PRANAY BHAKTHULA

Arlington, VA 22203 | bhakthulapranay@gmail.com | (202) 469-2775 | Github | Linkedin

EDUCATION

The George Washington University — Washington DC
Master's in Data Science GPA: 3.95 — Jan 2021 - Dec 2022
Sathyabama Institute of Science and Technology — Chennai, India
Bachelor's in Electronics and Communication Engineering GPA: 8.66 — Aug 2015 - May 2019

TECHNICAL SKILLS

Programming languages: Python, PySpark, SQL, R, Scala
Machine Learning and Data Analysis Tools: Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn, TensorFlow, PyTorch, Excel
Business Intelligence Tools: Tableau, Power BI, Looker, Google Data Studio, SAS Visual Analytics, Qlik
Data Engineering Tools: Apache Kafka, Apache Flink, Jenkins, Docker, Kubernetes, Cloudera, Snowflake
Database Management: MySQL, NoSQL (MongoDB), Big Data (Hive, Hadoop, Spark)
BI and Statistical tools: Excel, Tableau, SAS, SPSS, Microsoft Power BI, Looker Studio
Developer Tools: Git, Jupyter, RStudio, BitBucket, GitLab, Terraform, Postman, AWS Data Pipeline, Databricks
Cloud Services: Google Cloud Platform (GCP), Amazon Web Services (AWS), Microsoft Azure, Databricks

WORK EXEPRIENCE

Data Engineer | **RPA Technology** — Feb 2023 – Present
- Automated end-to-end ETL and data analysis pipeline using Cloud Data Fusion, GCS, Databricks, BigQuery, and Looker Studio, reducing processing time by 40% and improving data accuracy by 30%.
- Improved Spark application performance by 25% through optimization of batch interval time, parallelism levels, and memory tuning, leading to a 35% reduction in runtime.
- Developed six insightful dashboards in Looker Studio, enabling stakeholders to track key metrics, resulting in a 50% improvement in decision-making efficiency.

Solutions Architect | **Amazon Web Services (AWS)** — May 2022 – Aug 2022
- Led the migration of 15 on-premise SQL databases to AWS using AWS DMS, Amazon RDS, and AWS Glue, resulting in a 20% reduction in operational costs and a 40% improvement in data accessibility and reliability.
- Reviewed and optimized over 50 SQL queries and database components, including stored procedures, functions, indexes, views, and triggers. This comprehensive optimization effort led to a 35% improvement
- Developed 10 AWS QuickSight dashboards, enhancing reporting capabilities and increasing actionable insights by 30%.

Data Analytics–Research Assistant | **George Washington School of Public Health** — Nov 2021 – Dec 2022
- Conducted quantitative analysis on 221K patient records, revealing critical insights into disease trends, correlation, and patterns, resulting in a 15% improvement in understanding of epidemic causation.
- Designed and implemented an ETL pipeline using efficient SQL queries, increasing data quality by 30%. Utilized predictive modeling to identify high-risk patients with a 72% accuracy rate using Random Forest and XGBoost classifiers.

Data Analyst |**Centre for Rural Studies and Development** — Jun 2019 – Jul 2021
- Led a team to develop and execute 100+ complex SQL queries analyzing government data, resulting in 25+ data-driven reports and facilitating a 20% increase in policy advocacy effectiveness.
- Fostered a 30% improvement in decision-making processes by collaborating with non-technical stakeholders on KPI development and presenting insights via Tableau, leading to more informed policy decisions.

PROJECT WORK

Uber Data Analytics *GCP data engineering* — Jan 2024 – Feb 2024
- Conducted comprehensive data analytics on Uber data using GCP Storage, Python, Compute Instance, Mage Data Pipeline Tool, BigQuery, and Looker Studio.
- Enabled data-driven decision-making, leading to a 20% increase in operational efficiency.

Covid-19 Analysis *SQL & Tableau* — May 2023 – Jun 2023
- Analyzed a 313k records of covid 19 dataset using SQL queries to assess the global impact of the pandemic, examining mortality rates, infection rates, vaccination coverage, hospitalization trends and economic indicators.
- Published two interactive Tableau dashboards to effectively communicate key insights from the analysis.

Loan Prediction *Python,SQL,Data Mining* — Nov 2022 - Dec 2022
- Predicted loan status with an accuracy of 82% using Random Forest classifier, Naïve Bayes, KNN, Logistic Regression, XGBoost classifier models, with XGBoost classifier having highest accuracy.
- Enhanced the model's efficiency by 20% through data pre-processing techniques including handling missing data, deduplication, class balancing using SMOTE, Exploratory Data Analysis (EDA), and outlier removal.