**CHESTHA SHARMA**

**Contact:** +91 – 9953816491 **| Email:** [chesthasharmadsda@gmail.com](mailto:chesthasharmadsda@gmail.com) **| Location:** Delhi, India **|** [**LinkedIn**](https://www.linkedin.com/in/chestha-sharma/?originalSubdomain=in) **|** [**GitHub**](https://github.com/sharmachestha)

**DATA SCIENTIST PROFILE**

*~ Driving Business Insights through Advanced Data Analytics and Visualization ~*

**Core Skills:** Data Science  Data Visualization  SQL Database Management  Machine Learning  Python Programming  Azure Services   NLP  End User Training  Stakeholder Management

**Certifications:** Python and SQL certification from HackerRank  Applied Machine Learning from IBM

# PROFILE OVERVIEW-----------------------------------------------------------------------------

⮩ Dedicated Data Scientist with over 6 years of experience translating complex business requirements into actionable data- driven insights. Expertise in data architecture, infrastructure design, and data visualization aimed at optimizing business performance.

⮩ Successfully worked on multiple projects, namely - **Barry-Wehmiller – Dimension Model, Church and Dwight**

## - Chatbot Project, One Call - Azure SQL Database Configuration, Telecom Micro-Credit Defaulter, and School Construction Authority - SSRS Reports and SQL Development

⮩ Proven ability to streamline processes, enhance data quality, and improve decision-making.

⮩ Adept at leveraging technical proficiency and analytical rigor to solve business challenges effectively.

⮩ Highly skilled in enhancing operational efficiency and user satisfaction as well as managing the setup and optimization of Azure SQL databases, ensuring robust data storage and retrieval for critical business operations.

⮩ Proficient in developing compelling data visualizations using Tableau, facilitating clear communication to stakeholders.

⮩ Adept at collaborating with cross-functional teams to implement innovative data solutions, contributing to continuous improvement in data engineering practices.

# CAREER REVIEW---------------------------------------------------------------------------------

**DATA Scientist**⯈ **Heineken International Aug 2024 – Present**

## Key Deliverables:

* **Data Pipeline Development and Maintenance**: Implemented and maintained harmonized data pipelines to streamline the integration and transformation of business-critical data from various sources.
  + Set up and managed Git repositories for source control.
  + Updated the Mercury brand P&L data source for the Brazil OpCo in .yml and .sql files.
* **Cross-functional Collaboration**: Worked closely with stakeholders to resolve Jira tickets related to data updates and issue resolution.
  + Regularly collaborated with teams to ensure seamless project updates and task completion.
* **Technology Utilization**: Leveraged a range of technologies, including Git, Python, SQL, Databricks, and Jira, to drive efficiency and support data-driven decision-making.
* **Problem Solving**: Resolved complex issues in data pipelines and ensured data accuracy and integrity, contributing to improved business operations.

**DATA Scientist** ⯈ **Insight Enterprises May 2022 – Aug 2024**

### Key Deliverables:

* **Business Translation and Opportunity Identification:** Clearly translate business needs into actionable data analysis plans and technical solutions. Regularly analyze business processes to identify areas for improvement through data-driven insights.

## Developed a Python chatbot to improve user interaction with data.

* **Data Architecture and Infrastructure Design:** Analyze data requirements and recommend optimal storage solutions based on performance, scalability, and cost considerations. Develop efficient data pipelines and data streams to automate data movement, transformation, and integration. Develop and maintain data warehouse schemas and architectures that support efficient data storage, retrieval, and analysis. Ensure adherence to data governance standards for data quality, security, and access control.

## Managed setup of Azure SQL databases.

* **Data Visualization and User Enablement**: Develop compelling data visualizations using tools to effectively communicate insights to both technical and non-technical audiences.
* **Data Management and Quality Control**: Develop and implement data quality monitoring processes to identify and address data errors and inconsistencies. Analyze data usage patterns to identify opportunities for improving the data analysis experience for users. Develop recommendations for improvement and track their adoption rate and impact on user satisfaction

**DATA Scientist** ⯈ **SVAM International Jan. 2020 – Apr. 2022**

### Key Deliverables:

* **Data Acquisition:** Analyzed data sources (structured/unstructured) to ensure completeness, accuracy, and consistency. Implement data profiling techniques to identify anomalies and potential issues. Collaborated with business stakeholders to define and document data requirements.

## Conducted EDA to understand data distribution and identify potential issues.

* **Data Modeling and Visualization**: Designed and implemented data pipelines for efficient data movement and transformation. Performed regular database tuning and optimization to ensure efficient data storage and retrieval. Developed and maintained data models that meet business needs. Implemented best practices for data visualization to enhance data communication and insights.

## Developed SQL queries and stored procedures to optimize data retrieval for SSRS reports, improving data processing efficiency.

* **Innovation and Continuous Improvement**: Participated in architectural discussions and propose innovative solutions for data management. Contributed to the continuous improvement of data pipelines and infrastructure.
  + **Built a robust classification model using Random Forest Classifier in Python with 91% accuracy.**

# PROJECTS-----------------------------------------------------------------------------------------

**Church and Dwight - Chatbot Project:** Implemented a Python-based chatbot integrated with SQL databases using Jupyter Notebook. Enhanced functionality with @dsr and @docsearch tools leveraging Azure Vector Search Index. Validated Power BI report data through SQL queries, improving data integrity and operational insights.

**Skills Used:** Python, Gen AI, Azure Open AI, LLM,

**Ratings Prediction Project**: Developed a Ratings Prediction system by scraping product reviews and ratings from Flipkart using Selenium, followed by text preprocessing with TF-IDF. Built and evaluated multiple machine learning models, with Random Forest achieving the best performance for predicting product ratings.

**Skills Used:** Python, Selenium, Machine Learning, NLP

**School Construction Authority - SSRS Reports and SQL Development**: Developed 60+ SSRS reports using Visual Studio, utilizing advanced SQL queries and stored procedures. Enhanced data retrieval and analysis capabilities, supported QA activities to ensure accuracy and efficiency. Contributed to improved data-driven decision-making processes within the organization.

**Skills Used:** SQL, SSRS, Visual Studio

**Telecom Micro-Credit Defaulter**: Built a 91% accurate model using Python to predict loan defaulters. Performed EDA, addressed data issues, and achieved 91% accuracy with Random Forest (Jupyter Notebook).

**Skills Used**: Python, Machine Learning

# CREDENTIALS-------------------------------------------------------------------------------------

Dec. 2020 – Oct. 2021 PG Diploma in Data Science, Machine Learning and Neural Networks from DataTrained Education

2017 – 2019 Master’s Degree in Applied Mathematics from Amity University

2014 – 2017 B.Sc. (Hons.) in Mathematics from Mata Sundri College for Women

# IT SKILLS-----------------------------------------------------------------------------------------

Python ◾ SQL ◾ Tableau ◾ Web Scraping ◾ Visual Studio 2017 ◾ Azure Services ◾ NLP ◾ Microsoft Excel ◾

Pandas ◾ Numpy ◾ Seaborn ◾ Matplotlib ◾ Databricks ◾ Azure Open AI ◾ Azure Gen AI ◾ Statistics ◾ Machine Learning ◾ Azure Bot Service ◾ Joins ◾ Windowing Functions ◾ OOPS Concept ◾ Pyspark

◾ Linear Regression ◾ Logistic Regression ◾ Random Forest

# PERSONAL DETAILS-----------------------------------------------------------------------------

**Date of Birth**: 25th October 1996 **| Languages Known:** English and Hindi