# SUMMARY

Jiaxu (Joshua) Luo

NEW YORK, NY [jl4298@columbia.edu](mailto:jl4298@columbia.edu) (917)399-7490 <https://www.linkedin.com/in/jiaxuluo> <https://github.com/Joshuacourse>

A lifelong learner with experience in data science research, product development, and big data engineering on cloud architecture. Problem solver with research stamina, a can-do attitude, and cross-functional influencing skills. My career goal is to become a data scientist (engineer) with strong engineering skills to scale (with cloud technologies) analytical solutions to business problems.

# EDUCATION

**Udacity** - Data Engineering Nanodegree Program **In Progress**

**Udacity** - Machine Learning Engineer Nanodegree Program **Aug 2019 NYC Data Science Academy** | New York, NY - A three month, full-time program on data science **Sept 2016 Columbia University** | New York, NY - MS, Chemical Engineering (Data Science Concentration) **Dec 2015**

**Wuhan University of Technology** | Wuhan, China - BS, Material Science and Engineering **Jun 2014**

# EXPERIENCE

**Broker Genius** | New York, NY (startup team with two members) **Oct 2018 - Present**

*Data Scientist*

Tools: Python, BigQuery, AI platform, Machine Learning, Container

* Devised methodology in distinguishing fake sales from sales and implemented engineering solutions for automated modeling re- training, data validation, and hourly batch prediction (the company’s first ML model).
* Create production-ready data warehouses with scheduler and automate data pipelines and educate stakeholders in Sales, Support, Pricing, and Finance to utilize enriched data to inform data-driven decision making.

*Business Intelligence Engineer* **May 2017 - Present**

Tools: Python, R, Shiny, Crontab, Hive, Airflow, ElasticSearch, Tableau

* Designed, built, deployed, and keep monitoring 20+ production-ready reporting pipelines for the company’s $2 billion ticket inven- tory management, 200+ clients’ activities monitoring, 40+ pricers productivity evaluation, etc.
* Reduced 80% operating costs and improved accounting accuracy for the financial department by designing and developing an au- tomated invoicing, delivery, and recording system on monthly billing and accounting of 200+ clients.
* Validate new ideas by researching data integrity and developing fast prototypes with end-to-end solutions.

**Beijing Genomics Institute |** Shenzhen, China (Remote part-time)

*Machine learning consultant* **Feb 2017 - May 2017**

* Researched, validated and implemented risk prediction modeling techniques for the detection of HPV infection based on genotyping data, which will increase early detection of cervical cancer for adult women.

# PROJECTS

[Customer Segmentation for Arvato Financial Services](https://github.com/Joshuacourse/Udacity_Data_Engineer)

Tools: Python, AWS, Machine Learning (PCA, Clustering, Lightgbm) **Aug 2019 - Oct 2019**

* Characterized customers segment of using demographics information and compare against that of the general population, identify- ing the parts of the which that best describe the core customer base of a company for subsequent target marketing.
* Built supervised machine learning model that predicts whether or not each individual will respond to the campaign and made pre- dictions on the campaign data as part of a Kaggle Competition. [(Udacity Capstone with Kaggle Leaderboard #9)](https://www.kaggle.com/c/udacity-arvato-identify-customers/leaderboard)

[Yelp Review Mining](https://nycdatascience.com/blog/student-works/yelpquest-begin-journey/) - (Team leader of four members)

Tools: R, Shiny, Tableau, Machine Learning (Topic Modeling, Xgboost) **Sept 2016**

* Extracted hidden topics and menus from Yelp customers reviews (positive & negative) for each type of restaurant by LDA(Latent Dirichlet allocation) to shed insights into the business operation of different restaurant types.
* Detected the key business operational metrics with predictive modeling for highly-rated restaurants and built up an application to deliver actionable insights. [(App)](http://joshuaneverstop.com/yelp_jiaxuluo/)

# SKILLS

Certification: [GCP Professional Data Engineer](https://www.credential.net/lt295qij?key=efd06b052b07e0f9259a0ba89b130e02b145b541fa274fecab5e1b1bd2bec6e7) | [Data Science Bootcamp](https://www.slideshare.net/slideshow/embed_code/key/2N6rFIkvqAJPBF) | [Machine Learning Engineer](https://graduation.udacity.com/nd009t) Language & Tools: Python | R | SQL | Spark | Hive | Airflow | ElasticSearch | Git | Tableau | Looker | Linux | H2O Machine Learning: Clustering | PCA | GLM | Xgboost | Random Forest | Neutral Network | NLP

Cloud: Google Cloud Platforms (BigQuery, AI Platform, Cloud Storage, Dataflow) AWS (Sagemaker, S3, Redshift, Lambda Fucntion) | Docker | Kubernetes