

# VRUNDA SHAH (U.S CITIZEN)

Data Analyst | Data Scientist | SQL | Python | SAFe Agile | Azure Databricks

+1(603)6740027 | shahvrunda231296@gmail.com | <https://www.linkedin.com/in/VShah231296/> | <https://github.com/VShah2312> | Dallas, Tx, 75019

## SUMMARY

Data Analyst with 3 + years of experience in data modeling and building analytical solutions with skills in Mathematics, Statistics and Machine Learning. Seeking to leverage my passion and skills to drive innovation and solve complex analytical problems at your company.

## PROFESSIONAL EXPERIENCE

### Data Analyst, Tata Consultancy Services (Client: Johnson and Johnson / Kenvue )

03/2022 – Present

- **Tech Led** for semantic team, driving initiatives like Shipment Tracker, Cost to Deliver, Cost to Receive and Cost to Service. Partnered with **cross-functional teams** to gather requirements, designed robust **data models**, and built scalable **ETL** pipelines using **Python** and **SQL**.
- Successfully delivered a shipment tracker integrated with **global ERP and J&J Data Lakes** within tight timelines, enabling real time monitoring of order status across key milestones. Leveraged **strong problem-solving, collaboration, and communication** skills to efficiently manage large-scale datasets and data lakes.
- Optimize existing data pipelines to improve performance and reliability and implemented **data quality** checks to ensure data accuracy. This implementation of **data quality** checks resulted in a 15% improvement in data accuracy, reducing the risk of errors in downstream analytics and reporting.
- Implement continuous integration and delivery (**CI/CD**) steps using Bitbucket and Azure Data Factory to automate data pipelines ensuring timely and accurate delivery of data to downstream systems by incorporating automated testing, version control, and release management into the development process.

### Data Science Mentor, Great Learning

09/2024 – Present

- **Mentored** students by leveraging knowledge of Python, SQL, Statistics, Exploratory Data Analysis and Machine Learning.
- Generated detailed reports using **advanced SQL** queries and integrated data visualization tools like Tableau and Power BI to present key insights, enabling data-driven business strategies.
- Demonstrated expertise in statistical modeling, **feature engineering**, and **predictive analytics** by analyzing real-world case studies to vividly illustrate complex concepts and practical applications.

## EDUCATION

University of Texas at Arlington

M.S Mathematics (Data Science)

08/2019 – 05/2021

University of Texas at Arlington

B.S Mathematics (Pure and Applied)

08/2015 – 05/2019

## PROJECTS

### Amazon Recommendation System

- Leveraged Python to implement advanced machine learning techniques including similarity-based, rank-based, and matrix-factorization based recommendation systems to predict potential customers.
- Enhanced targeted marketing strategies, achieving a 15% increase in campaign conversion rates by applying predictive insights to identify and engage likely customers.

### Predicting Potential Customers

- Employed Decision Trees and Random Forest classification models, optimized through rigorous hyperparameter tuning, to predict potential customers effectively.
- Validated models using statistical hypothesis testing to ensure reliability in varying market conditions, enhancing the robustness of predictive analytics.
- Achieved a 20% increase in customer acquisition rates by applying machine learning insights to marketing strategies, directly impacting the bottom line.

## CERTIFICATIONS

- **MIT: Data Science and Machine Learning** 07/2024
- **Microsoft Certified: Azure Fundamentals (AZ - 900)** 12/2023
- **DAT228x: Developing Big Data Solutions with Azure Machine Learning** 10/2019

## SKILLS

- **Programming & Tools:** Python (Scikit-Learn, Pandas, NumPy, TensorFlow), R, SQL, Git, GitHub, Bitbucket, JIRA, Databricks, Azure Data Factory
- **Data Analysis & Visualization:** Exploratory Data Analysis (EDA), Data Cleaning, Feature Engineering, Data Visualization (Seaborn, Matplotlib, ggplot, Power BI, Tableau)
- **Machine Learning:** Predictive Modelling, Statistical Analysis (Bayesian statistics, Regression), Decision Trees, Random Forests, Clustering, Neural Networks, Hyperparameter Tuning.
- **Cloud Platforms:** Azure, Experience with LLM Ops and MLOps
- **Advanced Concepts:** Large Language Models (LLM) fine-tuning, agentic workflows, document retrieval systems, RAG chatbots
- **Version Control & CI/CD:** Expertise in GitHub, automated pipelines, and CI/CD practices