Munich, Germany
[linkedin.com/in/vaidyasejal](https://linkedin.com/in/vaidyasejal)

# SEJAL VAIDYA

[github.com/sejalv](https://github.com/sejalv)
vaid.sej@gmail.com

## SUMMARY

- <u>Data/ML Engineer</u>, with 10 years of software development experience, and proficiency in Data Platforms, Infrastructure & Cloud architecture
- Recent experience with productionizing models, services, and workflows in ML & analytics
- Core skills: <u>Python, AWS, Serverless, Airflow, NoSQL, SQL, Spark, Flask, Docker</u>

## WORK EXPERIENCE

**Temedica GmbH** | Healthcare | Munich, Germany                                      *June 2020 – Present*
***Data Engineer***

- Designed a centralized Ingestion system & developed pipelines to track <u>patient journeys & clinical data</u> using <u>AWS (Lambda, S3, Batch, ECR), Python, Flask & Docker</u>
- Developed an Event Management System with Schema Registry for monitoring new tracking events from PostHog, using <u>AWS, Postgres, Docker, Kubernetes & Tableau</u>

**Fineway GmbH** | *Travel* | Munich, Germany                                          *Jan 2020 – May 2020*
***Data Engineer***

- Developed services/APIs for the Conversation engine to generate faster Trip Plans, redesigned them with horizontal scaling, and using <u>Python, MongoDB, Flask, Celery, Docker</u>
- Collaborated with Data Scientists for <u>model deployment</u> and versioning, orchestration of training jobs, as well as pipelines for analytics & monitoring using <u>Airflow, Python, PostgreSQL, Docker, Kubernetes</u>

**Freeletics GmbH** | Fitness | Munich, Germany                                        *Oct 2018 – Dec 2019*
***Data Engineer***

- Redesigned <u>data platform</u> to enable analytics on training journeys of 35+ mil users, with <u>event-driven microservice architecture in AWS</u>, and a fault-tolerant infrastructure (IaC) with Terraform, Jenkins & Travis
- Developed scalable data pipelines to asynchronously ingest & process 1k+ events per minute, using <u>Python, AWS (Lambda, Kinesis, DynamoDB, S3, Glue, Athena) & serverless</u> (framework)
- Contributed to requirements gathering for product/marketing analytics to align with long-term data strategy
- Conducted hands-on <u>workshop on building a serverless app</u> in Python & AWS for 20+ diverse engineers

**Servify** | Consumer Services | Mumbai, India                                        *Apr 2017 – Nov 2017*
***Senior Data Engineer***

- Enabled cross-functional <u>analytics for post-sales & service lifecycle</u>, by <u>designing a centralized Data Lake</u>
- Led team to build distributed pipelines to ingest & process streams using <u>Kafka, Spark, HDFS/Hive,  Tableau</u>
- Created a <u>demand forecasting engine</u> for inventory management, to predict SKU usage and reorder parts on crossing MOQ. Implemented ML models (EWMA, scikit-learn, pandas, numpy) with Python & Flask APIs

**LexisNexis Risk Solutions** (Reed Elsevier/RELX) | Insurance | Mumbai, India      *Apr 2015 – Sep 2016*
***Senior Software Engineer***

- Developed data products to enable <u>Fraud Analytics</u> & Risk Management in <u>Life & Health Insurance</u> space
- Designed and developed a <u>search engine with Information Retrieval</u> for Identity Linkage of customers
- Facilitated <u>Anomaly Detection</u> to identify outliers within claims. Developed services for Named Entity Recognition, risk profiling, and user segmentation
- *On-Site Visit (Atlanta, USA)*: Collaborated on architectural decisions with business teams to develop POCs

\*\*\*

## WORK EXPERIENCE (cont.)

**JP Morgan Chase Services** | Finance | Mumbai, India                    *Feb 2013 – Mar 2015*
  *Applications Developer*
  - Designed logical data models & BI reporting mechanisms for <u>Finance, Credit Risk & Marketing Analytics</u>
  - Developed ETLs for Incremental & Historical data, using PL/SQL & Shell scripts, and utilities for auditing
  - Redesigned DWH schemas with dimensional modelling, optimised query & database performance in Oracle

**HERE Technologies** (Formerly NAVTEQ) | GIS | Mumbai, India            *Dec 2009 – Feb 2013*
  *Geographic Technician*
  - Data management & Quality Analysis for Maps. Optimized SQL queries to improve data retrieval by 37%
  - <u>Golden Impulse Award, Sep '12</u>: Developed test automation framework to reduce quality overheads by 45%

---

## EDUCATION

**University of Mumbai**, India | 2006 –2009 | Bachelor of Science – Information Technology (IT) | First Class

---

## INDEPENDENT PROJECTS

- [Containerized ML deployment with AWS Lambda](#) using AWS Lambda, Docker & ECR
    - Workshop for [DataTalks.Club](#): [Serverless Machine Learning Deployment with AWS](#)
    - Repo: [github.com/sejalv/serverless-ml-workshop](#)
    - Medium: [sejalv.medium.com/containerized-ml-deployment-with-aws-lambda](#)
- [Data Engineering Confessions](#): Podcast interview for [Pipeline Data Engineering Academy](#)
- [SQL in a Nutshell](#): Article published in [Towards Data Science](#) (Medium)
- [Foundational projects for Big Data / ML with PySpark](#)
- [Movie Recommendations Website](#): with CF & K-Means using Python, PostgreSQL, scikit-learn

◇◇◇◇◇◇