

# Krishna Apurva

(469)-880-8745 | krishnaapurva469@gmail.com | <https://www.linkedin.com/in/krishnaapurva469/> | <https://github.com/krishnaapurva>

## EDUCATION

### Master of Science, Business Analytics and Artificial Intelligence

May 2025

The University of Texas at Dallas

GPA: 3.9/4.0

Coursework – AWS Cloud Analytics, Big Data, Predictive Analytics, Machine Learning, Advanced Statistics

### Bachelor of Technology, Information Technology

May 2020

Medicaps University

GPA: 3.8/4.0

## SKILLS

**Languages:** Python, R, SQL, PySpark, Linux

**Databases:** PostgreSQL, MySQL, MongoDB, NoSQL, Teradata, Azure Data Lake Storage (ADLS)

**ETL:** Apache Airflow, Talend, Data Build Tool (dbt), Azure (Data Factory, Synapse Analytics)

**ML Algorithms:** Regression, Classification, Clustering, Neural network, ANN

**Data Analytics:** Tableau, Power BI, MS Excel, Statistical Analysis, A/B Testing

**DevOps Tools:** Docker, Kubernetes, Jenkins, Git

**Certifications:** AWS Academy Graduate Data Engineer

## PROFESSIONAL EXPERIENCE

### Data Engineer | Innova Solutions

May 2021 - June 2023

- Optimized ETL pipeline efficiency by implementing a CTE-based stored procedure that filtered zero-payment claims pre-staging, cutting processing time by 37.5% per run
- Engineered ETL workflows in Talend with parallel processing, optimizing resource utilization to handle peak loads of 40M records per run, reducing processing time per million records from 8 minutes to 2.5 minutes
- Automated processes using PL/SQL procedures to streamline daily tasks such as data backups, indexing, and performance monitoring, resulting in an 18% improvement in overall operational efficiency within the database
- Integrated lookup validation framework in Talend, enforcing data consistency by cross-referencing predefined reference tables, preventing invalid records, and reducing data discrepancies
- Containerized the pipeline with Docker, improving deployment efficiency, ensuring consistency across virtual environments, and enabling seamless scalability to handle future data growth

### Data Engineer | JJ Enterprises

June 2020 - May 2021

- Designed Tableau dashboards to monitor KPIs across 700K+ sales transactions, refining inventory planning and pricing strategy, driving a 1.2% rise in business revenue
- Applied K-means clustering to segment 6k+ customers based on purchase patterns, leading to a 12% growth in repeat purchases through personalized marketing
- Collaborated with sales and marketing teams to analyze customer feedback and conducted data-driven market research to introduce new products, driving an \$11,000 increase in quarterly sales
- Built an Airflow ETL pipeline to transform and load data into a Data warehouse, creating structured views that provide fast access to 24 months of historical data, accelerating business decision-making

### ML Engineer Intern | Medicaps University

June 2019 - August 2019

- Leveraged OpenCV for accurate identification and tracking of individuals and moving objects, enabling real-time monitoring of suspicious activities with 75% accuracy

## PROJECTS

### Olist E-commerce Azure Data Pipeline (ADF, ADLS, Databricks, Synapse Analytics, MongoDB, MySQL)

- Designed and implemented a scalable Azure data pipeline leveraging Medallion architecture to efficiently ingest, transform, and analyze 1M+ records, optimized data storage with ADLS, enabling real-time insights through Azure Synapse and Tableau

### MediPrice Insight (PySpark, PowerBI, Flask, HTML, PostgreSQL)

- Processed and analyzed large-scale hospital price transparency data based on CPT codes using PySpark, providing actionable insights on healthcare costs and enabling users to make cost-effective decisions through dashboards

## HONORS AND ACHIEVEMENTS

**Dean's Meritorious Scholarship:** Awarded for exceptional academic excellence and leadership in management

**High Distinction Scholar:** Graduating with outstanding performance throughout the master's program

**Top 1% of 500 groups:** Finalist in the national-level Smart India Hackathon