Siddhesh Sheth

✉ shethsiddhesh268@gmail.com 📞 8127784657 in LinkedIn ⊙ GitHub 🔗 Portfolio ◆ Google Scholar

EDUCATION

Indiana University Bloomington | GPA: 3.97/4.00   Aug 2022 – May 2024 | Bloomington, USA
Master of Science in Computer Science
 Applied Algorithms | Information Visualization | Advanced Database Technologies | Applied Machine Learning | Data Mining

Pune University | GPA: 4.00/4.00   Jul 2017 – Jul 2021 | Pune, India
Bachelor of Engineering in Computer Science & Engineering
 Data Structures and Algorithms | Computer Networks | Object-Oriented Programming | Software Engineering | Web Technologies

PROFESSIONAL EXPERIENCE

Data Analyst   May 2024 – present | Indianapolis, USA
Hoosier Community Network
- **Ingested CDC healthcare data** into pandas DataFrames using **Python** and **sodapy** library for efficient access to the CDC's **Socrata** open data **RESTful API**, enabling comprehensive analysis of **healthcare metrics** and trends.
- **Deployed a PostgreSQL database on Heroku**, leveraging **SQLAlchemy** for **ORM** (Object-Relational Mapping) to load **DataFrames** into the database.
- Created interactive dashboards by connecting the **PostgreSQL** database to **Metabase**, enhancing data exploration through **SQL queries**. This **integration** reduced reporting time by **30%**, enabling swift data-driven decision-making.

Data Analyst   Mar 2023 – May 2024 | Bloomington, USA
Indiana University, Bloomington
- Led the initial design and indexing of **SQL database schema**, including transactions, orders, **inventory management**, and department tables. This design reduced data storage by **20%** and improved **query performance**, facilitating efficient analysis of sales activities.
- Extracted inventory data from **Excel** files using **openpyxl** and transaction data from the **SQL database** using **SQL queries**.
- Implemented advanced transformations, including **data type conversion** and **normalization**, while conducting rigorous data wrangling and validation to ensure **data integrity** and **accuracy** for downstream reporting.
- Utilized **SQL** for extraction and transformation and created interactive **dashboards** in **Power BI**, applying **DAX** for metrics that visualized sales trends. This effort led to a **15% increase in decision-making efficiency** and improved strategic planning through **real-time insights**.

Database Engineer Associate   Aug 2021 – Aug 2022 | Pune, India
Accenture
- Oversaw the migration, updates, and backups of **AT&T databases** for **Financial Billing Operations** through **SSMS** from **SQL Server** to **Azure SQL Database**, guaranteeing no data loss and a **15%** reduction in recovery time.
- Orchestrated the seamless transfer of packages, overseeing 100+ **SQL Server** databases with **Azure DMS**. Established streamlined handling protocols, resulting in a **20%** reduction in migration time.
- As the lead **RSA Archer Admin**, managed and configured **access roles**, handling **50+ weekly queries** on permissions and security. I implemented advanced workflows and **customizations** to meet **GRC** requirements, reducing GRC-related issues by **75%**.
- Monitored and scaled **Azure VMSS** based on usage, cutting resource wastage by **15%** and maintaining **99.9% uptime**, while optimizing costs by **10%**. Managed **Astra alerts** and implemented **real-time threat detection**, reducing security risks by **20%**, with weekly audits and **vulnerability patching** improving system resilience and cutting **GRC incidents by 30%**.

PROJECTS

Comprehensive Car Sales Analysis and Visualization Using Power BI   May 2024 – Jul 2024
- Used **Power Query** to extract, transform, and load (ETL) over **250k records**, optimizing data models for analysis. Applied advanced **DAX** functions for dynamic **KPI** calculations, **time intelligence**, and **complex measures** to support business requirements.
- Developed interactive dashboards with **drill-through reports**, **cross-filtering**, and custom visuals, including **geospatial analysis** using ArcGIS maps, allowing detailed insights into regional sales performance.
- Integrated **Python** scripts within Power BI for advanced data manipulation and **forecasting models**, enabling accurate **prediction of sales trends** and enhanced analytical capabilities.

Netflix Trend Analysis and Recommendation System   Dec 2023 – Jan 2024
- Implemented a movie **recommendation system** using **TF-IDF vectors** and cosine similarity. Developed models for **predicting** ratings ('Adults,' 'Kids,' 'Teens') and types ('Movies', 'TV-Show') using **Naive Bayes, Logistic Regression, and Decision Tree models**.
- Performed EDA on **50,000+ records** with **Pandas**, visualizing genre trends using **Matplotlib** and **Seaborn**. Identified the **top 10% of genres**, leading to a **30% increase** in interactions with top-rated content.
- Utilized **Matplotlib** and **Tableau** for advanced data visualizations, including **dynamic dashboards** and **parameter controls**, which facilitated **interactive filtering** and real-time data exploration.

Analyzing Stock Market data in Real-Time using Kafka-driven ETL.   Jun 2023 – Jul 2023
- Developed a **Python** script using the **yfinance library** to fetch **1 year of NASDAQ stock data**, iterating through **ticker symbols** and extracting data to **flat files**. Then pushed the data in **S3** and launched an **EC2 Linux instance** for **Kafka setup**.
- **Instantiated a Zookeeper**, created a topic in **Kafka**, and delivered over **1,000 data messages** to a **consumer** and **S3** for ingestion into **Snowflake** using **Kafka Connect**.
- Integrated **Snowflake** with **Snowpipe** for **real-time data ingestion** from **S3**. This enabled enhanced **SQL capabilities** and **materialized views**, aiding data analysis efficiency and **investment decision-making**. Finally, connected **Power BI** to Snowflake for effective **data visualization**.

TECHNICAL SKILLS

Programming Languages — Python, SQL, R, C++, Java, **Data Analysis Tools** — Power BI, Tableau, Metabase, AWS Quicksight, Azure Synapse Analytics, SAS, Google Analytics 4 , **Databases** — PostgreSQL, MySQL, Oracle, NoSQL, RedShift, BigQuery, Snowflake, MongoDB, DynamoDB, ,
Data Engineering — PySpark, Kafka, Airflow, Hive, Hadoop, Databricks, SSIS, SSAS, SSMS, Data Lakes, Data Extraction, Data Warehouse,
Platforms & Security — RSA Archer, SecureID, SSMS, Jupyter, Apache, Anaconda, VS Code, Shell, Bash, **Cloud and OS** — AWS, Microsoft Azure, Heroku, Google Cloud, Windows, Linux, MacOS, Unix, **Other Tools and Technologies** — JIRA, Flask, SaaS, Agile, REST API, Docker, CI/CD, Git

CERTIFICATIONS

Amazon Web Services Certified Cloud Practitioner ☐ — 57H7X4627EQEQ8W8