

ARPITA DESHMUKH

Dallas, TX | H1-B Visa

arpitapdtamu@gmail.com | [linkedin.com/in/arpitapd](https://www.linkedin.com/in/arpitapd) | <https://arpitapd.github.io/>

ABOUT ME

I'm a results-driven data engineer with 5 years of experience delivering end-to-end healthcare analytics data projects—from building automated ETL pipelines (Python, Alteryx), analyzing large datasets (SQL) to creating high-quality dashboards (PowerBI). I've also built cloud-based pipelines using AWS (S3, Glue, Athena, Redshift) in personal projects. I focus on streamlining processes, data validation, and deriving actionable insights.

SKILLS

Languages & Skills: SQL | Python | R | API Data Extraction | ETL pipelines | Git | Data Modeling

Tools: Alteryx | PowerBI (DAX, M query, data visualization & data modeling) | RDBMS | SQL Server | SSIS | SSRS | SSAS | AWS (S3, Glue, Redshift, AWS Athena)

Other skills: Neo4j | NoSQL (MongoDB) | Airflow

EXPERIENCE

Behavioral Health Group, LLC (*BI Data Engineer*) Dallas, TX Jan 2022 – Present

- Collaborated with executives to capture requirements, identify relevant **KPIs** & trends, leading to data-driven decisions across for 120+ global locations by delivering end-to-end data solutions from ETL to reporting
- Streamlined **Call Center JSON** data extraction via **APIs** using **Python**, making process **100%** automated
- Automated **API** data extraction, transformations in **Python**, backing up and uploading report to HR system via SMTP using **Alteryx** and **batch** scripting, eliminating manual efforts completely
- Created & optimized **SQL** queries using complex joins and window functions, improving data retrieval by **15%**
- Implemented Row Level Security (**RLS**) in **PowerBI**, enabling tailored data access based on employee location and job title, providing personalized, role specific insights for different user groups

Behavioral Health Group, LLC (*Revenue Data Analyst*) Dallas, TX July 2020 – Jan 2022

- Expedited **ETL process** by replacing normal functionality to In-DB in Alteryx, reducing execution time by **80%**
- **Patient No Show Project** – Used **SQL** (CTEs, window functions, joins) to analyze **40M+ Patient records**, built report to track missed doses and follow-up calls; reduced patient No shows by **10%**

Strategic Materials, Inc (*Data & Business Intelligence Analyst Intern*) Houston, TX May 2019 – Aug 2019

- Conducted requirements elicitation from C-suites, and stakeholders for creating interactive dashboards
- **Tracking on-Time shipments** - Created BI dashboard to compare on-time shipments percentage across global facilities; incorporated email alert system if on-time percentage falls below 70% for the current month.

TECHNICAL PROJECTS

YouTube Data Analysis AWS S3 | Glue | Athena | Lambda | IAM | Quicksight | ETL | SQL

- Built serverless ETL pipeline using AWS services to analyze YouTube data, automating ingestion of JSON and CSV data files, transformations, querying in Athena and visualized in QuickSight

Predict Presence of Heart Disease in patients R Language

- Built and tuned models (Logistic Regression, KNN, Random Forest, SVM) in R Studio using UCI Heart Disease dataset, achieving 87.4% accuracy and 12.5% FNR with optimized Random Forest and KNN models

EDUCATION

• **Texas A&M University** (*MS in Management Information Systems*)

College Station, TX Aug 2018 - May 2020

• **University of Pune** (*Bachelor's in Information Technology*)

Pune, India Aug 2014 - May 2018