

Pooja Patel

4086909923 | poojapatel.3010@gmail.com | [Linkedin](#) | [GitHub](#)

SKILLS

Languages / Frameworks: Python, Java, Flask, Django, Apache Airflow, Spark, Kafka, REST APIs, SQL, GraphQL, ML, Snowflake

Databases: MySQL, Postgresql, MongoDB, Oracle

Cloud: AWS (S3, Redshift, Athena, Glue, Crawler, SNS, RDS, Lambda, CloudFormation, EC2, CloudWatch, IAM, Secrets Management), GCP

Tools: Tableau, QlikSense, QlikView, JIRA, Jenkins, Git, Google Analytics, Google Data Studio / Looker

Other: Data Warehouse, Data visualization, Business Intelligence, Data Modeling, ETL (Extract, Transformation, Loading)

PROFESSIONAL EXPERIENCE

Student Assistant, San Jose State University, San Jose, CA

Aug 2023 - May 2024

- Designed mail tracking system using **SQL** and **Snowflake**, applying external stage integration with **S3**, simplifying package monitoring on **Tableau** for staff of 10.
- Developed pdf based chatbot using **Langchain** and **OpenAI LLM API**, to provide instant answers to student queries at the department.
- Launched **MERN (MongoDB, Express, React, Node)** stack project management system for 7+ faculties and 50+ students, streamlining progress tracking.

Senior Software Engineer, Infosys Ltd., Pune, India

Dec 2020 - Jun 2022

- Spearheaded **ETL data pipeline** development using **AWS S3, Athena, Python Glue** jobs for batch processing and **data warehousing** in **Redshift**, to enable automated approvals for Pricing tool, preventing Gross Profit Margin Leakage of **3M USD** per year.
- Authored **SQL** scripts to perform **dimensional data modeling** leveraging data SerDe in Parquet, optimized and fine-tuned ETL data flows and SQL queries by using temp/staging tables reducing runtime by 45%.
- Created **SQL views** to provide tailored monthly rolling 5-year sales data involving **6M+** records for cross-functional teams.
- Initiated event based alerting solution with **Lambda** and **SNS** notifications in **Python**, and enabled CloudWatch logging to notify production support team, facilitating error resolution within 30 minutes.
- Engineered **data visualizations** in **QlikSense, Tableau dashboards** for digital campaign, ecommerce involving **AB tests, Salesforce SFMC**, for 6+ countries, achieving cost savings of **120K USD** per year per country, resulting in 40% increase in campaign efficiency.
- Improved infrastructure provision by writing architecture scripts using **CloudFormation YAML**, and executed deployments through **GitLab** and **Jenkins CI/CD**, boosting delivery time by 50%.

Software Engineer, Infosys Ltd., Pune, India

Aug 2019 - Nov 2020

- Led a team of 3, orchestrated data pipelines using **Airflow** on **EC2**, built Flask **REST APIs** to prepare data sources for marketing, sales teams, saving 6+ hours of manual work per week.
- Conducted Segmentation and Churn analysis on customer accounts using **Logistic Regression** and **K-Means Data Mining** and **Machine Learning** techniques utilizing **Jupyter, Pandas, Scikit-learn** to determine strategies to retain high risk customers.
- Designed Postgres cumulative tables for **time series** analysis over extensive time frames of over 1000+ days derived from daily metrics.
- Developed **Python Flask API** for pest detection using **GCP AutoML Vision** to integrate in an android app for growers.
- Collaborated and developed custom data pipelines utilizing Selenium, BeautifulSoup, **REST** and **GraphQL APIs** for Google Ads, YouTube and Facebook data; cut down reliance on paid data connectors by 70%.

EDUCATION

Master of Science, Computer Software Engineering, San Jose State University, California

Aug 2022 - May 2024

Bachelor of Technology in Computer Engineering, NMIMS Mumbai, India

Jul 2015 - Apr 2019

PROJECTS

Real-Time Voting System with Kafka

- Architected voting system with Python, **Kafka, PySpark** for **Stream Processing**, integrated with Postgres for data storage and retrieval.
- Devised functionalities encompassing **Kafka topic** creation, data enrichment, **Spark** aggregations, and **Postgres** DB management orchestrated via **Docker**, facilitating handling of voting data streams for analysis and display through **Streamlit**.

CreditCard Parser

- Devised **Java** application using **OOPs** concepts in parsing diverse file formats containing credit card information, enabling precise authentication based on card types such as AmEx, Visa, etc., employing **Factory** and **Singleton design patterns**.
- Implemented automated test cases to ensure the application's reliability using **JUnit**.

Emergency News Delivery Application

- Composed **data extraction** and **MongoDB** load scripts, developed **Flask APIs** with throughput 92.98 RPM for emergency events. Established **Firebase cloud messaging** for **real-time** notifications and deployed this solution on **EC2**.

Airport Management System (AMS)

- Conceptualized scalable backend for AMS leveraging Python's **Django** Rest Framework and ReactJS in a team of 4.
- Enhanced low/high-level design, and database architecture; containerized app with **Docker** increasing delivery time by 40%.