk & high-cost population and adjusted risk management to supply targeted communication
  - 1.2 Predicted future costs using models of PCA, K-means, Support Vector Machine, Random Forest, Gradient Boosting
2. Designed and developed statistical software scripts to detect and diagnose bugs and created targeted cost savings analysis for sales, operations and cross-functional areas
3. Supplied troubleshooting, analysis, and solutions for 2nd.MD MySQL database and back-end application issues
4. Created effective business requirements documentation and technical reporting
5. Developed and managed 2nd.MD RFP database
6. Provided technical support for 2nd.MD consultations
7. Provided QA testings on Web Applications and APIs
8. Designed new internal interfaces and metric dashboards for performance monitoring
9. Visualized and optimized operations workflow in a migration from Salesforce to in-house software

## MolecularMatch

Software Engineer Internship

July 2014 - December 2014 (6 months)

1. Structured data queries and data mining via a gene-based FDA clinical trial & targeted drug MongoDB big database

2. Consolidated data from pharmaceutical & clinicaltrials.gov databases and enabled streamlined updated access based on FDA regulations
3. Researched and Refined data for medical affiliations and clinical locations phases with JavaScript and Python
4. Assembled a user interface prototype that incorporates accurate search algorithms

---

## Education

### Rice University

Master of Engineering (MEng), Bioengineering and Biomedical Engineering · (2013 - 2014)

### University of Electronic Science and Technology

Bachelor of Engineering (BEng), Optoelectronic Science and Engineering · (2009 - 2013)