ANDREW JOHNSON

Mobile: (214)-457-3338 | Email: andrewjohnson9393@gmail.com | GitHub | LinkedIn | Portfolio | Cover letter

SUMMARY

Data Analyst with strong experience in Python, SQL, R, and data visualization, specializing in building end-to-end analytics solutions for finance, fintech, and government use cases. Proven ability to generate, clean, validate, and analyze large datasets to support fraud detection, risk analysis, audit compliance, and operational decision-making. Skilled in developing KPI-driven dashboards using Tableau, Power BI, and Excel (Power Query, pivot tables), and translating complex data into clear, executive-ready insights. Adept at data modeling, ETL workflows, anomaly detection, and ensuring data quality across analytical pipelines.

EDUCATION

B.S. in Information Technology – Software Engineering
Middle Georgia State University, Macon, GA
Relevant Courses: System Analysis, & Design, Software Engineering, Operating Systems, Database Principles

CERTIFICATIONS

Data Analytics Certification – Introduction to Data Analytics

Simplilearn | 2025

Skills covered: Data Analysis, SQL, Python, Data Visualization, Business Analytics Fundamentals

SKILLS

1. **Programming Languages:** Python, R, SQL, Java (Spring Boot), JavaScript, TypeScript, Django
2. **Python Libraries:** Pandas, NumPy, Matplotlib, openpyxl, psycopg2, scikit-learn
3. **Database Systems & Warehousing:** PostgreSQL, MongoDB, **Snowflake (data warehousing concepts & analytics workloads)**
4. **Data Modeling & ETL:** Star schema design, fact & dimension tables, ETL pipelines, data transformation, data validation
5. **Data Visualization & BI:** Tableau, Power BI
6. **Data Analysis Tools:** Advanced Excel (Power Query, Pivot Tables)
7. **Version Control:** Git
8. **Cloud & DevOps:** AWS (S3, EC2, Amplify, Elastic Beanstalk), Docker
9. **Other:** API & Webhook Development, Automation, SaaS Platforms, Agile (Scrum, Sprint Planning, Jira, Slack, Code Reviews), Security & Data Privacy Best Practices
10. **AI & Machine Learning:** Applied machine learning concepts, anomaly detection, feature engineering, predictive modeling, model evaluation, and data-driven decision support
11. **AI Tools & Automation:** AI-assisted analytics, intelligent reporting workflows, and automated data insights

EXPERIENCE

Full Stack Developer & Software Tester Internship
Next Play Nation | Apr 2025 – Jun 2025 | Web & Mobile Internship App |Internship recommendation letter

1. Developed and maintained React micro-frontends integrated with Node.js and Spring Boot APIs in a
2. microservice-style architecture.
3. Implemented RESTful and GraphQL APIs supporting data integrations and internal tooling.
4. Applied Test Driven Development (TDD) with Jest, React Testing Library, and Supertest to validate
5. features.

6. Optimized system performance and resolved technical issues, applying debugging and system
7. architecture troubleshooting skills.
8. Participated in code reviews and Agile ceremonies, contributing to a collaborative engineering culture.

PROJECTS

Finance & Fintech Data Analytics Project

1. Built an end-to-end finance and fintech analytics pipeline by generating, cleaning, and enriching **customer, account, transaction, and loan datasets** using **Python (Pandas, NumPy)**, **SQL schemas**, and **R (tidyverse)** to simulate real-world financial operations.
2. Designed **automated data workflows** to produce **analysis-ready datasets** exported to **CSV and Excel (.xlsx)** formats, enabling seamless integration with **Tableau, Power BI, and Excel** for executive-level reporting.
3. Analyzed **key financial performance and risk metrics**, including fraud detection, customer account balance aggregation, loan exposure, credit risk segmentation, and recovery rate analysis to support operational and strategic decision-making.
4. Modeled and structured datasets using **BI-optimized schemas**, validated KPIs, and reusable transformations, ensuring **scalability, consistency, and Tableau-ready reporting** across analytics tools.

Large-Scale Government Audit & Anomaly Detection System

1. Engineered an end-to-end analytics pipeline processing **500,000+ simulated government financial transactions** (procurement, payroll, benefits, inspections) using **Python, SQL, and R**.
2. Designed structured **data validation and cleansing workflows** (duplicate detection, missing value imputation, error flagging, schema controls) to ensure audit-grade data integrity.
3. Implemented **machine learning–based anomaly detection (Isolation Forest)** to identify 4,976 high-risk transactions, representing **$1.03B+ in potential financial exposure**.
4. Developed **budget vs. actual variance analysis framework**, flagging material deviations (>20%) to detect potential mismanagement or misuse of public funds.
5. Integrated anomaly results, KPI metrics, and fiscal-year filtering (2021–2025) into a unified **Tableau-ready reporting dataset (CSV & Excel)** for executive dashboards.
6. Built reproducible project architecture using virtual environments (venv), modular Python scripts, and structured directory design for scalable audit analytics.
7. Produced executive-ready dashboards and financial impact summaries translating technical findings into actionable risk insights for non-technical stakeholders.

Data Warehouse Analytics: Global Sales Data Warehouse & Dashboards

1. Designed and implemented an end-to-end **data warehouse analytics pipeline** using Python and PostgreSQL, modeling global sales data with a **star schema** (fact and dimension tables) to support scalable analytical queries.
2. Built a Python-based **ETL pipeline** to simulate, clean, transform, and enrich multi-year (2021–2025) sales data, calculating total price, profit, and sales metrics, and exporting analytics-ready datasets in **CSV and Excel formats**.
3. Performed **feature engineering** to derive time-based attributes (year, month abbreviations), regional segmentation, and KPI-ready fields optimized for Tableau and BI dashboards.
4. Loaded transformed data into PostgreSQL using psycopg2, populating **customer, product, date dimensions, and sales fact tables**, and ensuring relational integrity for warehouse analytics.
5. Wrote analytical SQL queries to calculate **revenue by product, country, and month**, total profit, and business KPIs using joins and aggregations across fact and dimension tables.
6. Developed interactive **Tableau dashboards** showcasing global sales performance, monthly trends, product revenue, and customer distribution with shared filters and KPI views.
    a. Implemented a **baseline linear regression model** in Python to predict sales revenue based on quantity and unit price, demonstrating applied analytics beyond reporting.

b. Conducted supplementary exploratory analysis and visualization in **R**, validating sales distributions and price–quantity relationships.
c. Version-controlled the entire project using **Git/GitHub**, including ETL scripts, warehouse schemas, SQL queries, and BI-ready datasets.

Healthcare Data Analytics: Billing, Visits & Outcomes

1. Built an end-to-end healthcare analytics pipeline by cleaning, enriching, and modeling **10,000+ rows of visit and billing data** using Python (openpyxl, NumPy) and Excel-compatible workflows.
2. Developed automated Python scripts to **remove missing billing values**, preserve raw data integrity, and **enrich datasets with multi-year (2021–2025), monthly, regional, state, and insurance-based financial metrics** optimized for analytics and visualization.
3. •ngineered **Tableau-ready data structures** including company-wide yearly billing totals, department-level revenue, insurance-based sales aggregation, and global "ALL" filters to support KPIs, trend analysis, and circle charts.
4. •Performed exploratory data analysis (EDA) in Jupyter Notebook to calculate **total, average, and peak billing**, department visit distributions, and validate enriched financial metrics.
5. Designed and queried a relational healthcare dataset in PostgreSQL using **CTEs, window functions, and aggregations** to analyze department performance, provider visit volume, and billing trends.
6. Developed interactive Tableau dashboards visualizing **billing trends, monthly sales (0–60M range), department revenue, regional performance, and insurance-based sales**, published as a unified analytics dashboard.
    i. Ensured full **macOS Excel compatibility** for stakeholder analysis and dashboard ingestion by delivering cleaned and enriched .xlsx outputs.
    ii. Version-controlled the complete project using Git/GitHub, managing Python scripts, SQL queries, notebooks, and analytics-ready datasets.