





Samuel Tseng

Data Engineer/AWS/GCP

Berlin, Germany 

+49 177 3203241 

samuel.tseng.de@gmail.com 

linkedin.com/in/samuel-tseng 

I'm a Data Engineer with Mathematics background who loves building data solutions with commercial value. My latest project provides ETL pipelines which aggregated several APIs' endpoints with Python, Fivetran, Snowflake and respectively cloud infrastructure designing.

Key stacks: Python, PostgreSQL, AWS, Snowflake, Container, asynchronous request/process

EXPERIENCE

Homeday GmbH, Berlin, DE - Data Engineer

March 2021 - PRESENT

Mainly supported the Business Intelligence team and its stakeholders in the development and maintenance of data infrastructure, ELT, and data warehouses.

- Maintained Homeday Property Price Estimation models and APIs which provide the backbone price estimation and trends in [Preisatlas Homeday](#).
- Developed ETL pipelines with asynchronous requesting/processing to extract and load daily property listings.
- Redesigned in-house API data aggregator to quadruple processing speed with asynchronous http requests.
- Automated press release document for quarterly price fluctuation via Google APIs, Snowflake, deployed on AWS ECS.

Freelancing, Volunteer, Berlin, DE - Cloud Solutions Consultant

May 2020 - January 2021

Provided professional perspectives about Data Engineering (infrastructure design, DWH, pipeline) and Cloud solutions.

digitty.io GmbH, Berlin, DE - Data Science Engineer

February 2019 - April 2020

Worked on digitty.io AI platform and Intelligence solution for customer and business strategy.

- Built scalable data pipeline and REST API to support the backend of digitty Human Resource platform, using Python, Flask, Docker, Celery, MongoDB, Redis, and Google Firebase, BigQuery.
- Developed candidates' CV classifiers, using Python, Pandas, Scikit-Learn, spaCy, Twitter API, and web-scraper.
- Analyzed and got insight out of user data, visualized, and reported them to C-levels and stakeholders, using Jupyter, Matplotlib, Google BigQuery, and Google Data Studio.

Freelancing, Berlin, DE/Taipei, TW - Data Science Consultant

August 2017 - January 2019

- **CodeDoor.org** - Tutor (Data Science), Oct 2018 - Jan 2019
- **DSP.Inc.** - R language Mentor @ BenQ Corp., Apr 2018 - Jul 2018
- **Zipjet GmbH** - Consumer demands forecast

SKILL SET

Programming Language

Python ●●●●●
Linux/Bash ●●●●○
SQL ●●●●○
R ●●●○○

On-premise/Cloud Data Solutions

RDBMS & DWH - PostgreSQL, Snowflake
AWS - EMR, Batch, ECS, Redshift, RDS, S3, Lambda, CloudWatch
GCP - Cloud SQL, BigQuery, Dataflow, Pub/Sub, Cloud Run, Cloud Build
Other modern stacks - dbt, Tableau, Rundeck, Prefect, Airflow

EXTRAS (Awards/Certificates/Activities)

Google Cloud Certified Professional Data Engineer ([Credential](#))

Jul 2020, issued by Google Cloud
Certification ID: C82TQ2

Taiwanese Professionals in Germany (TPG) - Co-Organizer

May 2019 - PRESENT

- 15+ keynote speeches/panel talks, career workshop
- 1000+ followers and event participants on Facebook, LinkedIn page

EDUCATION & TRAINING

Udacity Data Engineering Nano Degree, distancing learning

Jan 2022 - Mar 2022

Taiwan Artificial Intelligence Academy, Taipei, TW

Jan 2018 - May 2018

Complete theoretical courses and practical projects hand-by-hand with instructors, teaching assistants, and mentors from various industries. The content of courses were including Machine Learning, Neural Network, Natural Language Processing, and Computer Vision.

National University of Kaohsiung, Kaohsiung, TW - B.Sc

Major in Applied Mathematics, solid trained in mathematical, statistical thinking, hands-on experience of Matlab, and R language.

PET PROJECTS

FastAPI implementation - Robots vs Dinos ([document](#))

A service that provides a REST API to support simulating an army of remote-controlled robots to fight the dinosaurs, supports asynchronous processing. Stacks: Python/FastAPI, GCP Cloud Build, GCP Cloud Run

Candidates' resume parser ([Link](#))

One of the featured projects @ digitty.io. Implemented customized Name Entity Recognition (NER) by spaCy. Recreated 11 classes of name entities, trained the model, wrapped into API, and deployed to the cloud. Stacks: Amazon S3, Dataturks, Heroku, Apache tika, Docker, Python/Flask, spaCy.