

Surabhi Tiwari

surabhit544@gmail.com — +91-9336696791 — Hyderabad (Open to Relocate)
GitHub: SurabhiTiwari — LinkedIn: Surabhi Tiwari

Objective

Data-driven M.Sc. Statistics student with practical exposure through one internship and six applied data science projects. Skilled in Python, R, SQL, and Tableau with experience handling large-scale datasets. Seeking opportunities to apply analytical and visualization skills to deliver actionable business insights.

Work Experience & Projects

Intern, Ministry of Statistics and Programme Implementation (MoSPI)

28 May 2025 – 25 Jul 2025

- Analyzed **ASUSE 2025 survey data** with over 50,000 records using R, extracting sectoral insights that supported policy-related reporting.
- Developed automated cleaning workflows which streamlined repetitive preprocessing tasks, resulting in a **30% reduction in processing time**.
- Designed and presented **5+ interactive visualizations** using ggplot2, helping senior statisticians communicate results effectively.

World Population Data Visualization Project — Python

- Conducted exploratory analysis on **10,000+ population records** and built charts and dashboards to illustrate age and gender distributions across countries.
- Integrated bar plots, histograms, and time-based trends to showcase global demographic shifts.

Titanic Dataset EDA — Python (pandas, seaborn, matplotlib)

- Cleaned the dataset by imputing missing values, detecting outliers, and restructuring categorical variables.
- Examined relationships among **age, gender, and passenger class** with survival probability, uncovering key trends in survival outcomes.

Decision Tree Classifier on Bank Marketing Dataset — scikit-learn

- Built and tuned a decision tree classifier that achieved **82% accuracy** in predicting customer subscriptions.
- Applied feature engineering, hyperparameter optimization, and validation for model robustness.

Sentiment Analysis of Social Media Data — Python, NLP

- Collected and preprocessed **20,000+ tweets** via APIs and web scraping, creating a structured dataset for NLP tasks.
- Implemented sentiment classification with TextBlob and NLTK to categorize tweets into positive, negative, and neutral classes.
- Developed **6+ visualizations** tracking sentiment shifts and emerging topics, providing insights for **brand monitoring and public opinion analysis**.

Opinion Survey on Plastic Usage — Academic Project (BHU)

- Designed a structured survey and collected responses from **500+ participants**.
- Applied the **Chi-Squared Test** to test associations between demographic variables and plastic usage behavior.
- Analyzed results in MS Excel and documented findings comprehensively in MS Word.

Skills

Statistical Analysis: Regression, Hypothesis Testing, Sampling, Multivariate Analysis
Programming: Python (pandas, NumPy, scikit-learn), R (tidyverse, ggplot2), SQL
Data Visualization: Tableau, ggplot2, Matplotlib, Seaborn
Data Handling: Cleaning, Feature Engineering, Large-scale Data Processing
Productivity Tools: MS Excel (Pivot Tables, VBA), PowerPoint, Word

Education

M.Sc. Statistics, University of Hyderabad

Jul 2024 – May 2026 (Ongoing)

Coursework: Time Series, Multivariate Analysis, Machine Learning with R

B.A. (Hons.) Statistics, Banaras Hindu University

Aug 2020 – May 2023

Subjects: Probability, Linear Algebra, Real Analysis, Sampling

Certifications

- Deloitte Data Analytics Virtual Internship — training in data cleaning, visualization, and storytelling.
- SQL Workshop — TVASTR'25, University of Hyderabad.
- Basic R Programming Workshop — Banaras Hindu University.
- 100 Days of Code: The Complete Python Pro Bootcamp — Udemy.