

PURNA SHAH

Tucson, AZ | (480) 330-1815 | pvs25@cornell.edu | github.com/pvsaz

Accomplished Data Engineer with proven record of writing scalable pipelines, managing production-level infrastructure, and developing data-intensive applications to drive efficiency and revenue.

EXPERIENCE

2U – Lantham, MD

Data Engineer (remote)

01/2023 – 01/2024

- Built and maintained Python-based pipelines (e.g. Airflow, Prefect) providing terabyte-level volumes of data to stakeholders through Snowflake/Redshift warehouses.
- Improved and maintained legacy codebases in dbt, Terraform, Argo/Kubernetes, and Jenkins.
- Engineered workflow infrastructure in AWS, configuring Lambdas, EC2 instances, Cloudwatch alerts, VPCs, and logging.
- Devised Javascript-based frontends for internal stakeholders to consume data.
- Solved high-acuity issues as part of regular on-call shifts for production workflows.

LATCH – New York, NY

Analytics Engineer (remote)

09/2021 – 06/2022

- Scripted in Python, Bash, dbt, and Snowflake to build ETL pipelines, model and orchestrate data, migrate frameworks, etc.
- Built cloud reporting in Looker, eliminating hundreds of hours of manual data analysis and manipulation for stakeholders across supply chain, operations, support, and finance.
- Designed internal Looker environment, driving reliability and adoption through collaboration with other analysts, data engineers, and data architects.
- Performed company-wide live trainings and created written reference materials for internal Looker environment, enabling business stakeholders to analyse data and draw insights on-demand.

GOOGLE – New York, NY

Data Analyst, Customer Support

11/2019 – 04/2021

- Scripted in Python, LookML, and SQL to support technical Looker users.
- Provided white-glove, consultative support to technical Looker users on topics including business analysis, data modelling, embedding, APIs, etc. driving client retention and product adoption.
- Served as “voice of the customer” in working cross-functionally with product and engineering teams to drive bug fixes to completion and advocate for feature requests.

PORTFOLIO

- Wrote Python-based ETL pipeline that batch extracts sample car data, stages in AWS S3, transforms to a star schema and inserts/updates rows to AWS RDS database (code [here](#))
- Contributed to Python-based open-source tool used to generate dbt models during employment at 2U: implemented handling of personally identifiable information (code [here](#)) and automatic prefixing of table names to prevent naming collisions (code [here](#))
- Created professional portfolio frontend, featuring custom CSS and animations, served by an AWS-deployed Docker container with CI/CD through Github Actions (website [here](#) and code [here](#))

SKILLS

Main languages - Python (advanced), SQL (advanced), Javascript (intermediate), C++ (intermediate)

Python - general scripting, OOP, data engineering, backend development (experienced in libraries including Django, Flask, SQLAlchemy, pytest, pandas, Airflow, Matplotlib, Boto3, etc.)

General programming - Git/Github, REST APIs, Postman, CI/CD, Unix/Linux, Bash, AWS (S3, EC2, Cloudwatch, Lambdas), GCP, Docker, Argo, Kubernetes, NoSQL (Elasticsearch/Redis), Jenkins, Jira, agile/scrum methodology

Data - dbt, Fivetran, SQL, NoSQL (Elasticsearch/Redis), REST APIs, Airflow, Prefect, Terraform, Argo, Kubernetes, Jenkins, Jupyter, Datadog

EDUCATION

CORNELL UNIVERSITY – Ithaca, NY

2010

B.S., Biology and Society | *magna cum laude*