

## JOHANNES JOHNSON

Houston, TX | +1 (832) 682-8814 | [johannes.shelson12@gmail.com](mailto:johannes.shelson12@gmail.com)

<http://www.linkedin.com/in/johannes-johnson-882636257>

Portfolio | GitHub

Data Analyst with hands-on experience in SQL, Python, Tableau, Snowflake, AWS, and Big Data technologies. Early-career professional who has built end-to-end analytical solutions, including data cleaning, ETL pipelines, and interactive dashboards for datasets of 20,000+ records. Improved data consistency by 30% and developed classification and regression models using Scikit-learn to support research trend analysis.

### TECHNICAL SKILLS

**Programming:** Python, SQL, R

**Data Analysis:** Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn, SciPy, EDA, Statistical Analysis, Data Cleaning

**Visualization & BI:** Tableau, Power BI, Excel, Power Query, DAX

**Databases & Big Data:** PostgreSQL, MySQL, Snowflake, Hadoop (HDFS, MapReduce concepts), Apache Spark (PySpark)

**Cloud:** AWS (S3, EC2)

**Concepts:** ETL Pipelines, Data Modeling (Star Schema), Data Warehousing, Batch Processing

**Tools:** Git, Jupyter Notebook, VS Code, Google Colab

**Soft Skills:** Analytical Thinking, Problem Solving, Stakeholder Communication, Team Collaboration, Agile/Scrum Environment

### PROFESSIONAL EXPERIENCE

**University of Houston–Clear Lake - Houston, TX**

**Graduate Teaching Assistant** Jan 2026 – May 2026

- Assisted in teaching Registry & Internet Forensics and Cybersecurity courses for 60+ students.
- Reviewed and graded 100+ technical submissions, delivering structured feedback that enhanced student analytical accuracy and documentation clarity by ~20%

**University of Houston–Clear Lake (Neumann Library) - Houston, TX**

**Student Assistant, Circulation Desk** | Oct 2024 – Dec 2025

- Assisted 100+ patrons weekly and managed 200+ daily circulation transactions.
- Decreased average wait time by ~15% through efficient workflow coordination.

**University of Houston - Houston, TX**

**Data Science Intern** | Apr 2025 – Jun 2025

- Analyzed structured datasets containing 20,000+ records to identify research trends and patterns.
- Cleaned and transformed raw datasets using Python (Pandas), improving data consistency by ~30%
- Built and evaluated classification and regression models using Scikit-learn and performed exploratory data analysis.

**PwC India - Chennai, India**

**Technology Trainee (Software & Data)** | Jan 2023 – Feb 2024

- Completed a 12-month structured training program in SDLC, Agile/Scrum, and system design across 15+ technical modules.
- Enhanced SQL queries (joins, aggregations) to analyze 5,000+ row datasets for reporting simulations.
- Solved 10+ structured consulting case problems, translating business requirements into data-driven insights.

### PROJECTS

**Job Title Data Aggregator** - *Python, SQL, Pandas, NLP, PostgreSQL*

- Aggregated and dealt with 50,000+ job records from BLS, O\*NET, and online sources.
- Standardized 5,000+ inconsistent job titles using NLP preprocessing and fuzzy matching, improving dataset consistency by 35% and examined salary ranges, job demand, and skill trends across 200+ standardized roles.
- Lowered manual job comparison effort by ~50% using automated SQL aggregation queries.

**Pizza Sales Analytics Dashboard** - *SQL, Power BI, Power Query, DAX*

- Cleaned and scrutinized 12,000+ transactional records using SQL joins and aggregations.
- Built interactive KPI dashboards tracking Revenue, Order Volume, Category Performance, and Sales Trends.
- Reduced manual Excel reporting effort by ~40% through computerized BI dashboards.

**Car Wash & Repair Management System** - *PostgreSQL, Django, Bootstrap*

- Created database-driven web application managing 100+ monthly service transactions.
- Designed normalized relational schema to improve billing accuracy and lessened data redundancy.
- Systematized scheduling and invoicing workflows, improving operational efficiency by 50%

**Spectral Mineral Classification Project** - *Python, ENVI Software, GIS Tools*

- Classified 5+ mineral types using spectral reflectance modeling techniques.
- Boosted classification accuracy by ~20% through iterative preprocessing and validation.

### EDUCATION

- University of Houston–Clear Lake, Houston, TX**  
M.S. Computer Information Systems | GPA: 3.60 / 4.00
- Rajalakshmi Engineering College, Chennai, India**  
B.E. Electronics & Communication | GPA: 8.79 / 10.00