

NITISH GOVIND BELAGALI

Boston, MA

(857)869-3085 | nitish.belagali@gmail.com | github.com/nitishbelagali/nitishbelagali | linkedin.com/in/nitish-belagali-392646158/

Data Analytics Scientist with 3.5+ years of experience in machine learning, BI, and cloud pipelines. Skilled in Python, SQL, and GenAI, delivering analytics solutions that enhance customer experience and drive business impact

EDUCATION

- Northeastern University** : *MS in Information Systems* | Boston, USA | GPA: **3.7 / 4.0** Sep 2023 - Dec 2025
- Courses: Data Warehousing and Business Intelligence, Gen AI and Prompt Engineering, Data science, ML
- KLE Technological University** : *BE in Computer Science* | Hubli, India Aug 2016 - Jun 2020

SKILLS

Scripting Languages	Python (NumPy, Pandas, Matplotlib), Spark, Java, R
Business Intelligence	Tableau, PowerBI, MS Excel, HDFS, Alteryx, Airflow, Data Storytelling
Statistics	Regression, Clustering, Classification, t-test, Chi-square, AB Testing, time series analysis
Database systems	PostgreSQL ,Oracle , SparkSQL, MySQL, NoSQL, Redshift
Cloud and Other tools	BigQuery, Azure Data Factory, GitHub, Jupyter , JIRA, Confluence

EXPERIENCE

- Boehringer Ingelheim**, *IT Data Analyst* | Danbury, CT, USA Jan 2025 – Jul 2025
- Automated weekly export of 50M+ rows from Oracle into distributed data pipelines using Python and SQL, integrating outputs into Amazon Redshift and Power BI/Tableau; improved data integrity, reduced manual effort by 90%, and ensured faster, reliable access to high-volume datasets for analytics
 - Delivered stakeholder-approved PowerBI dashboards with dynamic filters, automated refresh, and tailored insights for scientists and regulatory teams; accelerated reporting cycles by 70%, reduced manual review by 30%, and improved KPI tracking for preclinical studies
 - Modularized a complex R Shiny application by separating UI, server, and scoring logic into reusable components; built a scoring engine across 5+ biological domains using z-score and incidence logic, enhancing scalability, precision, and cross-study comparability for scientific decision-making
 - Performed exploratory and statistical analysis on toxicology datasets, uncovering trends and anomalies that improved review accuracy by 25% and reduced rework in regulatory submissions by 15%; presented insights through executive summaries and visualizations for leadership
- IBM**, *Technical Services Specialist* | Bengaluru, India Aug 2021 – Jun 2023
- Optimized ETL scripts and workflows for a major energy utility client using Informatica and automated pipelines with Apache Airflow, achieving a 45% performance boost and 40% reduction in processing time, which saved the team 100+ hours weekly. Ensured delivery excellence by tracking tasks in JIRA and maintaining documentation in Confluence
 - Designed and implemented real-time data collection strategies using Google BigQuery and SQL, improving data organization and accessibility. This reduced data errors by 50%, increased client acquisition by 18%, and streamlined operational efficiency across the utility's data ecosystem
 - Supported the adoption of Agile methodologies, collaborating with cross-functional teams to cut project timelines by 15%. Delivered Tableau dashboards and actionable insights that improved application quality, enhanced user experience, and accelerated delivery of business-critical features
- Accenture**, *Associate Software Engineer* | Bengaluru, India Feb 2021 - Aug 2021
- Implemented Tableau dashboards to analyze client-facing data, driving a 25% increase in customer satisfaction and a 30% improvement in decision-making accuracy for a biomedical client
 - Utilized advanced data analysis methods within Tableau visualizations to uncover critical trends and patterns, resulting in a notable 20% surge in client revenue and a substantial 35% reduction in operational costs

PROJECTS

- Spotify data pipeline** Personal Project | Boston, USA Jun 2025 - Jul 2025
- Built a distributed ETL pipeline using Spark to process Spotify streaming data from HDFS into AWS (S3, EMR, Redshift), improving pipeline efficiency by 30% and reducing query latency by 50%
 - Architected scalable AWS infrastructure (S3, EC2, EMR) and applied Spark transformations, reducing query latency by 50%
- GenAI RAG Chatbot for Policies & FAQs** Northeastern University | Boston, USA Nov 2024 - Dec 2024
- Developed retrieval-augmented chatbot using FastAPI, Chroma, Streamlit with accurate query and citation retrieval
 - Dockerized and deployed the GenAI pipeline achieving production-ready workflow with evaluation metrics on accuracy and latency
- End-to-End BI Project** | Northeastern University | Boston, USA Jan 2024 – Feb 2024
- Standardized and integrated 1M+ multi-format accident datasets (Austin, Chicago, NYC) using Alteryx, ER Studio, and Apache Airflow; designed a Star Schema data model in Azure SQL Data Warehouse to support efficient reporting and analytics.
 - Automated ETL load into Azure SQL and developed 5+ interactive dashboards in Power BI & Tableau with CI/CD pipelines in Azure DevOps, reducing reporting effort and improving data accessibility.

HIGHLIGHTS

- Served as a Teaching Assistant for Fall 2024 in Application Engineering and Design, overseeing and supporting a class of approximately 200 students
- Independent research on AI-augmented BI pipelines with real-world e-commerce data, focusing on KPIs, forecasting, and automated insights
- Familiarity with distributed storage concepts and pipeline optimization (HDFS, Spark, SSD performance analytics)