

Priyanka Gopi

(321) 444-2758 priyankagopi015@gmail.com
[Portfolio](#) github.com/priyankagopi [in/priyankagopi](https://in.priyankagopi)

SUMMARY

Experienced data engineer with three years of expertise in programming, data modeling, ETL, big data technologies, and cloud platforms, coupled with strong problem-solving and collaborative communication skills. Proven track record in building and maintaining data pipelines and infrastructure.

EDUCATION

Master of Science, Computer Science

University of Central Florida, Orlando, FL

May 2024

GPA: 3.8/4.0

Coursework: Data Structures & Algorithms, Data Mining, Machine Learning, Software Engineering, Natural Language Processing

SKILLS

Languages: Python, SQL, Shell Script, R

Cloud: Amazon Web Services (AWS), Microsoft Azure, Informatica

Database: Oracle, MySQL, Microsoft SQL Server, Azure SQL Database, Sybase, IBM DB2, Snowflake

Data Science: Business Intelligence, Data Wrangling, Data Visualization, Quantitative Analysis, Statistical Modeling, Data Modeling

Data Engineering: Data Integration, Extract, Transform and Load (ETL), Data Warehousing, Hadoop (HDFS), Azure Databricks

Software and Tools: Hadoop, Power BI, UNIX, Postman, Azure Data Factory, Tableau, Airflow, MS Excel, JIRA, Docker, Git, Jenkins

EXPERIENCE

Graduate Research Assistant, University of Central Florida

Aug 2023 - Present

- Investigated *students' behavioral factors* impacting exam performance with statistical tests and quantitative analysis.
- Extracted and **analyze data on over 600k student grades**, executing 2M+ queries resulting in a more *informed decision-making process* for study material selection.
- Identified key factors* from USA CDC Healthcare datasets that influence dental health outcomes, resulting in targeted interventions achieving a **15% reduction in dental health disparities** among underserved populations.
- Mentor 75+ students** and *designed assignments, quizzes and projects* on Visio, SQL, and MongoDB for the course Healthcare Database Management (HIM6217).

Data Engineering Intern, Coding it Forward

Jun 2023 - Aug 2023

- Leveraged agile methodologies* to collaborate with cross-functional teams at *City of Charlotte* to implement robust data models and adopt best practices in test design, data validation, reporting and analysis.
- Developed scripts to automate repetitive tasks leading to **20% reduction in workload**.
- Implemented quantitative analysis* to derive comprehensive predictions based on large extensive historical and current data.
- Streamlined data cleaning and transfer process by developing automated scripts, **reducing manual effort by 40%**.
- Deployed data pipelines in *Airflow* and *built a scalable database on AWS S3*, resulting in a **40% increase in operational output**.
- Demonstrated the ability to handle high-pressure situations by efficiently remediating escalations and *supporting critical business applications*, consistently meeting strict SLAs.

Senior Analyst, Capgemini

Aug 2019 - Aug 2022

- Participated in end-to-end pipelines for a Fortune 100 banking client, encompassing requirement gathering, *redesigning 700+ data models*, implementation of ETL processes and integration with *DevOps CI/CD* tools, causing a **20% rise in efficiency**.
- Developed a data framework utilizing *SQL stored procedures*, seamlessly orchestrating, and managing Data Pipelines in *Autosys*, resulting in a remarkable **60% reduction in manual workload**.
- Crafted *Perl scripts with SQL* for data extraction from *IBM Db2*, mapping 300+ unstructured data to *Amazon RDS* with *Athena*, **enhancing query performance by 50%**.
- Improved the ETL pipeline by adding various *data governance* checks to make the pipeline more resilient.
- Built test plans to *streamline complex data queries*, monitor database hygiene, and manage risk in *QA and production environments*.
- Received a **Rising Star Award** for "quick-learning abilities & subject matter expertise."

PROJECTS

- YouTube Data Analysis Using AWS:** Constructed an ETL pipeline for YouTube data analysis, seamlessly transitioning to *AWS infrastructure* and creating insightful data visualizations. **Tech** - AWS S3, Lambda, Athena, Quick Sight
- Extractive and Abstractive Text Summarizer:** Trained 4 ML models to generate summaries yielding a *ROUGE* score ranging from 0.1-0.5 when comparing against the original summaries for CNN Daily news articles. **Tech** - Scikit-Learn, Pandas, TensorFlow

PUBLICATIONS AND CERTIFICATIONS

- Publishing a paper titled '**A Comparison of Supervised Abstractive and Extractive Text Summarization**' in IEEE Xplore ([Link](#))
- Oracle SQL Database Certified Associate** ([Link](#))
- IBM Data Science Professional Certificate** ([Link](#))