

# Rahheb Shaikh

rah014@gmail.com | 857-230-6232 | [linkedin.com/in/rahheb-shaikh/](https://www.linkedin.com/in/rahheb-shaikh/) | Portfolio | Github | Open to relocate

## EDUCATION

**Northeastern University, Boston** *Master's of Science in Informatics - Data Analytics and Science* May 2024  
Relevant coursework: Decision Support and Business Intelligence, Data Management and Big Data, Data Mining Techniques and Applications, Predictive Analytics, Communication and Visualization for Data Analytics, Machine Learning, Data Validation & Quality Assurance.

**University of Mumbai, India** *Bachelor's of Engineering - Information Technology* March 2021  
Relevant coursework: Artificial Intelligence (AI), Data Structures, Cloud Computing, Database Management Systems, Engineering Mathematics.

## SKILLS

**Programming & Analysis** : Python(Pandas, NumPy, TensorFlow, Scikit-Learn), SQL(MySQL, PostgreSQL), R.

**Data Visualization & BI** : Tableau, Power BI, MS Excel, Looker Studio, Matplotlib, Seaborn, Plotly.

**Cloud & Big Data** : AWS, Google Cloud Platform, Microsoft Azure, Apache Spark, Snowflake, DBT, Salesforce.

**Data Science & ML** : Predictive Modeling, Regression Analysis, Hypothesis Testing, Clustering Techniques, Time Series - Forecasting, Statistical Analysis, Segmentation Strategies, Supply Chain Analytics.

## EXPERIENCE

**Data Analyst** October, 2024 – Present  
*Rebecca Everlene Trust Company - Chicago, IL, USA*

- Extracted and analyzed data from over 4,000 companies using Python (Beautiful Soup, Pandas) and SQL (MySQL, PostgreSQL), achieving 98% data accuracy and improving data integrity across critical workflows.
- Enhanced decision-making for 10,000+ students annually by developing Excel-based pivot tables and actionable visualizations utilizing Tableau, optimizing the ASPECTS app for college relocation and cost assessments.
- Standardized and cleaned datasets for 2,500 Illinois-based and 1,500 Hawaii-based companies using Python (NumPy, Pandas), increasing research usability and accuracy by 95%.
- Collaborated with cross-functional teams, including developers and researchers, to implement data-driven strategies, resulting in a 25% improvement in STEAM support for 500+ students across 10+ initiatives.
- Automated data reporting workflows using Python and SQL, reducing manual work by 30% and cutting report generation time by 50%, accelerating project delivery timelines and enhancing decision-making.

**Data Analyst and Scientist (Co-op internship)** October 2023 – May 2024  
*Taxgpt.ca - Boston, Ma, USA*

- Analyzed 11,500+ tax-related queries using NLP (LDA, NMF, LSA), segmenting data into 7 key tax topics, improving information retrieval efficiency by 35% for users.
- Developed and trained machine learning models, achieving a coherence score of 0.712, optimizing tax query categorization accuracy by 25% and reducing misclassification errors.
- Automated data preprocessing workflows with Python (pandas, NumPy, NLTK), cutting manual data cleaning time by 40% and improving dataset usability for model training.
- Designed 10+ interactive data visualizations in Tableau & Seaborn, enhancing stakeholder decision-making and improving executive reporting efficiency by 30%.

**Data and Business Analyst** March 2020 – August 2022  
*Textile Manufacturing Company - Mumbai, MH, India*

- Migrated legacy paper-based inventory and sales records to a cloud-based SQL database (AWS RDS, Google Cloud BigQuery) and developed interactive PowerBI dashboards, enabling real-time tracking and improving decision-making efficiency by 40%.
- Implemented Snowflake as a scalable data warehouse, optimizing inventory forecasting and order management, which reduced inventory waste by 30% and cut order processing times by 25% through data-driven insights.
- Utilized Apache Spark and DBT to analyze customer purchasing trends and supply chain logistics, leading to a 70% reduction in costs by transitioning to local suppliers, reducing delivery times from 1–2 weeks to 1–2 days.
- Utilized Salesforce CRM to manage customer interactions, expanding the client base from 17 to 23 during the pandemic and increasing profit margins by 2% per unit sold through targeted analysis and cost optimization.

## PROJECTS

**Climate Justice Capstone Project** - Web Crawling and Visualizations *Northeastern University* [Link](#)

- Orchestrated an end-to-end data collection project of 2,500+ faculty profiles across 9 Northeastern University schools using Python (BeautifulSoup, Selenium), and developed a Tableau dashboard with 12 interactive visualizations for key insights.

**Mental Health Project** - Machine Learning Project *Personal Project* [Link](#)

- Developed machine learning models, achieving an accuracy of 83%, to predict depressive disorders using a dataset with imputed missing values and feature engineering diverse variables like physical health, stress, and social isolation.

**Retail Customer Segmentation & Sales Analytics** - End to end data analysis *Freelance Project* [Link](#)

- Clustered 90K+ transactions, identifying five customer tiers to optimize marketing and boost targeted promotions.

## Leadership

**Event Management Secretary and Digital Media Marketing Secretary** *CESS & ABIT - RGIT* 2018 - 2021

Led the organization through 27 technical seminars resulting in over 200 engaged participants; generated a comprehensive feedback report that identified 3 key areas of improvements, enhancing seminar effectiveness and participant satisfaction..

## CERTIFICATIONS

Google Advanced Data Analytics Professional Certification *Pursuing* March 2025

Azure Administration Essential Training [Link](#) February 2024