SHAIK FAYAZ

shaikfaya9@gmail.com ● 7569797615 ● Hyderabad ● www.linkedin.com/in/fayazshaik2001 ● www.github.com/Shaik-Fayaz9

PROFESSIONAL SUMMARY

Analytical and motivated Data Analyst with hands-on training and internship experience in data analysis and visualization. Proficient in **Python, SQL, Power BI**, and **Machine Learning** techniques. Successfully completed multiple real-time projects involving data cleaning, analysis, dashboard creation, and predictive modeling. Strong foundation in transforming raw data into actionable insights to support data-driven decision-making. Eager to apply analytical and technical skills in a professional environment to solve real-world business problems.

SKILLS

Programming
Python (Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn)

Databases
SQL

Microsoft Excel (Advanced)
Pivot Tables, VLOOKUP, XLOOKUP, Power Query, Macros

Machine Learning
Feature Engineering, Model Training, Hyperparameter Tuning, Regularization, Supervised & Unsupervised Learning

Business Intelligence Tools
Power BI, Tableau, Jupyter, Google Sheets

Statistical Concepts
A/B Testing, Correlation Analysis, Regression (basic understanding)

EDUCATION

Master of Business Administration (MBA), Analytics & Data Science,   03/2025 – present │ Jaipur
Manipal University Jaipur

Professional Training - Data Analyst Diploma, *Innomatics Research Labs*   10/2023 – 08/2024 │ Hyderabad

INTERNSHIP EXPERIENCE

Data Analyst Intern, *Innomatics Research Labs*
- Collected, cleaned, and analyzed **10,000+ records** using Excel and Python (Pandas) to identify trends and generate insights for business decision-making.
- Wrote SQL queries to extract and transform data from relational databases, reducing report generation time by **30%**.
- Designed and presented **4+ interactive dashboards** using Power BI/Tableau to visualize KPIs across **3 departments**, supporting strategic decisions.
- Collaborated with cross-functional teams to understand data requirements and deliver weekly reports with actionable insights.

Machine Learning Intern, *The Sun Technologies*
- Performed data preprocessing and exploratory analysis on a **1,000+ record** heart disease dataset, including null value handling, statistical summaries, and class distribution checks.
- Built and trained a **Logistic Regression model** using scikit-learn, achieving **85%+ accuracy** in predicting heart disease on unseen data.
- Evaluated model performance using **accuracy score and confusion matrix**, and validated results with sample patient records.
- Developed a lightweight prediction system to accept **13 input features** and return real-time classification on heart disease risk.

PROJECTS

Banking Customer Churn Analysis (EDA Project)
Tools & Tech: Python, Pandas, Matplotlib, Seaborn, Jupyter Notebook
- Analyzed **10,000+ customer records** to uncover patterns and trends in banking churn behavior.
- Investigated key drivers of churn including **age, tenure, geography, credit score**, and **account balance**.
- Created **10+ visualizations** (bar charts, heatmaps, histograms, pie charts) to support data storytelling and stakeholder understanding.
- Identified that customers in **Germany**, with **low balances or short tenure**, had higher churn risk, and proposed **targeted retention strategies**.

Retail Sales & Customer Insights Dashboard
Tools: Power BI, Excel, DAX, Power Query
- Processed and analyzed **5,000+ rows** of retail sales data to uncover insights on **product performance**, **customer behavior**, and **regional trends**.
- Designed a **Power BI dashboard** showcasing KPIs such as **total sales, profit, customer count**, and **monthly trends** to support business reviews.
- Implemented **filtering by region, customer segment, and product category**, enhancing user-driven data exploration and decision-making.
- Identified **underperforming regions** and **top-growth product categories**, contributing to data-backed retail strategy recommendations.

Heart Disease Prediction Model
Tools & Tech: Python, Pandas, Scikit-learn, NumPy, Logistic Regression, Jupyter Notebook
- Collected and preprocessed a **clinical dataset with 1,000+ records**, handling missing values and analyzing feature distributions for model readiness.
- Built and trained a **Logistic Regression model** using scikit-learn to predict heart disease based on **13 patient health attributes**.
- Achieved **85%+ accuracy** on both training and test sets, demonstrating strong model generalization.
- Developed a **real-time command-line prediction system** that accepts user input and outputs heart disease risk classification.

CERTIFICATES

- Google Data Analytics Professional Certificate – Coursera 🔗
- Data Analysis Certification – Innomatics Research Labs
- Certification in Oracle SQL-PLSQL (SQL queries, stored procedures, database optimization)
- EDA with Python (Pandas, NumPy, Visualization)

LANGUAGES

- English
- Telugu
- Hindi