SHRUTI HOUJI

812.778.4724 • Louisville, KY, 40223 • sghouji@iu.edu • LinkedIn • GitHub • Portfolio

Data Analyst

EDUCATION

Indiana University, Bloomington
Master of Science – **Data Science** (GPA - 3.770)
Pune University
Bachelor of Engineering – **Computer Engineering** (GPA - 4.0)

SKILLS

Programming Languages: C++, Java, Python, R, DAX
Data Analysis Tools: Microsoft Power BI, MS Excel VBA, Tableau, Looker, Matplotlib, Seaborn, ggplot
Databases and Web Technologies: MySQL, NoSQL, PL/SQL, T-SQL, Neo4j, BigQuery, Snowflake, HTML5, JavaScript
Cloud Platforms: AWS (S3, EC2, Lambda, SageMaker), Google Cloud Platform (GCP), Microsoft Azure
Statistical Technologies/Tools: Regression, Hypothesis testing
Misc: Data Structures, Spark, PyTorch, Git, Flask, Pandas, NumPy, Scikit-learn, Hadoop MapReduce, APIs, Power Automate

WORK EXPERIENCE

Project 990 - *Bloomington, IN* 06/2024 – Present
Data Scientist

- Gathered insights and identified data quality issues from the Form 990 tax filing large dataset by doing **automated EDA** using python libraries like **Autoviz Widget** and **Ydata Profiling**, leading to a 25% reduction in data processing time.
- Created a robust text pipeline using **MySQL** and **Python** to **ETL** data from the database, enhancing data throughput by 20%.
- Key contributor to client deliverables through streamlined reporting, advanced analytics and dashboard creation, achieving a 30% efficiency enhancement with **Tableau**.

Danfoss Power Solutions - *Cleveland, OH* 05/2023 – 08/2023
Data Analyst Intern

- Interpreted complex datasets, performed exploratory data analysis, and statistical techniques to uncover trends, patterns, driving informed decision-making and strategic planning initiatives with a 5% increase in accuracy.
- Integrated diverse data sources into data warehouses, built **data pipelines** to **Extract, Transform, and Load** (ETL) data, and used **SQL** databases to enable data modeling and processing to streamline workflow, reducing data processing time by 10%.
- Provided automation solutions for production planning reports using **Python**, **R**, and **Excel VBA macros**, cutting creation time by 30 minutes.
- Crafted **KPIs**, 5 data visualizations, and productivity reports in **Power BI**, using **Python scripts**, and the **DAX** language, to deliver recurring business intelligence solutions, deliver insights and inform decisions.

Cognizant Technology Solutions - *Pune, India* 12/2019 – 05/2022
Programmer Analyst

- Utilized **Python**, **Power BI**, **HTML5**, and **JavaScript** to clean input files for data analytics and develop a web application.
- Collaborated with clients on **POC** feedback, enhancing communication skills and aligning deliverables with business strategies, resulting in a 10% increase in customer service satisfaction.
- Streamlined input feature decision-making by 5% using project management, critical-thinking skills, and team collaboration.
- Boosted business efficiency by 20% with Python **OCR scripts** and data solutions for financial data processing using **.NET**, **Oracle**, **SQL**, **Git**, **Azure**, and **agile methodologies** within the SDLC framework.
- Coordinated with a 10-person cross-functional team and stakeholders on **Jira** tasks, using organizational, problem-solving, and troubleshooting and interpersonal skills, presenting technical materials for imaging data and analysis with validation tests.

PROJECTS

Skin Lesion Classification for Melanoma Detection

- Spearheaded an end-to-end skin cancer detection system leveraging machine learning and Amazon web services like S3, SageMaker, Lambda, IAM, and API Gateway to streamline analysis, achieving 10% faster processing.
- Trained a neural network model with MONAI Pytorch and AWS cloud infrastructure to reduce resource utilization by 15%.

Datawage Navigator

- Structured a robust MySQL database design, allowing users to access job listings in specific locations within milliseconds.
- Integrated CRUD operations and dashboard page using Flask to get real-time updates with a 25% decrease in data latency.

Energy analytics and prediction in the USA

- Employed time series models such as ARIMA, VAR, and LSTM for forecasting optimal energy resources for the upcoming decade by applying data-driven recommendations.

Real-time Intrusion Detection Systems for IOT Networks using ML

- Built a dataset by self-generating an IoT node using the Tshark script to get instantaneous network logs of around 1 GB of data.
- Applied data mining techniques to develop an IDS using a Random Forest model with an accuracy of 99% and frontend notification to detect real-time attacks through websockets.