

SURAJ NAG

 [LinkedIn](#) |  +91 9902734500 |  surajnag.com |  suraj11nag@gmail.com |  [GitHub](#)

Skills

- SQL | Python | Excel | Looker | Metabase | Tableau | Power BI | Google Data Studio | Probability | Statistics | PostgreSQL | MongoDB
- Big Query | Redshift | Snowflake | Databricks | dbt | Spark | Docker | Hevo | Airflow | Airbyte | CI/CD | GA 4 | AWS (S3, Lambda, EC2)
- Machine Learning | Deep Learning | Scikit-Learn | Pandas | Numpy | Tensorflow | NLP | Look ML | Amplitude | Mixpanel | Clevertap | Git

Experience

Data Engineer

TouchNote

Remote

02/2024 - Current

- Managed data pipelines using Airbyte and Hevo to **migrate and sync data** from PostgreSQL RDS to Amazon Redshift.
- Developed and maintained **20+ Airflow DAGs** to orchestrate data workflows, **run transformations, trigger ML model, generate reports.**
- **Implemented dbt from scratch** to build key **data marts, automate testing,** and enable **log tracking,** enhancing data transformation.
- Built and optimized **LookML code** for **data models and dashboards,** modernizing legacy structures and improving system performance.
- Monitored and optimized **AWS Redshift clusters,** analyzing **query performance and workloads** for efficient resource utilization.
- Revamped **transformation scripts** by **refactoring legacy code and removing obsolete tables** to improve efficiency and data accuracy.
- Performed **data audits to validate** and ensure the accuracy of **data sent to external tools** like Optimove and Amplitude.
- **Developed scripts** to support the **Finance team during audits,** validating data integrity and troubleshooting any discrepancies.
- Assisting **Finance, Operations, and Audit teams** with queries related to **Sales, Credit, and Finance ledgers** for accurate reporting.

Data Engineer

Open Financial Technologies

Bangalore

12/2022 - 02/2024

- Automated **reports and alerts** using **Looker & Metabase** to streamline data analysis and keep teams informed of real-time data changes.
- **Optimized SQL queries, reducing the runtime to 90%,** and improving the efficiency of data retrieval by using models in Metabase.
- Migrated large-scale datasets to **BigQuery using Google Cloud Platform,** enhancing data accessibility and expediting analysis.
- Developed **python scripts** for various tasks such as **data transformation and automation initiatives** for the product and finance teams.

Data Analyst

Mojocare

Bangalore

09/2021 - 12/2022

- **Automated tasks such as daily reports and data extraction,** resulting in significant time savings and increased efficiency.
- Improved data fetching efficiency by approximately **30% through query optimization techniques.**
- **Designed the entire event architecture for analytical platforms** like Mixpanel, Clevertap, and GA 4 for both website and App.
- Designed, documented, implemented, and monitored a complete marketing **data warehouse,** enabling data-driven decision-making.
- **Automated the manual data reconciliation process** between different sources, **increasing data accuracy and efficiency by 90%.**
- Implemented **data pipelines using Hevo,** an end-to-end data pipeline platform, and developed **Python scripts** for data transformation.

Data Analyst

Eloelo

Bangalore

12/2020 - 09/2021

- Developed and reported KPIs for a mobile application with **500K users using PostgreSQL, MongoDB, Excel, Google Sheets, & Python.**
- Built data and reporting infrastructure from scratch using SQL to provide insights into the **product, funnels, and business KPIs.**
- Utilized **Google Analytics and Facebook Analytics** to create dashboards for user funnel analysis.

Engineering Intern

Bosch Rexroth

Bangalore

07/2020 - 08/2020

- Performed analysis on various business channels such as consumer experience and reporting to identify growth strategies and actions.

Education

Bachelor of Engineering

RajaRajeswari College of Engineering

Bangalore

08/2016 - 09/2020

- Major in Mechanical Engineering
- CGPA: 7.17

Projects

- **Machine Learning:** Developed **machine learning models** for predicting heart disease based on medical attributes.
- **Deep Learning:** Created an end-to-end multi-class image classifier **deep learning model** using TensorFlow 2.0 and TensorFlow Hub.
- **Data Engineering:** Created a **forex data pipeline** to facilitate efficient data management and analysis for financial insights.
- **Visualization:** Utilized Kaggle datasets to build a local database, perform EDA using SQL, and build data visualizations using **Tableau.**
- **Python:** Proficient in web scraping, data analysis, and scripting using Python.
- **Web App:** **Spam Detector, Book Recommender** and **NLP Business Assistance** app, powered by machine learning and Streamlit.
- **AI:** Developed a **domain-specific chatbot** using LLMs and Langchain with external sources of computation and data.

Blog

- **Medium:** Writing data science articles on **Medium** to share insights and promote knowledge sharing within the community.