

JYOTIREKHA BEHERA

+91 8260989516 | beherajyotirekha18@gmail.com | [LinkedIn](#) | [GitHub](#)

SUMMARY

Data professional transitioning into Data Science, with a strong foundation in SQL (joins, window functions, CTEs, query optimization), Python (Pandas, NumPy, Scikit-learn), and Power BI built across two internships and 10+ real-world datasets. Applies statistical analysis, exploratory data analysis, and feature engineering to uncover patterns in data, and is currently deepening Machine Learning, Deep Learning, NLP, and LLM skills through structured, hands-on upskilling to build end-to-end predictive modeling capability.

TECHNICAL SKILLS

- Programming & Libraries: Python (Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn), SQL (joins, window functions, CTEs, query optimization)
- Machine Learning (building foundation): Regression, Classification, Clustering, Feature Engineering, Train/Test Evaluation
- Data Analysis & Visualization: Power BI (DAX, Data Modeling, Power BI Service), Advanced Excel, Exploratory Data Analysis, Metabase (Docker)
- Data Modeling & Databases: Microsoft SQL Server, Fact/Dimension Tables, AWS (Learning)
- Generative AI: Prompt Engineering, LLM Fundamentals, ChatGPT, Claude, Gemini, AI-Assisted Analytics
- Core Skills: Data Cleaning & Validation, Statistical Analysis, Dashboard Development, KPI Tracking, Data Storytelling, Stakeholder Reporting

PROJECTS

Insurance Charges Prediction (Python, Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn)

- Analyzed a healthcare insurance dataset covering age, BMI, smoking status, region, and other demographic and medical attributes to uncover key drivers of medical insurance charges
- Performed exploratory data analysis, including distribution analysis, outlier detection, and correlation analysis, to identify the most influential predictors of cost, such as smoking status and BMI
- Cleaned the dataset and encoded categorical variables to prepare it for a regression modeling pipeline
- Built a baseline regression model to estimate insurance charges from the shortlisted features, forming the basis for further model tuning as Machine Learning skills develop

Spam Message Detection System (Python, Pandas, NumPy, Scikit-learn, NLTK, Streamlit)

- Addressed the problem of identifying spam messages from 1,338 real-world SMS records by developing an automated classification system using NLP and machine learning.
- Preprocessed and transformed text data using NLTK and TF-IDF, then trained and evaluated classification models to distinguish spam from legitimate messages with 80% accuracy.
- Developed a Streamlit web application that enables users to enter messages and receive real-time spam/not-spam predictions.

Sales Data Analysis & Dashboard (SQL, Excel, Power BI)

- Collected and cleaned multi-source sales data using SQL and Excel, ensuring consistency and accuracy across 3,616 orders
- Wrote SQL queries using joins and aggregations to run category-wise and region-wise performance analysis, identifying Clothing and Bikes as the top-performing categories (50K+ units sold)
- Designed and built an interactive Power BI dashboard to track \$78M in sales and \$999K in profit, with drill-downs by category, region, and time period
- Modeled monthly trends, profit margins, and target-achievement metrics to support data-driven decisions on inventory and marketing focus

EXPERIENCE

Data Science & AI Upskilling | Analytixlabs Nov 2025 - Present

- Building foundational skills in Data Science Using Python, Machine Learning, and LLM fundamentals through structured coursework and hands-on practice
- Practicing advanced SQL and exploring Metabase (Docker deployment) and MinusX to extend BI workflows into AI-assisted analytics

Data Analyst Intern | Elevate Labs Sep 2025 - Nov 2025

- Completed 15+ data analysis tasks across 10+ sales, customer, and operational datasets (50K+ records) using SQL, Excel, Python, and Power BI
- Performed data cleaning, EDA, and KPI reporting; built 8+ interactive dashboards to support stakeholder decision-making

Platform Analyst Intern | Inovaare Cloud Solutions Pvt. Ltd., Bhubaneswar Jul 2024 - Sep 2024

- Evaluated 5+ low-code/no-code platforms, identifying 8+ integration and UX gaps, and proposed improvements that led to 20% faster feature implementation

EDUCATION

Siksha 'O' Anusandhan MCA 2023 - 2025 (CGPA: 9.5)

BJB Autonomous College B.Sc. in Bioinformatics 2020 - 2023 (77.5%)

CERTIFICATES / ACHIEVEMENTS

- Certified Data Analyst | Analytixlabs | 2026
- Data Visualization and Analytics | NASSCOM | 2026
- Programming with SQL | Regional Institute of Computer Education | 2024