

Armaan Thapar

athapar.2017@gmail.com | (914)-486-0735 | github.com/athapar | Mableton, GA 30126

Experience

Georgia-Pacific, Atlanta, GA

Data Engineer

05/2022-Present

- Lead cross-functional data engineering projects for electrical asset health, serving as the primary point of contact for engineers and business teams to ensure alignment with project requirements and deadlines
- Led automation of asset monitoring model creation using Python, reducing manual team effort by 66%
- Developed a data pipeline to automate ingestion, perform calculations using AWS Lambda, and forecast data for a chemical process model used in live monitoring across 4 facilities
- Developed a Python program to parse XML-based process flow diagram data—including node metadata and flow relationships—to generate automated pipelines for material tracking
- Write SQL queries to extract data, build AWS Redshift tables, and design PowerBI dashboards

Reckitt, Belle Meade, NJ

Analyst

01/2021-12/2021

- Analyzed spectroscopy data and conducted statistical error calculations using MS Excel
- Maintained and updated database of raw materials inventory

The Cooper Union, New York, NY

Laboratory Assistant

09/2019-05/2020

- Supervised and instructed students conducting small-scale chemical engineering experiments, including distillation, heat exchange, and filtration
- Maintained, calibrated, and troubleshoot lab instruments used in chemical engineering experiments

Diamond S Shipping (now International Seaways), Greenwich, CT

Data Analyst Intern

06/2016-08/2016

- Analyzed oil tanker data in MS Excel to analyze fuel consumption trends
- Built simulation model to recommend optimal speeds projected to reduce fuel consumption by 5-15%

Projects

Master's Thesis

Simulating Molecular Clusters in Python ([Full Paper](#)) ([GitHub Link](#))

01/2021-12/2021

- Designed and implemented object-oriented Python library to simulate and study pollutant molecules
- Used numpy, pandas, scipy, and matplotlib to store, analyze and visualize simulation data
- Evaluated overall cluster properties via statistics on individual molecules

Modeling Differential Equations

Reactor Modeling in Python ([GitHub Link](#))

06/2016-05/2016

- Created vectorized models using numpy to optimize system performance
- Solved problems posed as systems of differential equations using RK4 and Euler's method
- Plotted and visualized results using matplotlib and seaborn

Skills

Programming: Python (pandas, NumPy, Tensorflow, matplotlib, seaborn, plotly), SQL, VBA, C#

Data Engineering: AWS (Lambda, Redshift, SQS), Apache Spark, ETL

Visualization: Power BI

Machine Learning: scikit-learn, Tensorflow

Other Tools: Git, PI Historian/AF, Azure DevOps, MS Excel (VBA, Lookups, PivotTables)

Education

The Cooper Union, New York, NY

Master of Engineering, Chemical Engineering

Dec 2020

Bachelor of Engineering, Chemical Engineering

May 2018

Relevant Online Coursework:

The Complete Oracle SQL Certification Course (Udemy: [Certificate Link](#))

Jan 2022

Stanford University & DeepLearning.AI Machine Learning Specialization (Coursera: [Certificate Link](#))

May 2025