

KATIE ASZKLAR

data science and engineering

✉ k.aszklar@columbia.edu

in www.linkedin.com/in/katieaszklar

🔗 kaszklar.github.io

SUMMARY

- Skilled in python, sql, etl, cloud platforms, statistical modelling
- Passionate about using data to provide actionable insights
- Motivated to design architectures and analyses that are reproducible, portable, and scalable

EXPERIENCE

DATA ASSOCIATE

Institute for Social and Economic Research and Policy

📅 Dec 2019 – May 2020

📍 NYC

- Lead python curriculum initiatives for GIS and Modern Data Structures courses
- Develop **Zipmigo**, a package aimed at early stage analysts working with zipfiles in cloud notebooks (see Projects)

QUANT. APPROACHES TO SOCIAL RESEARCH SOCIETY

Columbia University

📅 Fall 2019 – Spring 2020

Treasurer

- Spearheaded covid-19 response by creating and leading remote workshops and events
- Developed and documented SOPs to ease change of leadership transitions

Alumni Co-Chair

- Pioneered new blog feature "Alumni Interview Series"
- Facilitated QASR- and QMSS-hosted events

PROJECTS

The Anatomy of Activism

🔧 GCP Storage and Compute, NetworkX, Gephi, Seaborn, Scipy

Characterized two activist groups, over 50k Twitter users and 100k relationships, as social or informational based on network structure. Built a custom dataset of users using various API endpoints in order to construct a follow graph. Tested significance of clustering, reciprocity, 2hop neighborhood size, and measured correlations of assortativity.

Zipmigo

🔧 Requests, os, zipfile, tqdm

Python package aimed at early stage analysts working in Google Colab. Prevents recursive directory structure and provides insight into the working directory using formatted console output.

EDUCATION

Masters Quantitative Research Methods

Columbia University

📅 Feb 2021

Certificate in Data Science

Metis bootcamp

📅 Summer 2018

Bachelors Cognitive Science, minor Computer Science

New York University

📅 Jan 2014

SKILLS

Python

Scikit-learn

git

conda

SQL: Postgres, SQLAlchemy

NoSQL: PyMongo

AWS: ec2, s3

Databricks: mlFlow

GCP: Compute, Storage, BigQuery

Data Analysis and Statistical Modelling

Graph and Network Analysis

Supervised and Unsupervised Learning

Time Series

Web scraping

Geospatial Analysis

Natural Language Processing

VOLUNTEERING

Open Data Ambassador

📅 Fall 2019

📍 Queens Public Library

PyData Conference

📅 2018, 2019

📍 NYC