

# Senqi Zhang

New York City, NY | 585-503-4035 | [sz3016@columbia.edu](mailto:sz3016@columbia.edu) | [LinkedIn](#) | [Website](#)

## EDUCATION

### Columbia University

New York City, NY

*Master of Science in Data Science*, GPA: 3.72/4.00

*Anticipated December 2022*

- Related Coursework: Algorithms for Data Science, Applied / Statistical Machine Learning, Applied Deep Learning

### University of Rochester

Rochester, NY

*Bachelor of Science in Data Science*, GPA: 3.88/4.00

*May 2021*

- Related Coursework: Data Mining, Time-series Analysis, Database System, Data Science at Scale, Statistical Methods

## PROFESSIONAL EXPERIENCE

### BlocPower

New York City, NY

*Data Associate (Python, PySpark, AWS)*

*Jun 2022 – Current*

- Automated pipeline for generating quality assurance table using PySpark in AWS Glue. Enabled the program to run across tables and specific locations. Improved the program speed by 20%.
- Coordinated with other data scientists to design and deploy a quality assurance dashboard using AWS QuickSight.
- Performed ELT using DBT to store AWS Redshift query logs and designed a dashboard to monitor query performance.
- Complied data from multiple sources to perform preprocessing, EDA, and ingestion into S3 Data Lake.

### Columbia University

New York City, NY

*Course preparation Assistant (R, D3, Data visualization)*

*Jan 2022 – Aug 2022*

- Compiled bi-weekly assignments over a variety of topics for two semesters.
- Narrated [edav.info](#) to provide reference on data mining and data transformation. Provided guidance on best practices for graphing in R on various data types including multidimensional data, time series, and geospatial data.
- Contributed to the [D3 book](#) and provided a better learning experience.

## RESEARCH & PROJECTS

### Recommender system for Amazon products

*Sep 2021 – Dec 2021*

- Developed a text-based recommender system using Amazon review data and product metadata.
- Transformed data into a ten-core dataset to reduce sparsity and conducted feature engineering to select features.
- Implemented content-based filtering using review texts and product descriptions. Evaluated Bag-of-words and TFIDF models' performance using a set of products that were bought and viewed together in addition to human evaluations.
- Applied collaborative filtering using KNN, SVD, NMF models from the Surprise library. Evaluated performance for each model by calculating the RMSE metric and SVD achieved the lowest RMSE of 0.9467.

### Deep Learning for Natural Language Processing

*Jun 2022 – Aug 2022*

- Built a Neural Machine Translation system with bi-directional LSTM and attention to translate between Cherokee and English. Trained using AWS EC2 GPU instance and achieved a BLEU score of 12.9.
- Implemented sequence masking algorithm required for pretraining. Conducted pretraining and finetuning on GPT model for question answering tasks and achieved an accuracy of 38.8%.

### Mental Health Concerns on Twitter in the COVID-19 Pandemic

*Feb 2021 – Aug 2021*

- Collected 7TB COVID-19 mental-health related tweets uninterrupted on AWS EC2 instance.
- Applied topic modeling (Latent Dirichlet Allocation) to monitor the trend of discussions surrounding mental health concerns. Extracted ten topics and concluded underlying causes with example tweets.
- Leveraged machine vision and NLP algorithms (Face++/Ethnicolor) on over ten thousand users with a human profile picture and name to identify users' gender, age, and race. Conducted statistical analysis to compare derived user demographics with the general Twitter user population.
- Published one [paper](#) as the first author to help address the most common mental health conditions and identify groups that are most vulnerable to these conditions during the pandemic.

## SKILLS

- **Programming:** Python (PySpark, PyTorch, Sklearn, Boto3, Pandas, Numpy), SQL, R, Java, HTML, CSS, JavaScript
- **Product & Others:** A/B Testing, Data modeling, Data visualization, ETL, GCP, Azure
- **AWS Service:** Redshift, S3, EC2, Glue, QuickSight, SageMaker
- **Certificate:** AWS Certified Cloud Practitioner