

YUVA KRISHNA THANNERU

+1 (945)-208-0433 | t.yuvakrishna@gmail.com | [LinkedIn](#) | [Github](#) | [Portfolio](#)

SUMMARY

Passionate Data Engineer with over **3+** years of experience in designing scalable data pipelines, optimizing ETL workflows, and managing large datasets on cloud platforms such as AWS and GCP. Certified in Snowflake, proficient in SQL, Python, and Spark, with hands-on expertise. Adept at ensuring data quality, automation, and creating insightful visualizations using Power BI and Tableau to drive business decisions.

TOOLS & TECHNOLOGIES

Languages: SQL, SnowSQL, Python, C/C++, HTML/CSS, Java

Cloud Technologies: Google Cloud Platform, AWS, Microsoft Azure

Databases: PostgreSQL, MySQL, MS SQL server, Hadoop, MongoDB

Tools: MS Excel Advanced, Tableau, Power BI, Informatica, Snowflake, Alteryx, SAS

Certifications: Google Data Analytics Professional Certificate, Image Processing of Satellite Data, Data Analysis and Visualization with Power BI, Snowflake Snowpro Core Certification

EXPERIENCE

Data Analyst | Northern Illinois University

Aug 2023 - Present | Chicago, IL

- Designed and implemented a real-time data pipeline utilizing **AWS** to collect mental health-related data from Reddit.
- Implemented AWS Lambda Functions to automate real-time data collection from Reddit using API, processing approximately 10,000+ posts per day.
- Integrated real-time data processing with AWS Lambda and **Apache Airflow** to automate data collection, preprocessing, and model evaluation, maintaining up-to-date insights on mental health trends.
- Analyzed user behavior and language patterns using NLP libraries like **spaCy** and **NLTK**, applying techniques such as sentiment analysis and topic modeling to identify early signs of mental health concerns in social media users.
- Developed **Neural Network Models** using **Tensorflow** and Crafted Interactive Visualizations using **Power BI** to present mental health trends and insights, enhancing data interpretation by 36% through real-time updates.

Data Engineer | Wipro, Client: Citi Bank

June 2020 - June 2022 | India

- Designed and implemented scalable data warehousing solutions, managing over **5TB** of financial data for Citi Bank, improving data retrieval speed by 40%.
- Optimized SQL queries to accommodate **ad hoc** business requests, enhancing query performance by 36%.
- Analyzed over 1 million rows of financial data using SQL and Python to generate actionable insights and enhance decision-making processes. Proficient in designing ETL processes using Informatica to ensure smooth data integration.
- Conducted detailed financial analysis of 500k+ transactions in an agile environment to identify and rectify discrepancies and anomalies, reducing payment-related defects by 15%.
- Performed **A/B testing** on financial workflows, reducing transactional errors by 18% and increasing transaction throughput by 12%, enhancing overall process efficiency.
- Executed ETL process to extract client data from various sources, transformed it for analysis, ingested into **Snowflake**.

PROJECTS

YouTube Data Analysis | AWS, Python, ETL (Extract, Transform, Load) Pipeline

April 2024

- Designed and implemented a comprehensive data engineering pipeline for YouTube ad campaigns: Ingested 500,000 video records into Amazon S3, performed ETL processes with AWS Glue, and used AWS Lambda for data normalization. Reduced query times by 70% with AWS Athena. Developed an interactive dashboard in QuickSight to visualize YouTube metrics, significantly enhancing ad campaign effectiveness.

Research paper Search Engine | NLP Techniques, Machine Learning

Jan 2024 - Mar 2024

- Developed a search engine for 1500+ research papers, boosting search efficiency by 30% using NLP techniques (Stemming, Tokenization, TF-IDF), web scraping, and an Inverse Indexer. Leveraged scikit-learn and scikit-multilearn for text processing, and Streamlit for the interface. Enhanced user engagement by 20% with interactive Tableau dashboards, improving navigation and analysis.

E-commerce Data Analysis | Python, faker, K-Means, PCA, Machine Learning

Nov 2023 - Dec 2023

- Generated a realistic dataset of 5,000 transactions with 100 customers and 150 products for robust e-commerce analysis. Applied the Apriori algorithm to uncover product associations for targeted marketing, and implemented K-Means clustering for customer segmentation, enabling personalized recommendations.

EDUCATION

Masters in Computer Science, Merit Scholar

Jan 2023 - May 2024

Northern Illinois University, DeKalb, IL

CGPA - 3.71/4

Bachelor of Engineering in Electronics and Communication

Aug 2017 - May 2021

Sathyabama University, Chennai, India

CGPA - 8.86/10