

KEO CHAN

(917) 453 0278 ◇ 242 E 19th St Apt 3B, New York, NY, 10003
keozchan@gmail.com ◇ linkedin.com/in/keozchan ◇ github.com/thecolorkeo

EDUCATION

Princeton University

2014 - 2018

BA in Geosciences

Selected Coursework: Algorithms and Data Structures; Modeling the Earth System: Assessing Strategies for Mitigating Climate Change

University of Otago, Dunedin, New Zealand, Semester Abroad

2017

EXPERIENCE

Insight Data Engineering, Fellow
github.com/thecolorkeo/InsightWiki

1/2019 - Present
[editwindow.wiki](https://github.com/thecolorkeo/InsightWiki)

- Built a scalable data pipeline, EditWindow, to analyze user editing behavior on Wikipedia using AWS S3 and EC2 instances.
- Parsed raw XML data from Wikidump into dataframes using Spark distributed computing.
- Leveraged TimescaleDB to reduce search time by 5x on database queries for time series.
- Automated execution of Spark and TimescaleDB tasks using Airflow.
- Created a website with Dash and CSS to allow users to explore data interactively.

Xiamen University, China, Research Fellow

6/2018 - 9/2018

- Merged 5 global plankton biomass data sets in multiple formats into one unified database.
- Measured correlation between plankton growth and environmental variables using linear models in R and Matlab.

Sesame Workshop, Intern, Strategy Department

Summer 2016

- Assessed and researched Sesame Street's YouTube account strategy using YouTube Analytics to maximize views and profit.
- Compiled and analyzed competitors' strategies for preschool education and YouTube markets.
- Recommended strategies to promote new videos and designed an A/B test.

PROJECTS

Senior Thesis, Princeton University

2017 - 2018

- Created an index to measure environmental health by analyzing global ecosystem trends.
- Built a database of marine isotope data compiled from 200+ scientific publications to identify global patterns.
- Analyzed data in R and Matlab using statistical model-fitting, image processing and interactive visualizations.

SKILLS

Programming Languages: Python 2 & 3, R, SQL (PostgreSQL, MySQL), Matlab

Data Engineering: Spark, TimescaleDB, Postgres, Airflow, Flask

Services & Tools: AWS (S3, EC2, VPC), Git, Linux, LaTeX